19 *Proto Austronesian verbal morphology: a reappraisal*

MALCOLM ROSS

1 Introduction

In this paper I suggest that the system of verbal morphology hitherto reconstructed for Proto Austronesian (PAn) did not yet exist in PAn. Instead, the PAn system more closely resembled the pre-PAn system reconstructed by Ross (1995:749, 2002:40). Evidence in support of this suggestion is drawn mainly from the Formosan language Puyuma (Teng 2008a), which reflects the alleged pre-PAn system rather than the system previously reconstructed for PAn. Additional support is found in Tsou and Rukai, two other Formosan languages whose verbalsystems are more readily derived from the pre-PAn system than the PAn system.

A corollary of demoting the reconstructed PAn system to a lower node in the Austronesian tree is that the languages that reflect it belong to a subgroup which excludes Puyuma, Tsou and Rukai. This subgroup, which I dub 'Nuclear Austronesian', includes all other Austronesian languages. That is, I claim (somewhat tentatively) that PAn underwent a primary four-way split into Puyuma, Tsou, Rukai and Proto Nuclear Austronesian (PNAn). This claim entails only a minor conflict with the subgrouping proposals made by Robert Blust.[1] Blust (1999) classifies the Formosan languages into nine subgroups. The proposal here calls into question one of these subgroups, Tsouic, as it treats one of its member languages, Tsou, as a single-member off-shoot of PAn but assigns the other two members, Kanakanavu and Saaroa, to Nuclear Austronesian.[2]

[1] Bob Blust played a major role in introducing me to Austronesian historical linguistics when I first visited Canberra in 1976. He has remained a source of inspiration and has become a good friend, and it is a real pleasure to write this paper in his honour.

[2] I am grateful to Stacy Fang-ching Teng for Puyuma data, to Daniel Kaufman for discussion which stimulated the writing of this paper, and to Andrew Pawley, Lawrence Reid, Stacy Teng, John Wolff and Elizabeth Zeitoun for comments on earlier drafts.

Alexander Adelaar and Andrew Pawley, eds
Austronesian historical linguistics and culture history: a festschrift for Robert Blust, 295–326.
Pacific Linguistics, 2009.

2 Proto Nuclear Austronesian verbal morphology

Table 1 gives an overview of PNAn verbal morphology together with the reconstructed forms of the verb **kiRim* 'seek, look for'. It resembles the PAn morphology table presented in Ross (1995:739) and reproduced with a few changes in Ross (2002b:33).

Table 1: Proto Nuclear Austronesian verbal morphology
(= Proto Austronesian verbal morphology as previously reconstructed)

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

Some of the differences between Table 1 and the earlier tables concern labeling and presentation. I try here to stick to terms and frameworks used by typologists. Instead of positing four voices, I follow Himmelmann (2005) in analysing what he calls 'Philippine-type' languages as having two voices, actor voice (AV) and undergoer voice (UV).[3] Actor voice is intransitive in a number of these languages, whilst UV is transitive in all of them and is usually the default choice in discourse. The grammatical roles of a Philippine-type language are thus ergatively aligned (Starosta 1999; Reid and Liao 2004).[4] As well as default patient-subject UV (henceforth UVP) verb forms, a Philippine-type language has one or two sets of applicative-like forms which promote a location (UVL) or a circumstance role (UVC: instrument, theme or beneficiary) to transitive subject (Starosta 1986, Ross and Teng 2005).[5]

[3] Philippine-type languages include the majority of languages found in Taiwan, the Philippines, northern Borneo and northern Sulawesi. See Himmelmann (2005) for a definition.

[4] Their verbal morphology is not ergatively aligned, since most intransitive verbs are marked by the same morphemes as AV (Ross and Teng 2005).

[5] Philippine-type languages have long been regarded as typologically odd, but Peterson (2007:191–193, 217–219) comments that ergatively aligned languages with applicatives which place a referent in the highly topical subject position are relatively common, at least among languages with applicatives.

In order to compare form–function pairings across languages I have adopted the set of function terms and definitions listed in Appendix A, applying these in the analysis of Formosan verbal morphologies presented in summary form in Appendix B. Sources of language materials are listed in Appendix C.

The reconstruction in Table 1 represents an abstraction at two levels. First, it is inferred from the morphologies in Appendix B. Second, these morphologies are themselves abstractions. In Formosan languages—and in PAn and PNAn—a verb has two forms ('principal parts' in the language of Latin teachers until the mid-twentieth century), neither of which is predictable from the other but from which all other forms of the verb are usually predictable. Verbs fall into five classes on the basis of these two forms, as shown in Table 2 and illustrated from Puyuma.[6] The verb **kiRim* in Table 1 belongs to Class 1.

One of the two forms is the stem, which in PAn and PNAn and in a majority of Formosan languages is the AV imperative or dependent form.[7] A PAn/PNAn simple stem consisted of either a plain root or the root prefixed by **ka*-. There was a strong tendency for verbs with stems in **ka*- to be stative (cf. L.M. Huang 2000; Zeitoun and Huang 2000). There were also stems consisting of a root with a prefix other than **ka*- or of two roots, but these are not shown in Table 2.

The second of the two forms is the AV realis (in Tsou the AV dependent). In this form the morpheme *M*-is applied to the stem. In PAn and PNAn **M*- took three forms: the infix *‹um›*, the prefix **ma*-, or zero. Table 2 provides Puyuma examples from the five classes and illustrates how predictability works.[8] Thus the AV irrealis form **Ca*-STEM is predictable from the AV imperative (STEM) and the AV imperfective form **M*-*Ca*-STEM is predictable from the AV realis form (**M*-STEM). All other forms of the Puyuma verb can be predicted once the two basic forms are known, and the same was evidently true for all the forms of a PAn or PNAn verb.

Table 2: Proto Austronesian, Proto Nuclear Austronesian and Puyuma verb classes

Class		1	2	3	4	5
PAn, PNAn:						
AV imperative =	STEM	*ROOT	*ROOT	*ROOT	**ka*-ROOT	**ka*-ROOT
AV realis =	**M*-STEM	**‹um›*ROOT	**ma*-ROOT	*ROOT	**ma*-ROOT	*ROOT
Puyuma:		'bathe'	'tell'	'give'	'scold'	'be hot'
AV imperative =	STEM	*dirus*	*reŋay*	*beray*	*ka-ḍeki*	*ka-bias*
AV irrealis =	*Ca*-STEM	*da-dirus*	*ra-reŋay*	*ba-beray*	*ka-ḍa-ḍeki*	*bias*
AV realis =	*M*-STEM	*d‹em›irus*	*ma-reŋay*	*beray*	*ma-ḍeki*	*ka-ba-bias*
AV realis imperf. =	*M*-*Ca*-STEM	*d‹em›a-dirus*	*ma-ra-reŋay*	*ba-beray*	*ma-ḍa-ḍeki*	*ba-bias*

6 Tsukida (2005:315) gives a similar table for Seediq.

7 In Kanakanavu it is the UVP dependent form. In Tsou, Ishbukun Bunun and Siraya it does not occur in isolation, but several affixed forms transparently reveal the stem.

8 Puyuma verb classification is more complex than is shown here, but all forms of a verb are predictable from the two basic forms.

There are a few differences between the 1995 analysis and the present one. Some of these arise from a difference in method. The earlier analysis was largely based on a comparison of verbal affixes found in language descriptions. The present analysis is based on tabulations of forms found in the different verb classes of each language and on cross-linguistic comparison not of affixes but of whole verb forms, i.e. forms like those listed in Table 2.

Zeitoun et al. (1996) show that there is a primary division in most Formosan languages between realis mood, encoding realised events and states—present, past and sometimes habitual—and irrealis mood, encoding future and otherwise unrealised events and states. This points back to a similar division in PNAn and PAn. Three sets of realis forms are reconstructed Table 1: a set unmarked for aspect and labelled 'realis' (formerly 'neutral'), a perfective aspect set encoding completed events, and an imperfective aspect set encoding incomplete, ongoing events or changes of state.

In PNAn, unmarked realis, perfective realis and irrealis forms served both as verbs and as gerundive nominalisations. This is annotated in Table 1 by '(V/N)'.

Two comments on the imperfective are pertinent. First, it was evidently marked by **Ca*- reduplication, i.e. by reduplication of the initial syllable and replacement of its vowel by *-a-*. **Ca*- is reflected as *Ca*- in Puyuma, Kanakanavu, Saaroa, Thao, Amis and Siraya but replaced by CV- in Saisiyat, Pazih, Bunun, Paiwan, Yami and Bisayan languages. Second, the imperfective contrasted with a durative (not shown in Table 1), marked by *CVCV-reduplication (*CV- with monosyllables), which apparently encoded iterativity with telic verbs and an enduring event with atelic verbs. The contrast is reflected in Kanakanavu, Saaroa, Pazih and Siraya. There is also a CVCV- durative, but no *Ca*-imperfective, in Amis, Yami and Manobo. The earlier analysis confused matters by labelling as 'durative' what is here labelled imperfective, and Ross (2002b) compounded this confusion by suggesting that both **Ca*- and *CV- were reconstructable as markers of the 'durative' (=imperfective).[9]

Reid (2007) argues on phonological grounds that **Ca*- reduplication must be derived from earlier *CV- reduplication. This is true in principle. However, the fact that **Ca*- and *CVCV- and/or *CV- reduplication are in contrast in some Formosan languages (Zeitoun and Wu 2006) supports the reconstruction of this contrast in PNAn. Further, Puyuma is an external witness to the reconstruction of PNAn **Ca*- and supports its reconstruction in PAn (see below, Table 5). I infer that PAn **Ca*- imperfective reduplication reflects a *CV-reduplication which occurred at a pre-PAn stage for which we have no witnesses, whereas PNAn *CVCV-/*CV- durative reduplication reflects a later innovation, one which took place after the earlier *CV- had become PAn **Ca*-. **Ca*-reduplication was replaced by CV- reduplication in Saisiyat, Pazih, Bunun, Paiwan and Proto Malayo-Polynesian because of its formal and functional similarity to CVCV-/CV- durative reduplication.

**Ca*-reduplication also marked the irrealis, and Ross (1995:751–752) suggested that the irrealis ('future') was simply a functional extension of the imperfective. I am now less sure of this. I tentatively reconstruct a contrast between PNAn realis imperfective AV **‹um› Ca*-STEM and irrealis AV **Ca*-STEM on the basis of Puyuma, an external witness. I also reconstruct a contrast between the corresponding PNAn UVC forms, imperfective **Sa-/Si-Ca*-STEM (Pazih *sa-Ca*-STEM, Paiwan *si*-CV-STEM) and irrealis **Ca*-STEM (reflected in

[9] Puyuma, Kanakanavu and Pazih each have an alternative pattern whereby with certain verbs (membership in the category is morphologically or lexically determined) *Ca*-is replaced by *‹a›* infixation, either between the morphemes of a compound stem or in Pazih after the initial consonant of the stem.

Saisiyat, Atayal and Seediq). Crucially, UVC nominalisations—which reflect irrealis rather than imperfective forms—also reflect **Ca*-STEM, never **Sa-/Si-Ca*-STEM. There was, however, no contrast between UVP or UVL imperfective and irrealis forms. How this came to be is discussed in §6.

A number of Formosan languages have optative forms encoding volition or definite intention, and hortative forms encoding a command addressed to self and hearer ('Let us …' or 'Shall we …?'). The evidence suggests that a single set of forms, labelled optative/hortative (formerly 'projective') in Table 1, had both functions in PNAn. Their reflexes have both functions in Mayrinax Atayal, are optative in Paiwan and Amis, hortative in Seediq, and imperative in Kanakanavu, Saaroa, Ishbukun Bunun, and irrealis in Pazih and Siraya. Since it is fairly clear that imperative and irrealis are encoded by other forms in Table 1, it is a reasonable inference that these meanings reflect extensions in the functions of optative/hortative forms.

Among the imperative and dependent (formerly 'atemporal') forms in Table 1, only UVP forms differ from one another. The assignment of functions to forms in **-u* and **-a* is tenuous. Forms reflecting **STEM-u* occur in Puyuma (an external witness), Saaroa, Paiwan and outside Taiwan in Lun Dayeh. All are UVP imperatives, but the Paiwan form also serves as AV imperative. Forms reflecting **STEM-a* occur as UVP dependent in Tsou (an external witness) and in Yami. In the Bisayan dialects, Manobo, Timugon, Kimaragang and Eastern Kadazan, reflexes are used both as imperatives and in dependent clauses, whilst in Ishbukun Bunun, Bonggi and Lun Dayeh they apparently occur only as imperatives (in Bunun AV, Lun Dayeh UVL). The Kavalan reflex of **STEM-a* is a general UV irrealis, whilst a Javanese reflex serves as imperative and subjunctive and a Proto Malayic reflex as subjunctive (Adelaar1992:148). The hypothesis that best accounts for these data is that PNAn **STEM-u* was imperative but has been displaced in a number of languages by an extension in function of the reflex of the dependent **STEM-a*, probably because AV, UVL and UVC imperative and dependent forms were already identical.

Dependent forms were used after certain preverbs ('auxiliaries' in conventional Formosanist terminology), including negators, as they still are in a range of Formosan and Philippine languages (Ross 1995:744–747). They are also used for foreground events in narrative in Mantauran Rukai, Paiwan, Kimaragang, Eastern Kadazan and Timugon Murut, although in this function the AV form was apparently STEM (rather than **M*-STEM). This was probably an application of their use with preverbs, as certain coordinators meaning 'and then' functioned as preverbs. A sentence thus began with an independent clause with (presumably) a realis verb, followed by one or more clauses each introduced by 'and then' and having a dependent verb. This created a coordinate-dependent form of clause linkage which still occurs in Mantauran Rukai and Paiwan (A.H. Chang 2006).[10] It is not clear whether the 'and then' preverb remains obligatory in the other languages.

The UVC forms in Table 1 require comment. Relevant data are in Table 3. First, alternant forms of the realis circumstance-subject and circumstance-nominaliser prefix, **Sa-* and **Si-*, are reconstructable. No language has regular reflexes of both (the expected Kavalan reflex of **Si-* would be *si-*, not *ti-*). Notably, there are no unambiguous reflexes of **Sa-* with the perfective infix **‹in›*, but reflexes of **S‹in›i-* occur in Saisiyat and Paiwan.

[10] The term 'coordinate-dependent' was coined by Foley (1986) and is synonymous with 'cosubordinate' (Foley and Van Valin 1984). The Papuan languages described by Foley are verb-final and have strings of dependent clauses ending in an independent clause. In Paiwan the pattern is reversed.

Pazih *s‹in›u-* reflects either **S‹in›a-* or **S‹in›i-* irregularly. Perhaps (i) an original **Sa-* became **S‹in›i-* in the perfective then (ii) in some languages a form looking like a reflex of **Si-* spread by analogy to non-perfective realis slots in the paradigm.

In my earlier analysis I doubted whether PAn had circumstance-subject UV forms, as the data then available suggested that most Formosan languages reflected **Sa-* or **Si-* in nominalisations but not in finite verbs (Ross 1995:756–758). I added UVC forms tentatively to my 2002 presentation Ross (2002b:33, 42). However, the reflexes listed in Table 3 suggest quite strongly that in PNAn **Sa-* and **Si-* both formed verbs and nouns.

Table 3: The morphology of circumstance-subject

PNAn	**Sa*-STEM		**Si*-STEM		**Ca*-STEM	
Kanakanavu	—	—	*si*-STEM	N	—	—
Saaroa	*sa*-STEM[*-a*]	V	—	—	—	—
Saisiyat	—	—	*si*-STEM	V	*Ca*-STEM	V, N
Pazih	*sa-/s‹in›u*-STEM	V, N	—	—	*Ca*-STEM	N
Mayrinax Atayal	—	—	*si*-STEM	V	*Ca*-STEM	V, N
Seediq	*se*-STEM	V, N	—	—	*Ce*-STEM	V, N
Thao	—	—	—	—	*Ca*-STEM[*-an*]	N
Ishbukun Bunun	—	—	*is*-STEM	V	…	…
Paiwan	—	—	*si*-STEM	V, N	—	—
Amis	*sa*-STEM	V, N	—	—	—	—
Kavalan	*sa*-STEM[*-an*][a]	N	(*ti*-STEM	V, N)	—	—
Puyuma	—	—	*i*-STEM	N	*i*-*Ca*-STEM	N
Proto Rukai	**sa*-STEM[*-ane*]	N	—	—	—	—

Another fact emerges from Table 1, namely that PNAn **Sa-* and **Si-* occurred only in realis forms (unlike **-en* UVP and **-an* UVL, which also occur in irrealis forms, and unlike **M-*, also found in the optative/hortative form). The PNAn irrealis UVC form is the same as the irrealis AV form: **Ca*-STEM. It has no voice or applicative marker, as **Ca-* characterises the imperfective and the irrealis. It too formed both finite verbs and nominalisations.

The facts in the previous paragraph allow a fresh interpretation of the material presented in Blust (1998). Blust takes **Ca*-STEM to be a template just for instrument (i.e. circumstance) nominalisations, and argues that **Si*-STEM probably had the basic function of forming verbs. Almost the opposite is true: PNAn **Ca*-STEM was a member of the set of irrealis verb forms, whilst PNAn **Sa-/Si*-STEM and its perfective and imperfective variants formed realis verbs. Both could simultaneously serve as nominalisations in PNAn. Ironically, the evidence indicates that in PAn **Ca*-STEM was verbal and **Sa-/Si*-STEM was originally nominal (see §3). This undermines Blust's conclusion that Starosta et al. were wrong to derive PAn (my PNAn) voice and applicative morphology from nominalisation .

Finally, the optative/hortative and imperative/dependent UVC forms in Table 1 need comment. I have assumed on the basis of the Wulai dialect of Atayal (L.M. Huang 1994) that **an-* was a preverb which took UVL suffixes and was followed by the stem form of the

verb.[11] In all other languages that reflect these forms, however, reflexes of **anay* and **ani* are suffixed to the stem: Mayrinax Atayal UVC optative/hortative STEM-*anay*, Paiwan UVC imperative/dependent STEM-*an* (with unexpected loss of **-i*), etc. (see Appendix B).

Where I have **-ani* Wolff (1973) reconstructs **-án*, with invariable stress,[12] contrasting with the UVL suffix **-an*, which was only stressed when suffixed to a stem with final stress. In my earlier analysis I reconstructed stress, but it seems to me that Blust (1997) is correct in maintaining that we lack decisive evidence for the reconstruction of PAn stress and I err here on the side of caution.

3 Nominalisations into verbs

The PAn forms reconstructed by Ross (1995, 2002) were similar to those reconstructed by Wolff (1973). The main advance lay in filling gaps in what are here the imperative, dependent and optative/hortative paradigms: **-aw*, **-anay* and **-u* were added and UVC **-án* was amended to **an-i*. This revealed the paradigmatic pattern in the lower part of Table 1 (Ross 1995:763, 2002b:40), repeated in (1). The middle line expands the optative/hortative suffixes into the morpheme sequences from which they are historically derived: each optative/hortative suffix consists of **-a* plus zero (AV) or an imperative suffix marking UVP or UVL. For convenience I will refer to the optative/hortative set as *a-* grade suffixes and the imperative set as zero-grade suffixes.

(1)

	AV	UVP	UVL
Optative/hortative	**-a*	**-aw*	**-ay*
	**-a-Ø*	**-a-u*	**-a-i*
Imperative	**-Ø*	**-u*	**-i*

The patternedness of this paradigm stands in sharp contrast with the hotchpotch of morphemes in the upper part of Table 1, which includes two infixes, two suffixes, one prefix and **Ca-* reduplication. Of this collection, just one member, **‹um›* AV, also appears in the lower half of the table (in the optative/hortative form).

It was suggested more than thirty years ago (by Andrew Pawley in lectures at the 1977 Institute of the Linguistic Society of America) that this jumble arose because at an earlier stage these morphemes all formed argument nominalisations,[13] which were then reanalyzed as verbs. This hypothesis was elaborated in Starosta, Pawley and Reid (1981), published in abbreviated form as Starosta, Pawley and Reid (1982).[14] The jumble of forms was accounted for by the plausible supposition that a collection of nominalisers might have disparate origins. Supporting the hypothesis is the fact that all the forms in the upper part of Table 1 except the imperfectives (which were probably a PNAn innovation; see §6) are widely reflected as nominalisers in modern languages and the fact that with undergoer verb forms the actor, whether a pronoun or a full noun phrase, in Philippine-type languages is

---

11 In Wulai Atayal the form following the preverb actually reflects **Si-*STEM. I assume this to be an analogical development.

12 Malayo-Polynesian reflexes also support **-an*.

13 Argument nominalisations are nominalisations encoding an actor, patient, location etc, as opposed to action or state nominals, the early Austronesian forms of which have not been reconstructed.

14 As Starosta, Pawley and Reid (1981) point out, various scholars had suggested that undergoer verb forms in Philippine-type languages should be analysed as nominals (Lopez 1941; Capell 1964; Egerod 1966).

almost universally encoded by the genitive (i.e. possessor) case. The hypothesis continues to find support today (Kaufman 2007).

The reanalysis envisioned by Starosta, Pawley and Reid (1981) is illustrated by reconstructed PNAn sentences in (2).[15]

(2) a. **qaLup-en ka babuy*
   i. hunt-NMLZP NOM pig
   'the pigs are something to be hunted'

   ii. hunt-UVP NOM pig
   'the pigs are hunted' > 's/he hunts the pigs'

   b. **qaLup-en na aLak ka babuy*
   i. hunt-NMLZP GEN child NOM pig
   'the pigs are the child's prey'

   ii. hunt-UVP GEN child NOM pig
   'the pigs are hunted by the child'> 'the child hunts the pigs'

The sentence in (2a) is a non-verbal clause with a nominal predicate consisting of the patient nominalisation *qaLup-en* 'something to be hunted, prey'. This is reanalyzed in (ii) as a verbal predicate 'is hunted'. In (2b) the nominal predicate *qaLup-en na aLak* 'the child's prey' includes a possessor, reanalysed in (ii) as the actor.

The sentences in (3) illustrate the corresponding reanalyses of location, circumstance and actor nominalisations.

(3) a. **qaLup-an na aLak [Ca babuy] ka bukij*
   i. hunt-NMLZL GEN child [OBL pig] NOM interior
   'the interior is the child's [pig-]hunting place'

   ii. hunt-UVL GEN child [OBL pig] NOM interior
   'the interior is hunted [pigs] in by the child'
   > 'the child hunts [pigs] in the interior'

   b. **Sa-qaLup na aLak [Ca babuy] ka asu*
   i. hunt-NMLZC GEN child [OBL pig] NOM dog
   'the dog is the child's means of [pig-]hunting'

   ii. hunt-UVC GEN child [OBL pig] NOM dog
   'the dog is used-to-hunt [pigs] by the child'
   > 'the child hunts [pigs] with the dog'

   c. **q<um>aLup [Ca babuy] ka aLak*
   i. hunt-NMLZA [OBL pig] NOM child
   'the child is the one who hunts [pigs]'

   ii. <AV>hunt [OBL pig] NOM child

[15] There are numerous pitfalls in reconstructing phrasal units in a protolanguage, but it is useful to illustrate morphosyntactic structures in this way, and no better alternative comes to mind. The lexical items are drawn from Blust (1995), paying quite careful attention to reconstructed meanings. The case-markers are from Ross (2006).

The nominalisations in (2) and (3) are gerundive (clausal) nominalisations which took arguments of their own,[16] otherwise reanalysis as transitive verbs could not have occurred. An indefinite patient was permitted as an oblique argument of the nominalisation/verb.

Unfortunately gerundive nominalisations in most Formosan languages are not well described, but they clearly encode—or encoded in the past—not only voice but also aspect and mood.[17] Mayrinax Atayal and Puyuma nominalizations have realis/irrealis and perfective/unmarked aspect distinctions (L.M. Huang 2002; Teng 2008a), although it is not entirely clear how many combinations of voice, mood and aspect actually occur in these nominalisations. These distinctions probably also occur in other Formosan languages and evidently occurred in PAn. One mood distinction that has left its mark in lexical nominalisations is the UVC distinction between realis **Sa-/Si-*STEM, reflecting an earlier nominalisation, and irrealis **Ca-*STEM, reflecting the PAn irrealis verb (*pace* Blust 1998). Table 3 indicates that only Pazih and Seediq retain the formal contrast. Whether this reflects a semantic contrast remains undetermined.

It seems improbable that a language would make distinctions in the voice, mood and aspect of nominals that it did not make in its verbal system, and we can thus infer that in the period before nominalisations underwent reanalysis as verbs, i.e. the PAn period prior to PNAn, the verbal system must also have had an actor/undergoer voice contrast, patient-, location- and circumstance-subject verb forms, and the perfective/unmarked-aspect and realis/irrealis distinctions. The question is, what were the verb forms in this period? Starosta, Pawley and Reid (1981) assumed they were the forms that I label as dependent, but the reconstruction in Ross (1995) of what are here called optative/dependent forms suggests that they were partly wrong. All the evidence points to PAn dependent forms only having followed preverbs. This leaves forms ancestral to the PNAn optative/hortative forms as candidates for independent verbs.

For convenience I will use the cover terms 'first-generation verbal affixes' for the optative/hortative (*a*-grade) and imperative and dependent (zero-grade) suffixes in the lowersection of Table 1 and 'second-generation verbal affixes' for the affixes in verbal forms that reflect renalysed nominalisations, i.e. those in the upper section of the table.

4 Puyuma

Table 4 presents a summary of Puyuma verbal morphology, based except as indicated on Teng's (2008) reference grammar, which has substantially advanced knowledge of this language. The striking feature of this table is that reflexes of the second-generation affixes **‹in›*, **-en*, **-an* and **Si-* turn up only in nominalisations. Verb forms proper, below the line, reflect only the suffix array in (1), i.e. first-generation affixes.

There are two alternative explanations of this state of affairs. Either (a) Puyuma has innovated by undoing the reanalysis of predicate nominalisations as verbs which had allegedly occurred by PAn times, or (b) Puyuma continues unchanged the state of affairs reconstructed for pre-PAn. If (b) is true, then the reanalysis of predicate nominalisations as verbs had not occurred in PAn, nor had it occurred in any interstage ancestral to Puyuma.

---

[16] Reflexes in modern languages are often lexical nominals. Gerundive nominals must have undergone lexicalisation throughout the history of Philippine-type languages.

[17] Languages that make multiple distinctions in nominalisations are probably not very common, but they do exist: Turkish nominalisations, for example, distinguish tense and voice (Comrie and Thompson 1985).

Table 4: Puyuma verbal morphology

	ACTOR VOICE (intransitive)	UNDERGOER VOICE		
		Patient subject	Location subject	Circumstance subject
Realis (N only)[a]	*M*-STEM	‹*in*›STEM	‹*in*›STEM-*an*	*i*-STEM
Irrealis (N only)[a]	*Ca*-STEM	*Ca*-STEM-*en*	*Ca*-STEM-*an*	*i*-*Ca*-STEM
Realis	*M*-STEM	STEM-*aw*	STEM-*ay*	STEM-*anay*
Optative/hortative	*M*-STEM-*a*			
Realis imperfective	*M*-*Ca*-STEM	*Ca*-STEM-*aw*	*Ca*-STEM-*ay*	*Ca*-STEM-*anay*
Negative hortative		(no undergoer voice forms)		
Imperative	STEM	STEM-*u*	STEM-*i*	STEM-*an*
Negative	*M*-STEM	STEM-*i*		
Irrealis	*Ca*-STEM	*Ca*-STEM-*i*		*Ca*-STEM-*an*

[a]Undergoer voice forms are from the Ulivelivek, Katipul and Tamalakaw dialects (Stacy Teng, pers. comm.). Nanwang has ‹*in*›STEM[-*an*] for all undergoer voices in the realis and *Ca*-STEM-*an* for all undergoer voices in the irrealis.

Answer (a) requires that precisely the verbal functions of second-generation forms which were gained in PAn were lost again in Puyuma. This is unlikely: we would expect Puyuma to preserve some reflex of the alleged intervening PAn stage, but it doesn't. Answer (a) also requires that PAn (first-generation) undergoer-voice optative/hortative forms have extended their function in Puyuma to include the realis, displacing the PAn second-generation forms—a step which seems quite implausible. Answer (b) on the other hand requires no innovations.

If (b) is true, then, as I anticipated in §1, the pre-PAn state of affairs now needs to be reconstructed for PAn, and the alleged PAn state of affairs is reconstructed only for PNAn (as in Table 1), the interstage ancestral to Atayalic, East Formosan, Paiwan, Bunun, Western Plains, NW Formosan and Malayo-Polynesian. A strong caveat is necessary here. If it can be shown that predicate nominalisations have still not been reanalysed as verbs in some modern languages, as some scholars have suggested (Lopez 1941; Capell 1964; Egerod 1966; Himmelmann 1999; Kaufman 2007), then the innovatory feature of PNAn was that predicate nominalisations replaced first-generation verb forms in main clauses without reanalysis.

5 Matters of method

Before I turn to the details of PAn reconstruction (Table 5), however, some matters of method need to be addressed.

There is little reason to doubt the broad outlines of the reconstruction in Table 1. First, it resembles the systems found in Philippine-type languages. The assumption underlying the table is that these languages are morphosyntactically more conservative than other Austronesian languages. This assumption appears to be justified, as the verbal systems found in other Austronesian languages can be derived from a Philippine-type system, but

not vice versa (Ross 2002a:52–56; Lynch, Ross and Crowley 2002:57–63). Secondly, Starosta, Pawley and Reid (1981) and Ross (1995) assume the upper nodes of Blust's (1977) subgrouping, which makes a primary division into several Formosan subgroups and a single Malayo-Polynesian subgroup embracing all Austronesian languages outside Taiwan. The evidence for this sub-grouping is independent of the verbal system. As the system reconstructed in Table 1 is reflected in more than one Formosan subgroup and in Malayo-Polynesian languages spoken in the Philippines and northern Borneo and northern Sulawesi, it must be attributed to a language ancestral to all the subgroups in which it occurs. Until now this language has been assumed to be PAn.

The claim made here is that the language to which a system like the one in Table 1 is attributable was not PAn but PNAn, a somewhat later interstage language. Puyuma, Rukai and Tsou do not reflect the PNAn system and also have systems that have little in common with each other. One may ask why so little attention has been paid to these languages in past reconstructions of the PAn verbal system. A minor reason is that only a sketchy account of Puyuma grammar (Cauquelin 1991) was available. But the major reason was an (unconscious?) adherence to the 'majority wins' principle. The Philippine type was so common among both Formosan and Malayo-Polynesian languages that it was easy to conclude that PAn was also a Philippine-type language. But within the comparative method of historical linguistics there is no maxim which encourages one to reconstruct on the basis of a majority of witnesses. *A priori* it is just as possible that the PAn system is more nearly reflected by that of Puyuma, Tsou or the Rukai dialects, and that the system in Table 1 should instead be reconstructed for an interstage ancestral to the six groups listed above plus Malayo-Polynesian. The hypothesis I put forward here is that Puyuma verbal morphology more nearly reflects that of PAn than does the verbal morphology of any other Austronesian language.

Only Starosta (1995, 2001) has previously presented a subgrouping based on the inference that the reanalysis of nominalisations as verbs had not occurred in PAn. According to Starosta, PAn split into Rukai and an unnamed subgroup containing all other Austronesian languages; the latter split into Tsou and another unnamed subgroup; the latter into Saaroa and yet another unnamed subgroup, and so on. Puyuma is absent from the 1995 version of Starosta's subgrouping and present in the 2001 version in a lower-level group which also includes Paiwan, Bunun, Siraya, Kavalan, Amis and Proto Malayo-Polynesian. Reasons for this placement are not given. The subgrouping is based on shared innovations in morphology, with some resemblances to those presented here, but it has attracted little attention, because of the obscurity of its presentation and faults in its execution noted by Blust (1999:63–66).

There is, of course, a risk in reconstructing PAn in the way I propose. Four primary branches are now attributed to Austronesian: Puyuma, Tsou, Rukai and PNAn. Whereas PNAn is reconstructed on the basis of a large number of languages, the PAn reconstruction in Table 5 relies heavily on a comparison of Puyuma and PNAn (Tsou and Rukai play smaller roles: see below). If Puyuma has undergone substantial unrecognised innovations since PAn times, then this may distort our reconstruction. For this reason it is important to attend to the details of the reconstruction and the changes which turned PAn into PNAn.

Table 5: A tentative reconstruction of Proto Austronesian verbal morphology

	ACTOR VOICE	UNDERGOER VOICE		
		Patient subject	Location subject	Circumstance subject
Realis (N only)	**M*-STEM	(*STEM-*en*)	(*STEM-*an*)	(**Sa-*/)**Si*-STEM
Realis perfective (N only)	**M*-‹*in*›STEM	*‹*in*›STEM	*‹*in*›STEM-*an*	(*‹*in*›*Si*-STEM)
Irrealis (N only)	C*a*-STEM	C*a*-STEM-*en*	C*a*-STEM-*an*	(**Sa-*/)**Si*-C*a*-STEM
Realis Optative/hortative	**M*-STEM **M*-STEM-*a*	*STEM-*aw*	*STEM-*ay*	**an-ay* + STEM
Realis imperfective	**M*-C*a*-STEM	*C*a*-STEM-*aw*	*C*a*-STEM-*ay*	**an-ay* + C*a*-STEM
Imperative Dependent	*STEM **M*-STEM	*STEM-*u* *STEM-*a*	*STEM-*i*[a]	**an-i* + STEM
Irrealis	*C*a*-STEM	(*C*a*-STEM-*a*)	*C*a*-STEM-*i*	**an-i* + C*a*-STEM

Formatting: Parentheses indicate that a reconstruction is not reflected in Puyuma, Tsou or Rukai. Note that Puyuma reflects **M-‹in›*STEM only in fossilised form. Puyuma *m-inaṭay* 'die' is synchronically AV-die. Diachronically, however, it reflects **M-‹in›aCay* AV‹PF›die, i.e. the AV perfective form of **paCay* 'die', attesting to PAn **M-‹in›*STEM.

[a] The evidence is quite strong that *STEM-*i* was not only UVL, as expected, but also UVP. Only Paiwan has reflexes of both *STEM-*u* and *STEM-*i*, and here the distinction is between imperative and hortative, not between UVP and UVL.

6 Proto Austronesian verbal morphology

A tentative reconstruction of PAn verbal morphology is presented in Table 5—tentative because of the risk just mentioned. Forms which are reflected in PNAn but not in Puyuma, Tsou or Rukai are shown in parentheses.

The principal difference between PAn and PNAn is that in PAn second-generation affixes (above the line in Table 5) only formed nominalisations but in PNAn their function was expanded so as also to encode finite verbs. It is here in Table 5 that three out of four sets of parentheses occur (excluding those around **Sa-*), because Puyuma has undergone aspect syncretism in realis nominalisations, such that the distinction between PAn unmarked-aspect and perfective has been lost and Puyuma retains only a single set of realis nominalisations.

First-generation affixes are assumed to have undergone little change in Puyuma. Puyuman ergative[18] STEM-*i* and irrealis *Ca*-STEM-*i* serve as both UVP and UVL, and expected PAn UVP irrealis **Ca*-STEM-*a* is reflected nowhere, except perhaps in Saaroa UVP imperfective STEM-*a* ('perhaps' because it is not clear whether Saaroa *-a* reflects PAn **-a* or isan irregular reflex of **-en*). PAn UVP dependent *STEM-*a* is not reflected in Puyuma, but it is present in Tsou. Tsou has developed quite differently from Puyuma (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. If the preverb is realis, it agrees with the verb in voice but distinguishes only AV and UV, as in (4).[19]

[18] Puyuma negative forms follow the negative preverb *aḍi* and reflect PAn dependent forms.

[19] Interlinear glosses in examples are modified to reflect the terms used in the text. Abbreviations are: AV – actor voice, DEF – definite, GEN – genitive, IMPF – imperfective, INDEF – indefinite, IRR – irrealis, ITR – intransitive, NMLZ – nominaliser, NMLZA – actor nominaliser, NMLZC – circumstance nominaliser, NMLZL – location nominaliser, NMLZP – patient nominaliser, NOM – nominative NPERS – non-personal, OBL – oblique,

(4) *i=si* *an-a* *'o* *tac~~u~~m~~u~~* *to* *amo*
PREVERB:UV:R=3S eat-UVP NOM banana OBL father
'Father's banana has been eaten.'

The transfer of voice and mood distinctions from the verb to the preverb has meant that Tsou has no reflexes of PAn verbs other than the dependent set, as shown in Table 6. The only exceptions are fossilised forms in lexical nominalisations.

Table 6: Tsou verbal morphology

	ACTOR VOICE	UNDERGOER VOICE		
		Patient subject	Location subject	Circumstance subject
Dependent	*M*-ROOT	ROOT-*a*	ROOT-*i*	ROOT-[*n*]*eni*

If PAn was like Puyuma, then realis and optative/hortative UV forms were identical, using **a*-grade suffixes (presumably in practice the difference was marked by particles or clitics). Only the AV forms differed: the realis form was **M*-STEM, the optative/hortative form **M*-STEM-*a*. When in PNAn the realis was replaced by forms reflecting realis nominalisations,the function of the UV forms was radically narrowed to optative/hortative.

One set of PNAn forms, the realis imperfective, formed with **Ca*- reduplication and second-generation affixes, did not occur in PAn. Their putative ancestors would be PAn realis imperfective nominalisations, but no language reflects such nominalisations. Puyumahas realis imperfective verbs formed with **Ca*- reduplication and **first**-generation *a*-gradesuffixes. For example, PAn **Ca*-STEM-*aw* (realis imperfective UVP) is reflected in Puyuma, where PNAn innovated **Ca*-STEM-*en*. I surmise that the PNAn imperfective forms arose by an analogy with the unmarked-aspect realis forms whereby the new second-generation verbal affixes replaced first-generation suffixes. The basis of the analogy was that PAn AV **M*-STEM served both as a realis verbal form and a realis nominalisation, i.e. **M*- was simultaneously a first-and second-generation affix. Once PAn UV nominalisations had become PNAn realis verbs, the paradigmatic relationship between PNAn realis AV **M*-STEM and realis UVP *STEM-*en* provided a basis for the analogical creation of UVP **Ca*-STEM-*en* from AV **M*-*Ca*-STEM, replacing **Ca*-STEM-*aw*. The same process applied to each of the UV imperfectives, resulting in the conflation of UVP and UVL realis imperfective and irrealis forms shown in Table 1.

Indeed, the presence of **M*- among both first-and second-generation affixes may have provided the trigger for the reanalysis of nominalisations as finite verbs, since only the syntactic context determined whether a form in **M*- was being used in a noun phrase or a verb phrase. This situation continues in Puyuma. Thus in (5), where the nominalisation is undergoer voice, the predicate is marked as nominal both by the determiner *a* and by the suffix *-en*.

If this clause were verbal, we would have no determiner and the finite verbal suffix *-i*, as in (5b).

P – plural, PF – perfective, PSR – possessor, R – realis, S – singular, UV – undergoer voice, UVC – undergoer voice/ circumstance subject, UVL – undergoer voice/location subject, UVP – undergoer voice/patient subject.

(5) a. **Katipul Puyuma** (Stacy Teng pers. comm.):
a *ka-kezeng-en* *ini* *na* *hung*
NOM:INDEF *Ca*-pull-UVP:NMLZ this:NOM NOM:DEF ox
'This ox is (something) to be pulled away.'

b. (manufactured example):
tu=ka-kezeng-i *ini* *na* *hung*
AGT:3S=*Ca*-pull-UVP:IRR this:NOM NOM:DEF ox
'This ox will be pulled away (by someone).'

In (6), on the other hand, only the determiner *na* in (6a) tells us that *na s‹em›a-senay* is a nominal. The form is the same as the verb in (6b).

(6) a. *amau* *kuiku* *na* *s‹em›a-senay*
COPULAR free:1S NOM:DEF ‹AV›IMPF-sing
'The one who was singing is me.' (Ross and Teng 2005)

b. *s‹em›a-senay* *i* *Walegan*
‹AV›IMPF-sing NOM:SG Walegan
'Walegan is/was singing.' (Teng 2008a)

It is only a few short steps from here to the reanalysis of a form like *ka-kezeng-en* in (5) as a finite verb by analogy with the construction in (6a).

The functional load of nominalisations in Puyuma and certain other Formosan languages is high because they are used in relative-clause-like modifier constructions. Teng (2008a) analyses a noun phrase as a series of one or more 'small NPs', each beginning with a determiner. By default, determiners agree in case and definiteness. There is no syntactic marking of head or modifier, and the head small NP may occur anywhere in the noun phrase. Thus in (7) there are three small NPs, all nominative and definite:

(7) *[na* *suan]* *[na* *ma-ʈina]* *[na* *uʈeuʈem]*
[NOM:DEF dog] [NOM:DEF] ITR-big [NOM:DEF ITR:black]
'the [big] [black] [dog]' (Teng 2008a)

The Puyuma equivalent of a relative clause behaves in much the same way: it is a gerundive nominalisation:

(8) *[na* *teɭu-a]* *[na* *kipiŋ]* *[na-ntu*
[NOM:DEF three-NPERS] [NOM:DEF clothes] [NOM:DEF-PSR:3S

d‹in›away *kan* *nanali]*
‹PF:NMLZ›make OBL:SG my.mother]
'the [three] [(pieces of) clothing] [that my mother made]' (Teng 2008a)

Again, when a small NP is in actor voice, there is no distinction between the form of a realis verb and the form of a nominalisation. Teng (2008a) analyses an AV verb in a small NP as a finite verb, as in (9), but, given the fact that UV verbs in this context are all nominalisations, one could also analyse it as an AV nominalisation that is homophonous with the AV realis form. It is precisely this ambiguity which provided the template for post-PAn speakers to reanalyse PAn nominalisations as finite verbs in PNAn.

(9) *[ɖa ma-ʔiɖangan] [ɖa ma-sangal ɖa basak]*
OBL:INDEF old.person OBL:INDEF AV-carry OBL:INDEF sack
'… older people who carry sacks (on their shoulders) …'

The syntactic ambiguity of AV forms in small NPs does not carry over into Puyuma independent clauses because a predicate nominal is always preceded by a nominative determiner, as in (5a), but a verb isn't. Starosta, Pawley and Reid (1981) reconstruct PAn (my PNAn) predicate nominals without a case-marker, as in (2) and (3), because there is usually no case-marker with a Formosan predicate nominal. Puyuma is exceptional in having both definite and indefinite case-marked determiners and in using a determiner with a predicate nominal. If PAn resembled Puyuma in this respect, then loss of the determiner in this context must have preceded reanalysis of nominalisations as verbs. Conversely, the retention of determiners with predicate nominals has prevented reanalysis of Puyuma nominalisations.

The Puyuma small NP construction exists in other Formosan languages, with two differences. First, small NPs other than the first in a sequence are introduced by an invariable linker. Second, whilst a UV verb in a relative-clause-like small NP reflects the same PAn forms—i.e. nominalisations—as in Puyuma, unlike in Puyuma the same verb form also occurs in an independent clause. These points are illustrated by the Paiwan examples in (10), from A.H. Chang (2006).

(10) a. *[a za vatu] [a ku=k‹in›eɭem katiaw], macay=aŋa*
[NOM that dog] [LNK AGT:1S=‹UVP:PF›hit yesterday] AV:die=COMPL
'[That dog] [which I hit yesterday], it's dead.'

b. *ku=k‹in›eɭem a za vatu katiaw.*
AGT:1S=‹UVP:PF›hit NOM that dog yesterday
'I hit that dog yesterday.'

The difference between Puyuma, with agreeing case-marked determiners between small NPs, and most Philippine-type languages, both Formosan and Philippine, with an invariable linker, again suggests that Puyuma is more conservative than other Philippine-type languages. The fact that Philippine-type languages do not agree on the form of the linker (Ross 2006, *pace* Starosta, Pawley and Reid 1981) suggests that linkers have evolved independently at various interstages in their histories, probably from determiners. Indeed, Nanwang Puyuma also allows a linker *na*, apparently reflecting the definite nominative determiner *na*, to be used between small NPs. The simplest inference from these facts is that Puyuma is again uniquely conservative and retains an NP construction prevalent in PAn.

One difficulty remains in the PAn verbal morphology presented in Table 5. The infix **‹in›* is reconstructed in perfective nominalisations on Rukai and Nuclear Austronesian evidence, yet no perfective aspect finite verbs are reconstructed in PAn. This seems odd, and perhaps indicates that Puyuma does not reflect PAn well in this regard. Two facts are relevant here.

First, Puyuma nominalisations with *‹in›* are simply realis and may also encode an imperfective sense. Second, Puyuma encodes perfective aspect with finite verbs with the enclitic *=la*. This situation allows the alternative sets of inferences below.

1. Like Puyuma, PAn marked the perfective in finite clauses with a clitic. A candidate reconstruction is PAn *=*(a)ŋa* (Rukai =*ŋa* PF, Paiwan =*aŋa* COMPLETIVE).
2. Puyuma at some point lost the perfective/unmarked-aspect distinction in its verbal morphology, and later innovated the perfective enclitic =*la*. This would imply that PAn had a now lost set of perfective finite verbs, perhaps combining *‹*in*› with first-generation affixes, but there is no evidence of this.[20]

The data do not allow us to decide between the two possibilities.

Peterson (2007:161–169) has recently offered an alternative account of the origin of second-generation forms. He rejects Starosta, Pawley and Reid's account (§3) on typological grounds, namely that it is unusual for anything except an action or state nominalization to be reanalysed as a main-clause verb form. Instead, he infers that *-*an* UVL arose by the capture of an erstwhile preposition and that **Si-* UVC represents a functional extension of **Si-* 'have, possess, wear'. Thus *-*an*, at least, arose in much the same manner as most applicative affixes in the world's languages. These forms, he suggests, are likely to have arisen in relative clauses, where the object or prepositional object is gapped. He also suggests that they were reanalysed as nominalisations when they were still restricted to relative clauses. Their use was then extended to main indicative clauses.[21]

There are several difficulties with this account. First, because it is framed within the diachronic typology of applicatives, it provides no account of *-*en* UVP, ignoring the fact that the UVP form is equal in morphological complexity with the UVL and UVC forms. Second, the claim that relative-clause use was prior to nominalisation is based on a generalisation from diachronic typology, but this generalisation is contradicted by DeLancey's (1986) account of the history of Lhasa Tibetan relative clauses, in which nominalisations have come to serve as relative-clause-like modifiers in a manner almost exactly parallel to that entailed in Teng's (2008) analysis of Puyuma. Third, as noted earlier, the difference between Puyuma verbal morphology and corresponding morphology in Nuclear Austronesian Formosan languages is more readily accounted for under the supposition that nominalisations became finite verb forms than that finite-verb uses of these forms were lost in Puyuma. Finally, Peterson's account offers no explanation of the fact that the genitive serves as the agentive case.

In fact first-generation suffixes are more promising candidates for captured morphemes, as noted by Starosta (1995:703–704).

7 The non-Nuclear Austronesian languages

There are three non-Nuclear Austronesian languages: that is, languages that do not reflect the innovations of PNAn verbal morphology. They are Puyuma, Tsou and Rukai. Puyuma verbal morphology has been dealt with at length above.

---

[20] The infix *‹*um*› was both a first-and second-generation affix, and it would not be surprising if this were also true of the infix *‹*in*›.

[21] Peterson (2007:167) also proposes an alternative scenario whereby 'there simply is no true direct diachronic relationship between relativizations/nominalizations and the focus constructions: they simply share related source elements.'

Tsou reflects only the PAn dependent forms (Table 6), and this entails a problem noted by Starosta (1985). Not only does Tsou lack second-generation verbal forms: it also appears to lack nominal reflexes of second-generation affixes. Starosta rightly makes the point that if these forms ever occurred in the language, they ought to be reflected at least in fossilized form in lexicalized nominals, and yet they are apparently not found.[22] This, Starosta suggests, means that second-generation affixes had not yet been grammaticised as nominalisers when Tsou broke away from the language ancestral to the rest of Austronesian, a suggestion that merits further investigation.[23]

Table 7: Proto Rukai verbal morphology

NOMINALISATIONS	
Agentive	…
Patient	**a*-STEM-*anə*
	**‹in›*STEM-*anə*
Location	**ta*-STEM-*anə*
Instrument	**sa*-STEM
Realis	**u-a*-STEM
Subjunctive	**‹u›*STEM
Imperative	**‹u›*STEM-*a*
Dependent	*STEM
Passive	**ki-[a]*-STEM

The Rukai system has moved in a direction quite different from Puyuma or Tsou. Rukai has six recognised dialects. Because of differences among them, it is simpler to work here with the Proto Rukai forms set out in Table 7, based mostly on material from Zeitoun (2003).

Proto Rukai has lost the undergoer voice and has acquired a passive reflecting grammaticisation of the PAn lexical prefix **ki-*'get, obtain', also reflected with the same function but a much lower functional load in Puyuma and Paiwan (Zeitoun and Teng 2006).[24] In other Formosan languages the default choice in narratives is an undergoer voice. In Rukai it is the actor voice. The history of Rukai is complex and still something of a mystery (Ross 2003), but Table 7 allows several observations and inferences. Reflexes of the PAn second-generation affixes **‹in›*, **-an* and **Sa-* are alive and well in Rukai nominalisations but not among finite verbs—the same situation as in Puyuma. Like Tables 1 and 5 and the table in Appendix B, Table 7 is a summary of forms across the PAn verb classes illustrated in Table 2, and it is only when we examine the Rukai verb classes in Table 8 that certain probable historical facts emerge.[25] The correspondence between the five

[22] Szakos (n.d.) records numerous placenames and family names in *-ana*, apparently reflecting PAn **-an*, but the absence of other lexical items in *-ana* opens up the possibility that these are all borrowings. Szakos (1994:73–76) records nominalising affixes, the most frequent of which is *le-*, prefixed to a verb inflected forvoice to form agent, location and instrument nominals. I have found no cognates.

[23] In his 1995 subgrouping, however, Starosta places Rukai—which does reflect second-generation affixes as nominalisers—at a node above Tsou. No reason is given for the abandonment of the 1985 position.

[24] There is a second, non-agentive, passive reflecting a prefix **ku-*, but it is missing from Mantauran and Maga and may be a later development.

[25] The passive is omitted from the table, as its form is **ki-*[*a-*]ROOT with all verb classes (**-a-*is reflected in all dialects except Mantauran)

PAn/PNAn verb classes in Table 2 and the five Proto Rukai classes is imperfect. Proto Rukai class 5 clearly corresponds with PAn class 4. In Proto Rukai classes 2 and 4, **m-* ROOT means that the root either has an initial vowel or an initial **p-* or **k-*, replaced by **m-*.

Table 8: Proto Rukai verb classes

Class	PAn source	1	2	3	4	5
Proto Rukai:						
Realis	?	**u-a-*ROOT	**u-a-*ROOT	**m-*ROOT	**a-m-*ROOT	**ma-*ROOT
Subjunctive	**M-*STEM	* ROOT	**<u>*ROOT	**m-*ROOT	**m-*ROOT	**ma-*ROOT
Imperative	**M-*STEM-*a*	* ROOT-*a*	**<u>*ROOT-*a*	...	**m-*ROOT-*a*	**ma-*ROOT-*a*
Dependent	* STEM	* ROOT	* ROOT	* ROOT	* ROOT	**ka-*ROOT

Crucially, the prefix **ma-*in class 5 corresponds with **ma-* in PAn class 4, and we know from Table 2 that **ma-*ROOT is a manifestation of PAn AV **M-*. There is a complication, however: in Table 8 class 5 **ma-* corresponds both with the realis and with the subjunctive and imperative forms in the other classes. The most straightforward inferences here are as follows.

1. The subjunctive and imperative sets reflect PAn AV **M-*STEM,**‹u›* reflecting **‹um›* with nasal loss.
2. The imperative reflects PAn **M-*STEM-*a* (AV optative/hortative).
3. Realis **u-a-*in classes 1 and 2 and **a-m-*in class 4 reflect a combination of PAn **M-*STEM and a prefix or infix **(-)a-*, the origin of which remains unknown.[26]
4. Dependent STEM reflects the PAn **M-*STEM AV dependent.

The fact that Rukai retains second-generation morphemes only in nominalizations meshes with the fresh reconstruction of PAn verbal morphology proposed here. It could of course be argued that if second-generation UV morphemes had been reanalysed as finite verbs, they would then have been lost in any case. However, we would expect to find some fossil record of their verbal use somewhere in the language, and we don't.

Kanakanavu and Saaroa have previously been subgrouped as 'Tsouic' along with Tsou. However, both languages reflect second-generation affixes in finite verbs (Appendix B), indicating that they are Nuclear Austronesian and do not subgroup with Tsou. Confusingly, both languages also appear to have a first-generation affix in one or more UVP forms where, on the basis of the reconstructed PNAn system in Table 1, it should have been replaced by a reflex of second-generation **-en*. The critical forms are Kanakanavu STEM-*ai* and Saaroa UVP forms in *-a*.

Kanakanavu STEM-*ai* is labeled 'OF2' (object focus 2) by Mei (1982:212–214) and 'special focus' by Tsuchida (1976). According to Mei (1982:227–228) the UVP form STEM-*ene* (his 'OF1') only occurs in certain subordinate clauses, while STEM-*ai* is the default UVP form.

[26] Starosta (1995:701–702) and Ross (1995:746–747) both suggested that Proto Rukai realis **u-a-*reflected a PAn preverb, a grammaticisation of **ua* 'go'. Zeitoun (2003) has questioned this, and it is clear from inspection of the full range of verb classes that she was right to do so.

If Mei is correct, then STEM-*ai* retains its PAn realis status,[27] a fact which would be problematic for its Nuclear Austronesian status. But Tsuchida's analysis is quite different: STEM-*ene*, regularly reflecting *-*en*, is UVP realis, and STEM-*ai* occurs in circumstances that he does not properly understand, but always in narrative. The texts in Tsuchida (2003) bear this out, and this is the analysis adopted in Appendix B. Kanakanavu is thus straightforwardly Nuclear Austronesian.

Saaroa UVP forms in *-a* raise a different difficulty, namely that *-a* is found where a reflex of PNAn *-*en* is expected. Is Saaroa *-a* (i) a reflex of PAn UVP dependent *-*a*? Or (ii) an irregular reflex of PNAn *-*en*? Two facts favour (ii). First, *-a* co-occurs with *li-*, the Saaroa reflex of *‹*in*›*, and the latter never coocurs with first-generation affixes. Second, PAn *-*a* marked the dependent, and would have undergone a massive extension of function to occur in the realis, imperfective, perfective and irrealis as Saaroa *-a* now does. For the time being I assume that (ii) is true and that Saaroa is also straightforwardly Nuclear Austronesian.

8 The agent case question

In Philippine-type languages the subject NP is marked as nominative and in an undergoer voice clause the agent NP is marked as genitive. Both are referenced by pronominal clitics.[28] In most languages these are second-position enclitics, i.e. they follow the first constituent of the clause, which in many languages is usually the verb but in some is often a preverb. Only *one* set of clitics is reconstructable for PAn, referencing either subject or agent (Ross 2006:532), and this situation apparently still prevailed in PNAn.

Starosta, Pawley and Reid (1981) argued that agents in Philippine-type languages are in the genitive case—the case of the possessor—precisely because the verbs with which they co-occur were once gerundive nominalisations which would have treated their agents as possessors.[29] If this is so, then we would expect to find that agents in languages where nominalizations have not been reanalyzed as verbs—Tsou, Rukai and Puyuma—are not in the genitive case. This is arguably what we find in Tsou. Rukai, described in §7, has lost the undergoer voice and become an accusative language, and so the agent case question does not arise. The Puyuma situation is more complicated.

Tsou enclitics reflect the probable PAn situation: there is only one enclitic pronoun set (the enclitic is attached to the preverb) and it marks both nominatives and agents.

Puyuma pronominal clitics are shown in Table 9. The agent (AGT) and possessor in nominative-NP proclitics (P/NOM) and the nominative enclitics differ little from each other in form, reflecting the single set of PAn clitics. The procliticisation of the agentive forms is explained by Starosta, Pawley and Reid (1981) and Wolff (1996): originally an enclitic to the preverb (as in Tsou), with loss of the preverb the pronominal became proclitic to the following verb. The history of Puyuma possessor pronouns awaits explanation.

With regard to NP case-marking, Tsou makes no genitive/oblique distinction: there is only an oblique case. There is thus no genitive agent marking in Tsou. The same is true of the Nanwang dialect of Puyuma, with oblique common *kana*, personal *kan*. In Katipul and Ulivelivek Puyuma, this is true of indefinite noun phrases, but case-markers in both

---

[27] But with a change of undergoer subject from UVL to UVP—a common enough shift.

[28] There is some variation across languages as to whether a clitic is invariably present or is present only when there is no subject/agent noun phrase.

[29] Capell (1964) considered this a reason to analyse undergoer voice verbs as nominals.

dialects distinguish between definite genitive (common *na* or *nina*, personal *ni*) and definite oblique (common *kana*, personal *kani*). In Ulivelivek an agent is marked as oblique, in Katipul as genitive (Teng 2008b). It is a reasonable inference that Katipul reflects pre-Puyuma morphology, since the obliques appear to consist of a preposition *ka-* and a genitive case-marker, but we cannot tell whether the pre-Puyuma agent was case-marked as genitive (as in Katipul) or as oblique (as in Nanwang and Ulivelivek).

Table 9: Puyuma pronominals

	1S	2S	3S	1IP	1EP	2P	3P
PAN	**=ku*	**=Su*	–	**=(i)ta*	**=mi[a]*, **[S]ami*	**=mu*	–
NOM	*=ku*	*=yu*	∅	*=ta*	*=mi*	*=mu*	∅
AGT	*ku=, ti=*	*nu=*	*tu=*	*ta=*	*mi=*	*mu=*	*tu=*
P/NOM	*ku=*	*nu=*	*tu=*	*ta=*	*niam=*	*mu=*	*tu=*

The agent is thus not encoded by a distinct genitive case clitic in Tsou, and I take this to have been the situation in PAn. It would be convenient to claim that the same was true of agent noun phrases. It is true of Tsou and of two Puyuma dialects, but not of definite agents in the third.

We are left, in any case, with a puzzle. In Nuclear Austronesian languages of the Philippine type we would expect undergoer voice verbs with second-generation forms to be accompanied by an agent in the genitive case, but those with first-generation forms to be accompanied by an agent in the oblique case. To my knowledge this situation does not occur: the agent of an undergoer voice verb is in the genitive case regardless of that verb's form. We can only infer that, perhaps as early as PNAn, the new case alignment appearing with second-generation forms was generalised to first-generation forms.

9 Subgrouping

The set of innovations involved in the reanalysis of nominalisations as finite verbs (§3) is complex, and it is improbable that they occurred independently in different languages. Instead, they probably occurred once, in PNAn. No shared innovations have been found supporting a subgroup containing two or three of Puyuma, Tsou, Rukai and PNAn, and so each is assumed to form a primary subgroup of Austronesian in its own right. That is, Austronesian has four primary branches: Puyuma, Tsou, Rukai and Nuclear Austronesian. In this concluding section I compare this subgrouping briefly with three other current hypotheses: Tsuchida's (1976:9–15), which continues to be cited in descriptive works, Sagart's (2004), and Blust's (1999).

The Nuclear Austronesian hypothesis is compatible with neither Tsuchida's nor Sagart's subgroupings, as it cuts across a major subgroup in each case. It cuts across Tsuchida's Southern Formosan (consisting of all Formosan languages except Atayalic) and across both its member subgroups, Rukai-Tsouic and Paiwanic, the latter consisting of Puyuma, Siraya, Paiwan, Amis, Bunun, Thao, Saisiyat and Pazih.

Sagart has three primary subgroups: Saisiyat and Pazeh are single-language subgroups, whilst Pituish contains all other Austronesian languages. Pituish in turn consists of some single-language subgroups (Thao and four extinct languages) and Enemish, consisting in its turn of Siraya and Walu-Siwaish. The latter contains six subgroups: Puyuma, Rukai-

Tsouic, Amis, Bunun, Paiwan and Muish, which includes Kavalan and Malayo-Polynesian. Nuclear Austronesian cuts across Walu-Siwaish, so that the Nuclear Austronesian hypothesis and Sagart's are irreconcilable.

The incompatibilities between these three hypotheses reflect differences in method (Tsuchida's subgrouping is based on lexicostatistics, modified by certain shared innovations) and the differences in the weight given to different kinds of innovation. The Nuclear Austronesian hypothesis rests on innovations in verbal morphology, whilst Sagart's Pituish, Enemish and Walu-Siwaish, are based on innovations in terms for the numerals 5 to 9.

Blust (1999:44–53) uses phonological evidence to place Formosan languages into nine subgroups. Three are established on the basis of shared phonological innovations (East Formosan, Western Plains and Northwest Formosan), four have only a single language each (Puyuma, Rukai, Paiwan and Bunun), and two, Atayalic and Tsouic, are taken as established on the basis of research by other scholars. Of the nine subgroups, six (NW Formosan, Atayalic, Western Plains, Bunun, Paiwan, and East Formosan[30]) reflect the PNAn system in Table 1 and two (Puyuma and Rukai) do not. The ninth subgroup, Tsouic, appears to fall in both camps: two members, Kanakanavu and Saaroa, discussed in §7, reflect the system in Table 1 and the third, Tsou (Table 6) does not.

Blust's hypothesis and the Nuclear Austronesian hypothesis can thus be reconciled provided that Blust's Tsouic subgroup is broken into Kanakanavu and Saaroa on the one hand and Tsou on the other. Significantly, perhaps, Tsouic is one of the two subgroups[31] which Blust (1999:52) takes as established on the basis of the work of other scholars. Since he establishes other groups on the basis of phonological innovations, I infer that he has not found phonological innovations which uniquely associate the Tsouic languages. In his list of 'significant mergers' Tsou, Kanakanavu and Saaroa appear to share three innovations: (i) PAn **j* is lost; (ii) merger of PAn **S* and **s*; (iii) merger of PAn **k* and **g*. None of these is convincing as a uniquely shared innovation, i.e. one which occurred in a putative Proto Tsouic, nor does Blust claim that they are. Loss of **j* is categorical in Tsou, but occurs only adjacent to **i* in Kanakanavu and Saaroa. Mergers (ii) and (iii), on the other hand, are categorical in Kanakanavu and Saaroa but only partial in Tsou. There are grounds here for a subgroup comprising Kanakanavu and Saaroa, but not including Tsou.

Blust's case for Tsouic rests on Tsuchida's (1976) work, but Tsuchida does not provide a list of shared innovations. Blust perhaps assumes that the exclusively shared lexicon of Tsou, Kanakanavu and Saaroa (Tsuchida 1976:6–10), established lexicostatistically, is extensive enough for him to infer that the three languages share a significant collection of shared innovations. However, identifying lexical innovations entails distinguishing between them and shared inheritances. This is easier to do within an Austronesian subgroup like Oceanic, where one can appeal to non-Oceanic languages as external witnesses. It is difficult when one is dealing with the primary subgroups of Austronesian, as there are no external witnesses to help determine which items should be reconstructed for PAn—and are therefore inherited into daughter languages—and which items are innovations in primary subgroups. Tsuchida (1976:15), incidentally, considered Kanakanavu and Saaroa to form a subgroup within Tsouic.

---

[30] Blust's NW Formosan consists of Kulon-Pazih and Saisiyat and is only weakly supported, as he points out. The only diagnostic innovation is **C > s*, and this was also reflected in the extinct Central Western Plains languages Taokas, Papora and Hoanya.

[31] The other subgroup is Atayalic, whose members are so similar that their relationship is obvious by inspection.

In a recent paper, H.Y. Chang (2006) has also questioned the Tsouic subgroup, recognising that Tsou displays morphosyntactic features which do not occur in any other Formosan language. His observations are correct (and it is useful to have them in a single publication), but he takes them to be innovations relative to the earlier reconstruction of PAn morphosyntax (Wolff 1973; Ross 1995) which single Tsou out as a subgroup in its own right. However, as he says himself, these innovations do not necessarily speak against a Tsouic subgroup: they could have occurred after Tsou speakers had become separated from Kanakanavu and Saaroa. Refuting the existence of a Tsouic subgroup entails showing that alleged shared Tsouic innovations are not what they seem, as I have tried to do above. It also entails proposing an alternate subgrouping hypothesis and showing that Tsou, Kanakanavu and Saaroa cannot belong to the same subgroup: this is a spin-off of the Nuclear Austronesian hypothesis.

If the Nuclear Austronesian proposal is superimposed on Blust's subgrouping, the result is as shown below. Ten primary subgroups are reduced to four:

1. *Puyuma*
2. *Rukai*
3. *Tsou*
4. Nuclear Austronesian
    a. *Kanakanavu* and *Saaroa*
    b. Northwest Formosan: *Saisiyat, Kulon-Pazih*[32]
    c. Atayalic: the dialects of *Atayal* and *Seediq*
    d. Western Plains: *Thao, Taokas, Favorlang-Babuza, Papora, Hoanya*
    e. *Bunun*
    f. *Paiwan*
    g. East Formosan: *Basay-Trobiawan, Kavalan, Amis, Siraya*
    h. Malayo-Polynesian: all extra-Formosan Austronesian languages (including *Yami*, which lies within Taiwan's political boundary)

Appendix A: Functions terms and their definitions

In order to compare form–function pairings across languages the following function terms and definitions are used:

Realis:	non-future, used for present, past and sometimes habitual.[33]
Irrealis:	future and hypothetical non-future events.[34]
Subjunctive:	irrealis used only in subordinate clauses.[35]
Non-past:	future, habitual (in Amis only).
Perfective:	completed, usually past, realis event.

[32] The evidence for Northwest Formosan is weak, and I would prefer to treat Saisiyat and Kulon-Pazih as separate subgroups, but this issue lies beyond the scope of this paper.

[33] In Teruku also after one future auxiliary.

[34] In Saaroa also habitual and negative.

[35] In Mantauran Rukai also used as an imperative.

Imperfective: incomplete realis event regardless of tense, typically an ongoing atelic event (progressive, or, with a stative verb, a change of state), or sometimes an iterative or habitual event.

Durative: a process continuing for an appreciable time: 'keep/kept on doing something', a repeated or frequent action; contrasting in some languages with imperfective.

Imperative: command addressed to hearer.

Hortative: inclusive imperative: command addressed to self and hearer ('let us').

Optative: volition or intention.

Narrative: event in a narrative sequence (in Kanakanavu) or after a coordinator in a coordinate dependent clause (Paiwan, Mantauran Rukai, Kimaragang, Eastern Kadazan, Timugon Murut).

Dependent: after a preverb.

Negative: after a negative preverb (i.e. negative is a subset of dependent).

Timerative: in Amis only, 'I am afraid that'

Appendix B: Verb forms in Formosan languages

This appendix sets out the verb forms in Formosan languages on which the reconstructions in the paper are based. These forms are the result of an application of the analytic approach described in §2 to materials from the sources listed in Appendix C.

Verb forms in Formosan languages

	ACTOR VOICE	UNDERGOER VOICE		
		Patient subject	Location subject	Circumstance subject
Puyuma				
Realis nominal	*M*-STEM	<*in*>STEM	<*in*>STEM-*an*	*i*-STEM
Irrealis nominal	*Ca*-STEM	*Ca*-STEM-*en*	*Ca*-STEM-*an*	*Ca*-STEM-*an*
Realis	*M*-STEM	STEM-*aw*	STEM-*ay*	STEM-*anay*
Optative/hortative	*M*-STEM-*a*	STEM-*aw*	STEM-*ay*	STEM-*anay*
Imperfective	*M-Ca*-STEM	*Ca*-STEM-*aw*	*Ca*-STEM-*ay*	*Ca*-STEM-*anay*
Imperative	STEM	STEM-*u*	STEM-*i*	STEM-*an*
Negative	*M*-STEM	STEM-*i*	STEM-*i*	STEM-*an*
Irrealis	*Ca*-STEM	*Ca*-STEM-*i*	*Ca*-STEM-*i*	*Ca*-STEM-*an*
Proto Rukai				
Nominal	–	**a*-STEM-*anə*, *<*in*>STEM-*anə*	**ta*-STEM-*anə*	**sa*-STEM
Realis	**M-Ca*-STEM ?	–	–	–
Subjunctive	**M*-STEM	–	–	–
Imperative	**M*-STEM-*a*	–	–	–
Narr./Dependent	*STEM	–	–	–

	ACTOR VOICE	UNDERGOER VOICE		
		Patient subject	Location subject	Circumstance subject
Tsou				
Dependent	*M*-STEM	STEM-*a*	STEM-*i*	STEM-[*n*]*eni*
Kanakanavu				
Realis	*M*-STEM	STEM-*ene*	STEM-*ene*	–
Future	...	...	a-STEM-*ene*	–
Imperfective	*M*-*Ca*-STEM	...	...	–
Perfective	<*in*>*M*-STEM	<*in*>STEM	<*in*>STEM-*ane*	–
Nominal	–	<*in*>STEM	*ta*-STEM-*ane*	*si*-STEM[36]
Narrative	–	STEM-*ai*	–	–
Imperative	*M*-STEM-*a*	STEM-*au*/-*i*	STEM-*au*/-*i*	–
Dependent	–	STEM	–	–
Durative	*M*-CV-STEM	–	–	–
Saaroa				
Realis	*M*-STEM	STEM-*a*	STEM-*a*[*na*]	*sa*(*a*)-STEM[-*a*]
Imperfective	*M*-*Ca*-STEM	*Ca*-STEM-*a*	*Ca*-STEM-*a*[*na*]	–
Perfective	*ɬi*-*M*-STEM	*ɬi*-STEM-*a*	*ɬi*-STEM-*a*[*na*]	V
Irrealis	*a*-STEM	*a*-STEM-[*a*]	*a*-	STEM-*a*[*na*]
Imperative	*M*-STEM-*a*	STEM-*u*	STEM-*i*	STEM-*ani*[37]
Negative	STEM	–	–	–
Saisiyat				
Realis	*M*-STEM	STEM-*en*	–	*si*-STEM
Imperfective	CV-*M*-STEM	–	–	–
Perfective	<*in*>*M*-STEM	<*in*>STEM	<*in*>STEM-*an*[38]	<*in*>*si*-STEM[39]
Irrealis	(*ʔam M*-STEM)	*ka*-STEM-*en*	*ka*-STEM-*an*	*Ca*-STEM, *ka*-STEM
Nominal	*ka*-*ma*-STEM	<*in*>STEM, *ka*-STEM[-*en*]	*ka*-STEM-*an*	*Ca*-STEM
Imper./dependent	STEM	STEM-*i*	–	STEM-*ani*
Pazih				
Realis	*M*-STEM	STEM-*en*	STEM-*an*	*saa*-STEM
Future	CV-STEM-*ay*	CV-STEM-*en*	CV-STEM-*ay*	...
Imperfective	CV-STEM, <*a*>STEM	CV-STEM-*en*	CV-STEM-*an*	*sa*-CV-STEM
Durative	*M*-CVCV-STEM	...	...	...
Perfective	<*in*>M-STEM	<*in*>STEM	<*in*>STEM-*an*	*s*<*in*>*u*-STEM
Nominal	*ta*-STEM	<*in*>STEM, CV-STEM-*en*	<*in*>STEM-*an*, [*ta*-/*Ca*-]STEM-*an*	*saa*-STEM, *Ca*-STEM
Irrealis	*M*-STEM-*ay*	STEM-*aw*	...	...
Imper./hortative	STEM	STEM-*i*	…	…

[36] I am indebted to Stacy Teng for examples drawn from Tsuchida (2003).
[37] Tsuchida (1976:80) gives only one example each of STEM-*i* and STEM-*ani*, and assumes them both to be UVI.
[38] Rare.
[39] Rare.

	ACTOR VOICE	UNDERGOER VOICE		
		Patient subject	Location subject	Circumstance subject
Mayrinax Atayal				
Realis	*M*-STEM	STEM-*un*	STEM-*an*	*si*-STEM
Irrealis	*pa*-STEM	*Ca*-STEM-*un*	*Ca*-STEM-*an*	*Ca*-STEM
Perfective	*M*‹*in*›-STEM	STEM	‹*in*›STEM-*an*	*si*-STEM
Nominal	*M*-STEM	‹*in*›-STEM[-*an*], [*Ca*-]STEM-*an*, STEM	‹*in*›-STEM-*an*, *Ca*-STEM-*an*	*Ca*-STEM
Optative/hortative	*M*-STEM-*ay*	STEM-*aw*	STEM-*ay*	STEM-*anay*
Imperative	STEM	STEM	STEM-*i*	STEM-*ani*
Negative	BASE	STEM-*i*	STEM-*i*	STEM-*ani*
Seediq				
Realis	*M*-STEM	STEM-*un*[40]	STEM-*an*	*se*-STEM
Irrealis	*mpe*-STEM	–	–	[*Ce*-]STEM
Perfect	*M*-‹*in*›STEM	‹*en*›STEM-*an*	–	‹*en*›STEM
Nominal	*M*-[‹*en*›]STEM	‹*in*›STEM, STEM-*un*	[‹*in*›]STEM-*an*	*se*-STEM, *Ce*-STEM
Imper./hortative	STEM-*a*	STEM-*aw*[41]	STEM-*ay*	STEM-*anay*
Imper./dependent	STEM	STEM-*i*	–	STEM-*ani*
Thao				
Realis	*M*-STEM	STEM-*in*	STEM-*an*	–
Imperfective	*Ca*-*M*-STEM	–	–	–
Perfective	*M*-‹*in*›-STEM	‹*in*›-STEM[-*in*]	‹*in*›-STEM-*an*	–
Irrealis	*a*-*M*-STEM	*a*-STEM-*in*	–	–
Nominal	*M*-STEM	[‹*in*›]STEM-*an*, STEM-*in*, *Ca*-STEM-*an*	[‹*in*›]STEM-*an*	*Ca*-STEM[-*an*]
Irrealis	STEM	STEM-*a*	STEM-*a*	–
Imperative	STEM	STEM-*i*	STEM-*an*	–
Ishbukun Bunun				
Realis	*M*-STEM	STEM-*un*	STEM-*an*	*is*-STEM
Imperfective	*M*-CVCV-STEM	CVCV-STEM-*un*	CVCV-STEM-*an*	–
Perfective	‹*in*›*M*-STEM	‹*in*›STEM-*un*	‹*in*›STEM-*an*	*sin*-STEM
Nominal	–	–	*Ca*-STEM-*an*	–
Imperative	STEM-*a*	STEM-*av*	–	–
Paiwan				
Realis	*M*-STEM	STEM-*en*	STEM-*an*	*si*-STEM
Imperfective	*M*-CV-STEM	CV-STEM-*en*	CV-STEM-*an*	*si*-CV-STEM
Perfective	*na M*-STEM	[‹*in*›]STEM,	‹*in*›STEM-*an*,	‹*in*›*si*-STEM
Nominal	*M*-STEM	[‹*in*›]STEM, [*Ca*-]STEM-*en*	‹*in*›STEM-*an*, [*Ca*-]STEM-*an*	*s*‹*in*›*i*-STEM

[40] Used after certain preverbs. STEM-*an* is the unmarked predicate form.

[41] Pecoraro (1979:106). Tsukida (2005) does not record *-aw*. Instead she records *-ay* as UV.PAT and no UV.LOC form.

	ACTOR VOICE	UNDERGOER VOICE		
		Patient subject	Location subject	Circumstance subject
Optative	–	STEM-*aw*	STEM-*ay*	–
Imperative	STEM-*u*	STEM-*u*	–	STEM-*an*
Hortative	STEM-*i*	STEM-*i*	–	–
Narr./Dependent	STEM	STEM-*i*	STEM-*an*	STEM-*an*
Dependent imperf.	CV-STEM	…	…	…
Kavalan				
Realis	*M*-STEM		STEM-*an*	*ti*-STEM[42]
Realis perfective	–		‹*in*›STEM	–
Nominal	*pa*-STEM-*an*	[‹en›]STEM-*an*, STEM-*an*	*sa*-STEM[-*an*][43]	
Irrealis	STEM		STEM-*a*	STEM-*a*
Imperative	STEM		STEM-*i*	–
Haian Amis[44]				
Non-past	*M*-STEM	–	*M*-STEM-*an*	*ma*-s*a*-STEM
Future	–	ROOT-*en*		s*a*-STEM-*en*
Perfective	–	*ma*-STEM, [*ka*-]ROOT-*en*[45]	–	–
Imperative	STEM	ROOT-*en*	–	s*a*-STEM-*en*
Habitual	–	–	STEM-*an*	s*a*-STEM
Irrealis	*Ca*-*M*-STEM	*Ca*-ROOT-*en*	–	–
Optative 2	*sa*-STEM-*an*	–	–	s*a*-STEM-*aw*
Nominal	STEM	STEM-*an*, [*Ca*-]STEM-*en*	STEM-*an*	s*a*-STEM
Optative 1	*M*-STEM-*aw*	ROOT-*aw*	–	–
Timerative	*ma*-STEM-*aw*	*ma*-[*M*-]STEM-*aw*	–	–
Non-past negative	STEM	k*a*-STEM		
Siraya				
Realis	*M*-STEM	STEM-*en*	STEM-*an*	–
Imperfective	*M*-*Ca*-STEM	*Ca*-STEM-*en*	–	–
Realis past	‹*in*›*M*-STEM	‹*in*›STEM-*en*	…	–
Imperfective past	‹*in*›*M*-*Ca*-STEM	–	–	–
Durative	Apparently CVCV- applied to any realis form.			

[42] The expected reflex of **Si*- is *si*-.

[43] Elizabeth Zeitoun observes that *sa*-STEM occurs in Kavalan alongside the form *sa*-STEM-*an* reported by other sources.

[44] Amis displays a striking reorganisation of affixes. Discussion lies outside the subject matter of this paper. There are three stem categories, ‹*um*›STEM, *pi*-STEM and *ka*-STEM. The following rules provide for expansions of forms summarised in the table: *M*-+‹*um*›STEM > ‹*um*›ROOT, Ø+‹*um*›STEM > *ka*-‹*um*›ROOT, *ka*-+‹*um*›STEM > *ka*-ROOT, *sa*-+‹*um*›STEM > *sa*-*ka*-‹*um*›ROOT, *ma*-+*pi*-STEM > *ma*-ROOT, *ka*+*pi*-STEM > *ka*-ROOT.

[45] Agentive.

Appendix C: Data sources

Sources of data used in this work are listed below. Works in parentheses were mostly consulted only for minor points.

Formosan: Puyuma: Teng 2008a (Stacy Teng pers. comm.); Rukai: Zeitoun 2003, 2007; Tsou: Zeitoun (2005); Kanakanavu: Tsuchida 1976, 2003 (Mei 1982); Saaroa: Tsuchida 1976 (Radetzky 2003, 2006); Saisiyat: Yeh 1991, 2003; Pazih: Li and Tsuchida 2001, 2002 (Li 2002); Mayrinax Atayal: Huang 1995, 2000; Seediq: Tsukida 2005 (Pecoraro 1979); Thao: Blust 2003 (Wang 2004); Ishbukun Bunun: Yeh n.d.a,b (Qi 2000); Paiwan: A.H. Chang 2006 (Egli 1990); Kavalan: Tsuchida 1993 (H.Y. Chang 1997, pers. comm., Chang and Lee 2002, Lee 1997, Li and Tsuchida 2006); Amis: Wu 2006 (Y. Huang 1988, Liu 1999, Wu 2003); Siraya: Adelaar (1999).

Malayo-Polynesian: Yami: Rau 2004, 2006 (Shih 1997); Bisayan: Zorc 1977; Western Bukidnon Manobo: Elkins 1970; Bonggi: Boutin 2002; Lun Dayeh (Sarawak Murut): Clayre 2005; Kimaragang: Kroeger 2004; Eastern Kadazan: Hurlbut 1988; Timugon Murut: Prentice 1971; Brewis and Levinsohn 1991.

References

Adelaar, K. Alexander. 1992. *Proto Malayic: the reconstruction of its phonology and parts of its lexicon and morphology*. Canberra: Pacific Linguistics.

—— 1999. Siraya reduplication. *Oceanic Linguistics* 38:33–52.

Adelaar, K. Alexander and Nikolaus P. Himmelmann, eds. 2005. *The Austronesian languages of Asia and Madagascar*. London: Routledge.

Blust, Robert A. 1977. The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report. *University of Hawaii Working Papers in Linguistics* 9/2:1–15.

—— 1995. Austronesian comparative dictionary. Computer files.

—— 1997. Rukai stress revisited. *Oceanic Linguistics* 36:398–403.

—— 1998. *Ca*-reduplication and Proto-Austronesian grammar. *Oceanic Linguistics* 37:29–64.

—— 1999. Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. In Zeitoun and Li, 1999:31–94.

—— 2003. *Thao dictionary*. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.

Boutin, Michael E. 2002. Nominative and genitive case alternations in Bonggi. In Wouk and Ross, 2002:209–239.

Brewis, Richard and Stephen H. Levinsohn. 1991. Topic and emphasis in Timugon Murut. In Stephen H. Levinsohn, ed. *Thematic continuity and development in languages of Sabah*, 29–43. Canberra: Pacific Linguistics.

Capell, Arthur. 1964. Verbal systems in Philippine languages. *Philippine Journal of Sciences* 93:231–249.

Cauquelin, Josiane. 1991. The Puyuma language. *Bijdragen tot de Taal-, Land- en Volkenkunde* 147:17–60.

Chang, Anna Hsiouchuan. 2006. A reference grammar of Paiwan. PhD thesis. The Australian National University.

Chang, Henry Yungli. 1997. Voice, case and agreement in Seediq and Kavalan. PhD dissertation. Hsinchu, Taiwan: National Tsing Hua University.

—— 2006. Rethinking the Tsouic subgroup hypothesis: a morphosyntactic perspective. In Chang, Huang and Ho, 2006:565–583.

Chang, Henry Yungli, Lillian M. Huang and Dah-an Ho, eds. 2006. *Streams converging into an ocean: Festschrift in honor of Professor Paul Jen-kuei Li on his 70th birthday*. Taipei: Institute of Linguistics, Academia Sinica.

Chang, Henry Yungli and Amy Peijung Lee. 2002. Nominalization in Kavalan. *Language and Linguistics* 3:349–368.

Clayre, Beatrice M. 2005. Kelabitic languages and the fate of 'focus': evidence from the Kerayan. In I Wayan Arka and Malcolm Ross, eds *The many faces of Austronesian voice systems: some new empirical studies*, 17–57. Canberra: Pacific Linguistics.

Comrie, Bernard and Sandra A. Thompson. 1985. Lexical nominalization. In Timothy Shopen, ed. *Language typology and syntactic description*. vol. 3, 349–398. Cambridge: Cambridge University Press.

DeLancey, Scott. 1986. Relativization as nominalization in Tibetan and Newari. Paper presented at the 19th International Conference on Sino-Tibetan Languages and Linguistics.

Egerod, Sören. 1966. Word order and word classes in Atayal. *Language* 42:346–369.

Egli, Hans, 1990. *Paiwangrammatik*. Wiesbaden: Otto Harrassowitz.

Elkins, Richard E. 1970. *Major grammatical patterns of Western Bukidnon Manobo*. Norman: Summer Institute of Linguistics of the University of Oklahoma.

Foley, William A. 1986. *The Papuan languages of New Guinea*. Cambridge: Cambridge University Press.

Foley, William A. and Robert D. Van Valin. 1984. *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.

Himmelmann, Nikolaus P. 1999. Lexical categories and voice in Tagalog. Unpublished MS.

—— 2005. The Austronesian languages of Asia and Madagascar: Typological characteristics. In Adelaar and Himmelmann, 110–181.

Huang, Lillian M. 1994. *A study of Atayal syntax*. Taipei: The Crane.

—— 1995. *A study of Mayrinax Atayal*. Taipei: Crane Publishing Co.

—— 2000. Verb classification in Mayrinax Atayal. *Oceanic Linguistics* 39:364–390.

—— 2002. Nominalization in Mayrinax Atayal. *Language and Linguistics* 3:197–225.

Huang, Yajiun. 1988. Amis verb classification. MA thesis. Taipei: Fu-jen Catholic University.

Hurlbut, Hope M. 1988. *Verb morphology in Eastern Kadazan*. Canberra: Pacific Linguistics.

Kaufman, Daniel. 2007. The nominalist hypothesis in Austronesian. Paper presented at the Workshop 'Beyond focus and ergativity: towards a more comprehensive view of Austronesian morphosyntax', Zentrum für allgemeine Sprachwissenschaft, Humboldt-Universität Berlin.

Kroeger, Paul. 2004. Morphosyntactic vs. morphosemantic functions of Indonesian *-kan*. Unpublished MS.

Lee, Amy Peijung. 1997. The case-marking and focus systems in Kavalan. MA thesis. National Tsing Hua University, Hsinchu, Taiwan.

Li, Paul Jen-kuei. 2002. Nominalization in Pazih. *Language and Linguistics* 3:227–239.

Li, Paul Jen-kuei and Shigeru Tsuchida. 2001. *Pazih dictionary*. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.

—— 2002. *Pazih texts and songs*. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.

—— 2006. *Kavalan dictionary*. Taipei: Institute of Linguistics, Academia Sinica.

Liu, Dorinda Tsai-hsiu. 1999. Cleft constructions in Amis. MA dissertation. National Taiwan University.

Lopez, Cecilio. 1941. *A manual of the Philippine national language*. 3rd ed. Manila: Bureau of Printing.

Lynch, John, Malcolm Ross and Terry Crowley. 2002. *The Oceanic languages*. Richmond: Curzon Press.

Mei, Kuang. 1982. Pronouns and verb inflection in Kanakanavu. *Tsing Hua Journal of Chinese Studies* New Series 14:207–231.

Pecoraro, Ferdinando. 1979. *Eléments de grammaire taroko, précédés de la présentation dela vie et de la culture des taroko*. Paris: Association Archipel.

Peterson, David A. 2007. *Applicative constructions*. Oxford: Oxford University Press.

Prentice, D.J. 1971. *The Murut languages of Sabah*. Canberra: Pacific Linguistics.

Qi, Lisha [Elizabeth Zeitoun]. 2000. *Bùnóngyǔ cānkǎo yǔfǎ [Bunun reference grammar]*. Taipei: Yuanliu.

Radetzky, Paula. 2003. Grammaticalization of a definiteness marker in Saaroa. Unpublished MS, University of Oregon.

—— 2006. The semantics of the verbal complex with particular reference to Saaroa (Taiwan). Paper presented to the Tenth International Conference on Austronesian Linguistics, Puerto Princesa.

Rau, D. Victoria and Maa-neu Dong. 2006. *Yami texts with reference grammar and dictionary*. Taipei: Institute of Linguistics, Academia Sinica.

Rau, Der-hwa Victoria. 2004. Iconicity and temporal-aspectual-modality in Yami. Paper presented to the Ninth International Symposium on Chinese Languages and Linguistics, Graduate Institute of Linguistics, National Taiwan University, November 2004.

Reid, Lawrence A. 2007. Abductive and deductive change in reduplicative plurals of human nouns in Philippine languages. Paper presented to the Second Conference on Reduplication, University of Graz.

Reid, Lawrence A. and Hsiu-chuan Liao. 2004. A brief syntactic typology of Philippine languages. *Language and Linguistics* 5:433–490.

Ross, Malcolm, 1995. Reconstructing Proto Austronesian verbal morphology: evidence from Taiwan. In Paul Jen-kuei Li, Dah-an Ho, Ying-kuei Huang, Cheng-hwa Tsang and Chiu-yu Tseng, eds *Austronesian studies relating to Taiwan*, 727–791. Taipei: Institute of History and Philology, Academia Sinica.

—— 2002a. Takia. In John Lynch, Malcolm Ross and Terry Crowley, eds *The Oceanic languages*, 216–248. Richmond: Curzon Press.

—— 2002b. The history and transitivity of western Austronesian voice and voice-marking. In Wouk and Ross, 2002:17–62.

—— 2003. Notes on reconstructing Proto Rukai case-marking. Unpublished MS.

—— 2006. Reconstructing the case-marking and personal pronoun systems of Proto Austronesian. In Chang, Huang and Ho, 2006:521–564.

Ross, Malcolm and Stacy Fang-ching Teng. 2005. Formosan languages and linguistic typology. *Language and Linguistics* 6:739–781.

Sagart, Laurent. 2004. The higher phylogeny of Austronesian and the position of Tai-Kadai. *Oceanic Linguistics* 43:411–444.

Shih, Louise Yu-mien, 1997. Yami morphology. MA thesis. Providence University, Taijung, Taiwan.

Starosta, Stanley. 1985. Verbal inflection versus deverbal nominalisation in PAN: the evidence from Tsou. In Andrew Pawley and Lois Carrington, eds *Austronesian linguistics at the 15th Pacific Science Congress*, 281–312. Canberra: Pacific Linguistics.

—— 1986. Focus as recentralisation. In Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL I: Papers from the Fourth International Conference on Austronesian Linguistics*, 73–95. Canberra: The Australian National University.

—— 1995. A grammatical subgrouping of Formosan languages. In Paul Jen-kuei Li, Dah-an Ho, Ying-kuei Huang, Cheng-hwa Tsang and Chiu-yu Tseng, eds *Austronesian studies relating to Taiwan*, 683–726. Taipei: Institute of History and Philology, Academia Sinica.

—— 1999. Transitivity, ergativity, and the best analysis of Atayal case marking. In Zeitoun and Li, 1999:371–392.

—— 2001. Reduplication and the subgrouping of Formosan languages. Paper presented at the International Symposium on Austronesian Cultures: issues relating to Taiwan, Academia Sinica, Taipei.

Starosta, Stanley, Andrew Pawley and Lawrence A. Reid. 1981. The evolution of focus in Austronesian. Paper presented at the Third International Conference on Austronesian Linguistics, Bali.

—— 1982. The evolution of focus in Austronesian. In Amran Halim, Lois Carrington and S.A.Wurm, eds *Papers from the Third International Conference on Austronesian Linguistics, 2: Tracking the Travellers*. vol. 2, 145–170. Canberra: Pacific Linguistics.

Szakos, József. 1994. Die Sprache der Cou: Untersuchungen zur Synchronie einer austronesischen Sprache auf Taiwan. PhD thesis. Rheinische Friedrich-Wilhelms-Universität zu Bonn.

—— n.d. Cou-Wörterbuch. Unpublished computer file, Providence University, Taichung.

Teng, Stacy Fang-ching. 2008a. *A reference grammar of Puyuma, an Austronesian language of Taiwan*. Canberra: Pacific Linguistics.

—— 2008b. Case syncretism in Puyuma. Unpublished MS, Academia Sinica, Taipei.

Tsuchida, Shigeru. 1976. *Reconstruction of Proto-Tsouic phonology*. Tokyo: Institute for the Study of Languages, Cultures of Asia, and Africa.

—— 1993. Short sketch of Kavalan grammar. Unpublished MS.

—— 2003. *Kanakanavu texts (Austronesian Formosan)*. Osaka: Endangered Languages Pacific Rim, Osaka Gakuin University.

Tsukida, Naomi. 2005. Seediq. In Adelaar and Himmelmann, 2005:291–325.

Wang, Shan-shan. 2004. An ergative view of Thao syntax. PhD dissertation. University of Hawai‘i.

Wolff, John U. 1973. Verbal inflection in Proto-Austronesian. In Andrew Gonzalez, ed. *Parangal kay Cecilio Lopez*, 71–91. Quezon City: Linguistic Society of the Philippines.

—— 1996. The development of the passive verb with pronominal prefix in western Austronesian languages. In Bernd Nothofer, ed. *Reconstruction, classification, description: festschrift in honor of Isidore Dyen*, 15–40. Hamburg: Abera.

Wouk, Fay and Malcolm Ross, eds. 2002. *The history and typology of western Austronesian voice systems*. Canberra: Pacific Linguistics.

Wu, Joy J. 2003. Voice affixes and verb classification in Amis. Paper presented at the Second Workshop on Formosan Languages, Academia Sinica, Taipei, November 2003.

—— 2006. Verb classification, case marking and grammatical relations in Amis. PhD dissertation. State University of New York at Buffalo.

Yeh, Marie Meili. 1991. Saisiyat structure. MA thesis. National Tsing Hua University.

—— 2003. A syntactic and semantic study of Saisiyat verbs. PhD dissertation. Taipei: National Taiwan Normal University.

—— n.d.a. Bunun reference grammar. Unpublished MS.

—— n.d.b. Morphological causativization in Bunun. Unpublished MS, Lien Ho Institute of Technology.

Zeitoun, Elizabeth. 2003. Toward a reconstruction of Proto-Rukai morpho-syntax. Paper presented at AFLA X, Honolulu.

—— 2005. Tsou. In Adelaar and Himmelmann, 2005:259–290.

—— 2007. *A grammar of Mantauran (Rukai)*. 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* 39:391–414.

Zeitoun, Elizabeth, Lillian M. Huang, Marie M. Yeh, Anna H. Chang and Joy J. Wu. 1996. The temporal, aspectual and modal systems of some Formosan languages: a typological perspective. *Oceanic Linguistics* 35:21–56.

Zeitoun, Elizabeth and Paul Jen-kuei Li, eds. 1999. *Selected papers from the Eighth International Conference on Austronesian Linguistics*. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.

Zeitoun, Elizabeth and Stacy Fang-ching Teng. 2006. The passive *ki-*in Rukai, Paiwan and Puyuma: borrowing, shared in innovation parallel development. Unpublished MS, Institute of Linguistics, Academia Sinica.

Zeitoun, Elizabeth and Chen-huei Wu. 2006. An overview of reduplication in Formosan languages. In Chang, Huang and Ho, 2006.

Zorc, R. David. 1977. *The Bisayan dialects of the Philippines: subgrouping and reconstruction*. Canberra: Pacific Linguistics.