CHAPTER 32

Formosan Languages and Proto-Austronesian Morphology

Malcolm Ross and Elizabeth Zeitoun

32.1 Introduction*

The present chapter shows how the Formosan languages support the reconstruction of Proto-Austronesian (PAN) morphology.

* Each author read and edited the other's work. The introduction was written jointly; Zeitoun wrote §§ 32.3.2–32.3.4 and § 32.6 and Ross the remainder, to which Zeitoun contributed many of the tables.

 | DOI:10.1163/2772_5766_COM-103030

Research into PAN morphology is challenging for several reasons. With the exceptions of Siraya and Favorlang, the earliest written records of Formosan languages date from the Japanese occupation (1895–1945), and we rely solely on descriptions of the modern languages. Inevitably there is much that will escape us. Reconstruction is all the more complicated because many affixes in Formosan languages are homophonous and others exhibit dual or portmanteau functions. Luckily, the past 20 years or so have witnessed tremendous progress in our knowledge of the languages, and views of PAN morphology have evolved accordingly.

We are left, then, with the task of reconstructing PAN mainly from the data of present-day languages and from manuscript sources that are at the most 400 years old—still modern in relation to the age of the proto-language. PAN is at the top of the Austronesian tree. Dahl (1976) and Blust (1977) establish that all Austronesian languages other than the Formosan languages belong to a single subgroup, Malayo-Polynesian (MP). When we reconstruct Proto-Malayo-Polynesian (PMP), we draw not only on "internal" evidence from MP languages but also on "external" evidence from Formosan languages, as the latter can tell us what PMP inherited from PAN. The same principle applies to lower-order interstage languages like Proto-Oceanic (Dempwolff 1937). But this is obviously not true of PAN. Despite suggestions of external relationships (Sagart 2005 is the most plausible),[1] these are neither close enough nor well enough established to provide external evidence for the reconstruction of PAN morphosyntax. We are faced instead with a methodological chicken-and-egg situation. Reconstruction on the basis of internal evidence entails knowing the subgrouping of daughter languages. If, say, family X has three primary subgroups ABC, DE, and FGHJ, then a feature present in just one language from each of two subgroups can be reconstructed to Proto-X. But under the classical comparative method of historical linguistics, subgroup ABC is defined by the innovations that Proto-ABC has undergone relative to Proto-X, and similarly for DE and FGHJ. In other words, we need to know the family's subgroups in order to reconstruct the proto-language, but we need to have the reconstruction in order to define the subgroups.

The way that historical linguists deal with this situation is by abductive reasoning. From what we know of language change in general, which of the available hypotheses about the shape of the proto-language and the consequent innovations thereby attributed to subgroups is the most plausible? Unfortu-

1 Blust (2013, pp. 702–721) provides a useful overview of proposals, and Blust (2014) provides further critical analysis.

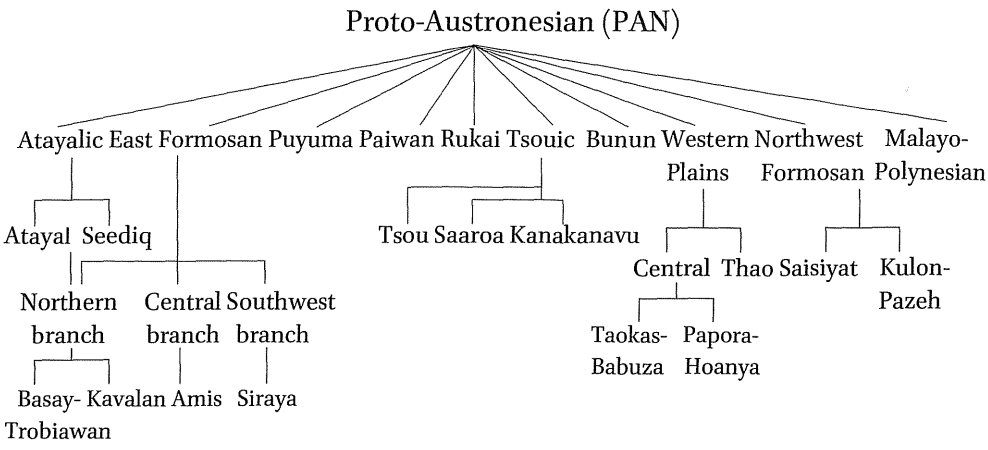

FIGURE 32.1 Blust's (1999b) phonological subgrouping

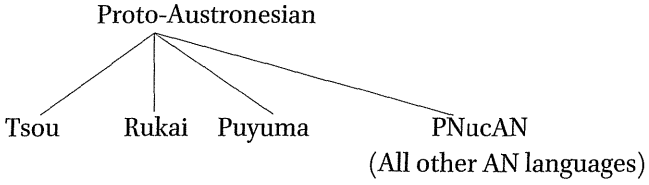

FIGURE 32.2 Ross's (2009) subgrouping hypothesis

nately, linguists do not always agree about plausibility, and this is manifest in two competing hypotheses about the subgrouping of the Formosan languages. The first of these is Blust's (1999a), which divides Formosan languages into nine primary subgroups, coordinate with MP, on the basis of inferred innovations from the "majority view" reconstruction of PAN phonology. The other is Ross's (2009, 2012) proposal that Austronesian has four primary branches: three are the single languages Tsou, Rukai, and Puyuma, and the fourth is Nuclear Austronesian (NucAN), which contains all other Austronesian languages, i.e., the remaining Formosan languages and MP. Thus Blust hypothesizes ten primary subgroups, as shown in Figure 32.1, and Ross four, as in Figure 32.2.[2]

2 As an aside, Ross's grouping dismantles only one of Blust's (1999) ten groups, namely Tsouic (Ross 2012), a grouping called into question early on by Ferrell (1969, pp. 68, 70). The others remain intact. There are two other extant subgroupings. Recent work by Aldridge (2015, 2016, 2021) suggests a revision of the NucAN hypothesis (§ 32.4.3). Sagart's (2004) numeral-based hypothesis is treated briefly in § 32.6.

Ross's (2009) NucAN hypothesis rests on the facts that the verbal morphologies of Tsou, Rukai, and Puyuma are quite different from one another and from the morphological framework shared by the remaining Formosan languages and, with modifications, by Philippine languages, the latter being reckoned to be the most conservative MP languages. This means that the conventional reconstruction of PAN verbal morphology, here labeled PAN1, and based on the work of Wolff (1973), is effectively reassigned to PNucAN. It also means reconstructing a somewhat different PAN morphology, here labeled PAN2, ancestral to Tsou, Rukai, and Puyuma as well as NucAN. There are just three sections of this chapter in which PAN1 and PAN2 are not distinguished and reconstructions are simply attributed to PAN, because they are reconstructed to the proto-language under both hypotheses: these sections are § 32.3, § 32.5.1, and § 32.6.

We note that the NucAN hypothesis has been questioned by Sagart (2010, 2014), by Jiang (2016), and by Blust & Chen (2017), who ask, "How can we argue for the past existence of some feature if no trace of it remains? The simple answer is: we cannot, but neither can we argue the contrary case, since we cannot logically exclude the possibility that the feature was once present but disappeared before it was observed" (p. 578). We beg to differ. The genesis or the disappearance of a feature can be assigned a probability based on abduction, and it is the probabilities that proponents of alternative hypotheses disagree on. It seems less than scholarly to assume that every potential PAN1 form that is missing from Tsou, Rukai, and Puyuma is just a loss. That said, we try here to give a dispassionate account of the reconstruction of both PAN1 and PAN2.

Indeed, examining the historical morphology of Formosan languages through the lens of the NucAN hypothesis confronts us with yet more questions about subgrouping at the top of the Austronesian tree. It is not our intent to tackle these here, but the reader may notice that data from Tsou in particular, and sometimes from Rukai and/or Puyuma, are missing from a particular reconstruction. Tsou is structurally and morphologically more different from other Formosan languages than the latter are from one another, so that its inclusion in a morphological comparison is sometimes not possible, and the same is in some respects true of Rukai and Puyuma. The only recent attempt to make sense of this is Aldridge (2021) (see § 32.4.3). Where one or more of these languages is missing from a dataset, this may mean that what we present as a PAN2 reconstruction is more strictly attributed to a node between PAN2 and PNucAN, or, in PAN1 terms, to a node just below the top of the tree.

There is a good deal of self-reference in this chapter, as, along with Robert Blust, the authors have been responsible for much of the recent work on

PAN morphology. Below we discuss verbal (§32.3 and §32.4) and nominal (§32.5) morphology. These entail verbal alignment and case-marking, and each depends on the other, so in §32.2, we provide an outline of their interrelationships, i.e., their syntactic context.

32.2 PAN1 Morphology in Context

Under the NucAN hypothesis, the reconstruction of PNucAN verbal morphology is assumed to be identical with PAN1, i.e., the reconstruction of PAN based on Wolff (1973). The difference between the two lies solely in the protolanguage to which reconstructions are assigned under each theory. Rather than repeatedly writing PAN1/PNucAN, we use "PAN1" to refer to both.

This section provides examples from the NucAN Formosan languages Paiwan and Seediq in order to provide a context for the reconstructions in the remainder of the chapter. These rest on the assumption that, like NucAN Formosan languages, PAN1 had an ergatively aligned clause structure, with two voices, actor (intransitive) voice (AV) and undergoer (transitive) voice (UV) (Himmelmann 2005). With some variations, this structure is maintained in NucAN Formosan languages.

A PAN1 UV clause was transitive with two core arguments: a nominative-marked undergoer (subject) and a genitive-marked actor, as in the Paiwan sentences in (1). The sentences begin with the verb, as they probably did in PAN1. In (1a) the case markers *ni* and *a* encode the case of the following noun and whether that noun is common or personal. In (1b), the core arguments are in the form of pronouns that are cliticized to the verb. In Paiwan, the genitive (actor) pronoun *ku=* precedes the verb, and the nominative (undergoer) ***=(e)sun*** follows it.

(1) Saichia Paiwan (Chang 2006)

a. *k⟨in⟩an ni zepulj a kinsa.*
⟨PFV.UVP⟩eat GEN.PN Zepulj NOM food
'Zepulj ate the food.' (p. 64)

b. ***ku***=*k⟨in⟩elem*=***esun*** *katiaw.*
1SG.GEN=⟨PFV.UVP⟩beat=2SG.NOM yesterday
'I beat you yesterday.' (p. 65)

In other NucAN Formosan languages, both clitics follow the verb, as in (2).

(2) Tgdayan Seediq (Holmer 1996)
*qta-un=**su**=**mu**.*
see-UVP=2SG.NOM=1SG.GEN
'I'll see you.' (p. 69)

The verb phrase may also include an auxiliary before the verb. In the clause below, the past auxiliary *wada* precedes the transitive verb *ata-un* and attracts both pronouns. That is, the clitics "climb" from the verb to the auxiliary.

(3) Tgdayan Seediq (Holmer 1996)
*wada=**ku**=**na** qta-un.*
PST=1SG.NOM=3SG.GEN see-UVP
'He saw me.' (p. 69)

An AV clause is intransitive with an actor subject in nominative case, as in (4). Actor voice is often signaled by the infix ⟨*em*⟩, as in (4a) and (5), and is also signaled in other ways, like the initial *m-* of *mangetjez* in (4b) (the Paiwan root is *pangetjez*). These are allomorphs of a morpheme that for convenience we call *M-* (§ 32.3.1). In (4b), the clitic climbs to the auxiliary.

(4) Paiwan (Chang 2006)
a. *dj⟨em⟩ava~davac=**aken**.*
⟨AV⟩RED~walk=1SG.NOM
'I am walking.' (Sandimen variety, p. 67)

b. *ki=**ken** a mangetjez.*
VOL=1SG.NOM LNK AV.come
'I am going to come.' (Saichia variety, p. 68)

There is also evidence that a PAN1 AV clause could have an indefinite undergoer in the oblique case, as in (5).

(5) Puljetji Paiwan (Huang 2012)
*na=k⟨em⟩an=aken **ta** demangasan.*
PFV=⟨AV⟩eat=1SG.NOM OBL goat
'I ate goat (meat).' (p. 6)

The undergoer voice subsumes three voices, according to the semantic role of the argument that occurred in the nominative: patient UV (UVP), location UV (UVL), and circumstance UV (UVC). These are again encoded in verbal morphology. "Circumstance" includes various roles but perhaps most often instrument.

The sentences in (1), (2), and (3) are all in UVP. The sentence in (6a) is in UVL, marked by the verbal suffix *-an* and making the subject the person spat upon. The sentence in (6b) is in UVC, marked by the verbal prefix *si-*, and the instrument *azua kisi* is the subject.

(6) Paiwan (based on Chang 2006)

a. *s⟨in⟩udilay-an ni zepulj ti kaljalju.*
⟨PFV⟩-spit-UVL GEN.PN Zepul NOM.PN Kalalu
'Zepul spat on Kalalu.' (Sandimen variety, p. 143)

b. *uri=ku=**si**-kan a-zua kisi.*
INTENT=1SG.GEN=UVC-eat NOM-that bowl
'I will use that bowl to eat with.' (Saichia variety, p. 329)

The terms here follow Formosanist parlance. A syntactician might argue that the system is only aberrantly ergative and prefer the term "Philippine-type". A typologist might prefer "absolutive" and "ergative" to "nominative" and "genitive", but this ignores the fact that the genitive is also the possessor case. Peterson (2007) has argued that in an ergative language, the applicative "promotion" of adjunct arguments is to subject (not to object as in accusative languages), and that UVL and UVC can thus be regarded as applicatives.

Schematically, the Paiwan clause structures with full core NP arguments can be summarized as in (7).

(7) Paiwan clause skeletons

AV	*M*-verb [NOM NP-ACTOR]
UVP	verb [GEN NP-ACTOR] [NOM NP-PATIENT]
UVL	verb-*an* [GEN NP-ACTOR] [NOM NP-LOCATION]
UVC	*si*-verb [GEN NP-ACTOR] [NOM NP-INSTRUMENT etc.]

If the arguments are pronominal clitics, then the structures are as follows:

(8) Paiwan clause skeletons

AV	[*M*-verb=PRON.NOM]
	[AUX=PRON.NOM] *M*-verb
UV	[PRON.GEN=verb=PRON.NOM]
	AUX [PRON.GEN=verb=PRON.NOM]

The term ACTOR in (7 AV) is not precise enough, as actors typically occur with dynamic verbs, but not all verbs are dynamic. The AV verbal morpheme is

shown as *M-* here and frequently in subsequent examples. These matters are handled in § 32.3.1 in the context of PAN1 verb classes.

One other morpheme that crops up in the examples is ⟨*in*⟩ PERFECTIVE (cf. Table 32.1). The morphemes ancestral to Paiwan ⟨*in*⟩, *M-*, *-an*, and *si-* are reconstructed in the context of verbal morphology in § 32.4.

The clause skeletons in (7) and (8) refer to Paiwan. As we noted in connection with (2) and (3), the position of Paiwan genitive clitic pronominals is not typical of NucAN Formosan languages. We return to this when we reconstruct PAN pronouns in § 32.5.2.

32.3 Stem-Forming Verbal Morphology

This section deals with how PAN formed verb stems from verb roots. This entails positing formal (morphological) verb classes, each with two stems (§ 32.3.1), and forming complex stems via valency-changing morphology (§ 32.3.2 and § 32.3.3) and verbalizing prefixes (§ 32.3.4).

32.3.1 *Verb Classes*

Zeitoun & Huang (2000) show that in Formosan languages, each verb has two stems, and that this must also have been true of PAN. Verbs fall into morphological classes, defined by the forms of two stems, which Ross (2015c) labels the Mstem and the Kstem.

Table 32.1 shows a partial paradigm of the Paiwan verb *cakav* 'steal'. The AV indicative cells, shown in darker gray, contain the infix ⟨*em*⟩ (§ 32.2), inserted into the first syllable of the simple or reduplicated root. This infix realizes the morpheme here labeled *M-*. It forms what Ross (2015) calls the "Mstem" of the verb. The other cells of the table are based on the "Kstem", consisting in this instance of the root *cakay* alone. A distribution of this kind occurs across all Formosan languages, but it is not always quite as neat as this. Pazeh, for example, forms the AV imperfective from the Kstem.

TABLE 32.1 Partial paradigm of Paiwan voice, mood, and aspect morphemes

	AV	UVP	UVL	UVC
INDICATIVE				
Neutral	*c⟨em⟩akav*	*cakav-en*	*cakav-an*	*si-cakav*
Imperfective	*c⟨em⟩aka-cakav*	*caka-cakav-en*	*caka-cakav-an*	*si-caka-cakav*
Perfective	*na=c⟨em⟩akav*	*c⟨in⟩akav*	*c⟨in⟩akav-an*	*s⟨in⟩i-cakav*

TABLE 32.1 Partial paradigm of Paiwan voice, mood, and aspect morphemes (*cont.*)

	AV	UVP	UVL	UVC
NON-INDICATIVE				
Subjunctive	*cakav*	*cakav-i*	*cakav-i*	*cakav-an*

AFTER HUANG 2012, P. 92

Such a distribution of two stems was evidently present in PAN1 and PAN2. The only Formosan language that does not fully reflect it is Tsou. Paiwan ⟨*em*⟩ reflects PAN *⟨um⟩. Other realizations of *M-, forming an Mstem, were *ma- and zero. The Kstem was either *ka- ROOT or the unaffixed root, as in Table 32.1. The morphological class of a verb is defined by its Mstem and its Kstem. For example, the Paiwan verb *cakav* in Table 32.1 reflects a verb of the class *⟨um⟩/zero.

Table 32.2 shows the five PAN verb classes with an example of each. There was no class *⟨um⟩/*ka-.

TABLE 32.2 Reconstructions of PAN formal verb classes

Kstem → / **Mstem ↓**	***ka-**	***zero**
*⟨um⟩		*C⟨um⟩aŋis/*Caŋis 'weep'
*ma-	*ma-/*ka-buSuk 'be drunk'	*ma-tawa/*tawa 'laugh'
*zero	*Sadu/*ka-Sadu 'be many'	*beRay 'give'; *zazaN 'be old'

AFTER ROSS 2015C

A minor variation on class *⟨um⟩/zero is a small group of verbs in which root-initial *p- was replaced by *m-, as in *maCay/*paCay 'die', where *maCay seemingly reflects earlier *p⟨um⟩aCay with loss of its first syllable (see also 4b). This pattern also occurs in reciprocals formed with *maR-/*paR- and *ma-Ca-/*pa-Ca- (§32.3.3).

Ross (2015c) finds an approximate correlation between the PAN formal verb classes and a semantic hierarchy based on Foley's (2005, p. 391) proto-

role hierarchy. At one end of this continuum are agentive verbs encoding an action by a volitional performer, at the other end stative verbs in which something or someone is in a certain state or is coming to be in that state. A number of other categories are between these ends. One can divide the continuum roughly into dynamic verbs and resultative or stative verbs. In the dynamic part of the continuum are *⟨um⟩/*zero verbs and a few *ma-/*zero and *zero/*zero verbs. In the resultative/stative part are numerous *ma-/*ka- verbs and a few *zero/*ka- and *zero/*zero verbs. *Zero/*zero verbs occur at both poles of the continuum but hardly at all in the middle, suggesting that these verbs originally belonged to two different classes but have undergone morphophonological changes that have united them into a single formal class. *Zero/*zero verbs at the active pole include causatives formed with *pa- (§ 32.3.2).

32.3.2 *Causatives with PAN *pa-*

Zeitoun & Huang's (2000) recognition that a verb had two stems arose out of work on Formosan and PAN causatives. Blust (1999b) observes that in Pazeh, causative verbs are formed from dynamic and stative verbs with the prefixes *pa-* and *paka-*, respectively. He reconstructs two ancestral PAN prefixes, *pa- and *paka-, with these functions. Zeitoun & Huang (2000) point out that this analysis is incomplete, as the *ka-* of *paka-* also occurs as part of a stative stem in other contexts, e.g., in certain languages (Rukai, Atayal, Paiwan), it follows a conjunction such as '(and) then', where it is a dependent form, one of the paradigmatic slots where a Kstem occurs (§ 32.3.1). Hence PAN *pa-ka- 'CAUS-STAT-' was bimorphemic; PAN made a formal distinction between dynamic and stative verbs, continued in Formosan languages; and PAN verbs had two stem forms, Mstem and Kstem. Thus in (9), the verbs are both dynamic and AV and occur in their Mstem, while in (10), the same verbs have the causative affix *pa-* attached to their (zero) Kstem.

(9) Zeitoun & Huang (2000, p. 399)

a. Pazeh

yaku ***mi-kita*** *isiw.*
1SG.NOM AV-see 2SG.ACC
'I see you.'

b. Mayrinax Atayal

m-aniq *cku' bunga'* *ku'* *'ulaqi'.*
AV-eat ACC sweet.potato NOM.REF child
'The child is eating a sweet potato.'

(10) Zeitoun & Huang (2000, p. 399)

a. Pazeh

kaakuxan ka asikis a punu mausay ***pa-kita***
have.heatstroke TOP painful LNK head AV.IRR:go CAUS-see
takarat.
doctor
'I have (had) a heatstroke, I have a headache (and I) go to the doctor.'

b. Mayrinax Atayal

pa-*qaniq cku' 'ulaqi' 'i' yaya.*
CAUS-eat ACC child NOM mother
'Mother is feeding the child.'

In (11), however, the verbs are stative. Pazeh *baged* is a zero/*ka-* verb and Mantauran Rukai *takolra* a *ma-*/*ka-* verb. Accordingly, *b⟨in⟩aged* (AV.PFV) occurs in (11a) and *ma-takolra* in (11b) (their Mstems), but in (12), their Kstems are attached to causative *pa-*, giving *pa-ka-baged* and *pa-ka-tolra.*

(11) Zeitoun & Huang (2000)

a. Pazeh

b⟨in⟩aged yaku.
AV.STAT-⟨PFV⟩fat 1SG.NOM
'I used to be fat.' (p. 404)

b. Mantauran Rukai

ma-takolra taotao ocao=ni.
STAT.FIN-bad Taotao man=3SG.GEN
'Taotao is (a) bad (person).' (p. 405)

(12) Zeitoun & Huang (2000)

a. Pazeh

ana pa-ka-baged.
NEG CAUS-STAT-fat
'Don't let (him/her become) fat!' (p. 404)

b. Mantauran Rukai

pa-ka-takolr-a!
CAUS-STAT.NFIN-bad-IMP
'Make it bad!' (p. 405)

Reflexes of the stative prefix *ka- exhibit different variants across Formosan languages as a result of sound change. The prefix *ka-* remains in Rukai, Puyuma, Kanakanavu, Paiwan, Bunun, Pazeh-Kaxabu, and Atayal. It is reflected as *k(a)-* in Saisiyat, e.g., *ma-skes* 'cold' ~ *pa-ka-skes* 'make ... cold, cool down', *bain* 'lazy' ~ *pa-k-bain* 'make lazy' (Zeitoun et al. 2015); Truku Seediq *k(u)-*, e.g., *paro* 'big' ~ *p-n-k-paro* 'make big, enrich (CAUS-PFV-KA)' (Pecoraro 1979); and as Kavalan *q(a-)*, e.g., *ma-yseng* 'dry' ~ *pa-qa-yseng* 'cause to dry (CAUS-KA)', *m-ipes* ~ *q-ipes* 'dislike' (Li & Tsuchida 2006).

Alongside the PAN causatives *pa- '(general) causative' and *paka- 'causative (of stative verbs)', Blust (2003a) suggests the reconstruction of *pu- 'causative of motion' and *pi- 'causative of location', associated with the non-causative prefixes *mu- 'motion' (< *m-u-) and *i- 'location' (see L. Li, this handbook, Chapter 19). As with *paka-, the question is whether morphemes, here putative *pu- and *pi-, should be reconstructed based on their reflexes in a majority of daughter languages, or whether they should be analyzed as bimorphemic *p(a)-u -a with later deletion of the parenthesized vowel. For example, Adelaar (2011, p. 118) notes that in Siraya "the prefix *pa-* is often shortened to *p-*, especially before the orientation prefix *u-/*äw-*", e.g., *p-u-xalap* (ix:16, xiv:28)/*pa-u-xalap* 'to cover with, put on to something' (cf. *m-u-xalap* (xxiii:35) 'to come upon (AV)'); *p-u-kua* (vi:30)/*pa-u-kua* (xx:2) 'to send' (cf. *m-u-kua* (viii:9) 'to go to (AV)'). In Tona Rukai, *pa-'i-* is found, cf. *pa-'i-baivi* 'put at home, take care of' (< *'i-a-baivi* 'be at home, rest'). Thus, it seems necessary, as suggested by Adelaar (pers. comm.), to extend the bimorphemic analysis to the reconstruction of *pa-ka- to *pa-u- and *pa-i-. Verbs formed with *pa- were zero-marked (§ 32.3.1).

32.3.3 *Reciprocal Prefixes*

Following Ross's (1995, p. 772) observation that "[i]t would be useful ... to define the extent and functions of PAN *maR-", Zeitoun (2002) reconstructs two pairs of reciprocal prefixes, *ma-Ca-/*pa-Ca- 'reciprocal/collective of dynamic verbs' (where *C- is the copy of the initial consonant of the verb stem + the vowel *-a-) and *maR-/*paR- 'reciprocal/collective of stative verb'. The *m- form is the AV (Mstem) variant, the *p- form the Kstem (§ 32.3.1). These forms have been variably retained in most Formosan languages (Table 32.3; see also L. Li, this handbook, Chapter 19) and in languages outside Taiwan.[3] Following the

3 Pawley (1973) reconstructs Proto-Oceanic *paRi- 'combined or repeated action by a plurality of actors', and a very detailed study by Lichtenberk (2000) analyzes the various functions of its reflexes. Blust & Trussel (2020) extend the reconstruction to Proto Eastern Malayo-Polynesian *paRi-. The idiosyncratic addition of *-i presumably functioned to prevent the disallowed sequence *R + consonant.

same reasoning as above, *ma-Ca-/*pa-Ca- and *maR-/*paR- might be treated as bimorphemic prefixes composed respectively of *ma-/*pa- plus either *-Ca- or *-aR-. Blust (2013, pp. 389–393) reconstructs three infixes, *⟨ar⟩, *⟨al⟩, and *⟨aR⟩. He attributes the gloss 'plural' to *⟨ar⟩, but is uncertain about the functions of *⟨al⟩ and *⟨aR⟩. Li & Tsuchida (2009, p. 345) reconstruct three infixes, *⟨al⟩, *⟨aR⟩, and *⟨aN⟩ based on Formosan languages, and disregard the prefix *⟨ar⟩ because it is found in only one Formosan language, Paiwan, and various Malayo-Polynesian languages. The morpheme *-aR- here is apparently not infixed but prefixed: *paR- < *pa- + *-aR-, with reduction of *a-a to *a. While we follow Li & Tsuchida's (2009) reconstruction, the function of *-aR-/*⟨aR⟩ remains to be settled. It must have presumably encoded collective/distributive meaning.

TABLE 32.3 Reciprocal forms in Formosan languages

Language/Dialect	**Dynamic verbs**	**Stative verbs**
	PAN *pa-Ca-/*ma-Ca-	**PAN *paR-/*maR**
Mantauran Rukai	*pa*-C*a*-/*ma*-C*a*- *paa*-/*maa*-	*pa'a-ka*-/*ma'a-ka*-
Nanwang Puyuma	*pa*-C*a*-/*ma*-C*a*- *p*a-CVCV-/*m*a-CVCV-	*mar-ka*-
Puljetji Paiwan	*pa*-C*a*-/*ma*-C*a*-	*pare-ka*-/*mare-ka*-
Bunun	*mapa*-	*mapa-ka*-
Tungho Saisiyat	C*a*-, *pa*-C*a*-/*ma*-C*a*-	*pa-ka-k*(*a*)-/*ma-ka-k*(*a*)-
Pazeh	*ma*(*a*)-	*ma*(*a*)*-ka*-
Thao	*mapa*-C*a*-, *mapa*-, *ma*-	*mapa-ka*-, *ma*-C*a*-
Mayrinax Atayal	*p*(*a*)-C(*a*)-/*m*(*a*)-C(*a*)-	*pa*-/*ma*-, *pa*-/*ma*- CVCV-
Truku Seediq	*p*-C-/*m*-C-	*p-k*-/*m-k*-
Kavalan	*ma*-, *ma*-C(a)-, (*sim*-)	(*sim-qa*-)
Central Amis	*pa*(C)*a*-, *mal*(*a*)-	*ma*(C)*a*-, *mal*(*a*)*-ka*-

The stative reciprocal prefixes are also used with kinship nouns to form predicates of kinship, e.g., Nanwang Puyuma *malru-wadi* '(be) brother and sister' (*wadi* 'younger sibling'), Thao *mapa-minlhafut* '(be) siblings' (*minlhafut* 'sibling'), or nouns expressing the relationship, e.g., Puljetji Paiwan *mare-kaka* '(a pair of) siblings' (*kaka* 'sibling').

32.3.4 *Verbalizers*

Three verbalizers are discussed here, PAN *ki- 'to get', *Si-/*si- 'to wear, to carry, to have', and *ku- 'to eat'. In Formosan languages, these three affixes all attach to nouns and transform them into verbs. The first two prefixes can be reconstructed to both PAN1 and PAN2; the third prefix to at least PAN1.

The prefix *ki-* can be reconstructed in PAN with the meaning 'to get, obtain' (Zeitoun & Teng 2009). It occurs productively in Rukai, Puyuma, Kanakanavu, Saaroa, Paiwan, Saisiyat, and Kavalan, PAN *k is reflected as *q* in Kavalan and as a glottal stop in Southern Paiwan.

(13) a. Tona Rukai *ki-a-(a)cilay* 'fetch water'
b. Nanwang Puyuma *ki-'aputr* 'pick up flowers'
c. Saaroa *ki-mairange* 'gather sweet potatoes'
d. Kanakanavu *ki-tammi* 'gather sweet potatoes'
e. Southern Paiwan *'i-vaqu* 'pick up flowers'
f. Tungho Saisiyat *ki-pongaeh* 'pick up flowers'
g. Basay *ki-zanum* 'fetch water'
h. Kavalan *qi-tamun* 'pick up vegetables'

Teng (2014) follows Ross (1995, p. 758) in reconstructing the PAN verbalizer *Si- 'wear, carry, have'. She shows that its Puyuma reflex *i- has further grammaticalized to express existence and instrumentality in Nanwang Puyuma as in ***m-i-riwanes*** *na lrangitr* [AV-I-rainbow NOM.DEF sky] 'There is a rainbow in the sky', ***m-i-pitaw***=*ku me-rebay* [AV-I-hoe=1SG.NOM AV-weed] 'I use a hoe to weed' (Teng 2014, p. 137). Teng also shows that a PAN doublet form *si- can be reconstructed (Table 32.4).

TABLE 32.4 Evidence supporting the reconstruction of PAN *Si- and *si- 'wear, carry' in Formosan languages

	*S reflex	*Si-	Gloss	Example
Tanan Rukai	s	*si-*	to wear	*si-ki'ing* 'wear clothes'
Puyuma	∅	*i-*	to wear, have	*m-i-pitaw* 'have a hoe'
Paiwan	s	*si-*	to carry	*ma-si-vagu* 'carry millet'
Pazeh	s	*si-*	to have	*si-pazeng* 'have thorns'
Seediq	s	*se-/s-*	to grow, have	*s-lukus* 'dress up'
Saisiyat	sh	*shi-*	to wear	*shi-potoeh* 'wear shorts'

TABLE 32.4 Evidence supporting the reconstruction of PAN *Si- and *si- (*cont.*)

	*s reflex	*si-	Gloss	Example
Pazeh[4]	z/t	*ti-*	to carry, wear	*ma-ti-kuribu* 'wear a skirt'
Kavalan	s	*si-*	to wear, have	*si-kun* 'wear trousers'
Amis	c	*ci-*	to wear	*ci-cokap* 'wear shoes'

AFTER TENG 2014, P. 150

The verbalizing prefix 'eat' is reconstructed as PAN *ku-, based on Formosan data. Note that PAN *k is lost in Tsou. In Bunun, there are two reflexes, *ku-* and *ik-* 'eat'. In Kanakanavu, *ko:-* represents the monophthongization of *ku-a-*, where *-a-* expresses imperfectivity.

(14) a. Tsou *o-cni* 'eat one' < *coni* 'one'
b. Saaroa *kua-maini* 'eat little' < *maini* 'little'
c. Kanakanavu *ko:-vang(ʉ)vang* 'eat all' < |*vang(ʉ)vang(ʉ)*| 'all'
d. Bunun *ku-s'an* 'eat once' < |*s'an*| 'once'
ik-tanam 'taste' < *tanam* 'try'
e. Thao *k⟨un⟩tanlhuan* 'eat dinner (AV)' < *tanlhuan* 'evening'
f. Saisiyat *k⟨om⟩si'ael* 'eat lunch (AV)' < *si'ael* 'eat'

32.4 Voice, Mood, and Aspect Markers

In a groundbreaking study, Wolff (1973, p. 71) suggests that in PAN1, the verb "was inflected for four voices (an active and what most descriptions of current languages call three passive voices), two modes (which we may term 'dependent' and 'independent'), and three tenses (non-past, past, and future or general actions). In addition, there was a subjunctive form of the verb used in exhortations. There was also an imperative form, which in the modern languages sometimes coincides with the dependent form and sometimes with the subjunctive form but in the protolanguage coincided with the dependent form." His reconstructed paradigm is reproduced in Table 32.5, adapted for comparison's sake to the format of the tables that follow. The abbreviations in parentheses refer to the current terminology explained beneath the table. Wolff notes against *⟨um⟩-√ 'active non-past' and * ⟨in-um⟩√ 'active past' that they occur

4 Lim & Zeitoun (2023) show that *ti-* in Pazeh-Kaxabu is not a reflex of PAN *Si-.

with "bases not beginning with *p* or *b*". For "bases beginning with *p* or *b*", he notes that the active non-past is formed by changing *p*/*b* to *m*, and the active past by changing *p*/*b* to *m* and infixing ⟨*in*⟩. This represents Wolff's awareness of two of the formal verb classes in § 32.2.1.

TABLE 32.5 PAN1 voice, mood, and aspect morphemes

	ACTIVE (AV)	DIRECT PASSIVE (UVP)	LOCAL PASSIVE (UVL)	INSTRUMENTAL PASSIVE (UVC)
INDEPENDENT				
Non-past	*⟨um⟩-√	*√-en	*√-an	*i-√
Past	*⟨in-um⟩√	*⟨in⟩√	*⟨in⟩√-an	*i⟨in⟩-√?
Future/General action	?	*RED-√-en	*RED-√-an	?
DEPENDENT				
Dependent	*√	*√-a	*√-i	*√-an ?
Subjunctive	*√-a	?	*√-ay	?

AFTER WOLFF 1973, PP. 71–72

Ross (1995, p739) replicates Wolff's material in Table 32.5 but switches *⟨in-um⟩√ to *⟨um-in⟩√. He writes that his "reconstructions ... are based on material from fifteen Formosan languages and various extra-Formosan languages". This leads to a relabeling of many functions, to filling the cells that Wolff left empty, and to slight revisions, shown in Table 32.6 (Table 32.5), of Wolff's "dependent" forms. It also led to the omission of Wolff's "Instrumental passive" because the available data were messy. The table in Ross (2002, p. 33) is almost identical to the 1995 version but with Wolff's instrumental voice restored and relabeled 'circumstantial' in view of its Formosan reflexes' varied semantics. Both the 1995 and 2002 versions of the table include paradigms of PAN *kaRaw 'scratch' and *kaRaC 'bite' with acute accents showing stress, borrowed from Zorc's (1993) PMP reconstructed accents. In light of Blust's (1997) finding that stress cannot be reconstructed for PAN these are removed in Ross's (2009) table, here reproduced as Table 32.6.

The debt owed by Ross (1995, 2002, 2009) to Wolff's insightful reconstruction of 1973 is obvious from a comparison of Table 32.5 with Table 32.6. The reconstructed paradigm is shown in schematic form, as PAN verbs belonged to six different formal classes defined by two parameters: how *M was manifested morphologically, and what form the stem took in cells that lack *M (§ 32.3.1).

Forms of the PAN verb *kiRim 'seek, look for' are given as an example of the most common PAN formal class of verb, where *M was manifested as infixed *⟨um⟩ and the plain stem (*kiRim) was used in cells that lack *M. Table 32.6 differs from the similar tables of Ross (2009, 2012) in one other small respect. The latter shows the UVP dependent form as *√-a, but this is reflected in Tsou alone; it is here amended to *√-i, reflected in this function in Puyuma, Saaroa, Kanakanavu, and six NucAN languages.

TABLE 32.6 PAN1/PNucAN voice, mood, and aspect morphemes

	Actor voice	**Undergoer voice**		
		Patient subject	**Location subject**	**Circumstance subject**
INDICATIVE				
Realis	*M-√ *k⟨um⟩iRim	*√-ən *kiRim-en	*√-an *kiRim-an	*Sa-/Si-√ *Sa-/Si-kiRim
Realis perfective	*M-⟨in⟩√ *k⟨um-in⟩iRim	*⟨in⟩√ *k⟨in⟩iRim	*⟨in⟩√-an *k⟨in⟩iRim-an	*⟨in⟩Si-√ *S⟨in⟩i-kiRim
Realis imperfective	*M-Ca-√ *k⟨um⟩a-kiRim	*Ca-√-ən	*Ca-√-an	*Sa-/*Si-Ca-√ *Sa-/Si-ka-kiRim
Irrealis ??	*Ca-√ *ka-kiRim	*ka-kiRim-en	*ka-kiRim-an	*Ca-√ *ka-kiRim
NON-INDICATIVE				
Imperative	*√ *kiRim	*√-u *kiRim-u	*√-i *kiRim-i	*an-i + √, (√-ani) *an-i kiRim
Dependent				
Optative hortative	*M-√-a *k⟨um⟩iRim-a	*√-aw *kiRim-aw	*√-ay *kiRim-ay	*an-ay + √, (√-anay) *an-ay kiRim

AFTER ROSS 2009, P. 296

For scholars who do not accept the Nuclear Austronesian hypothesis (§ 32.1), Table 32.6 reflects the organization of the PAN1 verb. For those who do, it reflects PNucAN. Space precludes a detailed recitation of the arguments for the hypothesis. For these the reader is referred to Ross (2009), but we discuss the hypothesis briefly in § 32.4.1.

32.4.1 *Verbal Morphology and the Nuclear Austronesian Hypothesis*

The Nuclear Austronesian (NucAN) hypothesis (Ross 2009) says that sometime early in the history of Austronesian, a set of changes occurred in what became PNucAN, resulting in the verb system in Table 32.6. The ancestors of Tsou, Rukai, and Puyuma did not participate in these changes. According to the NucAN hypothesis, the UV indicative verbs of Table 32.6 were not verbs in PAN2 but nominalizations, while the non-indicative forms were the verbs of PAN2, albeit with different glosses.

Basically, the argument for the hypothesis consists of two steps. The first was to adopt the argument from Andrew Pawley's lectures at the 1977 Institute of the Linguistic Society of America that the forms labeled "Indicative" in Table 32.6 were originally nominalizations. The argument was that the reconstructed indicative-voice forms were a crazy mixture of infixes, reduplications, suffixes, and prefixes unlike anything normally found in a verbal paradigm, but not unlike the various devices forming nominalizations in languages with agglutinative morphology. At some point earlier than PAN1, these nominalizations, used in clefts, had been reanalyzed as finite verbs, so that the latter had the same forms as nominalizations in a number of Formosan languages. This argument took written form with Starosta, Pawley & Reid (1981), which was not published until 2009 (a much abbreviated version appeared as Starosta et al. 1982). This was largely written by Starosta, we understand, taking inspiration from Pawley's lectures and Pawley & Reid (1980).

The second step was by Ross (2009), who argued that the verbal systems of Tsou, Rukai, and Puyuma could not be derived from the PAN1 system in Table 32.6, but that their nominalizing morphology nonetheless resembled the nominalizations and indicative verbs of other Formosan languages. That is, the ancestors of Tsou, Rukai, and Puyuma had split off from the Austronesian tree before Pawley's change, i.e., before the reanalysis of nominalizations as verbs had taken place and it occurred in PNucAN. This hypothesis is supported by the fact that nominalizations are used as predicates in Rukai and Puyuma in ways that are ripe for reanalysis, but this has still not happened. Li (1973, pp. 202–211) is at pains to show that in constructions like those in (15), the nominalization, here *wa-kane-li* 'my eaten (thing)', behaves as a nominal, but it is easy to see that when the determiner *ka* is omitted, as it is in (15b), the nominalization could easily be reanalyzed as a verb.

(15) Tanan Rukai (P. Li 1973, p. 202)

a. *kuani kaang ka wa-kane=li*
that fish DET PST-eat=1SG.GEN
'That fish was my eating fish.' = 'I ate that fish.'

b. *wa-kane=li kuani kaang.*
PST-eat=1SG.GEN that fish
'My eating was that fish.' = 'I ate that fish.'

The Puyuma examples in (16) are unambiguous predicate nominalizations, as *tr⟨in⟩ima-an* in (16a) and *ni-ladra-ladram* in (16b) are each preceded by a pronoun that can only function as a possessor but not an actor. Furthermore, the *⟨in⟩√-an* construction in *tr⟨in⟩ima-an* forms nominalizing constructions, and in (16b) *ameli* is the negator used with nominal predicates. If the case markers were to disappear, then reanalysis as a verb would become possible.

(16) Nanwang Puyuma (Teng 2008, p. 131)

a. *nanku tr⟨in⟩ima-an na tilril*
1SG.PSR.NOM ⟨PFV⟩buy-NMLZ NOM.DEF book
'The book was my buying.' = 'I bought the book.'

b. *ameli nantu ni-ladra~ladram ta=ngai*
NEG 3.PSR.NOM PFV-RED~know 1PL.INCL.PSR=language
'Our language is not their learning.' = 'They haven't learned our language.'

If the inference that nominalizations spawned verbs no earlier than PNucAN is correct, then it leads to the challenge of reconstructing PAN2 verbal morphology. Ross's 2012 version is shown in Table 32.7, and the resemblance of its verb forms to those of Teng's (2018) Proto-Puyuma in Table 32.8 is obvious. Its nominalizations resemble certain PNucAN verbs in Table 32.6. The reasoning behind Ross's Puyuma-like PAN2 reconstruction was that Puyuma displays a patterned affix paradigm of the kind that one might expect to find in a verb system, in contrast to the collection of affixes that Pawley deemed unparadigmatic (Ross 2012). The paradigm was easily accessible to internal reconstruction by Ross (1995, pp. 767–768), who set out the skeletal forms in (17):

(17)

	AV	UVP	UVL	UVC
INDICATIVE	*M-√-a	*√-a-u	*√-a-i	*√-an-a-i
NON-INDICATIVE	*√	*√-u	*√-i	*√-an-i

The suffixes *∅, *-u, and *-i encode voice. The UVC forms consist of a morph *-an, perhaps an auxiliary verb, with UVL suffixes.[5] The suffix *-a encodes indicative mood. These suffixes are self-evident in both the PAN2 verbal reconstructions in Table 32.7 and in Proto-Puyuma in Table 32.8.

There is a mismatch between Tables 32.7 and 32.8: PAN2 is reconstructed with contrasting realis and irrealis moods, but Proto-Puyuma is not. Teng (2018) points out that both imperfective aspect and irrealis mood forms begin with *Ca-* reduplication, and that the two dialects that she examines disagree about the slots in which irrealis occurs: Katripul has only AV irrealis, but Nanwang also has UV irrealis. Teng concludes that Proto-Puyuma had only the imperfective, but some of its (non-indicative) forms were hijacked to serve as (indicative) irrealis. Interestingly, crossovers between imperfective and irrealis also occur in cognate sets of *Ca-* forms in other Formosan languages, and Teng's conclusion that Proto-Puyuma was aspect-sensitive rather than mood-sensitive seems to have applied to PAN2 and PNucAN.

TABLE 32.7 PAN2 voice, mood, and aspect morphemes

	AV	UVP	UVL	UVC
NOMINALIZATION				
Realis	*M-√	*√-an		*Sa/*Si-√
Realis perfective	*⟨in⟩M-√	*⟨in⟩√	*⟨in⟩√-an	—
Realis imperfective	*Ca-√	*Ca-√-ən	*Ca-√-an	*Sa/*Si-Ca-√

5 This *(-)an- is shown as a suffix in (17), but as an auxiliary in Table 32.7, reflecting different sources. A reviewer justly questions this discrepancy and comments that if *(-)an was originally an auxiliary, it would be "very surprising to find it post-verbally in a head-initial language." We agree. Syntax suggests that it started life as a suffix, morphology that it was an auxiliary. This is a puzzle seeking a solution.

TABLE 32.7 PAN2 voice, mood, and aspect morphemes (*cont.*)

	AV	UVP	UVL	UVC
INDICATIVE				
Realis	*M-√	*√-aw	*√-ay	*an-ay √
Hortative	*M-√-a			
Realis imperfective	*M-Ca-√	*Ca-√-aw	*Ca-√-ay	*an-ay Ca-√
NON-INDICATIVE				
Imperative	*√	*√-u	*√-i	* an-i √
Dependent	*M-√	*√-a		
Irrealis	*Ca-√	—	*Ca-√-i	* an-i Ca-√

AFTER ROSS 2012, P. 1264

TABLE 32.8 Proto-Puyuma voice and aspect morphemes

	AV	UVP	UVL	UVC
INDICATIVE				
Realis	*M-√	*√-aw	*√-ay	*√-anay
Realis imperfective	*M-Ca-√	*Ca-√-aw	*Ca-√-ay	*Ca-√-anay
NON-INDICATIVE				
Imperative	*√	*√-u	*√-i	*√-an
Dependent	*M-√			

TABLE 32.8 Proto-Puyuma voice and aspect morphemes (*cont.*)

	AV	UVP	UVL	UVC
Dependent imperfective	*Ca-√	*Ca-√-i	*Ca-√-an	
Hortative	*M-√-a	—	—	—

AFTER TENG 2018, P. 322

32.4.2 *Verbal Morphology: A Strict Application of the Comparative Method*

There is a problem with the reconstructions in Table 32.7. Strict reconstruction of forms in the proto-language of a language group requires that each reconstructed form be reflected either (a) in at least two primary branches of the group, or (b) in at least one external witness and one primary branch of the group. If Table 32.6 is taken as a reconstruction of PNucAN, then, assuming Blust's (1999a) subgrouping, the reconstruction readily meets this criterion. In the case of PAN2, (b) does not apply (§ 32.1), and under the NucAN hypothesis, (a) requires that each reconstruction be attested in two of Tsou; Rukai; Puyuma; and either NucAN or a possible primary branch containing Saaroa, Kanakanavu, and PNucAN (see below). The reconstruction in Table 32.7 does not meet this criterion.

Before looking more closely at this deficiency, however, there is a matter of classification to attend to. On the basis of new data, Zeitoun & Teng (2016) show that Saaroa and Kanakanavu reflect some but not all of the innovations of NucAN languages. Table 32.9 summarizes their verbal morphology.

TABLE 32.9 Comparing Saaroa and Kanakanavu voice, mood, and aspect morphemes

	Saaroa	Kanakanavu	Saaroa		Kanakanavu
	AV		UVP	UVC	UV
INDICATIVE					
Realis perfective	*lhi-M-√*	*⟨in⟩M-√*	*√-a(na)*	*√-ani*	*⟨in⟩√*
Realis imperfective	*M-Ca-*R*-√*	*M-Ca-√*			*√-un*
Irrealis	*M-Ca-√*		–		

TABLE 32.9 Comparing Saaroa and Kanakanavu voice, mood and aspect morphemes (*cont.*)

	Saaroa	Kanakanavu	Saaroa		Kanakanavu
	AV		UVP	UVC	UV
NON-INDICATIVE					
Imperative	*M-√-a*	*M-√-a*	*√-u*	*√-ani*	*√-o*
Dependent	*M-√*	*M-√*	*√-i*		*√-e*

In the Saaroa realis imperfective cell, 'R' = root reduplication.
AFTER ZEITOUN & TENG 2016, PP. 194–195

Under the NucAN hypothesis (§ 32.4.1), NucAN languages reflect UV nominalizing morphology in UV indicative verbs. Are Saaroa and Kanakanavu by this criterion NucAn languages? Yes, except for Saaroa *√-ani* 'UVC', a comparison with Table 32.6 tells us that they are. But the two languages are different from other NucAN languages in what they lack: (a) both languages apparently lack UVL, and Kanakanavu also lacks UVC; (b) both languages reflect PAN *Si-√ in instrumental nominalizations (Teng & Zeitoun 2016) but, unlike most NucAN languages, not in UVC verbal morphology. Instead, Saaroa has generalized the non-indicative UVC suffix *-ani to the indicative and to circumstantial nominalizations.

The question is, are these languages daughters of PNucAN or not? Possibly. There is little evidence as to whether pre-Saaroa or pre-Kanakanavu ever had UVC verb forms reflecting PAN *Si-√, but there is a small piece of evidence that pre-Saaroa once distinguished UVP and UVL verb forms. The Saaroa indicative UVP *√-a(na)* has the apparent allomorphs *√-a* and *√-ana*. In the light of the nominalizations, one could infer that the indicative allomorphs reflect earlier *√-a* UVP and *√-ana* UVL (from PNucAN *√-ən UVP and *√-an UVL), which have collapsed into a single voice.[6] Beyond this, we cannot readily tell whether Saaroa and Kanakanavu (i) are aberrant daughters of PNucAN, aberrant perhaps because of intense Tsou influence (Ferrell 1969), or (ii), as Zeitoun & Teng (2016) propose, are daughters of a primary branch or branches of Austronesian that have undergone some but not all of the innovations characteristic of NucAN.

6 In his description of Saaroa, Pan (2012) distinguishes between UVP and UVL, but the few examples of putative UVL are not consistently distinct semantically from UVP.

We now return to the strict reconstruction of the forms in Table 32.10. In view of their somewhat uncertain status, for this purpose it is best to ignore Saaroa and Kanakanavu, as we do not know whether we should treat them as NucAN or not.

Table 32.10 shows the distribution of reflexes of each item in the paradigm. Under each form is a sequence of letters showing the languages in which the form is reflected. In the top portion of the table are forms that occur both as nominalizations and as indicative verbs, with separate letter sequences representing the reflexes of each. Cells with darker shading meet the criterion for a PAN2 reconstruction. Cells with lighter shading contain forms that are reflected in just one of Tsou, Rukai, and Puyuma. Note that there is no space to list supporting data. The table is not exhaustive. Forms that are only reflected, for example, in one NucAN language and nowhere else are omitted.

TABLE 32.10 Distribution of reflexes of possible PAN2 voice, mood, and aspect morphemes

	AV	**UVP**	**UVL**	**UVC**
NOMINALIZATION/INDICATIVE				
Nominalization	*ta-√ *TRKN*	*ta-√-an *R*	*ta-√-an *TRSKN*	
Realis nominalization indicative	*M-√ *KN* (see below)	*√-ən *KN* *N*	*√-an *TRPSKN* *SN*	*Sa/*Si-√ *RPSKN* *N*
Realis perfective nominalization indicative	*⟨in⟩M-√ *SKN* *SKN*	*⟨in⟩√(-an) *(RPS)KN* *KN*	*⟨in⟩√-an *PSKN* *N*	⟨in⟩Si-√ *N* *N*
Realis imperfective nominalization indicative	*M-Ca-√ *RPKN* (see below)	*Ca-√-ən *PSN* *N*	*Ca-√-an *TRSN* *N*	*Sa/*Si-Ca-√ *N* *N*
Irrealis nominalization indicative	*Ca-√ *P* (see below)	*Ca-√-ən *SN* *N*	*Ca-√-an *RPSN* *N*	

TABLE 32.10 Distribution of reflexes of possible PAN2 voice, mood, and aspect morphemes (*cont.*)

	AV	UVP	UVL	UVC
INDICATIVE				
Realis	*M-√ *PN*	*√-aw *P*	*√-ay *P*	√-an-ay *P*
Realis imperfective	*M-Ca-√ *RPSKN*	*Ca-√-aw *P*	*Ca-√-ay *P*	*Ca-√-an-ay *P*
Irrealis	*Ca-√ *P*			*Ca-√(an-i) *(S)N*
Hortative	*M-√-a *RPSKN*	*√-aw *KN*		*√-ay *N*
NON-INDICATIVE				
Imperative	*√ *PN*	*√-u *PSN*	*√-i *PN*	*√-an-i *PSN*
Dependent	*M-√ *TRPSK*	*√-i *PSN*	*√-i *TPSN*	*√-an-i *TPN*
Dependent	*√ *RPSN*	*√-a *T*		

Key: *T* Tsou, *R* Rukai, *P* Puyuma, *S* Saaroa, *K* Kanakanavu, *N* NucAN

For readability's sake, the PAN2 forms that meet the criteria for reconstruction in Table 32.10 are set out again in Table 32.11, without the lists of languages and without the putative forms that are not reconstructable, but retaining in lighter gray those for which there is a single Tsou, Rukai, or Puyuma reflex. The results are intriguing.

TABLE 32.11 Reconstructable PAN2 voice, mood, and aspect morphemes

	AV	UVP	UVL	UVC
NOMINALIZATION				
General	*ta-√	*ta-√-an	*ta-√-an	
Realis			√-an	*Sa/*Si-√
Realis perfective		*⟨in⟩√(-an)	*⟨in⟩√-an	
Realis imperfective	*M-Ca-√	*Ca-√-ən ??	*Ca-√-an	
Irrealis	*Ca-√		*Ca-√-an	
INDICATIVE				
Realis	*M-√	*√-aw	*√-ay	√-an-ay
Realis imperfective	*M-Ca-√	*Ca-√-aw	*Ca-√-ay	*Ca-√-an-ay
Irrealis	*Ca-√			
Hortative	*M-√-a			
NON-INDICATIVE				
Imperative	*√	*√-u	*√-i	*√-an-i
Dependent	*M-√			
Dependent	*√	*√-a		

As one might expect, a fairly full set of nominalizations in the top section of Table 32.11 meets the criteria for reconstruction. Apart from forms in *ta-, these are forms that later double as PNucAN indicative verbs. *Ca-√-ən 'UVP realis imperfective nominalization' is the only form with the suffix *-ən and is reflected only in two Puyuma dialects, Katripul (Stacy Teng, pers. comm.)

and Tamalakaw (Tsuchida 1980, pp. 203, 208). By the criteria established at the beginning of this section, it should be excluded from Table 32.11. It is included with question marks because the default assumption would be that *-ən was present in PAN, and that *-ən has merged with *-an as *-an in Tsou, Rukai, Saaroa, and the Nanwang dialect of Puyuma (as has happened in Kavalan). An alternative possibility is that it is an unexplained post-PAN2 innovation.

The non-indicative reconstructions in Table 32.11 are unproblematic. The indicative reconstructions, however, raise questions. Only AV forms can be reconstructed. Only Puyuma provides supporting data for UV indicative reconstructions. This perhaps results from interaction between the paradigmatic structures of Tsou, Rukai, and PNucAN and the criterion that a form must be reflected in two primary branches.

The NucAN hypothesis asserts that PNucAN replaced all PAN2 UV indicative verb forms with forms drawn from PAN2 nominalizations. If this was so, then PNucAN had no reflexes of PAN2 UV indicative forms.

In Tsou, all verb phrases consist of an auxiliary marking voice and mood and one of the dependent verb forms in (18). An example of this structure is shown in (19). It follows from this that Tsou has no indicative verbs.

(18) Tsou

AV	*UVP*	*UVL*	*UVC*
M-√	√-a	√-i	√-[n]eni

(19) Tsou

o=su *eobak-a* (*na*) *a'o.*
REAL.UV=2SG beat-UVP (NOM) 1SG
'You beat me.' (Zeitoun 2005, p. 277)

Rukai is an accusative language with an innovatory passive in *ki-* reflecting PAN *ki-N 'get N' (§ 32.3.4). Its AV voice forms fairly transparently reflect PAN AV forms, as shown in Table 32.12, and it reflects no PAN UV accusative reconstructions.

TABLE 32.12 Verbal morphology of Proto-Rukai[7]

	Rukai gloss	Proto-Rukai	PAN	PAN gloss
Active	**Realis**	*u-a-√	*M-Ca-√	AV realis imperfective
	Imperative	*⟨u⟩√-a	*M-√-a	AV hortative
	Subjunctive	*⟨u⟩√	*M-√	AV dependent
	Dependent	*√	*√	AV dependent
Passive		*ki-[a-]√	—	—

We are thus left with a situation in which the data permit no reconstruction of PAN2 UV indicative forms. Is this an artefact of method or a possible PAN2 reality? Section 32.4.3 suggests it is the latter.

32.4.3 *Alternative Hypotheses*

Ross (2009) surmised that PAN2 had ergatively aligned clause structure like Tsou, Puyuma, and the NucAn languages of Taiwan and the Philippines, and he filled the UV indicative gaps with forms based on Puyuma. As he notes (p. 305), this is a risky procedure. Taken alone, the Puyuma forms do not meet the criteria for PAN2 reconstruction. Other logically possible hypotheses are the following:

(a) The NucAN hypothesis is wrong. Either (i) the derivation of UV indicative forms from nominalizations had occurred prior to PAN, Tsou and Rukai have lost these forms, and Puyuma had replaced them with hortative forms or (ii) UV indicative forms are not derived from nominalizations at all (Blust & Chen 2017).

b) PAN was an accusative language like Rukai, and the UV indicative forms appeared later (Starosta 1995). This presupposes an initial two-way division of Austronesian into Rukai and "Proto-Ergative-AN" (PEAN), as Aldridge calls it (Aldridge 2015, 2016, 2021).

c) PAN was like Tsou: auxiliaries encoded voice, aspect, and mood, and there were thus no independent verbs (Aldridge 2021).

Aldridge (2021) proposes a hybrid of (b) and (c). She argues against Ross's assumption that Rukai was originally ergative but has lost UV indicative forms, on the grounds that this should lead to a situation like that in Chamorro, where (erstwhile AV) intransitive verbs continue to be marked by a reflex of PAN *M,

7 In reconstructions ⟨...⟩ marks an infix; (...) marks an element whose presence is doubtful; and [...] indicates that the item it occurs in can be reconstructed both with and without it.

while (erstwhile UV) transitive verbs become unmarked. Aldridge infers from this that Rukai cannot have once been ergative. In Rukai, however, all AV indicative verbs are marked by *u-*, which we do take to reflect PAN *M. This inference, however, overlooks evidence from Tsou and NucAN that early Austronesian AV verbs had both an actor and an undergoer argument in certain circumstances, i.e., the system was (near) symmetrical.

But Aldridge's proposal about PEAN meshes well with the finding that PAN2 lacked UV indicative forms. The proposal is that PEAN was like Tsou. Tsou auxiliaries appear to be derived from verbs, as those marking AV reflect *M-. That is, they are/were auxiliary verbs. If PEAN was like Tsou, then "main" verbs only appeared in dependent forms. If for the sake of argument we assume that PEAN = PAN2, then Aldridge's hypothesis explains the lacunae in Table 32.7. It explains both the Puyuma and the NucAN UV indicative forms as subordinate-clause forms that have been reanalyzed as main-clause forms with the loss of auxiliaries, thereby filling the UV indicative paradigm slots. Evidence for loss of auxiliaries in Puyuma and Paiwan lies in the fronting of genitive and nominative pronouns (§ 32.2). In PNucAN, nominalized subordinate clauses were reanalyzed as main-clause verbs, as originally assumed by Starosta et al. (1981/2009). Aldridge's account of the process that gave rise to Puyuma UV indicatives is rather more sketchy, and its mechanics are not entirely clear to us. The proposal in Aldridge (2021) is shown in (20a), where *-a(-) is interpreted as marking its verb as nonfinite and *-i as applicative. This, however, does not account for the Puyuma forms shown in (20c). They are better accounted for as shown against "Proposed" in (20b), where *-a(-) marked its verb as subordinate but finite and *-u and *-i marked UVP and UVL, respectively, as in Table 32.11.

(20)

Pre-Proto-Puyuma		AV	UVP	UVL
a. Aldridge:	indicative	*M-√	—	—
	non-indicative	*M-√-a	*√-a	*√-a-i
b. Proposed:	indicative	*M-√	—	—
	non-indicative finite	*M-√-a	*√-a-u	*√-a-i
	non-indicative nonfinite	*√	*√-u	*√-i
c. Puyuma:	realis	M-√	√-aw	√-ay
	dependent	√	√-u	√-i

This interpretation requires us to infer the function of *√-u rather liberally, as in each of the languages in which it is reflected—Puyuma, Saaroa, Paiwan, and Siraya—it marks its verb as imperative, with some variation in voice.

Both Ross's and Aldridge's hypotheses go beyond the bounds of the comparative method, using abductive reasoning to determine which of several alternatives is the most probable. Aldridge's reasoning is usually explicit, but her abductions are often based on the Minimalist Framework, which has its own theory-internal means of motivating change. Ross's abductions are, as Aldridge points out, often not made explicit. But if we go back to applications of the comparative method itself, then Aldridge's inference that PAN2 lacked UV indicative forms makes sense of the reconstructions in Table 32.11.

32.5 Nominal Morphology

Different changes can be discussed in the modern languages, in terms of both form and function. We will only mention a few examples, referring the reader to language-specific chapters for additional details.

As far as the forms are concerned, it is noteworthy that the nominative pronominal clitic is *=lrao* '1SG.NOM' in Mantauran Rukai (Zeitoun 2007), instead of the commonly found *=(a)ku* '1SG.NOM'. The innovated pronominal genitive pronoun is *=li* '1SG.GEN', exclusively shared by the Rukai dialects, a form that has been borrowed in Nanwang Puyuma and is used to denote kinship terms, e.g., *muli* 'my grandfather', *namali* 'father', *nanali* 'my mother', *baeli* 'my older sibling' (Teng 2008, p. 97). In Bunun, Thao, and Saisiyat, the directional prefixes *mun-/mon-* along with the causative form *pun-/pon-* are found along with *mu-/mo-* and *pu-/po-* (Blust 2003b, Zeitoun et al. 2015, L. Li 2018). In Thao, Bunun, Tona, and Maga Rukai, the prefix *pa-ka- 'causative of stative verbs' has been replaced by a noncognate morpheme, cf. *pia-* in Thao, *pi-* in Bunun, *pa-ti-* in Tona Rukai, *pa-te-* in Maga Rukai (see L. Li, this handbook, Chapter 19).

Major changes in some modern languages include case syncretism and case attrition in the case marking and in the pronominal systems and reductions in the voice system.

Reconstructions of PAN case markers and personal pronouns are given in §§ 32.5.1 and 32.5.2, respectively.

32.5.1 *Case Markers*

Noun phrases (NPs) in Formosan languages are often preceded by a case-marking morpheme of the form CV or V (see examples 1, 5, 6, 9b, 10, 15a, 16a), and this was apparently so in PAN. We make no attempt to distinguish PAN1 and PAN2 here.

There are two extant reconstructions of PAN case markers: Ross (2006; Table 32.13) and Blust (2015; Table 32.14). In both, the consonant encodes the

case of the NP, and the vowel indicates whether the head noun is common or personal, and for Blust, whether a personal NP is singular or plural. However, there are key discrepancies between the two reconstructions, most notably that for Ross, *k- marks nominative and *s- oblique, but for Blust, the roles are reversed. Ross is unsure how *-a is functionally distinct from *-u, whereas Blust has *-a marking a personal NP as plural.

One reason for the nominative/oblique discrepancy is that Blust also uses Philippine data. Proto-Philippine *su 'NOM.CN' can be reconstructed, but PAN *su cannot (Blust 2015, p. 449). Another reason is the striving for paradigmatic regularity. Ross and Blust both reconstruct *si 'NOM.PN', supported by the NOM-PN markers Paiwan and Thao *ti*, Saisiyat *hi*, and Amis *ci*. Blust takes *si as evidence for PAN *s- 'NOM', whereas Ross treats it as anomalous.

TABLE 32.13 Ross's (2006, pp. 525–527) reconstruction of case markers

	NEUT	NOM	GEN	ACC (= OBL?)	? OBL	LOC
PAN	*∅-	*k-	*n-	*C-	*s-	*d-
CN	*[y]a	*ka	*na	*Ca	*sa	*da
CN	*u	*ku	*nu	*Cu	*su	—
PN	*i	*ki, *si	*ni	*Ci	—	—

TABLE 32.14 Blust's (2015, p. 467) reconstruction of case markers

	NOM	GEN	OBL	LOC	
PAN		*s-	*n-		*d-
CN	*-u	—	*nu	*ku	—
PN-SG	*-i	*si	*ni	(*ki)	(*di)
PN-PL	*-a	*sa	*na	[*ka]	[*da]

Evidence for PAN *ka/*ku 'NOM.CN' vs. *ki 'NOM.PN', is plentiful: Tsou *'o* 'NOM.REF.INVIS', Tona Rukai *ko* 'NOM.CN' vs *ki* 'NOM.PN', Pazeh *ki* 'NOM.CN', Saisiyat *ka* 'NOM.CN', Mayrinax Atayal *ku'* 'NOM.CN', Seediq *ka* 'NOM.CN' and Amis *ku* 'NOM.CN', as well as Saaroa *a~ka* 'NOM.CN', and Takbanuaz Bunun *a~ka* 'NOM.CN'.

The last two items show phonologically conditioned alternation between *ka* and *a*, with the possibility that earlier functionally separate *a and *ka have

fallen together. The surmise that PAN had a *k-less set is supported by Katripul Puyuma *a* 'NOM.CN' vs. *i* 'NOM.PN', Nanwang Puyuma *a* 'NOM.CN.INDF', Paiwan *a* 'NOM.CN', Mayrinax Atayal *'a'* 'NOM.CN.NREF' vs. *'i'* 'NOM.PN', Favorlang [*i*]*ya* 'NOM.CN', and Kavalan [*y*]*a* 'NOM.CN'. Blust rejects Ross's *k-less "neutral" set because Amis is the only language to display a functional contrast between sets with and without *k- (Central Amis *u* 'CNTSV.TOP.CN' vs *ku* 'NOM.CN'), but the *k-less reflexes in nominative forms convince us to reconstruct them.

In contrast, the evidence for Ross's oblique *s- is thin: Paiwan *ta, tua* 'OBL.CN' and Kavalan *sa* 'LOC.CN'. Blust's oblique *k- is supported as a prepositional element *ka- that precedes a genitive case marker: Puyuma *ka-na* 'OBL.CN' and 'OBL.PN-PL' vs. *ka-n*[*i*] 'OBL.PN-SG', Saisiyat *ka-n* 'LOC.PN-PL', and PMP *ka-ni 'OBL-PN' (Reid 1978). Note also Siraya *ki* 'GEN/OBL.CN'.

At least two cognate sets reflect forms marking adjuncts in PAN. The first replaces Ross's (2006) *C- set: PAN *ta, *tu 'OBL.CN' vs. *ti 'OBL.PN'. Its reflexes are Paiwan *tjay* 'OBL.PN', *tjay-a* 'OBL.PN-PL', Thao *ta* 'LOC.CN', Kavalan *ta, tu* 'OBL.CN' vs *ti* 'OBL.PN', and Siraya *tu* 'LOC.CN'. Putative PAN *ti is reflected only in Kavalan and is thus suspect. The second is the uncontroversial *d- 'LOC': PAN *da 'LOC.CN', Puyuma *dra* 'OBL.CN', Thao *sa* 'OBL.CN', Favorlang *de* 'LOC.CN', and Saisiyat *ray* 'LOC.CN'. Pazeh has *di* 'LOC.CN'. PAN also had a locative preposition *i, reflected as *i* in Puyuma, Paiwan, Takbanuað Bunun, Thao, Favorlang, and Central Amis and as Mayrinax Atayal *'i'*.

Ross and Blust also disagreed over PAN *-a forms. For Ross, they marked common nouns, and there are indeed numerous *a-grade forms with common nouns among the case markers cited above. For Blust, PAN *-a marked personal plurals. Zeitoun (2009) shows that PAN *-a was a personal plural marker that occurred in various environments, such that PAN *si-a 'NOM.PN.PL' and *ni-a 'GEN.PN-PL' can be reconstructed, directly reflected by Paiwan *ti-a* and *ni-a* and in reduced form by Amis *ca* and *na*, Saisiyat *na* 'GEN.PN.PL', and Puyuma [*ni*]*na* 'GEN.PN.PL'.

The functional difference between Ross's *a-grade and *u-grade forms remains elusive, but there is evidence that *-a markers were definite and *-u markers indefinite. This appears to be reflected by the Tsou oblique agent forms *ta* vs. *to* (Szakos 1994, pp. 92–95, Zeitoun 2005, pp. 274–276), by Budai Rukai core *ka* vs. *ku* (Shih 2012, pp. 13–14), by the Paiwan genitives *na/nua* vs. *nu* (Tang 2006) and obliques *ta/tua* vs. *tu* (Chang 2006, p. 115, Tang et al. 1997), and in Kavalan's recent past by the accusatives *ta* vs. *tu* (Lee 1997, pp. 19–21).

The areas of agreement between Ross and Blust are that PAN had genitives in *n- and locatives in *d- and *si 'NOM.PN'.

Table 32.15 shows a tentative and revised reconstruction of PAN case markers. The reconstruction of CV forms is tricky, because the possibility of chance

resemblances is higher than with longer forms, but here it is balanced by the fact that only five PAN consonants, *k-, *s-, *n-, *t-, and *d-, are found in the reconstructions. This provisional reconstruction is a little more disorderly (and perhaps thus more plausible) than Tables 32.13 and 32.14. Reconstruction is an attempt to retrieve a real language, and paradigms in real languages are rarely as regular as those reconstructed by Ross and Blust.

TABLE 32.15 Tentative revised reconstruction of case markers

	"NEUTRAL"	NOM	GEN	OBL	OBL	LOC
PAN	*∅-	*k-	*n-	*ka-n-	*t-	*d-
CN-DEF?	*a	*ka	*na	*ka-na	*ta	*da
CN-INDF?	*u	*ku	*nu	—	*tu	—
PN-SG	*i	*si	*ni	*ka-ni	(*ti)	(*di)
PN-PL	—	*si-a	*ni-a	(*ka-ni-a)	—	—

32.5.2 *Pronominals*

Blust's (1977) reconstruction of the PAN pronominal system represents pioneering work, bringing together the partial reconstructions by Dempwolff (1938) and Dahl (1976, p. 122) and integrating them into nominative and genitive cases of PAN1, as shown in Table 32.16.

TABLE 32.16 Blust's reconstruction of PAN1 personal pronouns

	NOMINATIVE	GENITIVE	
1SG	*i-aku	*i-ku	*ni-ku
2SG	*i-Su (i)kaSu (polite)	*i-Su	*ni-Su
3SG	*si-ia	*i-a	*ni-a
1INCL.PL	*i-kita	*i-ta	*ni-ta
1EXCL.PL	*i-kami	*i-mi	*ni-mi
2PL	*i-kamu	*i-mu	*ni-mu
3PL	*si-ida	*i-da	*ni-da

BLUST 1977, P. 10

Blust's goal, however, was not to reconstruct all PAN1 pronouns but to establish that Malayo-Polynesian (MP) was a major subgroup comprising all Austronesian languages outside Taiwan. He argues (1) that *(i)kaSu was a polite 2SG nominative form generated by what he calls the "first politeness shift" and (2) that MP languages reflect a second politeness shift, whereby PAN1 genitive 2PL *-mu came to be used in the singular as a polite variant of 2SG *Su.

Ross (2006) includes the first attempt at a comprehensive reconstruction of PAN pronominals, revised in Ross (2015a) to accommodate the NucAN hypothesis. They owe a substantial debt to Blust's reconstructions of almost 30 years earlier. Ross's PAN2 and PNucAN reconstructions comprise neutral, nominative, accusative, and genitive pronominals, but omit third-person forms because the Formosan data do not form coherent cognate sets.

TABLE 32.17 PAN2 and NucAN Austronesian personal pronouns

	1SG	2SG	1INCL.PL	1EXCL.PL	2PL
PAN2 and PNucAN {...} = PNucAN only					
NEUT	*aku	*iSu, *Su[qu]	*([i])ta	*ami	*mu[qu], {amu}
NOM/GEN1	*=ku	*=Su	*=ta	*=mi[a]	*=mu
GEN2	*m-aku	*m-iSu	*m-ita	*mia	*m-amu
?? GEN3	*n-aku	*n-iSu	*n-ita	*ni-am *n-ami	*ni-mu[qu], {namu}
PNucAN additions					
NEUT/NOM	*i-aku	*iSu[qu]	*ita	*i-ami	*i-mu[qu], (*i-amu)
ACC	*i-ak-en	*iSu-n *suqu-n	*[i]ta-en	*[i]am-en	*imu-n *muqu-n

AFTER ROSS 2015A, P. 114

The NOM/GEN1 set consists of enclitics, ancestral to those in examples (2), (3), (4b), and (6b). The other three sets were apparently standalone forms. Sets GEN2 and GEN3 are reconstructed in order to account for sets of genitive pronominals with initial *m-* and *n-*. Their members are sometimes mixed

in modern paradigms, like the Thao genitives *nak* '1SG' but *mihu* '2SG', *mita* '1INCL.PL'. The origin of initial *m-* here is unknown. Initial *n-*, as Blust (1977) recognized, reflects PAN **ni*, the genitive personal case marker (Table 32.15).

PNucAN added two additional sets. The first prefixed *i- to the PAN2 neutral set: this became the new neutral set and also served as a standalone nominative. The second added *-[e]n to this new neutral set to form a PNucAN accusative set.

Ross (2015b) sketches how pronominals fit into the reconstruction of PAN2. Space does not permit us to repeat the argument in detail here. In (8), a clause skeleton was given to summarize the Paiwan examples in § 32.2. In (21), we give the corresponding skeleton for PNucAN, with reconstructed examples in (22). It differs from Paiwan in that all clitic pronominals are attached after the auxiliary if there is one and after the verb if there is not.

(21) PNucAN clause skeletons

a. AV [*M*-VERB=PRON.NOM]
b. [AUXILIARY=PRON.NOM] DEPENDENT.verb
c. UV [VERB=PRON.GEN=PRON.NOM]
d. [AUXILIARY=PRON.GEN=PRON.NOM] VERB

(22) PNucAN

a. **k⟨um⟩aRaw =**ku***
⟨AV⟩scratch =1SG
'I am scratching.'

b. **azi =**ku** kaRaw*
NEG =1SG scratch.AV.DEP
'I am not scratching.'

c. **kaRaw- an =**ku** =**Su***
scratch-UVL =1SG =2SG
'I am scratching you.'

d. **azi =**ku** =**Su** kaRaw-i*
NEG =1SG =2SG scratch-UVL.DEP
'I am not scratching you.'

There is also evidence that where there were two pronominal arguments in a PNucAN transitive (UV) clause, one could occur in its standalone form after the verb, as in (23).

(23) PNucAN
*azi =**ku** kaRaw-i **iSu**
NEG =1SG scratch-UVL.DEP NOM:2SG
'I am not scratching you.'

At the end of § 32.2, we mentioned that Paiwan is atypical of Formosan languages in that its genitive clitics are attached in front of a UV verb, not after it. Starosta et al. (1981/2009) suggest that the Paiwan (and Puyuma) fronted clitic pronouns are the result of "AUX-axing", i.e., at some point in their histories, auxiliaries were lost from constructions like (3) and (4b), and the stranded clitics adopted the verb that followed them as their new phonological host. It is reasonable to assume that the hortative prefix *ta-* in Saisiyat and Pazeh/Kaxabu, as in Saisiyat *ta-ra'oe:* 'let's drink! (AV)' (Zeitoun et al. 2015) and Pazeh/Kaxabu *ta-kan-i* 'let's eat' (Lim 2022), similarly reflect the loss of a hortative auxiliary, the prefix reflecting the 1INCL.PL clitic in Table 32.17.

Below are given examples from Formosan languages showing how these structures have developed over time.

The language that best preserves PAN1/PNucAN structures is Siraya, extinct apparently since the early 19th century. However, there is plentiful manuscript evidence, mostly from the Dutch presence in the 17th century, analyzed by Adelaar (2011). The examples in (24a, b, c, d, e) match structurally those in (22a, b, c, d) and (23).

(24) Siraya (Adelaar 2011)
a. *ni-m-upänäx=**kamu** tu puläx k⟨m⟩ita ki mang*
PST-AV-come.out=2PL.NOM LOC wilderness ⟨AV⟩see NOM what
'what did you come out to the wilderness to see?' (p. 96)

b. *asi=**kaw** hahey-ən m-äya*
NEG=2SG.NOM allow-UVP AV-take.as.wife
'you are not allowed to have her [as a wife]' (p. 100)

c. *...kalang-ən=**au**=**kaw***
know-UVP=1SG.GEN=2SG.NOM
'... I know you ...' (p. 73)

d. *asi=**mau**=**kamu** ni-kalang-ən*
NEG=1SG.GEN=2PL.NOM PST-know-UVP
'I never knew you' (p. 100)

e. *ataral-ey=**mau*** *tini-än*
forgive-UVL.OPT=1SG.GEN 3SG.OBL
'I'll forgive her.' (p. 72)

The structures in (21) have undergone modifications in various Formosan languages. One of the more conservative languages in this regard is Seediq, where these structures are largely maintained. The sentences in (2) and (3) above reflect the structures in (22c, d), but with the significant difference that the order of postverbal clitics has shifted from =GEN=NOM to =NOM=GEN. Three pieces of evidence speak for =GEN=NOM in earlier Seediq:

(a) Formosan languages that allow clitic sequences all have =GEN=NOM, suggesting that this was the PAN1 order.
(b) In (25a), it is the genitive clitic that stays attached to the verb.
(c) Importantly, Seediq has portmanteau clitics like *=misu* in (25) that each encode a genitive actor and a nominative undergoer. In each, the second syllable, here *su*, is identical to the appropriate undergoer clitic, i.e., they are fossils of earlier =GEN=NOM (Puyuma also has portmanteau clitics like these; Teng 2015).

(25) Tgdayan Seediq
a. *ini=**ku*** *tutuy* *heya.*
NEG=1SG.GEN wake.DEP 3SG.NOM
'I didn't wake her up.' (Holmer 1996, p. 202)

b. *qta-un=**misu.***
see-UVP=1SG.GEN:2SG.NOM
'I'll see you.' (Holmer & Billings 2014, p. 120)

Another Formosan language that allows a sequence of =GEN=NOM pronominals following the clause-initial verb is Kavalan. The AV clause in (26a) behaves like (22a), the UV clause in (26d) like (22c). But here the resemblance stops. Two phenomena conspire to prevent a clitic sequence following an auxiliary, as it does in Siraya (24d) and Seediq (25). First, a *nominative* clitic may indeed climb to an auxiliary (26b, e), but it need not (26c, f). Lee (1997, p. 41) says that climbing is considered the "more natural" construction. Second, erstwhile genitive (actor) clitics have become suffixed to the verb, and a third-person actor suffix also functions as an agreement marker, still occurring when there is a genitive (actor) noun phrase (Chang 1997, Lee 1997). As a suffix, it remains with the verb and does not climb (27e, f), preventing a sequence of two pronominals after an auxiliary. Instead, the two phenomena conspire to allow a sequence of two pronominals after the verb, even when there is an auxiliary.

(26) Kavalan

a. *m-nanum=ti=iku.*
AV-drink=ASP=1SG.NOM
'I drank the water.' (Lee 1997, p. 40)

b. *mai-pama=**iku** busuq.*
not-yet=1SG.NOM AV.get.drunk
'I haven't got drunk yet.' (Lee 1997, p. 41) ("more natural")

c. *mai-pama busuq=**iku.***
Not-yet AV.get.drunk=1SG.NOM
'I haven't got drunk yet.' (Lee 1997, p. 41)

d. *pukun-an=**ku=isu.***
beat-UV=1SG.GEN=2SG.NOM
'You were beaten by me' (Lee 1997, p. 44)

e. *mai=**iku** pukun-an=**na.***
NEG=1SG.NOM beat-UV=3.GEN
'I was not beaten by them/her/him.' (Yen 2012, p. 108)

f. *mai pukun-an=**na=iku.***
NEG beat-UV=3.GEN=1SG.NOM
'I was not beaten by them/her/him.' (Yen 2012, p. 108)

The exact structures of PAN2 are difficult to determine. Puyuma behaves similarly to Paiwan (§ 32.2). Tsou, as noted in § 32.4.2, has obligatory auxiliaries, with an apparent rule that allows just one clitic, that of the actor, to follow the auxiliary, while the undergoer has the form of a neutral pronoun, giving the patterns in (27), illustrated in (28).

(27) Tsou clause skeletons

a. AV [*M*-AUXILIARY=PRON.NOM] AV.VERB NEUT
b. UV [UV.AUXILIARY=PRON.GEN] UV.VERB NEUT

(28) a. Tsou (Zeitoun 2005, p. 77)
mo=su eobako a'o.
AV=2SG. NOM beat.AV 1SG
'You beat me.'

b. *o=su* *eobak-a* (*na*) *a'o.*
UV=2SG. GEN beat-UVP (NOM) 1SG
'You beat me.'

Tona Rukai is accusatively aligned and has a set of nominative pronouns that may be suffixed to the verb. Because the history of Rukai structure is contested (§ 32.4.3), it is not clear what it might tell us about PAN2 structure.

32.6 Numeral Morphology

There are two hypotheses regarding the PAN numeral system, and this contest is still to be settled. On the basis of data from the Western Plains, Sagart (2004, p. 415) claims that PAN displayed a quinary system, and that such a system is best represented with Pazeh (and Kaxabu) whereby numerals from 5 to 9 are additive forms, viz. 5+1, 5+2 etc. In Thao, Saisiyat and Favorlang, some of the numerals above 5 are multiplicative. In Thao, Taokas and Favorlang, 9 is substractive, viz. 10–11. It is not impossible that two systems existed, as they do today in parts of northwest Melanesia. It seems that there a quinary system was used for everyday counting tallying on fingers and toes, whilst an extensive decimal system was part of the knowledge of senior men who had the task of counting goods at ceremonial exchanges (Ross, in press).

Even if there was a quinary numeral system in ancient times, most extant Formosan languages exhibit a decimal system. Blust (1998) reconstructs two different sets (set A and set B), reproduced in Table 32.18. Set A is used in serial counting and in the counting of nonhuman referents. The derived set (Set B) consists of reduplication of the first consonant followed by *-a-* (C*a*- reduplication) in reference to humans.

TABLE 32.18 PAN simple and reduplicated numeral forms

Set A	Set B	Gloss
*pija	*pa-pija	'how much, how many?'
*esa/*isa	*a-esa	'one'
*duSa	*da-duSa	'two'
*telu	*ta-telu	'three'
*Sepat	*Sa-Sepat	'four'
*lima	*la-lima	'five'
*enem	*a-enem	'six'

TABLE 32.18 PAN simple and reduplicated numeral forms (*cont.*)

Set A	Set B	Gloss
*pitu	*pa-pitu	'seven'
*walu	*wa-walu	'eight'
*Siwa	*Sa-Siwa	'nine'
*puluq	*pa-puluq	'ten'

AFTER BLUST 1998, P. 31

The reconstruction in Table 32.18 is not without problems. At least two issues arise. The first is the paucity of inherited reflexes (as opposed to borrowings) of *puluq '10' in Formosan languages. The second concerns the occurrence and function of C*a*-reduplication. The first is a lexical matter outside this chapter's scope. The second does concern morphology.

On the one hand, reconstructing PAN *pa-puluq is problematic, since reflexes of this form are not found in Paiwan or Amis. On the other hand, C*a*-reduplication occurs across the board in Katripul Puyuma, for both human and nonhuman counting, while both Bunun and Kanakanavu exhibit suppletive forms to mark human referents.

TABLE 32.19 Occurrence of C*a* -reduplication for the numerals 'one', 'two', and 'three' in Atayal, Thao, Puyuma, Kanakanavu, and Bunun

	Atayal	Thao	Puyuma	Bunun	Kanakanavu	
'one'	-hum	*qutux*	*tata*	*sa-sa[y]-a*	*tasa*	*u-cani*
	+hum			*sa-sa*	*tatini*	*tacini*
'two'	-hum	*'usaying*	*tusha*	*za-zua[y]-a*	*dusa*	*u-rucin(i)*
	+hum	*ra-rusa'*	*ta-tusha*	*za-zua*	*da-dusa*	*tasusa*
'three'	-hum	*tugal*	*turu*	*ta-telru[w]-a*	*tau*	*u-tulu*
	+hum	*ta-tuu'*	*ta-turu*	*ta-telru*	*ta-tau*	*ta-tulu*

It is thus questionable whether C*a*-reduplication should be understood as referring to human participants, or whether its core meaning is much broader, encompassing 'plurality', a notion often associated with that of 'humanness' in Austronesian languages.

Over against the rare occurrence of Formosan reflexes of PAN *puluq, numerals from 20 to 90 in most Formosan languages are made up of the reflexes of the complex affix PAN *ma-...-N (< *ma- 'multiple of tens'; *-N 'recurrence') (Table 32.20). Note that the bound form *|puSa| '2', which is always followed by *-N, must also be reconstructed as the counterpart of *duSa 'two' (Zeitoun, Teng & Ferrell 2010).

TABLE 32.20 The numerals 1–9 and 10–90 in Isbukun Bunun

	Isbukun Bunun			PAN	
1	*tasa'*	10	*ma-s'a-n*	*esa/*isa	*ma-sa-N
2	*dusa'*	20	*ma-pusa-n*	*duSa	*ma-puSa-N
3	*tau*	30	*ma-tiu-n*	*telu	*ma-telu-N
4	*paat*	40	*ma-sipat-un*	*Sepat	*ma-Sepat-eN
5	*'ima'*	50	*ma-ima-un*	*lima	*ma-lima-N
6	*nuum*	60	*ma-num-un*	*enem	*ma-enem-eN
7	*pitu'*	70	*ma-pitu-un*	*pitu	*ma-pitu-N
8	*vau'*	80	*ma-vau'-un*	*walu	*ma-walu-N
9	*siva'*	90	*ma-siva-un*	*Siwa	*ma-Siwa-N

AFTER LI 1997, PP. 551–554 AND ZEITOUN, TENG & FERRELL 2010

Two other PAN affixes that occur on numeral roots are *paka- 'frequentative or multiplicative' and *Sika- 'ordinal' (Blust 2013, pp. 291–292). Blust (2013, p. 292) mentions that "frequentative or multiplicative numerals are formed with a reflex of the causative prefix *paka-: Chamorro *faha-unum* 'six times', Arosi *ha'a-hai* 'four times', Fijian *vaka-ono* 'six times', Rennellese *haka-ono* 'do six times'." There are reasons to believe, though, that *paka- is actually a bimorphemic prefix composed of two directional prefixes *pa- 'go (MVT)' and *ka- 'to (GOAL)' (Zeitoun 2018). In Formosan languages, the reflexes of *pa- and *ka- co-occur with different prefixes, cf. for instance *pa-sa-* 'towards' in Kuljaljaw Paiwan, Central Amis, and Kavalan, e.g., Kuljaljaw Paiwan *pa-sa-timur* 'towards Timur' < *timur*, *pa-sa-inu* 'to(wards) where?' (< *inu* 'where?'), Central Amis *pa-sa-fafaw* 'go up' (< *fafaw* 'up'), *pa-sa-'amis* 'go north' (< *'amis* 'north') (M. Wu 2013, p. 105), Kavalan *pa-sa-babaw* 'raise, throw up' (< *babaw* 'upper part, up') (Li & Tsuchida 2006, p. 231), *pa-sh-* (~ *ma-sh-*) 'toward (AV)' in Thao, e.g., *pa-sh-du* (~ *ma-sh-du-du*) 'pass sth. along to s.o. (AV)' < *du* 'right, good' (Blust 2003b, pp. 125, 152), *pay-* (< PAN *pa-i-) in Saisiyat, e.g., *payshiri:* 'to go from one place to another, travel', *ka-l-* 'go by, pass through, walk', *ka-sh-* 'walk,

go by foot, step on' in Saisiyat, e.g., *ka-l-'oepaeh* 'go by in vain' (< |*'oepaeh*| 'be empty'), *ka-sh-masak* 'walk barefoot' (< |*masak*| 'without anything') (Zeitoun et al. 2015).

Blust (2013, p. 291) mentions that PAN ordinal numerals were derived by prefixing *Sika- to the base and that this process is well preserved in many daughter languages, e.g., Paiwan *sika-tjelu* 'third' (< *tjelu* 'three'), Tagalog *ika-ápat* 'fourth' (< *ápat* 'four'), and Malay *ke-lima* 'fifth' (< *lima* 'five'). It seems that another prefix should be reconstructed, cf. *Saka-, found, for instance in Kavalan and in Rukai, e.g., Kavalan *sa-qa-u-zusa* 'second' (whereby PAN *k > Kav *q/_a*, Li & Tsuchida 2006, p. 20), Mantauran Rukai *'a-ka-dho'a* 'second' (whereby PAN *S- > Ferrell Mt Rukai '; Zeitoun 2007, p. 268). PAN *Sika- and *Saka- might be bimorphemic prefixes, though at this moment, there is no certainty about the meaning of *Si-, *Sa-, and *ka-. However, it is possible that *ka- forms the Kstem of stative verbs from numerals (§ 32.3.1) and that *Si- and *Sa- are circumstantial-voice prefixes (Table 32.6).

32.7 Conclusion

Even if one rejects Ross's NucAN hypothesis or Aldridge's ergative Austronesian hypothesis, it is difficult to avoid the fact that Tsou, Rukai, and Puyuma are morphologically different from other Formosan languages, and that the latter are morphologically more similar to one another than Tsou, Rukai, and Puyuma. This fact requires a historical explanation. It also requires explaining that Puyuma is more similar in its morphological structure to NucAN languages than Tsou and Rukai are. These facts are almost invisible if one engages in historical Formosan linguistics solely on the basis of lexicon and phonology.

These observations demand further investigation, the more so as the languages concerned form a geographic region of the greatest morphological variety within Taiwan, reminiscent of Sapir's (1916) famous advice that a language family's area of the greatest diversity might well be its homeland.

Acknowledgments

We thank two anonymous referees for their insightful comments.

References

Adelaar, Alexander (2011). *Siraya: Retrieving the phonology, grammar and lexicon of a dormant Formosan language*. Berlin and Boston: Gruyter De Mouton.

Aldridge, Edith (2015). A Minimalist approach to the emergence of ergativity in Austronesian languages. *Linguistics Vanguard* 1(1), 313–326.

Aldridge, Edith (2016). Ergativity from subjunctive in Austronesian languages. *Language and Linguistics* 17(1), 27–62.

Aldridge, Edith (2021). Syntactic conditions on accusative to ergative alignment change in Austronesian languages. *Journal of Historical Linguistics* 11(2), 214–247.

Blust, Robert (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 (1997). Rukai stress revisited. *Oceanic Linguistics* 36(2), 398–403.

Blust, Robert (1998). *Ca*-reduplication and Proto-Austronesian grammar. *Oceanic Linguistics* 37(1), 29–64.

Blust, Robert (1999a). Subgrouping, circularity and extinction: Some issues in Austronesian comparative linguistics. In Elizabeth Zeitoun & Paul Jen-kuei Li (Eds.), *Selected papers from the Eighth International Conference on Austronesian Linguistics* (pp. 31–94). Symposium Series of the Institute of Linguistics (Preparatory Office), No. 1. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.

Blust, Robert (1999b). Notes on Pazeh phonology and morphology. *Oceanic Linguistics* 38(2), 321–365.

Blust, Robert A. (2003a). Three notes on early Austronesian morphology. *Oceanic Linguistics* 42(2), 438–478.

Blust, Robert A. (2003b). *Thao dictionary*. Language and Linguistics Monograph Series A5. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.

Blust, Robert A. (2013). *The Austronesian languages*. Revised Edition. Asia-Pacific Linguistics Open Access Monographs A-PL 008. Canberra: The Australian National University. [https://openresearch-repository.anu.edu.au/bitstream/1885/10191/6/Blust-2013-AustronesianLanguages.pdf]

Blust, Robert (2014). Some recent proposals concerning the classification of the Austronesian languages. *Oceanic Linguistics* 53(2), 300–391.

Blust, Robert (2015). The case markers of Proto-Austronesian. *Oceanic Linguistics* 54(2), 436–491.

Blust, Robert and Victoria Y. Chen (2017). The pitfalls of negative evidence: 'Nuclear Austronesian', 'Ergative Austronesian', and their progeny. *Language and Linguistics* 18(4), 577–621.

Chang, Anna Hsiou-chuan (2006). A reference grammar of Paiwan. PhD dissertation. Canberra: The Australian National University.

Chang, Henry Yung-li (1997). Voice, case and agreement in Seediq and Kavalan. PhD dissertation. Hsinchu: National Tsing Hua University.

Dahl, Otto Christian (1976). *Proto-Austronesian*. Scandinavian Institute of Asian Studies Monograph Series No. 15. London: Curzon Press. (2nd ed.)

Dempwolff, Otto (1937). *Vergleichende Lautlehre des Austronesischen Wortschatzes, 2: Deduktive Anwendung des Urindonesischen auf austronesische Einzelsprachen* [*Comparative phonology of the Malayo-Polynesian lexicon, vol. 2: Deductive application of Proto-Malayo-Polynesian to individual Austronesian languages.*] Beihefte zur Zeitschrift für Eingeborenen-Sprachen 17. Berlin: Dietrich Reimer.

Dempwolff, Otto (1938). *Vergleichende Lautlehre des Austronesischen Wortschatzes, 3: Austronesisches Wörterverzeichnis* [*Comparative phonology of the Malayo-Polynesian lexicon, vol. 3: Malayo-Polynesian dictionary*]. Beihefte zur Zeitschrift für Eingeborenen-Sprachen 19. Berlin: Dietrich Reimer.

Ferrell, Raleigh (1969). *Taiwan aboriginal groups: Problems in cultural and linguistic classification*. Institute of Ethnology, Academia Sinica Monograph, No. 17. Taipei: Institute of Ethnology, Academia Sinica.

Foley, William A. (2005). Semantic parameters and the unaccusative split in the Austronesian language family. *Studies in Language* 29(2), 385–430.

Himmelmann, Nikolaus (2005). The Austronesian languages of Asia and Madagascar: Typological characteristics. In Alexander Adelaar and Nikolaus P. Himmelmann (Eds), *The Austronesian languages of Asia and Madagascar*, (pp. 110–181). London: Routledge.

Holmer, Arthur J. (1996). *A parametric grammar of Seediq*. Travaux de L'Institut de Linguistique de Lund 30. Lund: Lund University Press.

Holmer, Arthur and Loren Billings (2014). Clitic pronouns in Seediq. In I Wayan Arka and N.L.K. Mas Indrawati (Eds.), *Argument realisations and related constructions in Austronesian languages: Papers from 12-ICAL*, vol. 2 (pp. 131–141). Asia-Pacific Linguistics 13 / Studies on Austronesian Languages 002. Canberra: ANU Press.

Huang, Wei-chen (a.k.a. Ng, I-sin) (2012). A study of verbal morphology in Puljetji Paiwan. MA thesis. Hsinchu: National Tsing Hua University.

Jiang, Haowen (2016). Nominalization and possession in Formosan languages. PhD dissertation. Houston, TX: Rice University.

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

Li, Paul Jen-kuei (1973). *Rukai structure*. Institute of History and Philology Special Publication 64. Taipei: Institute of History and Philology, Academia Sinica.

Li, Paul Jen-kuei 李壬癸 (Ed.) (1997). *Gaoxiongxian nandao yuyan* 高雄縣南島語言 [*The Formosan languages of Kaohsiung*], (pp. 272–297). Gaoxiongxian wenxian congshu xilie 7 高雄縣文獻叢書系列 7 [Kaohsiung County Documents Series 7] Fengshan 鳳山市 [Fengshan]: Gaoxiongxian zhengfu 高雄縣政府 [Kaohsiung County Government].

Li, Paul Jen-kuei and Shigeru Tsuchida (2006). *Kavalan dictionary*. Language and Linguistics Monograph Series A19. Taipei: Institute of Linguistics, Academia Sinica.

Li, Paul Jen-kuei and Shigeru Tsuchida (2009). Yet more PAn infixes. In Bethwyn Evans (Ed.) *Discovering history through languages: Papers in honour of Malcolm Ross*, (pp. 345–362). Canberra: The Australian National University.

Lichtenberk, Frantisek (2000). Reciprocal without reflexives. In Zygmunt Frajzyngier (Ed.), *Reciprocals: Forms and functions*, (pp. 31–62). Amsterdam and Philadelphia, PA: John Benjamins.

Lim, Hong-sui 林鴻瑞 (2022). Gehawuyu cankao yufa: yige jidu binwei de taiwan nandao yu 噶哈巫語參考語法：一個極度瀕危的臺灣南島語 [A grammar of Kaxabu, a moribund Formosan language]. Boshi lunwen 博士論文 [PhD dissertation]. Nantouxian pulizhen 南投縣埔里鎮 [Puli, Nantou]: Guoli jinan guoji daxue 國立暨南國際大學 [National Chi Nan University].

Lim, Hong-sui and Elizabeth Zeitoun (2023). A note on morphological changes in Kaxabu. *Oceanic Linguistics* 62(1), 184–212.

Pan, Chia-jung (2012). A grammar of Lha'alua, an Austronesian language of Taiwan. PhD dissertation. Cairns: James Cook University.

Pawley, Andrew K. (1973). Some problems in Proto-Oceanic grammar. *Oceanic Linguistics* 12(1/2), 103–188.

Pawley, Andrew and Lawrence Reid (1980). The evolution of transitive constructions in Austronesian. In Paz Buenaventura Naylor (Ed.), *Austronesian studies: Papers from the Second Eastern Conference on Austronesian Languages*, (pp. 103–130). Ann Arbor: Center for South and Southeast Asian Studies, University of Michigan.

Pecoraro, Ferdinand (1977). *Essai de dictionnaire taroko-français* [*A preliminary Taroko-French dictionary*]. Cahier d'Archipel 7. Paris: Société pour l'Etude et la Connaissance du Monde Insulindien [Society for the Study and knowledge of the Insulindian World].

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

Reid, Lawrence (1978). Problems in the reconstruction of Proto-Philippine construction markers. In S.A. Wurm and Lois Carrington (Eds.), *Second International Conference on Austronesian Linguistics: Proceedings, Fascicle 1*, (pp. 33–66). Pacific Linguistics C-61. Canberra: Australian National University.

Ross, Malcolm (1995). Reconstructing Proto-Austronesian verbal morphology: Evidence from Taiwan. In Paul Jen-kuei Li, Cheng-hwa Tsang, Ying-kuei Huang, Dah-an Ho and Chiu-yu Tseng (Eds.), *Austronesian Studies Relating to Taiwan* (pp. 727–791). Symposium Series of the Institute of History and Philology No. 3. Taipei: Institute of History and Philology, Academia Sinica.

Ross, Malcolm (2002). The history and transitivity of Western Austronesian voice and voice-marking. In Fay Wouk and Malcolm Ross (Eds.), *The history and typology of Western Austronesian voice systems*, (pp. 17–62). Pacific Linguistics 518. Canberra: Australian National University.

Ross, Malcolm (2006). Reconstructing the case-marking and personal pronoun systems of Proto Austronesian. In Yung-li H. Chang, Lillian M. Huang and Dah-an Ho (Eds.), *Streams converging into an ocean: Festschrift in honor of Professor Paul Jen-kuei Li on his 70th birthday*, (pp. 521–563). Language and Linguistics Monograph Series W-5. Taipei: Institute of Linguistics, Academia Sinica.

Ross, Malcolm (2009). Proto Austronesian verbal morphology: A reappraisal. In Alexander Adelaar and Andrew Pawley (Eds.), *Austronesian historical linguistics and culture history: A festschrift for Robert Blust*, (pp. 295–326). Pacific Linguistics 601. Canberra: Australian National University.

Ross, Malcolm (2012). In defense of Nuclear Austronesian (and against Tsouic). *Language and Linguistics* 13(6), 1253–1330.

Ross, Malcolm (2015a). Revising the reconstruction of early Austronesian personal pronouns. In Malcolm Ross and Wayan Arka (Eds.), *Language Change in Austronesian Languages: Papers from 12-ICAL*, vol. 3, (pp. 109–122). Asia-Pacific Linguistics 018 / Studies on Austronesian Languages 004. Canberra: Australian National University.

Ross, Malcolm (2015b). The argument indexing of early Austronesian verbs: A reconstructional myth? In Dag T.T. Haug, with the assistance of Eiríkur Kristjánsson (Ed.), *Historical linguistics 213: Selected paper from the 21th International Conference of Historical Linguistics, Oslo, 5–9 August 2013*, (pp. 257–279). Amsterdam and Philadelphia, PA: John Benjamins.

Ross, Malcolm (2015c). Reconstructing Proto Austronesian verb classes. *Language and Linguistics* 13(3), 279–315.

Ross, Malcolm (in press). Counting: numerals and numeral classifiers. In Malcolm Ross, Andrew Pawley and Meredith Osmond (Eds.), *The lexicon of Proto Oceanic*, vol. 6: Society. Canberra: Asia-Pacific Linguistics.

Sagart, Laurent (2004). The higher phylogeny of Austronesian and the position of Tai-Kadai. *Oceanic Linguistics* 43(2), 411–444.

Sagart, Laurent (2005). Sino-Tibetan–Austronesian: An updated and improved argument. In Laurent Sagart, Roger Blench and Alicia Sanchez-Mazas (Eds.), *The peopling of East Asia: Putting together archaeology, linguistics and genetics*, (pp. 161–176). London: RoutledgeCurzon.

Sagart, Laurent (2010). Is Puyuma a primary branch of Austronesian? *Oceanic Linguistics* 49(1), 194–204.

Sagart, Laurent (2014). In defense of the numeral-based model of Austronesian phylogeny, and of Tsouic. *Language and Linguistics* 15(6), 859–882.

Sapir, Edward (1916). *Time perspective in aboriginal American culture: a study in method.* Department of Mines, Geological Survey, Anthropological Series No. 13, Memoir 90. Ottawa, Canada: Government Printing Bureau. Reprinted in Regna Darnell and Judith Irvine (Eds.) 1994. *The collected works of Edward Sapir*, vol. 4: *Ethnology*, (pp. 30–119). Berlin: Mouton de Gruyter.

Shih, Winslow Chia-Lin (2012). Are there case markers in Budai Rukai? MA thesis. Taipei: National Taiwan University.

Starosta, Stanley (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*, (pp. 683–726). Symposium Series of the Institute of History and Philology, Academia Sinica 4. Taipei: Institute of History and Philology, Academia Sinica.

Starosta, Stanley, Andrew Pawley and Lawrence Reid (1981). The evolution of focus in Austronesian. Honolulu: Department of Linguistics, University of Hawai'i at Mānoa. Republished in Elizabeth Zeitoun (Ed.) 2009. *Formosan linguistics: Stanley Starosta's contributions*, (pp. 329–481). Language and Linguistics Monograph Series. Taipei: Institute of Linguistics, Academia Sinica.

Starosta, Stanley, Andrew Pawley and Lawrence Reid (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*, vol. 2: *Tracking the travellers*, (pp. 145–170). Pacific Linguistics, C-75. Canberra: Australian National University. Republished in Elizabeth Zeitoun (Ed.) 2009. *Formosan linguistics: Stanley Starosta's contributions*, (pp. 297–328). Language and Linguistics Monograph Series. Taipei: Institute of Linguistics, Academia Sinica.

Szakos, József (1994). Die Sprache der Cou: Untersuchungen zur Synchronie einer austronesischen Sprache auf Taiwan. [The language of the Tsou: Synchronic research into an Austronesian language of Taiwan]. PhD thesis. Bonn: Rheinische Friedrich-Wilhelms-Universität zu Bonn.

Tang, Chih-Chen Jane (2006). Case-marking, reference and DP structure: A comparative study of Paiwan, Atayal and Chamorro. In Yung-li Chang, Lillian M. Huang and Dah-an Ho (Eds.), *Streams converging into an ocean: Festschrift in honor of Professor Paul Jen-kuei Li on his 70th birthday*, (pp. 143–183). Language and Linguistics Monograph Series W5. Taipei: Institute of Linguistics, Academia Sinica.

Tang, Chih-chen Jane, Yung-li Chang, Dah-an Ho (1997). On noun phrase structures in Paiwan. *Tsing Hua Journal of Chinese Studies, New Series* 27(2), 230–256.

Teng, Stacy Fang-ching (2008). *A reference grammar of Puyuma, an Austronesian language of Taiwan*. Pacific Linguistics PL 595. Canberra: The Australian National University.

Teng, Stacy Fang-ching (2014). Grammaticalization of predicative possession in Nanwang Puyuma and as a basis for reconstruction in PAN. *Oceanic Linguistics* 53(1), 136–154.

Teng, Stacy Fang-ching (2015). A comparative study of the personal pronominal systems in three Puyuma dialects. In Elizabeth Zeitoun, Stacy F. Teng and Joy J. Wu (Eds.), *New Advances in Formosan Linguistics*, (pp. 407–429). Asia-Pacific Linguistics 17 / Studies in Austronesian Linguistics 003. Canberra: ANU Press.

Teng, Stacy Fang-ching (2018). A reconstruction of the Proto-Puyuma aspectual and modal system. *Oceanic Linguistics* 57(2), 303–334.

Teng, Stacy Fang-ching and Elizabeth Zeitoun (2016). The noun-verb distinction in Kanakanavu and Saaroa: Evidence from pronouns. *Oceanic Linguistics* 55, 134–161.

Tsuchida, Shigeru 土田滋 (1980). Puyuma-go (Tamarakao hōgen) goi: fu gohōgaisetsu oyobi tekisuto. プユマ語(タマラカオ方言) 語彙 [Puyuma (Tamalakaw dialect) vocabulary: with grammatical notes and texts]. In Kuroshio Bunka no Kai 黒潮文化の会編 [Black Current Cultures Committee] (Ed.), Kuroshio no minzoku, bunka, gengo 黒潮の民族文化言語 [Ethnology, cultures and languages along the Black Current], (pp. 183–307). Tokyo: Kadokawa Shoten

Wolff, John U. (1973). Verbal inflection in Proto-Austronesian. In A.B. Gonzalez (Ed.), *Parangal kay Cecilio Lopez: Essay in honor of Cecilio Lopez on his seventy-fifth birthday*, (pp. 71–91). Philippine Journal of Linguistics Special Monograph Issue 4. Quezon city: Linguistic Society of the Philippines.

Yen, Doris Ching-jung (2012). Clitics in Kavalan (East Formosan, Austronesian). MA thesis. Puli, Nantou: National Chi Nan University.

Zeitoun, Elizabeth (2002). Reciprocals in Formosan languages. Paper presented at the *ICAL 9*, January 8–11. Canberra: The National Australian University.

Zeitoun, Elizabeth (2005) Tsou. In Alexander and Nikolaus P. Himmelmann (Eds.), *The Austronesian languages of Asia and Madagascar*, (pp. 259–290). London: Routledge.

Zeitoun, Elizabeth (2007). *A grammar of Mantauran (Rukai)*. Language and Linguistics Monograph Series A4–2. Taipei: Institute of Linguistics, Academia Sinica.

Zeitoun, Elizabeth (2009). Reassessing the reconstruction of plural affixes in PAn: Evidence from the Formosan languages. In Alexander Adelaar and Andrew Pawley (Eds.), *Austronesian historical linguistics and culture history: A festschrift for Robert Blust*, (pp. 348–361). Pacific Linguistics 601. Canberra: The Australian National University.

Zeitoun, Elizabeth (2018). From space to time: (Re-)examining the grammaticalization path of some spatial prefixes in Formosan languages. Invited talk at the First ILAS Annual Linguistics Forum. April 20, 2018. Academia Sinica, Taiwan.

Zeitoun, Elizabeth, Tai-hwa Chu and Lalo a Tahesh Kaybaybaw (2015). *A study of Saisiyat morphology*. Oceanic Linguistics Special Publication No. 40. Honolulu, HI: University of Hawai'i Press.

Zeitoun, Elizabeth and Lillian M. Huang (2000). Concerning *ka-*, an overlooked marker of verbal derivation in the Formosan languages. *Oceanic Linguistics* 39(2), 391–414.

Zeitoun, Elizabeth and Lillian M. Huang (2023). An overview of the case-marking and personal pronoun systems of Formosan languages. In Paul Jen-kuei Li, Elizabeth Zeitoun and Rik De Busser (Eds.), *Handbook of Formosan languages: The indigenous languages of Taiwan*. Leiden: Brill.

Zeitoun, Elizabeth and Stacy Fang-ching Teng (2009). From *ki*-N 'get N' in Formosan languages to *ki*-V 'get' V-ed' (passive) in Rukai, Paiwan and Puyuma. In Bethwyn Evans (Ed.), *Discovering history through language: Papers in honour of Malcolm D. Ross*, (pp. 479–500). Pacific Linguistics 605. Canberra: The Australian National University.

Zeitoun, Elizabeth and Stacy Fang-ching Teng (2016). Reassessing the position of Kanakanavu and Saaroa among the Formosan languages. *Oceanic Linguistics* 55(1), 162–198.

Zeitoun, Elizabeth, Stacy Fang-ching Teng and Raleigh Ferrell (2010). Reconstruction of '2' in PAN and related issues. *Language and Linguistics* 11(4), 853–884.

Zorc, R. David (1993). Overview of Austronesian and Philippine accent patterns. In Jerold A. Edmondson and Kenneth J. Gregerson (Eds.), *Tonality in Austronesian languages*, (pp. 17–24). Oceanic Linguistics Special Publications no. 24. Honolulu: University of Hawai'i Press.

Online Resources and Open Access Data

Blust, Robert and Stephen Trussel (2020). *Austronesian comparative dictionary*. Retrieved from www.trussel2.com/ACD.