

# Reconstructing Proto Austronesian Verb Classes\*

Malcolm Ross

*The Australian National University*

Language and Linguistics  
16(3) 279–345  
© The Author(s) 2015  
Reprints and permissions:  
sagepub.co.uk/journalsPermissions.nav  
DOI: 10.1177/1606822X15569162  
lin.sagepub.com



Although morphological innovations are usually regarded as important in reconstructing the histories of language families, relatively little attention has been paid to the reconstruction of Proto Austronesian (PAn) verbal morphology, and this paper aims to take a step in this direction. Morphological classes of verbs have been reconstructed for several Formosan languages, and this paper offers a preliminary reconstruction of PAn verb classes, examines the challenges inherent in that reconstruction and offers tentative responses to them. The final section takes a brief look at how the reconstruction of PAn verb classes permits the establishment of morphological innovations and casts light on the subgrouping of Austronesian languages.

Key words: Austronesian, Proto-Austronesian, verbal morphology, reconstruction, Formosan languages, morphological classes

## 1. Introduction

Why reconstruct the morphological classes of Proto Austronesian (PAn) verbs? Because their reconstruction is an essential part of reconstructing PAn morphology in general, and this larger project is part of the task of determining the history and subgrouping of early Austronesian. Without an understanding of PAn morphological systems we cannot ascertain the morphological innovations that may have occurred in Formosan (F) languages and Proto Malayo-Polynesian.

PAn verbs probably fell into six morphological classes. Two of these carried a much higher functional load than the others. One mainly encoded agentive verbs (§6.2), the other inchoative and state verbs (§6.4), but there was a semantic cline running from agentive activities to states. Between the two extremes, representatives of these and two other classes occurred, and principled reasons can be given for this distribution (§6.8).

In F languages the forms taken by the mood- and aspect-marked variants of the voices of a given verb and by nominalizations, are fully predictable if one knows which morphological class the verb belonged to. However, descriptions of F languages sometimes do not identify the language's

---

\* I am indebted to Anna Hsiou-chuan Chang, Henry Yungli Chang, Stacy F. Teng and Elizabeth Zeitoun for providing me with data on verb classes from various Formosan languages and to Wei-chen Huang and József Szakos for corrections to Paiwan and Tsou data, respectively. Much of the research on which this paper is based was conducted during periods of research at the Institute of Linguistics of Academia Sinica (ILAS), Taipei, from September to December 2003 and from September 2008 to January 2009. During the earlier period I was a Visiting Research Fellow under the project 'A morpho-syntactic study of Rukai' (Principal Investigator: Elizabeth Zeitoun), supported by National Science Council (NSC) of Taiwan grant NSC 92-2811-H-001-010. I am grateful to Elizabeth Zeitoun, Dah-an Ho and Jackson Sun for facilitating these visits and to the NSC and ILAS for their support. I am also very grateful to an anonymous reviewer who commented on the paper in detail.

morphological verb classes. Instead, there is a widespread tendency to give long lists of affixes and their functions. These can be rather confusing, as two or more affixes often appear to have the same function, and one affix may have several functions. For example, reflexes of PAN *\*-an* and *\*ka-...-an* occur among other things as markers of location nominalizations, for example *\*kaen-an* ‘eating place’ (< *\*kaen* ‘eat’), *\*ka-zawiL-an* ‘far away place’ (< *\*zawiL* ‘be far away’). The reflex of *\*ka-...-an* is sometimes listed as a circumfix. But PAN *\*ka-...-an* consists of affixes from two cross-cutting morphological paradigms. PAN *\*-an* certainly formed nominalizations, many of them locative, but *\*ka-* was part of one of the stems of two verb classes and its occurrence in the ‘circumfix’ *\*ka-...-an* simply reflects the fact that a nominalization was formed by suffixing *\*-an* to that stem (Zeitoun & Huang 2000:407–408).

The study reported here is preliminary. To perform an adequate reconstruction of PAN verb classes requires a good analysis of the verb classes of each F language, together with extensive listings of the members of each class in each language. The recognition of morphologically defined verb classes in F languages is due to Shigeru Tsuchida, who analysed the verb classes of Tsou, Saaroa and Kanakanavu in his 1973 doctoral thesis (published as Tsuchida 1976) and went on to analyse the verb classes of Tamalakaw Puyuma (Tsuchida 1980) and Siraya (Tsuchida 2000). The two latter were adopted and extended by Teng (2008:120–123) for Nanwang Puyuma and by Adelaar (2011:101–104) for Siraya. Tsukida (2005:315, 2008, 2009) provides a similar analysis of Teruku Seediq verb classes. Huang’s (2000) analysis of Mayrinax Atayal focuses on form–meaning correspondences, and achieves a classification similar to these. Other analyses of morphological verb classes are found in Zeitoun (2000a, 2007) for Mantauran Rukai, Zeitoun (2003) for Rukai dialects, Zeitoun et al. (2015) for Saisiyat, and Huang (2012) for Paiwan.

For other languages I have worked out the verb classes from the sources available to me, assigning verbs to classes on the basis of sometimes limited data, as noted in the subsections of §5, and future work will doubtless find errors.

In a number of F languages some verbs have derived stems, consisting either of a derivational prefix and a root or of more than one root. However, these are in especial need of further research, and this paper focuses on verbs formed from a monomorphemic root.

My strategy is first to describe the morphological verb classes of one F language, Puyuma, in §2 and §3, as an example of the analysis of F verb classes. Section 4 briefly describes my reconstruction strategy. This entails giving brief accounts in §5 of verb classes in other F languages, leading in §6 to constructionally and lexically-based reconstructions of PAN verb classes. Section 7 looks briefly at Malayo-Polynesian verb classes to show how an understanding of PAN verb classes may elucidate early Austronesian linguistic history.

The PAN orthography employed here is that used in most recent work (other than by John Wolff), with the additions proposed by Ross (2012a).

## 2. The morphological classes of Puyuma verbs

Puyuma provides a convenient example of F verb classes, as its classes have been described by Tsuchida (1980) and by Teng (2008:120–123).

**Table 1:** Forms of Puyuma *dirus/d<em>irus* ‘bathe, wash’ (after Teng 2008:112, 217)

	ACTOR VOICE (intransitive)	UNDERGOER VOICE (transitive)		
		Patient subject	Location subject	Circumstance subject
Realis	<i>d&lt;em&gt;irus</i>	<i>dirus-aw</i>	<i>dirus-ay</i>	<i>dirus-anay</i>
Optative/hortative	<i>d&lt;em&gt;irus-a</i>			
Realis imperfective	<b><i>d&lt;em&gt;a-dirus</i></b>	<b><i>da-dirus-aw</i></b>	<b><i>da-dirus-ay</i></b>	<b><i>da-dirus-anay</i></b>
Realis durative	<b><i>d&lt;em&gt;a-da-dirus</i></b>	<b><i>da-da-dirus-aw</i></b>	<b><i>da-da-dirus-ay</i></b>	<b><i>da-da-dirus-anay</i></b>
Imperative	<i>dirus</i>	<i>dirus-u</i>	<i>dirus-i</i>	<i>dirus-an</i>
Negative	<i>d&lt;em&gt;irus</i>	<i>dirus-i</i>		
Irrealis	<b><i>da-dirus</i></b>	<b><i>da-dirus-i</i></b>		<b><i>da-dirus-an</i></b>

## 2.1 Voice, mood and aspect morphology

Table 1 shows the voice, mood and aspect morphology of the Puyuma verb *dirus/d<em>irus* ‘bathe, wash’. Five boxes in the actor voice column<sup>1</sup> (shaded) include the base *d<em>irus*, formed with infix <em>, whilst all the other boxes include the base *dirus*. Forms in the realis imperfective, realis durative and irrealis rows contain *Ca-* reduplication<sup>2</sup> (bolded).

Table 2 shows the corresponding forms for the verb *ka-qeki/ma-qeki* ‘scold, condemn’. In slots where Table 1 has *d<em>irus*, Table 2 has *ma-qeki*. In slots where Table 1 has *dirus*, Table 2 has *ka-qeki*. Suffixes and *Ca-*reduplication are the same.

Thus the morphological organization of Table 2 matches that of Table 1. All Puyuma verbs display this organization. The verbs *dirus/d<em>irus* and *ka-qeki/ma-qeki* represent two of Puyuma’s six verb classes. What varies from class to class are the two bases (in the shaded and unshaded boxes). These are not predictable, but are lexically determined, and need to be shown for each verb in a Puyuma dictionary.<sup>3</sup>

Table 3 shows this organization in schematic form. I call the base in the shaded boxes ‘Mstem’, the base in the other boxes ‘Kstem’. The Kstem consists minimally of a root (like *dirus*), but sometimes of a root with *ka-* (like *ka-qeki*). Most of the affixes in Tables 1–3 have an invariable

<sup>1</sup> Because Puyuma actor voice forms are always intransitive and undergoer voice forms always transitive, Teng (2008) labels them ‘intransitive’ and ‘transitive’ respectively. I use the labels ‘actor voice’ and ‘undergoer voice’ here to facilitate comparison with other F languages.

<sup>2</sup> That is, a syllable consisting of the first consonant of the stem, plus the vowel *a*.

<sup>3</sup> Among F lexicons, only Adelaar’s (2011:296–400) lexicon of Siraya systematically gives the Kstem and Mstem of each verb (Kstem as headword, followed by Mstem). Nonetheless, several dictionaries provide information that allows one to identify the principal parts. Egli’s (2002) dictionary of Paiwan, Blust’s (2003) of Thao, Li & Tsuchida’s (2001) of Pazih, Egerod’s (1999) of Atayal, and Li & Tsuchida’s (2006) of Kavalan all provide under the headword a variety of forms and some examples. Ferrell’s (1982) dictionary of Paiwan provides a list of derived forms under each headword, and Szakos’ (1994, Vol. 3) dictionary of Tsou cross-references related derived forms. Less helpful are Fey’s (1986) dictionary of Amis with restricted examples and Cauquelin’s (1991) of Puyuma with no information except the headword.

**Table 2:** Forms of Puyuma *ka-deki/ma-deki* ‘scold, condemn’

	ACTOR VOICE (intransitive)	UNDERGOER VOICE (transitive)		
		Patient subject	Location subject	Circumstance subject
Realis	<i>ma-ɖeki</i>	<i>ka-ɖeki-aw</i>	<i>ka-ɖeki-ay</i>	<i>ka-ɖeki-anay</i>
Optative/hortative	<i>ma-ɖeki-a</i>			
Realis imperfective	<i>ma-ɖa-ɖeki</i>	<i>ka-ɖa-ɖeki-aw</i>	<i>ka-ɖa-ɖeki-ay</i>	<i>ka-ɖa-ɖeki-anay</i>
Realis durative	<i>ma-ɖa-ɖa-ɖeki</i>	<i>ka-ɖa-ɖa-ɖeki-aw</i>	<i>ka-ɖa-ɖa-ɖeki-ay</i>	<i>ka-ɖa-ɖa-ɖeki-anay</i>
Imperative	<i>ka-ɖeki</i>	<i>ka-ɖeki-u</i>	<i>ka-ɖeki-i</i>	<i>ka-ɖeki-an</i>
Negative	<i>ma-ɖeki</i>	<i>ka-ɖeki-i</i>		
Irrealis	<i>ka-ɖa-ɖeki</i>	<i>ka-ɖa-ɖeki-i</i>		<i>ka-ɖa-ɖeki-an</i>

**Table 3:** Puyuma voice, mood and aspect morphology

	ACTOR VOICE (intransitive)	UNDERGOER VOICE (transitive)		
		Patient subject	Location subject	Circumstance subject
Realis	Mstem	<i>Kstem-aw</i>	<i>Kstem-ay</i>	<i>Kstem-anay</i>
Optative/hortative	Mstem- <i>a</i>			
Realis imperfective	$\alpha$ -Mstem	$\alpha$ - <i>Kstem-aw</i>	$\alpha$ - <i>Kstem-ay</i>	$\alpha$ - <i>Kstem-anay</i>
Realis durative	$\alpha$ - $\alpha$ -Mstem	$\alpha$ - $\alpha$ - <i>Kstem-aw</i>	$\alpha$ - $\alpha$ - <i>Kstem-ay</i>	$\alpha$ - $\alpha$ - <i>Kstem-anay</i>
Imperative	Kstem	<i>Kstem-u</i>	<i>Kstem-i</i>	<i>Kstem-an</i>
Negative	Mstem	<i>Kstem-i</i>		<i>stem-an</i>
Irrealis	$\alpha$ -Kstem	$\alpha$ - <i>Kstem-i</i>		$\alpha$ - <i>Kstem-an</i>

shape, but two display variation. These are the affix that forms the Mstem ( $\langle em \rangle$  in *d $\langle em \rangle$ irus*, *ma-* in *ma-deki*), and  $\alpha$ -, expounded as *Ca-* reduplication or as infix  $\langle a \rangle$ .

## 2.2 Morphological classes

Table 4 provides an example from each of Puyuma’s six verb classes, showing each verb in three forms:

- (1) a. Kstem: AV imperative;
- b. Mstem: AV realis;
- c.  $\alpha$ -Kstem: AV irrealis.<sup>4</sup>

<sup>4</sup> Abbreviations in glosses follow the Leipzig Glossing Rules (<http://www.eva.mpg.de/lingua/resources/glossing-rules.php>), with the addition of ACAUS anticausative, AV actor voice, CN common, DEP dependent, IRR irrealis, LNK linker, PERS personal, RED reduplication, RLS realis, UV undergoer voice, UVC circumstance-subject UV, UVL location-subject UV, UVP patient-subject UV. Other abbreviations are F Formosan, PAn Proto Austronesian and PMP Proto Malayo-Polynesian. For abbreviations used in cognate sets see fn.26.

**Table 4:** Puyuma verb classes (after Teng 2008:120–123, 165–178, 181–183)

labels ↓	forms of Kstem ↓	labels → exponents of Mstem →	U ⟨em⟩ and its allomorphs	M <i>m-</i> replaces <i>p-</i>	MA <i>ma-</i> and its allomorphs	Z zero
Ø	Ø-ROOT	Kstem Mstem α-Kstem	‘drink’ <i>teke</i> <i>t⟨em⟩eke</i> <i>ta-teke</i>		‘tell’ <i>rejay</i> <i>ma-rejay</i> <i>ra-rejay</i>	‘give’ <i>beray</i> <i>beray</i> <i>ba-beray</i>
P	<i>p</i> -initial ROOT	Kstem Mstem α-Kstem		‘have children’ ( <i>pi-</i> ‘have, wear’) <i>pi-walak</i> <i>mi-walak</i> <i>pi⟨a⟩walak</i>		
K	<i>ka</i> -ROOT	Kstem Mstem α-Kstem			‘scold’ <i>ka-ɖeki</i> <i>ma-ɖeki</i> <i>ka-ɖa-ɖeki</i>	‘be hot’ <i>ka-bias</i> <i>bias</i> <i>ka-ba-bias</i>

In order to focus on form rather than function I prefer to use labels indicating form rather than more conventional labels like ‘finite’ or ‘non-finite’. The Mstem could reasonably be labelled  $\Delta V$  realis, but the Kstem and  $\alpha$ - are purely morphological, that is morphomic (Aronoff 1994), in that there is no morphosyntactic feature shared in common by the morphological structures of which each is a part.

The morphological classes in Table 4 are arranged by two cross-cutting categorizations. Vertically, verbs are categorized by the forms of Kstem, to which the various suffixes in Table 3 and also the nominalising infix ⟨*in*⟩ are added. These forms are labelled as in (2).

- (2) a. Ø: Kstem is unaffixed.  
 b. P: Kstem is *p*-initial (but not all *p*-initial stems belong to this class); in Puyuma all such stems are formed with a derivational prefix.  
 c. K: Kstem consists of the root prefixed with *ka-*, unpredictably sometimes *k-* before a vowel.

The prefix *ka-* is treated separately from derivational prefixes for two reasons. One is formal: with a derivational prefix the exponent of  $\alpha$ - is usually infixed ⟨*a*⟩ (as in (3b); also *pi⟨a⟩walak* in Table 4), but with *ka-* it is reduplication (as in (3a)).

- (3)      ROOT      Kstem      α-Kstem  
 a. *bias*      *ka-bias*      *ka-ba-bias*      ‘be hot’  
 b. *kuraw*      *ki-kuraw*      *ki⟨a⟩kuraw*      ‘catch fish’

Second, each derivational prefix has a (roughly) specifiable meaning, but *ka-* does not. Zeitoun & Huang (2000) show that reflexes of PAN \**ka-* typically occur in statives. Its Puyuma reflex *ka-* still

tends to occur on statives, but not regularly enough to analyse it synchronically as a stative marker (Teng 2008:123–126, 167–168). I suggest in §3.3 that its function was typically to mark verbs construed as non-agentive.

Horizontally verbs are categorized in Table 4 by their exponent of Mstem, labelled as follows:

- (4) **U**: Mstem contains an allomorph of *<em>*.<sup>5</sup>
- (5) **M**: in Mstem root-initial *p-* is replaced by *m-*.
- (6) **MA**: Mstem consists of *ma-* + root.
- (7) **Z**: Mstem is unaffixed.

The exponent *ma-* is invariable, but the exponent *<em>* has four allomorphs: *m-* with a root-initial vowel (e.g. *a[ak* ‘take’ > AV *m-a[ak*), *me-* with root-initial *l*, *l*, *r*, *n* or *ŋ* (*naʔu* ‘see’ > AV *me-naʔu*), *<en>* with root-initial *p* or *b*, and *<em>* otherwise (Teng 2008:25–27, 121).

The two dimensions of Table 4 allow us to classify verbs as, for example, class  $\emptyset/U$ , the class with a zero-prefixed Kstem and the *<em>* exponent of Mstem (e.g. *ʔekel/ʔ<em>ekel* ‘drink’), class  $\emptyset/Z$  (e.g. *beray* ‘give’), and so on.<sup>6</sup> It shows that for any given verb we need to know two forms, Kstem and Mstem, the ‘principal parts’ of the Puyuma verb,<sup>7</sup> neither of which is predictable from the other, except in  $\emptyset/U$  verbs, where the  $\emptyset$  form can be predicted from the U form. From the two principal parts we can generally predict all other potential forms of the verb in Table 3 (‘potential’ because verb paradigms are often defective: they lack some undergoer voice forms).

With a few exceptions the exponent of  $\alpha-$  can be predicted from the two principal parts. Verbs with  $\emptyset$  or K stems take the exponent *Ca-*, inserted before the penultimate syllable of the Kstem, where C- duplicates the onset of that syllable. Thus if the Kstem has more than two syllables, *Ca-* is infixal, as in *a-ra-remɛŋ* ‘become darker’, realis imperfective of *k-a-remɛŋ/are-mɛŋ* ‘dark’ (K/Z; Teng 2008:40, 121) and if it has just two syllables, *Ca-* is prefixed, as the  $\alpha$ -Kstem (AV irrealis) examples in Table 4 show. If the onset of the penultimate syllable is zero, then *Ca-* is realized as *a*, for example *paetɛŋ* ‘be long’ > irrealis *pa<a>etɛŋ*, *atel* ‘drop’ > irrealis *a-atel*, *a[ak/ma-a[ak* ‘take’ > AV realis imperfective *m-a-a[ak* (Teng 2008:38–39, 121). With derived stems the exponent of  $\alpha-$  is usually *<a>*, infixal between the prefix and the Kstem.<sup>8</sup>

<sup>5</sup> The label U chosen as *<um>* is the most widespread form among cognates of Puyuma *<em>*.

<sup>6</sup> Ross (2009) lists the five Puyuma classes in which non-derived stems occur, omitting P/M. The classes were labelled 1–5, corresponding to the labels here as follows: 1 =  $\emptyset/U$ , 2 =  $\emptyset/MA$ , 3 =  $\emptyset/Z$ , 4 = K/MA, 5 = K/Z.

<sup>7</sup> The term ‘principal parts’, used in traditional grammars of Latin and Greek, was (re)introduced into morphological theory by Blevins (2006) and Finkel & Stump (2007).

<sup>8</sup> There are exceptions: whilst *k<em>i-babayan* ‘have intercourse’ manifests  $\alpha-$  as *<a>* in *ki<a>babayan*, *k<em>i-qayar* ‘talk to’ has the *Ca-* exponent in *k<em>i-qa-qayar* (Stacy Teng, personal communication 2014). A more formal statement of Puyuma verbal morphology using default inheritance (like, say, Corbett & Fraser’s 1993 Network Morphology) would need to incorporate low-order constraints to account for verbs with exceptional forms.

As Teng (2008:56) points out, verbs of different classes can be derived from a single root. Thus *ka-bias/bias* ‘be hot’ is a stative intransitive K/Z verb and *bias/b<en>ias* ‘make hot’ a  $\emptyset$ /U verb derived from the root *bias*.

Table 4 is laid out as it is for the sake of clarity, but six boxes are greyed out because they do not occur in Puyuma, nor in most other F languages. Five belong to the row and column accommodating the P/M class. The sixth is located where K/U might be expected. Since PAn \**ka-* marked non-agentivity (§3.3), whilst PAn \**<um>* marked its verb as agentive, the two were semantically incompatible and did not co-occur.

Class P/M requires a row and column to itself in Table 4 because its Kstem is phonologically defined: it is *p*-initial. As there is no prefix, it is a  $\emptyset$  stem, but P/M verbs form a separate morphological class because they form the Mstem by *p*-replacement. Moreover, not all Puyuma *p*-initial stems occur in P/M verbs: the root/Kstem *pilaŋ* ‘bring’, for example, forms a  $\emptyset$ /U verb with AV *p<en>ilaŋ* (Teng 2008:121). This is true of all F languages with a P/M class. It happens that in Puyuma, all P/M verbs, like *pi-walak* ‘have children’ in Table 4, are derived, but this is not entirely true of other F languages. The PAn class of underived P/M verbs evidently had few members, the most frequently used of which was almost certainly \**paCay*/\**maCay* ‘die’. P/M verbs probably originated as a variant of class  $\emptyset$ /U. PAn avoided U verbs in initial \**p<um>*. . . (§4.3), and class P/M was probably a spin-off of this avoidance. As numerous scholars have pointed out, the PAn P/M verb \**maCay* ‘die’ almost certainly reflects earlier but unattested \**p<um>aCay*. That is, \**p*-replacement (P) verbs were once a phonologically defined subclass of U verbs, but in all F languages where class P/M is reflected, P/M verbs are lexically unpredictable and not determined by phonology alone.

Derived verbs, that is verbs whose Kstem is formed from a derivational prefix and a root (like *ki-kuraw* in (3b)), also belong to one of the three classes  $\emptyset$ /U, P/M or  $\emptyset$ /Z. Table 4 includes the P/M verb *pi-walak/mi-walak* ‘have children’. Class  $\emptyset$ /U is exemplified by *u-sabak/m-u-sabak* ‘go in’ (*u-* ‘go’) and class  $\emptyset$ /Z by *ki-kuraw* ‘catch fish’.

Causative and reciprocal verbs differ from other derived verbs in that their formatives co-occur both with *ka-* and with derivational prefixes. That is, their formatives are added to a verb’s Kstem, not to a root.

The causative prefix is *pa-*, or *p-* before *u* and sometimes before other vowels.<sup>9</sup> Verbs formed with causative *pa-* are in their turn  $\emptyset$ /Z verbs: the Mstem is the same as the Kstem form. Thus we find *pa-tekel* ( $\emptyset$ /Z) ‘cause to drink’ from *tekel/t<em>ekel* ( $\emptyset$ /U) ‘drink’, *pa-ka-tina* ( $\emptyset$ /Z) ‘make bigger’ from *ka-tina/ma-tina* (K/MA) ‘be big’, and *p-u-sabak* ( $\emptyset$ /Z) ‘make (someone (s.o.)) go inside’ from *u-sabak/m-u-sabak* ( $\emptyset$ /U with derived Kstem) ‘go inside’.

Reciprocal verbs are formed by prefixing *ma-Ca-* to a  $\emptyset$  Kstem and *mar-* to a K Kstem (Teng 2008:164–174) and are themselves P/M verbs—hence *pa-ba-be?el/ma-ba-be?el* ‘bite each other’ from *be?el/b<en>e?el* ‘bite’ and *par-ka-deki/mar-ka-deki* ‘scold each other’ from *ka-deki/ma-deki* ‘scold’. The Kstem to which a reciprocal prefix is attached may itself be a causative Kstem, as in *mar-pa-ragan* [AV.RECP-CAUS-erect] ‘build together’.<sup>10</sup>

<sup>9</sup> Before *a*, *i* or *e* there is unexplained variation between *pa-* and *p-*.

<sup>10</sup> Teng (2008:173) calls the reciprocal prefix a ‘plurality of relations’ marker, as it marks among other things collective subjects. There are a number of exceptional causatives and reciprocals, where *ka-* does or does not



Puyuma has two other valency-changing affixes, *mu-* ‘anticausative’ and *ki-* ‘passive’, but these lie outside the scope of the present study.

### 3. Puyuma verb classes as morphological constructions

#### 3.1 Relationships between morphology and meaning

Section 2 focused on verb classes as forms. Here we look at verb classes as morphological constructions, that is, as form–meaning pairings (Booij 2010, 2012). Most verb classes in F languages display at least a somewhat regular form–meaning correspondence. Puyuma class  $\emptyset/U$  consistently encodes dynamicity, in other words it encodes actions and activities. In Teng’s (2008: 123–126, 167–168) analysis, verbs in class K/MA are mostly stative, but some are dynamic.

There are two possible reasons for this seeming inconsistency. The first is that Mstem and Kstem formatives belong to derivational morphology, and derivational morphology is notorious for form–meaning irregularities arising from the fact that every lexical item has its own history. For example, the Puyuma verb *ka-taŋis/ma-taŋis* ‘weep’ is a K/MA verb, but its cognates in other F languages belong to class  $\emptyset/U$ . One may speculate that a class shift occurred because Puyuma speakers associated weeping with an emotion, and emotions are encoded as K/MA verbs in Puyuma (§3.3).

The second possible reason for such class discrepancies is that Huang’s (2000) and Zeitoun & Huang’s (2000) dynamic/stative division perhaps requires modification. There is a theoretical reason why this is probably true. We expect verb classes to differ, if at all,<sup>11</sup> in their lexical semantics, but the dynamic/stative opposition is ultimately a property of events, not of lexical semantics (Croft 2012:44–45). English *believe*, for example, appears to be stative, but *At what moment did you believe ...?* is a semantically coherent question in which *believe* encodes an inchoative, and therefore dynamic, event,<sup>12</sup> not a state. In Austronesian languages it is common for the same property verb to be used both statively and inchoatively, the difference being marked by aspect. In Puyuma the progressive aspect (*Ca-* reduplicated) form of a ‘stative’ verb, that is, one denoting a property, is dynamic and denotes a process (Teng 2008:40):

- |     |    |                       |                                   |
|-----|----|-----------------------|-----------------------------------|
| (8) | a. | <i>litek</i> ‘cold’   | <i>la-litek</i> ‘become colder’   |
|     | b. | <i>inaba</i> ‘good’   | <i>i-na-naba</i> ‘become better’  |
|     | c. | <i>ma-ʔidaŋ</i> ‘old’ | <i>ma-ʔa-ʔidaŋ</i> ‘become older’ |

---

occur as expected. This is due to the fact that verbal morphology is and has been somewhat productive. For example, we find both *ka-bias/bias* ‘be hot’ and *bias/bɛnvias* ‘make hot’ (Teng 2008:56), and Teng (2008:167–168) attributes these irregularities to the preservation of various forms from earlier such pairs. In reciprocals, *par-* is occasionally used where *pa-Ca-* is expected.

<sup>11</sup> Morphological classes often do not display a semantic difference. This is true of Romance verb classes, for example.

<sup>12</sup> An ‘achievement’ in Vendler’s oft-cited 1967 taxonomy.



Similarly if the perfect enclitic is attached to a property verb, it denotes a past event with a present consequence:

- (9) *litek=la na irupan*  
 cold=PRF DEF.NOM dish  
 ‘The dish has become cool.’ (Teng 2008:124)

The completive aspect enclitic has a similar effect in Paiwan: the first property predicate denotes a state, the second an event.

- (10) *na=ɲuaɲuaʔ timadju lakua kuyakuya=aja*  
 PRF=beautiful 3SG.NOM but ugly=COMPL  
 ‘She was beautiful but has become ugly.’ (Chang 2006:225)

In light of the fact that the same verbs encode both inchoative and stative meanings, it is not surprising that, where Zeitoun and Huang label *ka-* ‘stative’, other scholars label it ‘inchoative’ (Teng 2008:123).

### 3.2 A proto-role hierarchy

If the dynamic/stative distinction is not a lexical semantic distinction, then what is the relevant distinction made by  $\emptyset$  versus K? Foley (2005:386), building on Dowty (1991), points the way. The meaning of a verb determines the roles of the participants encoded by its arguments. For example, the verb *kiss* has two arguments, the ‘kisser’ and the ‘kissed’. Its meaning determines that the kisser be ‘an animate being, typically human, able to undertake the act of kissing, putting their lips to something’; ‘sentient, able to experience the sensation of kissing’; able ‘to move their lips to touch the body to be kissed’; and wanting to perform the act, that is, a volitional performer. The entailments for the kissed are fewer: some part of them (or it) is touched by the lips of the kisser, in other words is affected by the action of the kisser (Foley 2005:386–387). Foley organizes the properties of participant roles, drawn from Dowty, into a hierarchy with the most agent-like at the top and the most patient-like at the bottom, as shown in (11).

- (11) Revised macro-role hierarchy (Foley 2005:391)
- |            |   |                                       |  |  |
|------------|---|---------------------------------------|--|--|
| Actor:     | 1 | volitional performer                  |  |  |
|            | 2 | causing an event or a change of state |  |  |
|            | 3 | sentient                              |  |  |
|            | 4 | moving                                |  |  |
|            | 5 | stationary                            |  |  |
|            | 6 | causally affected                     |  |  |
|            | 7 | incremental theme                     |  |  |
| Undergoer: | 8 | undergoing a change in state          |  |  |

To simplify presentation I use the modified version of Foley’s hierarchy in Table 5. The modifications are superficial and are of two kinds. First, as Foley (2005:390–391) notes, his use of

**Table 5:** Proto-role hierarchy, adapted from Foley (2005:391)

		Argument	Verb
Agent:	1	volitional performer (agent) [1]	agentive
	1a	collective volitional performers (agents)	collective
	2	weather verb (no syntactic argument)	weather
	3	sentient perceiver [3]	perception
	4	moving relative to other participant [4]	movement
	5	stationary relative to other participant [5]	position
	6	sentient but non-volitional causer of an activity [2]	involuntary activity
	7	non-volitional causer of an activity [2]	inanimate activity
	8	sentient experiencer of emotion or cognition [3]	mental event
Patient:	9	causally affected [6]	resultative
	10	undergoing a change of state or being in a state [8]	state

*Note:* Numbers in brackets are the numbers of the categories in (11).

the terms ‘actor’ and ‘undergoer’ in (11) differs from their usage in Foley & Van Valin (1984). In the latter’s usage they are part of the mechanics of the grammar, and have been used in this way in Austronesian linguistics (initially by Himmelmann 2002, 2005). Foley (2005), however, uses them in a semantic sense, and to avoid confusion with their grammatical use I revert to Dowty’s terms ‘agent’ and ‘patient’.

Second, Foley’s intention is to show how languages encode participant roles that lie semantically between the prototypical agent at the top of the hierarchy and the prototypical patient at the bottom. Their encodings in F languages suggest some reorderings and some additions, listed in (12). Here and henceforth I show the numbers of the categories in Table 5 in curly braces.

- (12) {1a}: collective volitional performers: verb meaning, e.g. of ‘discuss’, entails a collective agent;<sup>13</sup>  
 {2}: weather verbs are agentive in form, but lack a subject in F languages;  
 {6}, {7}: Foley’s ‘causing’ an event or change of state is split into those with a sentient and those with an inanimate non-volitional causer.  
 {4}, {5}: the ‘moving’ and ‘stationary’ categories are shifted upward, as they tend to be treated as agentive in F languages;  
 {3}, {8}: Foley’s ‘sentient’ category is split into perceiver and experiencer (of emotion and cognition) as the perceiver is often treated as agent in F languages.

Foley’s ‘incremental theme’ is omitted because examples do not occur in the F data.<sup>14</sup>

<sup>13</sup> I number this ‘1a’ rather than, say, ‘2’, as it is orthogonal to the cline from prototypical agent to prototypical patient.

<sup>14</sup> An incremental theme is the participant of a telic predicate whose extent determines completeness: *Fred wrote a letter* (event is complete when the letter is finished); *Egbert climbed the mountain* (event is complete when the summit of the mountain is reached); *Ahab crossed the Pacific* (event is complete when Ahab reaches the other side of the Pacific) (Foley 2005:387).

A volitional performer must be encoded as an agent (A), a participant undergoing a change in state or being in a state as a patient (P). But as a participant becomes less prototypical, in other words is situated in the middle of the hierarchy, its encoding may go either way. Languages differ as to how they encode participants in the middle, drawing the line between A and P at different levels, and often allowing an overlap, as English does at {8}: *John frightens Michael* versus *Michael fears John*. Foley examines a small sample of Austronesian languages and shows that Acehnese encodes only {1} as A (everything else is P); Philippine languages encode categories 1 and 2 of (11) as A; Maluku languages categories 1–3 as A. In English the line divides arguments that are encoded as subject (in active voice) from those encoded as object. In F languages, on the other hand, all arguments in Table 5 may be encoded as subject, and the line divides arguments that are encoded as <sup>AV</sup> or intransitive subject of a  $\emptyset$  verb from those that are encoded as <sup>AV</sup> or intransitive subject of a K verb.  $\emptyset$  verbs are typically A-verbs, K verbs are typically P-verbs.<sup>15</sup>

The logic of Dowty's (and Foley's) approach requires definition of the argument entailed by the verb, but the discussion in the remainder of this paper concerns the verbs that make these entailments. In the interests of brevity, verbs are referred to below by the terms in the rightmost column of Table 5.

### 3.3 The Puyuma proto-role hierarchy

Puyuma has classes  $\emptyset/U$ ,  $\emptyset/MA$ ,  $\emptyset/Z$ ,  $K/MA$  and  $K/Z$ . There are no monomorphemic P/M verbs.

Puyuma  $\emptyset/U$  verbs are typically agentive. They also include weather and perception verbs, the verb 'go' and the occasional mental event. Weather verbs, which are subjectless, appear to reflect a strategy whereby the activity associated with a noun is verbalized as a  $\emptyset/U$  verb with the noun as its root. The undergoer voice forms of 'smell' and 'taste' have the locative-subject form, reflecting the fact that perception does not physically change the perceived entity.

(13)	Puyuma Ø/U			
	{1} agentive	‘bathe’	<i>dirus</i>	<i>d&lt;em&gt;irus</i>
		‘sing’	<i>senay</i>	<i>s&lt;em&gt;enay</i>
		‘eat’	<i>ekan</i>	<i>m-ekan</i>
		‘pull’	<i>keqey</i>	<i>k&lt;em&gt;eqey</i>
		‘drink’	<i>tekel</i>	<i>t&lt;em&gt;ekel</i>
	{2} weather	‘be cloudy’	<i>ku'em</i> ‘cloud’	<i>k&lt;em&gt;u'em</i>
		‘be sunny’	<i>kadaw</i> ‘sun’	<i>k&lt;em&gt;adaw</i>
		‘thunder’	<i>qerung</i> ‘thunder (N)’	<i>q&lt;em&gt;erung</i>
		‘rain’	<i>ʔudal</i> ‘rain (N)’	<i>ʔ&lt;em&gt;udal</i>
	{3} perception	‘see’	<i>naʔu</i>	<i>me-naʔu</i>
		‘smell’	<i>taul-ay</i> (UVL)	<i>t&lt;em&gt;aul</i>
		‘taste’	<i>tamtam-ay</i> (UVL)	<i>t&lt;em&gt;amtam</i>

<sup>15</sup> The labels 'A-verb' and 'P-verb' go back to Foley (1976), who notes that the categories reach across Austronesian from the Philippines to the central Pacific. As Foley notes, A-verbs are what Chafe (1970) calls 'action' and 'action-process' verbs, and P-verbs are process verbs. Alternatively, A-verbs are unergative, P-verbs unaccusative.

{4} movement	‘go’	<i>uka</i>	<i>m-uka</i>
{8} mental	‘be(come) angry’	<i>ŋerŋer</i>	<i>me-ŋerŋer</i>

Many movement verbs have stems derived with *u-/mu-* ‘go’, as in *u-sabak/mu-sabak* ‘go inside’, where the destination is incorporated into the verb stem. Other movement verbs are scattered across verb classes.

Class  $\emptyset$ /MA verbs are a much smaller category and include agentive and movement verbs.

(14) Puyuma  $\emptyset$ /MA

{1} agentive	‘tell’	<i>reŋay</i>	<i>ma-reŋay</i>
{4} movement	‘depart’	<i>berek</i>	<i>ma-berek</i>
	‘arrive’	<i>redek</i>	<i>ma-redek</i>
	‘get up’	<i>ragan</i>	<i>ma-ragan</i>

There is some evidence that agentive verbs with a collective subject also tend to be  $\emptyset$ /MA verbs.

(15) Puyuma collective agentive verbs

{1a} collective	‘discuss’	<i>qayar</i>	<i>ma-qayar</i>
	‘fight’	<i>qadikul</i>	<i>ma-qadikul</i>
	‘harvest’	<i>reʔani</i>	<i>ma-reʔani</i>

Puyuma  $\emptyset$ /Z verbs form a small category, and, like  $\emptyset$ /MA verbs, encode agentive and movement verbs. From (13), (14) and (16) one can infer that movement verbs are—unsurprisingly—treated as agentive verbs. Some of the verbs in (16) were perhaps originally  $\emptyset$ /U verbs. The roots *beray*, *pulaŋ* and *palu* are labial-initial, and reflect a tendency to avoid the sequences \*b<um> . . . and \*p<um> . . . (§6.2.2). F cognates of *alup* are  $\emptyset$ /U verbs.

(16) Puyuma  $\emptyset$ /Z

{1} agentive	‘give’	<i>beray</i>
	‘help’	<i>pulaŋ</i>
	‘demarcate’	<i>palu</i>
	‘close’	<i>aleb</i>
	‘hunt’	<i>alup</i>
{4} movement	‘come’	<i>qua</i>

The K/MA class is large and typically encodes states, whether inchoative or stative. However, it also encodes verbs at higher positions on the hierarchy. The common denominator of {6}, {8} and {10} is that they can be construed as non-agentive. Why movement and position verbs occur here doubtless has a reason in their individual lexical histories that is no longer accessible. The verb *ka-qeki/ma-qeki* ‘scold’ is included under involuntary activities in (17) because verbs for ‘scold’ are typically K verbs in F languages, but I do not know why. I can only speculate that scolding, like anger, is construed as involuntary.

Positional verbs are K/MA verbs, and are used both dynamically ('he sat down') and statively ('he was seated').<sup>16</sup>

(17) Puyuma K/MA			
{4} movement	'start out (on journey)'	<i>ka-sikasik</i>	<i>ma-sikasik</i>
{5} position	'sit (down)'	<i>ka-tenaḍaw</i>	<i>ma-tenaḍaw</i>
	'stand (up)'	<i>ka-retigir</i>	<i>ma-retigir</i>
{6} involuntary act	'kneel (down)'	<i>ka-retugun</i>	<i>ma-retugun</i>
	'cry'	<i>ka-ṭan̄is</i>	<i>ma-ṭan̄is</i>
	'sleep, lie'	<i>k-iedeṇ</i>	<i>m-iedeṇ</i>
	'scold'	<i>ka-ḍeki</i>	<i>ma-ḍeki</i>
{8} mental event	'miss, yearn for'	<i>ka-supen</i>	<i>ma-supen</i>
	'know, understand'	<i>ka-laḍam</i>	<i>ma-laḍam</i>
	'not know, not understand'	<i>ka-uḷid</i>	<i>ma-uḷid</i>
{10} state	'be(come) big'	<i>ka-ṭina</i>	<i>ma-ṭina</i>
	'be(come) strong'	<i>ka-keser</i>	<i>ma-keser</i>
	'be(come) drunk'	<i>ka-liay</i>	<i>ma-liay</i>
	'be(come) identical'	<i>ka-risan</i>	<i>ma-risan</i>
	'be(come) old'	<i>ka-ṭidaṇ</i>	<i>ma-ṭidaṇ</i>
	'die, be dead'	<i>k-inaṭay</i>	<i>m-inaṭay</i>
	'be(come) upset'	<i>ka-sepel</i>	<i>ma-sepel</i>

Mental event verbs are mostly K verbs. There is, however, a thin line to be drawn between state verbs that denote a mental state and verbs that appear to denote some mental activity on the part of the subject argument. The line is thin because it is easy to make unjustified inferences about the state/activity distinction on the basis of English glosses. Is 'be(come) angry' an event or a state? The morphology of the two categories is often the same or displays an overlap. The only objective distinction is that some mental event verbs take a second argument ('like someone', 'know something') whereas states typically do not. Where their morphology allows either categorization, I treat such verbs, for example *ka-sepel/ma-sepel* 'be(come) upset', as states.

Resultative verbs also belong to class K/MA.<sup>17</sup> Discussing the Mayrinax Atayal classes, Huang (2000) observes that if a state is construed as the result of an action, then it is far more likely to be a K/MA verb than a K/Z. Typically there is an agentive Ø/U verb with the same root which denotes the action that brings about the effect encoded by the K/MA verb. This is rare with K/Z verbs.

(18) Puyuma resultative verbs ({9})			
'be(come) cooked'	<i>ka-deru</i>	<i>ma-deru</i>	(cf. <i>ḍem&gt;eru</i> 'cook (ITR)')
'be(come) burned'	<i>ka-baṭiṭ</i>	<i>ma-baṭiṭ</i>	(cf. <i>b&lt;en&gt;aṭiṭ</i> 'burn (ITR)')
'be(come) lost'	<i>ka-san̄an</i>	<i>ma-san̄an</i>	(cf. <i>s&lt;em&gt;an̄an</i> 'go astray')

<sup>16</sup> There are also positional verbs that are used only statively: *ki-nugun/mi-nugun* 'bending the waist'; *ki-kakupu/mi-kakupu* 'to lie prone'; *ki-gagayang/mi-gagayang* 'to lie face up'. However, these have derived stems and lie beyond the scope of this paper.

<sup>17</sup> I owe the label 'resultative' to Huang (2012).

Puyuma K/Z verbs form a much smaller class, and partly cover the lower part of the range encoded by K/MA verbs. Note, however, that K/Z state verbs encode what Huang (2000) calls ‘inherent’ properties and thus do not include resultatives.

(19) Puyuma K/Z				
{6} involuntary act	‘sleep’	<i>k-alupe</i>	<i>alupe</i>	
	‘laugh’	<i>ka-saeru</i>	<i>saeru</i>	
{8} mental event	‘like’	<i>ka-sagar</i>	<i>sagar</i>	
{10} state	‘be(come) good’	<i>ka-inaba</i>	<i>inaba</i>	
	‘be(come) dirty’	<i>ka-riksis</i>	<i>riksis</i>	
	‘be(come) clean, beautiful’	<i>ka-ijdan</i>	<i>ijdan</i>	
	‘be new’	<i>ka-bekaʔ</i>	<i>bekaʔ</i>	

Inanimate activities ({7}) do not occur among the examples above. Typically they have stems derived with the anticausative prefix *mu-*, e.g. *mu-baʔit* ‘burn (ITR)’ (cf. *baʔit-aw* UVP ‘burn (ITR)’), *mu-kuatis* ‘break (ITR)’ (cf. *k-em>uatis* ‘destroy’, *kuatis* ‘bad’) (Teng 2008:180; personal communication 2014), and lie outside the scope of the research reported here (Thao and Mantaaran Rukai behave similarly; see §5.3 and §5.6.2).

A rough summary of the discussion above is given in Table 6. Puyuma appears to draw the line between A-verbs and P-verbs between {4} and {5}. Verbs with collective agents are ignored in placing the line, as the collective agents feature is not part of the agent-to-patient cline.

Himmelmann (2006, 2008) shows that in the Philippine languages Tagalog *ma-* verbs fall into different morphological classes. That is, their roots undergo different morphological derivations, and/or the same formal derivation may differ in meaning from class to class. It is likely that this is true of Puyuma K/MA verbs. As noted in association with (18), for example, *ka-deru/ma-deru* ‘be(come) cooked’ is the causal effect of the action denoted by *deru/d-em>eru* ‘cook (ITR)’. But for a good

**Table 6:** Semantic distribution of Puyuma verb classes

		Ø/U	Ø/MA	Ø/Z	K/MA	K/Z
{1}	agentive	x	x	x		
	{1a} collective		(x)			
{2}	weather	x				
{3}	perception	x				
{4}	movement	x (1)	x	x (1)	x (1)	
{5}	position				x	
{6}	involuntary activity				x	x
{7}	inanimate activity	<i>mu-</i>				
{8}	mental event	x (1)			x	x (1)
{9}	resultative				x	
{10}	state				x	x

many K verbs the corresponding causative is formed by prefixing *pa-* to the Kstem: *ka-tina/ma-tina* ‘be(come) big’ forms the causative *pa-ka-tina* ‘to make big’; *ka-inaba/inaba* ‘be(come) good’ forms the causative *pa-ka-inaba* ‘to make good’ (Teng 2008:166). Arguably *ka-deru* is derived from *deru*, that is, the derivation detransitivizes, whereas *pa-ka-tina* is derived from *ka-tina*, a transitivity derivation. This appears to place *ka-deru/ma-deru* and *ka-tina/ma-tina* in different morpholexical classes, but this is an issue that needs further research. In terms of Nichols et al.’s (2004) typology of lexical valence orientation, Puyuma is thus a ‘neutral’ language, in other words it has both transitivity and detransitivity morphology.

The proto-role hierarchy in Table 5 is at best a useful heuristic device. Semantic categories do not behave in a rule-governed manner, as Dowty’s (1991) discussion makes clear. Indeed, he assigns some verbs to more than one category and does not regard the categories as an ordered list. On the other hand, applying the heuristic here has enabled us to see which semantic categories in addition to states are coded with K, and to see that the Kstem division between  $\emptyset$  and K fits broadly into a crosslinguistic typology of verbal semantics. It roughly matches the semantic cline that Huang (2000) sees in Mayrinax Atayal verbs. It also shows that Zeitoun & Huang’s (2000) definition of \*ka- as stative is largely correct if we replace their dynamic/stative division with one between agentive and non-agentive events, where the majority of non-agentive verbs encode a state or change of state, but a minority encodes dynamic events.

#### 4. Two approaches to reconstructing PAn verb classes

The comparative method of historical linguistics compares form–meaning correspondences across languages and reconstructs ancestral form–meaning correspondences or ‘protoforms’. The forms in a form–meaning correspondence are typically words or morphemes, but they may also be morphological structures. There are thus two ways to approach the reconstruction of PAn verb classes. One is the conventional lexical approach, reconstructing the Kstem and Mstem of individual verbs to assign them to their PAn class. Thanks to the plethora of lexical and morphological changes in various F languages, the number of verbs for which both the Kstem and Mstem are securely reconstructable is rather few, making it necessary to turn to the second approach. This is a constructional approach, which takes the morphological structures of the Kstem and Mstem of each verb class language by language as §3.3 does for Puyuma, identifies the range of categories each class encodes in each language on the hierarchy in Table 5, and compares these to arrive at a set of putative PAn verb classes. The language-by-language analysis is presented, as far as the data permit, in §5.1–5.14. In §6 the two approaches are collated: the language-by-language analysis of §5 provides the basis for reconstructing the distribution of PAn verb classes across semantic categories, and this is supported where possible by reconstruction of the Kstem and Mstem of members of each verb class in §§6.2–6.7.

#### 5. Verb classes in Formosan languages

There is a good deal of variation among the verbal morphologies of F languages (for an overview see Ross 2009:317–320), but most of them have verb classes recognizably similar to Puyuma’s.



Probably the potential forms of the verb in all F languages can all be predicted from the Kstem and the Mstem.

In most languages the Kstem appears as an AV form: often the imperative, sometimes the dependent or negative form (i.e. the dependent of the negator), and in Kavalan the irrealis. In Tsou, Kanakanavu, Isbukun Bunun and Siraya the Kstem does not occur in isolation, but several affixed forms transparently reveal its form.

Most sources routinely provide the AV realis form of a verb. This is sufficient to assign a verb to class  $\emptyset/U$ , but where the Mstem marker reflects \*ma- or is zero, the Kstem is needed in order to assign it to its class. If the Kstem is unaffixed, it belongs to  $\emptyset/MA$  or  $\emptyset/Z$ ; if it reflects prefixation with \*ka-, it belongs to  $K/MA$  or  $K/Z$ . Given the thinness of some sources, I take as a pointer any form that uses the Kstem. Thus for a number of Thao verbs reflecting \*ma-, for example *ma-faḏaq* ‘know, understand’, the presence of a UVP realis form reflecting \*ka-, such as *ka-faḏaq-in*, is taken as evidence to assign them to class  $K/MA$ . A derived causative stem is also often a pointer. For instance, Pazih *ma-hatan* ‘laugh, be happy’ forms a causative *pa-ka-hatan*, indicating that the Kstem is *ka-hatan* and the verb is  $K/MA$ , but Pazih *ma-dawan* ‘bathe (self)’ forms a causative *pa-dawan*, indicating that the Kstem is *dawan* and the verb is  $\emptyset/MA$ . There are doubtless instances where this procedure distorts synchronic morphology, but it is probably reliable enough for reconstructive purposes.

Identifying the form–meaning correspondences of the semantic categories from {2} to {9} is a task that cannot be performed with particular accuracy, for three reasons.

First, in any language the vast majority of verbs (or verbs and predicate adjectives) are either agentive or state verbs. The in-between categories are much less richly represented.

Second, some F languages make extensive use of derived  $\emptyset$  stems, consisting of a prefix and a root.<sup>18</sup> The prefix may have a broad function like Puyuma anticausative *mu-* or it may have a fairly narrow lexical meaning. The root may be one that also forms a monomorphemic stem, or it may be a bound root that never occurs on its own. As this paper is concerned with monomorphemic stems, the prevalence of derived stems in some F languages leaves only restricted data for consideration. A majority of P/M verbs consist of a prefix in *pa.../m-a...* and a root, and thus lie outside the scope of this paper. Most movement verbs in several F languages have stems derived with reflexes of PAN \*u-/\*m-u-, for example Mantauran Rukai *o-valio/m-o-valio* ‘return home’ (Zeitoun 2007), Saaroa *u-sala* ‘go (on foot)’ (*sala* ‘road’), *u-cekele* ‘come’ (Pan 2012), Thao *u-nay/m-u-nay* ‘come’ (*nay* ‘here’) (Blust 2003:136–139). Liao (2011:854–855) also cites reflexes in Siraya, Tsou and Bunun. Amis verbs of movement have stems derived with *ta-* (e.g. *ta-lumaq* ‘go home’, *ta-yra* ‘go, arrive (there)’, *ta-yni* ‘come, arrive here’). Some languages form positional verbs similarly, for example Saaroa *tu-puru* ‘sit’, Thao *i-lhunqu/m-i-lhunqu* ‘sit’, *i-lhilhi/m-i-lhilhi*.

The third reason is that the data for several F languages are sparse, so we cannot be sure whether the verbs in the in-between categories reflect a morphological pattern corresponding to that semantic category, whether they are exceptions to a pattern, or whether there is no pattern at all. Where there is only one verb representing a form–meaning correspondence, this is indicated by ‘(1)’

<sup>18</sup> For descriptions, see Tsuchida (1985) for Tsou, Kanakanavu and Saaroa; Tsuchida (1990) for Tsou; Chang (2006) and Huang (2012) for Paiwan; Tsuchida (2000) for Siraya; Nojima (1996) for Bunun; Blust (2003) for Thao; Zeitoun et al. (2015) for Saisiyat.

in the summary table that ends most of §5.7–5.14. No summary table is given for Tsou or Isbukun Bunun, as the data are not well enough understood for this to be meaningful. The data for Tanan Rukai, Saaroa and Kananavu are also so sparse that the form–meaning correspondences for a majority of categories between {2} and {9} might be changed significantly by future research.

In view of the varying quality and quantity of the data from language to language, those languages for which a more thorough analysis of verb classes is possible are placed first. These are Paiwan, Kavalan, Thao, Amis, Saisiyat and Mantauran Rukai. As they are fairly well distributed geographically, they provide a reasonable cross-section of the distributions of verb classes in F languages. The verb class systems of Paiwan and Kavalan in particular appear more complex semantically than those of other F languages, but it may be that the latter have complexity that available data do not reveal.

For reasons of space, the verb classes of each F language are presented rather briefly here, and only a few examples of each class in each language are given. Examples are listed without comment where none is required. Certain abbreviatory conventions are used. As  $\emptyset/U$ ,  $\emptyset/MA$  and  $\emptyset/Z$  verbs often occur in the same semantic categories, I refer to them collectively as  $\emptyset$  verbs. Similarly,  $K/MA$  and  $K/Z$  verbs are referred to as  $K$  verbs.

## 5.1 Paiwan

Paiwan data are drawn from a range of sources (Chang 2006; Egli 1990, 2002; Ferrell 1982; Ho 1978; Huang 2012). The Paiwan verb classes are  $\emptyset/U$ ,  $P/M$ ,  $\emptyset/Z$ ,  $K/MA$  and  $K/Z$ . There is no class  $\emptyset/MA$ . I have found no monomorphemic positional verbs.

Class  $\emptyset/U$  verbs are mostly agentive, but weather, perception, movement, involuntary activity and mental event verbs are also found.

- (20) Paiwan  $\emptyset/U$
- |                 |  |
|-----------------|--|
| {1} agentive    | <i>k&lt;em&gt;an</i> ‘eat’, <i>k&lt;em&gt;ac</i> ‘bite’, <i>p&lt;em&gt;ana?</i> ‘shoot arrow’, <i>?&lt;em&gt;a[up</i> ‘hunt’, <i>m-a[ap</i> ‘take’ |
| {2} weather     | <i>q&lt;em&gt;udjal</i> ‘rain’, <i>s&lt;em&gt;ula</i> ‘snow’, <i>v&lt;em&gt;ali</i> ‘(wind) blow’  |
| {3} perception  | <i>[&lt;em&gt;aŋeda</i> ‘listen, hear’, <i>s&lt;em&gt;equ</i> ‘smell’, <i>tj&lt;em&gt;imutj</i> ‘taste’  |
| {4} movement    | <i>dj&lt;em&gt;alun</i> ‘arrive’, <i>z&lt;em&gt;uker</i> ‘reach (a place)’   |
| {6} invol. act. | <i>dj&lt;em&gt;[i</i> ‘laugh at’   |
| {8} mental      | <i>k&lt;em&gt;elan</i> ‘know, understand’  |

The roots of weather verbs also serve as nouns, and they seem to reflect a more general derivational template whereby a denominal  $\emptyset/U$  verb denotes the default activity that is performed with that noun, for example *p<em>aŋul* ‘beat with a stick’ (*paŋul* ‘stick’), *dj<em>iŋay* ‘trap animals’ (*d[iŋay* ‘trap’), *q<em>apaz* ‘fence in’ (*qapaz* ‘fence’), *dj<em>aum* ‘sew with needle’ (*djaum* ‘needle’) (Huang 2012:94).

There is just one diachronically monomorphemic  $P/M$  verb, the involuntary activity verb *pacay/ m-acay* ‘die’.

Class  $\emptyset/Z$  has the unusual property that it encodes verbs at both ends of the hierarchy, with a concentration towards the top, but encoding only states at the opposite extreme. This implies that it combines two diachronic sources (§6.6).

- (21) Paiwan Ø/Z
- |                 |   |
|-----------------|---|
| {1} agentive    | <i>kian</i> ‘follow’, <i>kaiv</i> ‘have dinner’, <i>aya</i> ‘say (to s.o.)’ |
| {3} perception  | <i>pacun</i> ‘see, look at’   |
| {4} movement    | <i>vaik</i> ‘go, depart’  |
| {6} invol. act. | <i>taqed</i> ‘be/fall asleep’   |
| {10} states     | <i>keđi-keđi</i> ‘be small’, <i>vece[e]</i> ‘(thing) be cold’               |

The main function of Class K/MA is to encode resultatives and states.

- (22) Paiwan K/MA
- |                 |   |
|-----------------|---|
| {6} invol. act. | <i>ma-ke[lu]</i> ‘fall over’  |
| {7} inan. act.  | <i>ma-cekas</i> ‘(skin) split’, <i>ma-gerger</i> ‘(cloth) tear’, <i>ma-gac</i> ‘(ground) split open’, <i>ma-kuli</i> ‘(river) dry up’   |
| {8} mental      | <i>ma-seŋac</i> ‘despise, not want to’, <i>tjeŋe[ay]</i> ‘love/like’, <i>sa[ŋa]</i> ‘want’, <i>mare-ka-siŋe[li]</i> ‘miss each other’, <i>siŋe[li]</i> ‘miss (s.o./something (s.t.))’   |
| {9} resultative | <i>ma-ŋaŋeda</i> ‘be(come) audible’ (cf. <i>ŋem&gt;ŋaŋeda</i> ‘hear’), <i>ma-kan</i> ‘be/get eaten, be edible’ (cf. <i>kem&gt;an</i> ‘eat’), <i>ma-rakac</i> ‘get pulled’ (cf. <i>rem&gt;akac</i> ‘pull’), <i>ma-panadip</i> ‘get caught on hook’ (cf. <i>m-anadip</i> ‘fish with hook’), <i>ma-ta[em]</i> ‘get planted’ (cf. <i>tem&gt;a[em]</i> ‘plant (s.t.)’) |
| {10} states     | <i>ma-rekutj</i> ‘be(come) afraid’, <i>ma-veŋe[ay]</i> ‘be(come) pregnant’, <i>ma-vetu</i> ‘be(come) full’, <i>ma-dudu</i> ‘be(come) angry’   |

Resultative verbs, like the corresponding verbs in other F languages, are related to a Ø/U or P/M verb with the same root. As in other languages, these verbs may be used statively or inchoatively. It is the inchoative, and therefore dynamic, use that has led to the application of labels such as ‘anticausative’ (Chang 2006) and ‘passive’ (Egli 1990; Ferrell 1982). At least some of these verbs may be accompanied by an oblique agent NP (Chang 2006:212–215), supporting the label ‘passive’. But ‘passive’ implies a paradigmatic opposition between active and passive within a single verb. Instead, Paiwan has two verbs derived from the same root, for example *ŋem>ŋaŋeda* ‘hear’, a Ø/U verb, and *ma-ŋaŋeda* ‘be heard, be audible’, a K/MA verb (Chang 2006:215–217). According to Chang, the dynamic verb in *ma-* implies a lack of intentionality on the part of the agent, and this justifies the label ‘anticausative’. The glosses ‘be audible’ and ‘be edible’ in (22{9}) hint at another fact, namely that at least some of these verbs may be used both anticausatively (‘be heard’) and potently (‘be audible’) (Huang 2012:134).

My analysis of K/MA verbs owes much to Huang (2012). However, he attempts to establish a function for the prefix *ma-* common to all dynamic K/MA verbs. This is unnecessary, as homophonous morphemes are not a linguistic rarity.

Class K/Z mainly encodes states, but—unlike K/MA—never resultatives.

- (23) Paiwan K/Z
- |                |   |
|----------------|---|
| {7} inan. act. | <i>ba[ecue]</i> ‘explode’   |
| {10} states    | <i>dja[aw]</i> ‘be(come) fast, quick’, <i>setukez</i> ‘be(come) diligent’, <i>sa[etu]</i> ‘be(come) sick’ |

**Table 7:** Semantic distribution of Paiwan verb classes

		∅/U	∅/Z	K/MA	K/Z
{1}	agentive	x	x		
	{1a} collective			(x)	
{2}	weather	x			
{3}	perception	x	x (1)		
{4}	movement	x	x (1)		
{6}	involuntary activity	x (1)	x (1)	x (1)	
{7}	inanimate activity			x	x (1)
{8}	mental event	x (1)		x	
{9}	resultative			x	
{10}	state		x	x	x

Verbs with a collective agentive subject ({1a}) belong to class K/MA. Often there is a corresponding ∅/U verb formed from the same root, typically with a non-collective agentive subject. According to Huang (2012:131), a collective *ma-* verb cannot be the base for further reciprocal derivation.

- (24) Paiwan collective agentive verbs ({1a}): *ma-ciur* ‘go/be together’ (cf. *c<em>iur* ‘take (s.o.) with you’), *ma-sudju-sudju* ‘court by group conversation’ (cf. *s<em>udju* ‘spend time with girlfriend’), *ma-zepzep* ‘discuss’ (cf. *z<em>epzep* ‘call together for discussion’), *ma-ʔiaw* ‘assist, support (in battle etc.)’ (cf. *ʔ<em>iaw* ‘call forth’), *ma-seŋ-seŋ* ‘work in field’, *ma-vacuk* ‘clear (land)’

A summary of Paiwan verb classes is presented in Table 7. The line between A- and P-verbs lies between {4} and {7}, with an overlap at {6}.

## 5.2 Kavalan

Data are drawn from a number of sources (Chang 1997 and personal communication 2003; Huang & Sung 2008; Lee 1997; Li & Tsuchida 2006; Tsuchida 1993). Kavalan has classes ∅/U, ∅/MA, ∅/Z, K/MA and K/Z. There are no monomorphemic P/M verbs.

The morphophonemics of Kavalan AV morphemes are tricky, as U and MA overlap at certain points.

- (25) a. U is manifested as *m-* before root-initial *bV*, *IV*, *rV*, *RV*, *zV*, *nV*, *wV* and *V*.  
b. MA is manifested as *m-* before root-initial *pV*, *tV*, *kV*, *qV*, *IV*, *rV*, *RV*, *zV*, *ŋV* and *V*.

The effect of these rules is that *m-* before *IV*, *rV*, *RV*, *zV* or *V* may manifest either U or MA. Perhaps encouraged by the rule overlap, there also seems to be a certain amount of ‘leakage’

whereby verbs where  $\langle m \rangle$  is expected occasionally occur with  $m-$ . Without the Kstem of such a verb, we cannot assign it to its class. When the Kstem is  $\emptyset$ , the verb remains ambiguous between  $\emptyset/U$  and  $\emptyset/MA$ .

Class  $\emptyset/U$  is large, mostly agentive, but also includes weather, perception, movement and inanimate activity verbs.

- (26) Kavalan  $\emptyset/U$
- |                |   |
|----------------|---|
| {1} agentive   | <i>qan</i> , <i>qaʔn/q&lt;m&gt;an</i> ‘eat’, <i>qaRat/q&lt;m&gt;aRat</i> ‘bite’, <i>pani/p&lt;m&gt;ani</i> ‘shoot’, <i>saraw/s&lt;m&gt;araw</i> ‘hunt’, <i>uti/m-uti</i> ‘vomit’  |
| {2} weather    | <i>u-bari</i> ‘(wind) blow’ ( <i>bari</i> (N)), <i>m-uzan</i> ‘rain’ ( <i>uzan</i> (N)), <i>s&lt;m&gt;uRna</i> ‘snow’ ( <i>suRna</i> (N)), <i>z&lt;m&gt;eŋzeŋ</i> ‘thunder’ ( <i>zeŋzeŋ</i> (N)), <i>sezaŋ/s&lt;m&gt;ezaŋ</i> ‘(sun) shine’ |
| {3} perception | <i>tayta/t&lt;m&gt;ayta</i> ‘see’, <i>siŋut/s&lt;m&gt;iŋut</i> ‘smell (ITR)’, <i>talam/t&lt;m&gt;alam</i> ‘taste (ITR)’   |
| {7} inan. act. | <i>buiq/m-buiq</i> ‘blossom, unfold (ITR)’, <i>lzas/mu-lzas</i> ‘fall off’  |

Classes  $\emptyset/MA$  and  $\emptyset/Z$  are small and include verbs from both ends of the hierarchy, like Paiwan class  $\emptyset/Z$ . Verbs of perception are all  $\emptyset$  verbs, but are divided between  $\emptyset/U$  and  $\emptyset/MA$ , raising the possibility that there has been leakage from one class into the other. The perception verb *ipir/ m-ipir* ‘hear’ and the mental event verb *rizaq/m-rizaq* ‘like, love’ may also belong in (27), but are morphologically ambiguous: they may be  $\emptyset/U$  verbs.

- (27) Kavalan  $\emptyset/MA$
- |                 |  |
|-----------------|--|
| {3} perception  | <i>qita/m-qita</i> ‘see’, <i>quriq/m-quriq</i> ‘find’  |
| {6} invol. act. | <i>tawa/m-tawa</i> ‘laugh’   |
| {7} inan. act.  | <i>tutun/m-tutun</i> ‘burn, catch fire (ITR)’, <i>bqiat/ma-bqiat</i> ‘split off (ITR)’, <i>buiq/m-buiq</i> ‘blossom, unfold (ITR)’ |
| {10} state      | <i>baŋsis/ma-baŋsis</i> ‘be fragrant’  |
- (28) Kavalan  $\emptyset/Z$
- |              |  |
|--------------|--|
| {1} agentive | <i>tmeq</i> ‘bake’, <i>satay</i> ‘sing’          |
| {4} movement | <i>wi</i> ‘go’                                   |
| {8} mental   | <i>supaR</i> ‘know’, <i>sizŋes</i> ‘think of’    |
| {10} state   | <i>qzan</i> ‘be old’, <i>Riŋu</i> ‘be unable to’ |

Class  $K/MA$  is large, and a majority of its members encodes states, but others are situated higher in the hierarchy. Li & Tsuchida (2006) list a number of state verbs in  $MA$  without their Kstem, so one cannot tell whether they belong to class  $K/MA$  or  $\emptyset/MA$ .

- (29) Kavalan  $K/MA$
- |              |  |
|--------------|--|
| {4} movement | <i>q-iRi/m-iRi</i> ‘stand up, set out’, <i>qa-tiw/ma-tiw</i> ‘go’, <i>qa-sarin/ma-saRin</i> ‘float away’, <i>qa-suat/ma-suat</i> ‘get up’, <i>qa-szeq/ma-szeq</i> ‘arrive’, <i>qa-qseq /ma-qseq</i> ‘arrive in’, <i>qa-wtu/ma-wtu</i> ‘come’ |
|--------------|--|

{5} position	<i>q-rikney/m-rikney</i> ‘stand still’, <i>qa-llel/ma-llel</i> ‘sit (down)’, <i>q-uRiRil/m-uRiRil</i> ‘sit (down on ground)’
{7} inan. act.	<i>qa-tama/m-tama</i> ‘set (sun, moon)’, <i>qa-wman/ma-wman</i> ‘change (ITR)’, <i>qa-siun/m-siun</i> ‘get less, decrease’, <i>qa-ttel/m-tel</i> ‘dry up, ebb’, <i>qa-tmeq/ma-tmeq</i> ‘get burnt, catch fire’, <i>qa-spit/ma-spit</i> ‘broken; crack (ITR)’, <i>q-imRus/m-imRus</i> ‘boil (ITR)’
{10} state	<i>qa-ytis/ma-ytis</i> ‘be afraid’, <i>qa-ηRu/ma-ηRu</i> ‘be tired’, <i>qa-rmes/ma-rmes</i> ‘be diligent’, <i>qa-rizaq/m-rizaq</i> ‘be happy’, <i>qa-Ranjazi/m-Ranjazi</i> ‘be rich’

The K/Z class includes only states.

(30) Kavalan K/Z

{10} state	<i>qa-Raya/Raya</i> ‘big’, <i>qa-nŋi/nŋi</i> ‘good’, <i>qa-zan/?zan</i> ‘old’, <i>qa-tasu/tasu</i> ‘new’, <i>qa-bsaR/busar</i> ‘white’, <i>qa-inpis/inpis</i> ‘thin’
------------	--

As in Paiwan, resultatives correspond with an agentive  $\emptyset$ /U verb. Unlike in Paiwan, however, some are  $\emptyset$ /MA verbs, and for others the Kstem is unknown.

(31) Kavalan resultative verbs ({9})

K/MA	<i>qa-pukun/ma-pukun</i> ‘get hit’ (cf. <i>p&lt;m&gt;ukun</i> ‘hit’), <i>qa-tpuq/ma-tpuq</i> ‘get wounded by knife’ (cf. <i>t&lt;um&gt;puq</i> ‘cut with knife’), <i>qa-tbeR/ma-tbeR</i> ‘get stepped on’ (cf. <i>tbeR</i> ‘step on’)
$\emptyset$ /MA	<i>bukas/ma-bukas</i> ‘be covered’ (cf. <i>u-bukas</i> ‘cover (ITR)’), <i>nukub/ma-nukub</i> ‘be covered’ (cf. <i>m-nukub</i> ‘cover (ITR)’)
?/MA	<i>ma-qan</i> ‘get eaten’ (cf. <i>q&lt;m&gt;an</i> ‘eat’), <i>m-qual</i> ‘get fined’ (cf. <i>q&lt;m&gt;ual</i> ‘be fined’), <i>m-saray</i> ‘get scalded’ (cf. <i>s&lt;m&gt;aray</i> ‘scald (ITR)’), <i>ma-tapu</i> ‘be blocked, obstructed’ (cf. <i>t&lt;m&gt;apu</i> ‘block, obstruct’)

Agentive verbs with collective subjects belong to class K/MA.

(32) Kavalan collective agentive verbs ({1a}): *qa-tanaq/ma-tanaq* ‘separate (ITR)’ (*m-tanaq* ‘separate from (ITR)’), *m-sarkiaw* ‘dance in group’ (*sarkiaw* ‘to dance’), *qa-kassa/ma-kassa* ‘divide (of family) (ITR)’, *qa-surun/ma-surun* ‘work together’, *qa-zmun/ma-zmun* ‘be many’

A summary of Kavalan morphological constructions is presented in Table 8. A- and P-verbs divide between {3} and {4}, whilst mental event verbs are out of place. They are  $\emptyset$  verbs where K is expected.

### 5.3 Thao

Data are drawn from Blust’s (2003) extensive dictionary. Thao has classes  $\emptyset$ /U,  $\emptyset$ /MA,  $\emptyset$ /Z and K/MA and appears to lack a K/Z category. There are no monomorphemic P/M verbs. No monomorphemic positional verbs have been found.

**Table 8:** Semantic distribution of Kavalan verb classes

		∅/U	∅/MA	∅/Z	K/MA	K/Z
{1}	agentive	x		x		
	{1a} collective				(x)	
{2}	weather	x				
{3}	perception	x	x			
{4}	movement			x (1)	x	
{5}	position				x	
{6}	involuntary activity		x (1)			
{7}	inanimate activity	x	x		x	
{8}	mental event			x		
{9}	resultative		x		x	
{10}	state		x (1)	x	x	x

Class ∅/U verbs are mostly agentive, but some are weather, perception and movement verbs.

(33) Thao ∅/U

- {1} agentive *θanit/θ<m>anit* ‘cry’, *sabu-i* (UVP)/*s<m>abu* ‘cover’, *kan/k<m>an* ‘eat’,  
*liklik-in* (UVP R)/*l<m>iklik* ‘saw’, *finaw/f<m>inaw* ‘wash’, *ku/kuf-i* (UVP)/  
*k<m>u/kuf* ‘scrape’, *ktun/k<m>utun* ‘break’
- {2} weather *qusað-in* (UVP) ‘be rained on’/*q<m>usað* ‘rain’, *rinuð-in* (UVP) ‘be  
shaken’/*m-rinuð* ‘(earth) quake’
- {3} perception *riqað/m-riqað* ‘see, look’, *faðk-in* (UVP)/*f<m>aðik* ‘smell’
- {4} movement *u/a/m-u/a* ‘go’

Most ∅/MA and ∅/Z verbs do not appear in the dictionary in a form that unambiguously reveals the Kstem form, but the absence of *ka-* from derivations with the same root suggests that they are indeed ∅ verbs. Thao has evidently lost state forms in *ka-*, so that the ∅/MA has become the main repository of state verbs. As a result, ∅/MA encodes meanings at both ends of the hierarchy.

(34) Thao ∅/MA

- {1} agentive *tunaw/ma-tunaw* ‘try one’s hardest’
- {3} perception *damdā-in* (UVP)/*ma-damdā* ‘taste, eat’
- {6} invol. act. *aθay/m-aθay* ‘die’, *ma-laθa?* (*pa-laθa?* (CAUS)) ‘fall from a height’
- {10} state *kamun/ma-kamun* ‘be spicy’, *cakaw/ma-cakaw* ‘be greedy’, *ma-  
qusum* ‘be black’, *ma-raʔin* ‘be big’, *ma-briθ* ‘be heavy’

The ∅/Z class is also small and resembles ∅/MA in its distribution of meanings.



- (35) Thao Ø/Z  
 {3} perception *tmaða-n* (UVP)/*tmaða* ‘hear, listen’  
 {4} movement *itiða* ‘come back’  
 {6} invol. act. *taubukbuk* ‘stumble’, *mafuafuaw* ‘yawn’  
 {10} state *faqlu* ‘be new’, *dadu* ‘be true’, *sansan* ‘be warm’, *ifqala* ‘be drunk’,  
*mataq* ‘be raw’

A search of forms in *ka-* finds plentiful UVP *ka-* forms of mental event verbs but none of state verbs. Because state forms in *ka-* do not occur, class K/MA is unusual among F verb classes in occupying only the middle of the hierarchy.

- (36) Thao K/MA  
 {6} invol. act. *ka-luθ/ma-luθ* ‘sleep’, *ka-θawa-i* (UVP)/*ma-θa-θawa* (AV ITER) ‘laugh’  
 {8} mental *ka-ania/ma-ania* ‘be clever, use wits’, *ka-salpu/ma-salpu* ‘worry’,  
*ka-a/kað-in* (UVP)/*ma-a/kað* ‘be jealous of’, *ka-difqun-in* (UVP)/  
*ma-difqun* ‘dislike (s.t.)’, *ka-skin-in* (UVP R)/*ma-skin* ‘despise’,  
*ka-faðaq-in* (UVP)/*ma-faðaq* ‘know, understand’

In a pattern cognate with Puyuma (§3.3), involuntary activity verbs are sometimes encoded as derived stems with the anticausative prefix *mu-*, for example *mu-qila* ‘split (inchoative)’ versus *ma-qila* ‘split (state)’, and thus lie beyond the scope of this paper.

Thao has no K/MA state or resultative verbs, but it does have a number of Ø/MA resultatives: *ma-kutbað* ‘(cooked rice) dried out’, *ma-pilðuq* ‘dry, brittle’, *ma-qalayqay* ‘dry, withered’, *ma-qaliw* ‘(clothes, riverbed) dry’, *ma-qila* ‘split’, *ma-kiwar* ‘bent, curved’, but none of these corresponds to a Ø/U verb. The most plausible hypothesis to account for this circumstance is that each of these forms once corresponded with a Ø/U verb, but, because Thao has a complex derivational morphology, the plain Ø/U verbs have been lost.

The form–meaning correspondences in Thao verb classes are summarized in Table 9. They present a rather confused picture, but the absence of K/MA state forms is striking and K/MA verbs

**Table 9:** Semantic distribution of Thao verb classes

		Ø/U	Ø/MA	Ø/Z	K/MA
{1}	agentive	x	x (1)		
{2}	weather	x			
{3}	perception	x	x (1)	x (1)	
{4}	movement	x (1)		x (1)	
{6}	involuntary activity		x	x	x
{8}	mental event				x
{9}	resultative		x		
{10}	state		x	x	

occur only in the middle of the hierarchy. The division between A- and P-verbs lies between {4} and {6}.

## 5.4 Amis

Amis data are drawn from Tsuchida (1981) (various dialects), Fey (1986), Huang (1988), Liu (1999), Wu (2003, 2006) (central dialects), Tsukida (2008) (Fata'an, a north-central dialect), Chen (1987) and Isabelle Bril (personal communication 2013) (Natauran, a northern dialect). The dialect classification is Tsuchida's (1981).

The morphemes that define Amis verb classes have cognates in other F languages, but their distribution in Amis is different from the latter. Amis has two classes that correspond semantically to classes Ø/U and P/M of other F languages.

The first includes a small number of frequently used agentive verbs (see (37)) in which the infix <um> appears to be (partially) fossilized, insofar as it occurs not only in the AV realis but also (i) in AV irrealis, optative, hortative and negative forms and (ii) in certain UV forms. It occurs in these morphological contexts in no other F language.

In the second of the two classes the Kstem is formed with *pi-*, the AV realis with *mi-* (see (37)). It is thus a P/M (sub)class, but has become the productive agentive class in Amis, with far more members than the class in <um>. Hence many of the roots that occur in class Ø/U of other F languages occur in the Amis class *pi-/mi-*. Both Wu (2006) and Isabelle Bril (personal communication 2013) suggest that there are semantic differences between the two classes, and ongoing research will hopefully cast light on these.

The function of *ma-* has split into two in Amis. It remains the marker of the Mstem form of class K/MA, but it also marks the UVP realis form of erstwhile Ø/U and *pi-/mi-*, where it replaces *-en* in certain contexts. The basis for this split was presumably that the K/MA Mstem marker *ma-* marked its verb as non-agentive, that is its subject was patient-like. For a good many agentive verbs there was a corresponding resultative K/MA verb, and it was a short step to reanalysing the Mstem of the resultative as the UVP form of the {1} verb. As a result there appears to be no separate resultative verb category.

Perhaps associated with this change, *ka-*, the Kstem marker of K/MA verbs, also became the Kstem marker of erstwhile Ø/U and *pi-/mi-* verbs in some morphological contexts as well. As a result, the distribution of *ka-* is complex and is no longer a straightforward verb class diagnostic. Amis thus has four verb classes: U, *pi-/mi-*, MA and Z (Tsukida 2008). The MA class may be a merger of earlier Ø/MA and K/MA classes, the Z class of earlier Ø/Z and K/Z classes. Verbs of the small class U are mostly agentive, but also include some positional meanings.

- (37) Amis U  
       {1} agentive       *kum>aʔen* 'eat', *tum>iɾeŋ* 'stand up', *r>um>akat* 'walk', *tum>aŋic*  
                               'cry', *c>um>ikay* 'run'  
       {5} position       *t>em>iɾeŋ* 'stand', *m-aruq* 'sit (down)'
- (38) Amis *pi-/mi-*  
       {1} agentive       *mi-qaca* 'buy', *mi-caliw* 'borrow', *mi-palu* 'beat', *mi-adup* 'hunt',  
                               *mi-bacaq* 'launder', *mi-aŋaŋ* 'call (s.o.)'

The large MA class includes not only states, but also everything else except agentive verbs. No monomorphemic movement verbs occur in the data. Weather verbs in other F languages are denominal  $\emptyset$ /U verbs, and one would therefore expect them to be either U or *pi-/mi-* verbs in Amis, but they are not: they are formed with MA.

- (39) Amis MA
- |                 |   |
|-----------------|---|
| {2} weather     | <i>ma-urad</i> ‘rain’, <i>ma-kutem</i> ‘be cloudy’, <i>ma-bali</i> ‘be windy’, <i>ma-piyat</i> ‘(lightning) flash’, <i>ma-betili</i> ‘thunder’  |
| {3} perception  | <i>ma-melaw</i> ‘see’   |
| {5} position    | <i>ma-rikarik</i> ‘lie’   |
| {6} invol. act. | <i>ma-palal</i> ‘wake up’, <i>ma-butiq</i> ‘fall asleep’, <i>ma-suwab</i> ‘yawn’, <i>ma-?etut</i> ‘fart’, <i>ma-uta</i> ‘vomit’, <i>ma-kerker</i> ‘shiver’, <i>ma-tawa</i> ‘laugh (at)’             |
| {7} inan. act.  | <i>ma-icanj</i> ‘dry up’, <i>ma-likat</i> ‘become bright’, <i>ma-babaw</i> ‘cool down’, <i>ma-pelej</i> ‘break (ITR)’, <i>ma-tepo</i> ‘fall’, <i>ma-korokor</i> ‘roll (ITR)’                        |
| {8} mental      | <i>ma-banaq</i> ‘know (s.o./s.t.)’, <i>ma-ulah</i> ‘like, love (s.o.)’, <i>ma-ŋalay</i> ‘want (to), like’, <i>ma-talaw</i> ‘fear’   |
| {10} states     | <i>ma-ruhem</i> ‘be ripe’, <i>ma-patay</i> ‘die, be dead’, <i>ma-usuy</i> ‘be slow’, <i>ma-puti</i> ‘be clean’, <i>ma-su?su?</i> ‘be fat’, <i>ma-urip</i> ‘be alive’, <i>ma-laluk</i> ‘be diligent’ |

The large class Z contains only states.

- (40) Amis Z
- |             |  |
|-------------|--|
| {10} states | <i>kimulmul</i> ‘be round’, <i>kuhtij</i> ‘be black’, <i>fa?det</i> ‘be hot’, <i>puti</i> ‘be white’, <i>?adidiq</i> ‘be small’, <i>si?naw</i> ‘be cold’, <i>faŋcal</i> ‘be beautiful’ |
|-------------|--|

The agentive verbs in (41) are arguably collective in meaning, and are formed with MA.

- (41) Amis collective agentive verbs ({1a})
- ma-tayal* ‘work’, *ma-umah* ‘work in field’, *ma-kru?* ‘dance’

A summary of the discussion above is given in Table 10. With a few exceptions, the division between A- and P-verbs lies between {1} and {2}.

## 5.5 Saisiyat

Sources of Saisiyat data are Li (1978), Yeh (1991, 2003) and Zeitoun et al. (2015).<sup>19</sup> Saisiyat verb classes are  $\emptyset$ /U, P/M,  $\emptyset$ /MA,  $\emptyset$ /Z, K/MA and K/Z.

The  $\emptyset$ /U examples in (42) reflect an incompletely understood piece of Saisiyat morphological history since, as Yeh (2003) recognizes, PAn \* $\langle um \rangle$  has two reflexes,  $\langle om \rangle$  and  $m-$ , and the

<sup>19</sup> Zeitoun et al. (2015) came to my attention too late for me to take full account of its data and analysis.

**Table 10:** Semantic distribution of Amis verb classes

		U	<i>pi-/mi-</i>	MA	Z
{1}	agentive	x	x		
	{1a} collective			(x)	
{2}	weather			x	
{3}	perception			x (1)	
{5}	position	x		x (1)	
{6}	involuntary activity			x	
{7}	inanimate activity			x	
{8}	mental event			x	
{10}	state			x	x

phonological contexts in which they occur overlap.<sup>20</sup> Thus we find *kaas/k<om>aas* ‘bite’ but *karmaʔ/m-armaʔ* ‘steal’ with apparent loss of the first syllable of expected \**k<om>armaʔ*. The change affects certain *k*-initial roots. It also affects perhaps all *p*-initial roots, thereby shifting them into class P/M alongside *paθay/m-aθay* ‘die’, a P/M verb by inheritance (see (43)). Thus the cognate of *panaeʔ/m-anaeʔ* ‘shoot’ in most other F languages is a  $\emptyset$ /U verb (§6.3). These changes may be associated with the fact that Saisiyat appears to have lost the agentive class  $\emptyset$ /MA. At least some verbs in *m-* may be its remnants.

As in other F languages, class  $\emptyset$ /U is a large one, embracing agentive verbs and stretching to include all semantic categories except resultative and state.

- (42) Saisiyat  $\emptyset$ /U
- |                |  |
|----------------|--|
| {1} agentive   | <i>haŋih/h&lt;æm&gt;aŋih</i> ‘weep’, <i>ʔalop/ʔ&lt;æm&gt;alop</i> ‘hunt’, <i>ayap/&lt;om&gt;ayap</i> ‘fly’, <i>ʔaraj/m-araj</i> ‘bring’, <i>kaaθ/k&lt;om&gt;aaθ</i> ‘bite’, <i>karmaʔ/m-armaʔ</i> ‘steal’            |
| {2} weather    | <i>ʔintawah/m-intawah</i> ‘(weather) clear’, <i>ʔæral/ʔ&lt;æm&gt;æral</i> ‘rain’, <i>hælaʔ/h&lt;æm&gt;olaʔ</i> ‘snow’, <i>hila:/h&lt;æm&gt;ila:</i> ‘be sunny’, <i>ʔa:ðaw/ʔ&lt;æm&gt;a:ðaw</i> ‘be bright and sunny’ |
| {3} perception | <i>kitaʔ/k&lt;om&gt;itaʔ</i> ‘see’   |
| {4} movement   | <i>waiʔ/m-waiʔ</i> ‘come’  |
| {7} inan. act. | <i>βθæh/mo-βθæh</i> ‘explode’  |
| {8} mental     | <i>f&lt;om&gt;inaʔ</i> ‘want’  |

- (43) Saisiyat P/M
- |                 |   |
|-----------------|---|
| {1} agentive    | <i>panaeʔ/m-anaeʔ</i> ‘shoot’, <i>paatol/m-aatol</i> ‘sing’ |
| {6} invol. act. | <i>paθay/m-aθay</i> ‘die’, <i>pæʔrem/m-æʔrem</i> ‘sleep’    |

<sup>20</sup> There is also an irregular verb, *βoay/mo-βay* ‘give’.

Class  $\emptyset/Z$  embraces both ends of the hierarchy, as it does in Paiwan and Kavalan.

- (44) Saisiyat  $\emptyset/Z$
- |                 |   |
|-----------------|---|
| {1} agentive    | <i>raawak</i> ‘dance’, <i>mari?</i> ‘take’, <i>potoy</i> ‘wrap’                         |
| {4} movement    | <i>rima?</i> ‘go’, <i>lobih</i> ‘return’, <i>potɲor</i> ‘arrive’, <i>ra?iw</i> ‘depart’ |
| {6} invol. act. | <i>θawa?</i> ‘laugh’, <i>ekla?</i> ‘awake’  |

In Saisiyat the prefix \*ka- has undergone a change in host, and before most KA verbs it surfaces as -k, attached to a preceding vowel (Zeitoun 2001). Hence the causative prefix is *pa-* with a  $\emptyset$  Kstem but *pak-* with a K Kstem. A linker intervenes between the preverbal negator *?izi?* and a following verb, with the form *?i* before a  $\emptyset$  Kstem but *?ik* before a K Kstem. The preverbal UV future marker is *ka-* before a  $\emptyset$  Kstem but *kak* before a K Kstem. These differences allow one to distinguish between  $\emptyset$  and K verbs, but only when one of these morphemes precedes the verb. Just a few KA verbs retain an alternation between *ka-* and *ma-*, for example *ka?ɲel/ma?ɲel* ‘be slow’ (Zeitoun et al. 2015:272). In the course of these changes Saisiyat has apparently also lost the K/MA resultative category.

- (45) Saisiyat K/MA
- |            |  |
|------------|--|
| {8} mental | <i>?aδ?aðem/ma-?aδ?aðem</i> ‘think’                            |
| {10} state | <i>ka-?ɲel/ma-?ɲel</i> ‘be slow’, <i>ntæ?/ma-ntæ?</i> ‘be raw’ |

- (46) Saisiyat K/Z
- |                |  |
|----------------|--|
| {3} perception | <i>βaðæ?</i> ‘hear’, <i>θaðek</i> ‘smell’  |
| {8} mental     | <i>raam</i> ‘know’, <i>farara?</i> ‘like’  |
| {10} state     | <i>fafo?</i> ‘be new’, <i>θoβææh</i> ‘be big’, <i>?æ:ʔæiw</i> ‘be dry’, <i>?æyæh</i> ‘be ill’, <i>?itofan</i> ‘be short’, <i>βalβalay</i> ‘be slow’, <i>hopay</i> ‘be tired’, <i>kayðæh</i> ‘be good’, <i>ma?ɲel</i> ‘be slow’ |

Class  $\emptyset/MA$  includes position verbs and states. The state verbs in (47) are presumably erstwhile K/M verbs that have lost *ka-* as explained above.

- (47) Saisiyat  $\emptyset/MA$
- |              |   |
|--------------|---|
| {5} position | <i>faŋay/ma-faŋay</i> ‘take a rest’, <i>faeŋ/ma-faeŋ</i> ‘sit’, <i>?iririi?/m-iririi?</i> ‘stand’ |
| {10} state   | <i>book/ma-book</i> ‘be rotten’, <i>kaθkeθ/m-aθkeθ</i> ‘be cold’, <i>?iθ?æh/m-iθ?æh</i> ‘be wet’  |

Collective agentive verbs belong to class P/M, for instance *pæ?æ?æ?/m-æ?æ?æ?* ‘quarrel’.

The form–meaning correspondences of Saisiyat verb classes are summarized in Table 11. Because of the changes mentioned at the beginning of this section, the distribution of form–meaning correspondences may well have shifted over time. The division between A- and P-verbs is between {7} and {8}.

**Table 11:** Semantic distribution of Saisiyat verb classes

		Ø/U	P/M	Ø/MA	Ø/Z	K/MA	K/Z
{1}	agentive	x	x		x		
	{1a} collective	P/M (1)					
{2}	weather	x					
{3}	perception	x (1)					x
{4}	movement	x (1)			x		
{5}	position			x			
{6}	involuntary activity		x		x		
{7}	inanimate activity	x (1)					
{8}	mental event	x (1)				x	x
{10}	state			x		x	x

## 5.6 Rukai

Rukai has six dialects, with considerable differences between some of them. Zeitoun (2003) contends on morphosyntactic grounds that Mantauran forms a subgroup with Budai, Labuan and Tanan. The other two Rukai dialects, Tona and Maga, form a second subgroup. Sufficient data are available to sketch the verb classes of only Tanan (Li 1973; Formosan Language Archive<sup>21</sup>) and Mantauran (Zeitoun 2000a, 2007), both in the first group. Both dialects lack class K/Z.

Unlike other F languages Rukai has an accusatively aligned system of active and passive, but this appears to have no impact on its verb classes. The history of Rukai AV affixes is discussed in §6.2.2.

### 5.6.1 Tanan Rukai

Both the Kstem and the Mstem are available for only a limited number of Tanan Rukai verbs.

Class Ø/U verbs are mostly agentive, but with a scattering of other categories, as (48) shows. Tanan appears to have four classes: Ø/U, Ø/MA, Ø/Z and K/MA.

(48) Class Ø/U

{1} agentive	<i>kane/wa-kane</i> ‘eat’, <i>oŋol/wa-oŋol</i> ‘drink’, <i>laob/wa-laob</i> ‘burn’, <i>a[a/am-a]a</i> ‘take’
{2} weather	<i>odal/wa-odal</i> ‘rain’
{3} perception	<i>ceel/wa-ceel</i> ‘see’
{4} movement	<i>oa/am-oa</i> ‘go’

<sup>21</sup> Until 31st December 2014 the archive was online at <http://formosan.sinica.edu.tw/>.

- {5} position *takaynen/wa-takaynen* ‘sit’  
 {6} involuntary activity *ʔacay/wa-ʔacay* ‘die’, *aʔec/w-aʔec* ‘sleep’

Classes  $\emptyset$ /MA and  $\emptyset$ /Z are not well represented, the former only by agentives, the latter by agentives and states, as such, embracing both ends of the hierarchy, as it does in Paiwan, Kavalan and Saisiyat.

- (49) Class  $\emptyset$ /MA  
 {1} agentive *havo[ol/ma-havo[ol* ‘cheat, lie’, *kawriva/ma-kawriva* ‘talk’

- (50) Class  $\emptyset$ /Z  
 {1} agentive *rivo* ‘scold’, *davac* ‘walk’, *ʔaswala* ‘do first’  
 {10} state *edekay* ‘be short’, *bekelay* ‘be little’, *tikia* ‘be small’

Class K/MA includes resultatives and states.

- (51) Class K/MA  
 {8} resultative *ka-ʔaθak/ma-ʔaθak* ‘broken’ (cf. *ʔ-aθak* ‘break (CAUS)’), *ka-leḍa ma-leḍa* ‘bright’ (cf. *leḍa* ‘lighten (ITR)’)  
 {10} state *ka-ḍaw/ma-ḍaw* ‘big’, *ka-lisi/ma-lisi* ‘bad, angry’, *ka-beleṅ/ma-beleṅ* ‘high, tall’, *ka-du/ma-du* ‘cooked’, *ka-gimuc/ma-gimuc* ‘be angry’

It is probably dangerous to draw any conclusion from Table 12 about the division between A- and P-verbs in Tanan, the more so as the corresponding Mantauran tabulation in Table 13 shows an overlap which the Tanan data are perhaps not rich enough to pick up.

### 5.6.2 Mantauran Rukai

Mantauran evidently has just three classes:  $\emptyset$ /U,  $\emptyset$ /Z and K/MA. It lacks class  $\emptyset$ /MA, which is present in Tanan.

**Table 12:** Semantic distribution of Tanan Rukai verb classes

				$\emptyset$ /U	$\emptyset$ /MA	$\emptyset$ /Z	K/MA
	Agent:	{1}	agentive	x	x	x	
		{2}	weather	x (1)			
		{3}	perception	x (1)			
		{4}	movement	x (1)			
		{5}	positional	x (1)			
		{6}	involuntary activity	x			
		{9}	resultative				x
	Patient:	{10}	state			x	x



Class  $\emptyset/U$  is the principal agentive class, but includes verbs for the full range of the semantic hierarchy other than resultatives and states.

- (52) Class  $\emptyset/U$
- |                 |   |
|-----------------|---|
| {1} agentive    | <i>ʔoŋolo/o-ʔoŋolo</i> ‘drink’, <i>coroko/o-coroko</i> ‘jump’, <i>a[a/om-a[a</i> ‘take’,<br><i>omoomo/o-omoomo</i> ‘kiss’, <i>pilipili/o-pilipili</i> ‘scratch’                         |
| {2} weather     | <i>ðaʔolo/o-ðaʔolo</i> ‘rain’, <i>ðeðere/o-ðeðere</i> ‘thunder’, <i>koliʔi/o-koliʔi</i><br>‘be sunny’, <i>ðamare/o-ðamare</i> ‘(moon) appear’, <i>veʔeveʔe/o-veʔeveʔe</i><br>‘be windy’ |
| {3} perception  | <i>ceŋele/o-ceŋele</i> ‘see’, <i>siala[a/o-siala[a</i> ‘hear’   |
| {4} movement    | <i>oa/om-oa</i> ‘go’, <i>ðaace/o-ðaace</i> ‘leave’  |
| {5} position    | <i>iki/om-iki</i> ‘be at’   |
| {6} invol. act. | <i>ha[aje/o-ha[aje</i> ‘yawn’   |
| {8} mental      | <i>lihoʔo/o-[ihoʔo</i> ‘know’, <i>iʔa/om-iʔa</i> ‘like’   |

Class  $\emptyset/Z$  is much smaller and includes some agentive and some involuntary activity verbs.

- (53) Class  $\emptyset/Z$
- |                 |  |
|-----------------|--|
| {1} agentive    | <i>aviʔi</i> ‘give’, <i>paðalroʔo</i> ‘do at once’, <i>patoʔo</i> ‘tell’ |
| {6} invol. act. | <i>ʔomaca</i> ‘wake up’, <i>ʔapece</i> ‘sleep’                           |

Like their equivalents in Puyuma and Thao, inanimate activities often have stems derived with the anticausative prefix *mo-*, for instance *mo-ðaolo* ‘get mixed together’ (Zeitoun 2007:244–246). No examples with monomorphemic stems have been found.

Class K/MA is, as expected, the class that includes resultatives and states. But it also includes verbs of other categories, thus overlapping with class  $\emptyset/U$  at {4}, {6} and {8}, as Table 13 shows.

**Table 13:** Semantic distribution of Mantaaran Rukai verb classes

		$\emptyset/U$	$\emptyset/Z$	K/MA
{1}	agentive	x	x	
{2}	weather	x		
{3}	perception	x		
{4}	movement	x		x
{5}	position	x (1)		
{6}	involuntary activity	x	x	x
{8}	mental event	x		x
{9}	resultative			x
{10}	state			x

- (54) Class K/MA
- |                 |  |
|-----------------|--|
| {4} movement    | <i>ka-avaʔi/ma-avaʔi</i> ‘come’  |
| {6} invol. act. | <i>ka-ʔete/ma-ʔete</i> ‘die’   |
| {8} mental      | <i>ka-haʔaoco/ma-haʔaoco</i> ‘scold’, <i>ka-ðalame/ma-ðalame</i> ‘love’, <i>ka-ʔamaðe/ma-ʔamaðe</i> ‘dislike’, <i>ka-[akas/ma-[akas</i> ‘dislike’, <i>ka-sarivo/ma-sarivo</i> ‘miss’, <i>ka-celekere/ma-celekere</i> ‘think, ponder’   |
| {8} resultative | <i>ka-kahici/ma-kahici</i> ‘mixed up’ (cf. <i>o-kahici</i> ‘mix (ITR)’), <i>ka-ðalese/ma-ðalese</i> ‘malleable’ (cf. <i>o-ðalese</i> ‘knead’), <i>ka-[otoko/ma-[otoko</i> ‘broken’ (cf. <i>o-[otoko</i> ‘break (ITR)’), <i>ka-ŋotoʔo/ma-ŋotoʔo</i> ‘cut off, short’ (cf. <i>o-ŋotoʔo</i> ), <i>ka-vecake/ma-vecake</i> ‘broken in two’ (cf. <i>o-vecake</i> ), <i>ka-kilisi/ma-kilisi</i> ‘tight’ (cf. <i>o-kilisi</i> ) |
| {10} states     | <i>ka-pacai/ma-pacai</i> ‘be dry’, <i>ka-tako[a/ma-tako[a</i> ‘be bad’, <i>ka-[akase/ma-[akase</i> ‘be lazy’, <i>ka-caʔeme/ma-caʔeme</i> ‘be ill’, <i>ka-caleme/ma-caleme</i> ‘be rotten’, <i>ka-poli/ma-poli</i> ‘be white’   |

## 5.7 Tsou

Tsou is grammatically exceptional among F languages in a number of respects (Ross 2012b), including perhaps its verb classes. Tsou examples are culled from Tsuchida (1976), Szakos (1994), Huang (2002), Zeitoun (2005) and Chang (2009). They show no trace of the two classes reflecting PAn \*ka-, leaving the four classes Ø/U, P/M, Ø/MA and Ø/Z.

Kstem forms appear with the addition of either *-a* UVP or *-i* UVL. The statives *m-aski* ‘salty’ and *m-eecu* ‘thirsty’ appear to belong to class Ø/U. If so, this is the only instance among F languages of state verbs belonging to the Ø/U class. However, without corresponding Kstems, this remains uncertain.

- (55) Tsou Ø/U
- |                |   |
|----------------|---|
| {1} agentive   | <i>chú-a/c&lt;m&gt;úhu</i> ‘roast (s.t.)’, <i>án-a/bóni</i> ‘eat’, <i>ím-a/m-imo</i> ‘drink’, <i>tnís-i/m-ónsi</i> ‘weep (for)’, <i>pn-áa/póno</i> <sup>22</sup> ‘shoot’, <i>sóf-i/s&lt;m&gt;ófî</i> ‘thatch’, <i>skun-a/s&lt;m&gt;uhnu</i> ‘pinch’, <i>fóq-a/m-ufqo</i> ‘bury’ |
| {3} perception | <i>tali-i/t&lt;m&gt;ali</i> ‘hear’, <i>ait-i/baito</i> ‘see’  |
| {8} mental     | <i>tausn-i/t&lt;m&gt;ausni</i> ‘know about’, <i>cohiv-i/bochio</i> ‘know’   |
| {10} states    | <i>m-aski</i> ‘salty’, <i>m-eecu</i> ‘thirsty’  |

The available data do not allow an adequate analysis by semantic categories,<sup>23</sup> but verbs from the remaining classes are set out below.

- (56) Tsou P/M
- |                 |  |
|-----------------|--|
| {1} agentive    | <i>páv-i/m-ávo</i> ‘open (s.t.)’                       |
| {6} invol. act. | <i>m-coi</i> ‘die’ (cf. <i>o-pcoz-a/o-pcoi</i> ‘kill’) |

<sup>22</sup> Before *p-*, U is zero.

<sup>23</sup> Szakos (1994) includes a substantial dictionary, but the absence of morphological labelling and the manner of glossing render the material hard to analyse.

- (57) Tsou  $\emptyset$ /MA  
{1} agentive *chí-a/mó-chi* ‘put’, *fí-i/mo-fi* ‘give’, *rímc-a/mí-rmici* ‘swallow’, *voz-néni* (UVC)/*mó-voi* ‘draw water’, *srúr-i/mú-sruri* ‘thread (needle)’
- (58) Tsou  $\emptyset$ /Z  
{1} agentive *eobak-a/eobako* ‘beat’  
{10} states *mon?i-a/mon?i* ‘be soon’, *osni-a/osni* ‘be immediate’, *poha?v-a/poha?o* ‘be slow’, *a-asoez-a/a-asoye* ‘be stealthy’, *kaeb-a/kaebi* ‘be happy’

Most statives belong to class  $\emptyset$ /Z and many are prefixed with *a-*, but this *a-* does not reflect \*ka-, the expected reflex of which is *?a-*. In any case, *a-* occurs on both UV and AV forms, whereas one would expect \*ka- to occur only in UV forms.

Chang (2009) shows that a stative form may occur in three main contexts in Tsou: as a predicate, as in (59a); as the first lexical verb in a serial verb construction, as in (59b) and (59c); and as the second part of a compound verb, as in (59d).

- (59) a. **poha?o** ?e Pasuya ho mi-ta yahioa  
AV.slow NOM Pasuya SUB AV-3SG AV.work  
‘Pasuya is slow at his work.’
- b. mi=ta **a-upo~poha?o** coeconu  
AV.RLS=3SG ADV-RED~AV.slow AV.walk  
‘He walks slowly.’
- c. te=ko **a-upo~poha?v-a** hio-a  
IRR=2SG ADV-RED-slow-UV work-UV  
‘You should do your work slowly!’
- d. mi=?o **pasu-po~poha?o**  
AV=1SG AV.sing-RED-slow  
‘I sang slowly.’

A Tsou verbal predication is said always to begin with an auxiliary, which, if realis, agrees in voice with its lexical verb(s) as in (59b) and (59d) (Zeitoun 2005:267–269). The verbs in a serial construction also agree in voice as in (59c) and (59d). Neither fact is true of other F languages. Note that *a-* only occurs in (59b) and (59c), when the stative verb is the first lexical verb in a serial verb construction. Furthermore, in (59a), where *poha?o* is the sole predicate, there is no auxiliary. Either a stative verb as a sole predicate requires no auxiliary, or the *poha?o* is not a verb and *a-* functions as a verbalizer. These are matters for investigation.

On current knowledge, Tsou verb classes look strikingly different from those of other F languages, and this is a topic that needs careful research.

## 5.8 Saaroa

Saaroa data are culled from Tsuchida (1976) and Pan (2012). Saaroa verbs belong to five classes:  $\emptyset$ /U, P/M (one member),  $\emptyset$ /MA,  $\emptyset$ /Z and K/MA. No Saaroa movement or positional verbs with monomorphemic stems have been found, nor has a resultative category.

- (60) Saaroa Ø/U  
 {1} agentive *cavu/c<um>avu* ‘wrap’, *cuhu/c<um>uhu* ‘burn, start fire’, *aala/um-aala* ‘take’, *alava/m-alava* ‘bring’, *kirimi/u-kirimi* ‘search’, *pana/u-pana* ‘shoot’  
 {2} weather *usale/um-usale* ‘rain’  
 {3} perception *kita/k<um>ita* ‘looks, see’
- (61) Saaroa P/M  
 {3} invol. act. *paci?i/m-aci?i* ‘die’
- (62) Saaroa Ø/MA  
 {3} invol. act. *alusape/m-alusape* ‘sleep’, .../m-auaua ‘yawn’
- (63) Saaroa Ø/Z  
 {1} agentive *pu?a* ‘sell’, *avavu* ‘cook’  
 {3} perception *timala* ‘hear’  
 {10} states *varu?u* ‘new’, *silaje* ‘bright’, *tekele* ‘wrong’, *seesema* ‘dark’

In (63) *timala* reflects PAn \*timaLa ‘hear, listen’, and there is evidence that PAn verbs with trisyllabic roots were never infixed with \*<um>, that is, can be regarded as the trisyllabic variant of class Ø/U (§6.6.1).

What I take to be class K/MA has its Kstem in *a-* instead of expected *ka-*. Its distribution differs from that of Tsou *a-*, as the latter also occurs on the Mstem form of the verb. Three apparent K/MA verbs in (64) have vowel-initial roots, and I take them to be K/MA verbs from which the affixal *a-* has been deleted. An alternative analysis is that they represent the missing class Ø/MA, but one would then expect the class to have some members with a stem-initial consonant.

- (64) Saaroa K/MA  
 {8} mental *a-rumuku/ma-rumuku* ‘like (doing s.t.)’, *a-catia/ma-catia* ‘know’  
 {10} states *a-ηane/ma-ηane* ‘dry’, *a-lavai/ma-lavai* ‘drunk’, *a-tava?iu/ma-tava?iu* ‘red’, *a-vacuku/ma-vacuku* ‘full (stomach)’, .../ma-sulu ‘cooked’, *eleηese/m-eleηese* ‘long’, *ilii/m-ilii* ‘full’, *isu?u/m-isu?u* ‘thirsty’

The rather restricted findings of this section are summarized in Table 14. There seems to be a dividing line between A- and P-verbs between {6} and {8}, but more data may well change this picture.

## 5.9 Kanakanavu

Kanakanavu data are drawn from Tsuchida (1976, 2003), Mei (1982), Ho (1997), Wu (2014) and Liu (2014). No examples of Ø/MA or K/Z verbs have been found, but it is uncertain whether the data represent all Kanakanavu verb classes, as the Kstem of a number of verbs is not recorded.

**Table 14:** Semantic distribution of Saaroa verb classes

		Ø/U	P/M	Ø/MA	Ø/Z	K/MA
{1}	agentive	x			x	
{2}	weather	x (1)				
{3}	perception	x			x (1)	
{6}	involuntary activity		x (1)	x		
{8}	mental event					x
{10}	state				x	x

## (65) Kanakanavu Ø/U

- {1} agentive *kaʔn-/kʌm>áʔn* ‘eat’, *ála-/l[u]m-álaʔ* ‘take’, *itar-/m-itár* ‘wait’,  
*panáʔ-/mu-panáʔ* ‘shoot’, *tʌm>aŋi* ‘weep’, *ʔʌm>ánupu* ‘hunt’, *vuá-/*  
*mu-vua* ‘give’  
 {2} weather *um-ucanʌ* ‘rain’  
 {3} perception *cʌʔúr-/cʌm>úʔura* ‘see’  
 {5} position *mu-ciri* ‘stand’

## (66) Kanakanavu P/M

- {1} invol. act. *.../m-acai* ‘die’

## (67) Kanakanavu Ø/Z

- {1} agentive *tarakari* ‘order (to)’, *puʔa* ‘buy’  
 {4} movement *ivatu* ‘come’, *puaiʔi* ‘return (ITR)’<sup>24</sup>  
 {8} mental *tavareʔe* ‘know, be able to’  
 {10} state *tatia* ‘big’

## (68) Kanakanavu K/MA

- {10} state *ka-manʌŋ/ma-manʌŋ* ‘good’, *ka-casʌ/ma-caasʌ* ‘long’, *.../ma-*  
*nateru* ‘strong’, *.../ma-reŋeca* ‘tall’

## 5.10 Siraya

Siraya is extinct, but its grammar has been reconstructed from seventeenth century texts by Adelaar (2011), according to whom there are no Ø/MA or K/Z verbs in the texts. Examples are few, largely because Siraya verbs tend to have derived stems. There are no positional verbs with monomorphemic stems, and no forms that I can identify as resultatives. The Siraya texts include weather nouns, but apparently no weather verbs.

<sup>24</sup> An additional movement verb, *aranai/m-aranai* ‘come from’, probably has a derived stem.

- (69) Siraya Ø/U  
 {1} agentive *kan/k<m>an* ‘eat’, *īt/m-īt* ‘drink’, *talix/t<m>alix* ‘ask’, *xail/x<m>ail* ‘testify’, *likux/l<m>ikux* ‘do again’, *rau/r<m>au* ‘deny’, *saplat/s<m>aplat* ‘scourge’  
 {3} perception *kīta/k<m>īta* ‘see’, *īliḡix/m-īliḡix* ‘hear’  
 {4} movement *daraṅ/d<m>araṅ* ‘go away’
- (70) Siraya Ø/Z  
 {1} agentive *tūbo* ‘chop off’, *tuul* ‘grow’, *lupux* ‘kill’, *īlūb* ‘shut’  
 {4} movement *lūmad/lūmad* ‘arrive’, *irua/irua* ‘arrive’, *thabul/thabul* ‘leave, travel’
- (71) Siraya K/M  
 {6} invol. act. *ka-patey/ma-patey* ‘die’, .../ma-irip ‘sleep’, .../ma-tawa ‘laugh’  
 {7} inan. act. .../m-atūb ‘collapse’, .../ma-rkāx ‘dry up’  
 {8} mental *ka-muy/ma-muy* ‘want’, *ka-vāaṅo/ma-vāaṅo* ‘love’  
 {10} state *ma-takut* ‘be afraid’, *kma-kuptix* ‘be pure, clean’, *ma-pule* ‘white’, *kma-iraṅ* ‘big’, *ma-darikax* ‘be healthy’, *ma-varaw* ‘be sinful’, *ma-vañey* ‘be rich’

Collective agentive verbs appear to be K/MA, for example *ka-ayam-en/ma-ayam* ‘reap the harvest’.

Siraya form–meaning correspondences are summarized in Table 15. The boundary between A- and P-verbs falls between {4} and {6}.

### 5.11 Isbukun Bunun

This account of Isbukun verb classes is based on Yeh (n.d.a, n.d.b) and Zeitoun (2000b). The posited classes are shown in (72), but their analysis requires diachronic comment.

There are no reflexes of PAn \*<um> or PAn \*ka- in the Isbukun data. Instead, we find Ø/ma- verbs both where we expect Ø/U verbs and where we expect K/MA verbs. Scholars working

**Table 15:** Semantic distribution of Siraya verb classes

		Ø/U	Ø/Z	K/MA
{1}	agentive	x	x	
	{1a} collective			(x)
{3}	perception	x		
{4}	movement	x	x	
{6}	involuntary activity			x
{7}	inanimate activity			x
{8}	mental event			x
{10}	state			x

on other Bunun dialects note that there are two distinct *ma-* prefixes, one agentive, the other stative (De Busser 2009:332; Jiang 2012:82), and the same is evidently true of Isbukun. Causative derivation reveals the distinction: the causative of an agentive *ma-* verb is formed with *pa-*, of a stative *ma-* verb with *pi-*. It is reasonable to infer that earlier  $\emptyset/U$  and (agentive)  $\emptyset/MA$  verbs have in Isbukun been collapsed into a single class, here labelled  $\emptyset/MA_1$ , whilst earlier  $K/MA$  verbs have lost their reflex of PAN \*ka- and become the class here labelled  $\emptyset/MA_2$ . The resulting form–meaning correspondences of Isbukun verb classes are summarized in (72) and illustrated below. The  $\emptyset/Z$  category is barely attested.

- (72)  $\emptyset/MA_1$  <  $\emptyset/U$  agentive  
 $\emptyset/Z$  agentive  
 $\emptyset/MA_2$  <  $K/MA$  stative

The available data are insufficient to allow a thorough analysis by semantic categories, beyond the fact that agentive verbs are  $\emptyset/MA_1$  and  $\emptyset/Z$  verbs and state verbs are  $\emptyset/MA_2$ .

- (73) Isbukun  $\emptyset/MA_1$   
 {1} agentive *ludah-un/ma-ludah* ‘beat’, *budað-un/ma-budað* ‘break’, *patað-un/ma-patað* ‘kill’, *pasah-un/ma-pasah* ‘wash (clothes)’, *kaun-un/m-aun* ‘eat’, *sayv-un/ma-sayv* ‘give’  
 {8} mental *kadima-un/m-adima* ‘like’
- (74) Isbukun P/M  
 {6} invol. act. *m-atað* ‘die’ (cf. *ma-patað* ‘kill’)
- (75) Isbukun  $\emptyset/Z$   
 {1} agentive *hud-un/huud* ‘drink’, *tajis* ‘weep’, *kudip* ‘go’
- (76) Isbukun  $\emptyset/MA_2$   
 {10} state *ma-naskal* ‘be happy’, *ma-kazav* ‘be cold’, *ma-dahpa* ‘be sick’, *ma-piskav* ‘be quick’

Isbukun resembles Puyuma, Thao and Saisiyat in forming both movement and inanimate activity verbs with the prefix *mu-* (I do not know the Kstem form), for example *mu-dan* ‘walk’ (*daan* ‘road’), *mu-halhal* ‘fall’ (cf. *ma-halhal* ‘drop (ITR)’), *mu-pudað* ‘break (ITR)’ (cf. *ma-pudað* ‘break (ITR)’), *mu-latlat* ‘wear out (ITR)’ (cf. *ma-latlat* ‘tear (ITR)’)

## 5.12 Pazih

Pazih is a recently extinct F language, and the data on which the analysis here is based were collected when the language was moribund. Data were taken from Blust (1999a) and checked against Li & Tsuchida (2001), then a sweep was made through the latter’s dictionary entries.



The Mstem of Pazih  $\emptyset/U$  verbs always has a prefix *m-* or *mV-*. Perhaps because the language was moribund, there is variation in the form of this prefix. It was *mu-* by default, but *me-* before *-Ce...* and *m-* before a vowel. Sporadically *mu-* becomes *mi-* before *Ci...* and *ma-* before *Ca...* or *Cu...* Li & Tsuchida record a number of Blust's *ma-* verbs with *mu-*, and others with both *ma-* and *mu-*. This leads to ambiguity, as it is difficult to decide whether this *ma-* is an exponent of U or of MA. In fact, the accumulated data contain no unambiguously  $\emptyset/MA$  verbs, namely with *ma-* before *Ce...* or *Ci...*, and it seems that Pazih has probably collapsed the  $\emptyset/U$  and  $\emptyset/MA$  classes. In light of this, Pazih has classes  $\emptyset/U$ ,  $\emptyset/Z$ , K/MA and K/Z.

The Kstem form of the verbs of the classes in (77) is identified by an imperative form: either the AV form (plain Kstem) or the UVP form (Kstem-*i*).

A majority of verbs in class  $\emptyset/U$  are agentive, but its members stretch down the hierarchy as far as {7}.

- (77) Pazih  $\emptyset/U$
- |                 |  |
|-----------------|--|
| {1} agentive    | <i>baked-i/mu-baked</i> 'hit', <i>kan-i/me-ken</i> [sic] 'eat', <i>pana-i/ma-pana</i> 'throw (at target)', <i>anjid-i/m-anjid</i> 'cry', <i>ara-i/m-ara</i> 'take', <i>baxa-i/mu-baxa</i> 'give', <i>bazu-i/mu-bazu</i> 'wash (clothes)' |
| {2} weather     | <i>ma-kuras</i> 'thunder'/ <i>kuras</i> 'thunder (N)', <i>ma-bari</i> '(wind) blow'/ <i>bari</i> 'wind (N)', <i>ma-hahela</i> 'snow'/ <i>hahela</i> 'snow (N)', <i>ma-laliw</i> '(earth) quake'/ <i>laliw</i> 'earthquake (N)'           |
| {3} perception  | <i>talam-i/mu-talam</i> 'taste', <i>sazek-i/mu-sazek</i> 'smell', <i>kita[-i]/mi-kita</i> 'see, look at'   |
| {4} movement    | <i>usa/m-usa</i> 'go (away)', <i>puzah/mu-puzah</i> 'come back'  |
| {5} position    | <i>ituku-i/m-ituku</i> 'sit', <i>itul/m-itul</i> 'stand up' (perhaps originally derived forms in <i>i-/m-i-</i> )  |
| {6} invol. act. | <i>xatis-i/mu-xatis</i> 'sneeze', <i>idem-i/m-idem</i> 'sleep'   |
| {7} inan. act.  | <i>m-anat</i> 'burn (ITR)' (cf. <i>pa-anat</i> 'burn (ITR)'), <i>ma-taja</i> 'open (ITR)' (cf. <i>pa-taja</i> 'open (ITR)')  |

The remaining verb classes are non-agentive, and the data generally do not include an imperative form. Here the causative is used to show the presence or absence of *ka-*.

Class  $\emptyset/Z$  has few members.

- (78) Pazih  $\emptyset/Z$
- |                |   |
|----------------|---|
| {3} perception | <i>tumala</i> 'hear, listen'                          |
| {7} inan. act. | <i>pa-talubig</i> (CAUS)/ <i>talubig</i> 'fall, drop' |

Pazih *tumala* reflects PAN \*timaLa 'hear, listen', which adheres to the apparent PAN rule that verbs with trisyllabic roots were never infixed with \* $\langle um \rangle$  (§6.6.1).

A majority of state verbs are in class K/MA, but its members extend up the hierarchy to level {6}.

- (79) Pazih K/MA
- |                 |  |
|-----------------|--|
| {6} invol. act. | <i>pa-ka-suhaw/ma-suhaw</i> 'yawn'                             |
| {7} inan. act.  | <i>pa-ka-urit</i> (CAUS)/ <i>ma-urit</i> 'collapse, fall down' |

{8} mental	<i>ka-baza-en</i> (UVP)/ <i>ma-baza~mu-baza</i> ‘know (s.t.)’, <i>ka-ŋesell/ma-ŋesell</i> ‘fear’
{10} state	<i>pa-ka-azih/ma-azih</i> ‘ripe, cooked’, <i>pa-ka-karit/ma-karit</i> ‘dry’, <i>pa-ka-kelem/ma-kelem</i> ‘salty’, <i>pa-ka-kuris/ma-kuris</i> ‘thin, skinny’, <i>pa-ka-akux/m-akux</i> ‘hot (weather)’, <i>pa-ka-ader/m-ader</i> ‘hard, tight’

The data include only one resultative verb, supported by a corresponding agentive verb: *ma-telej* ‘caught in trap’ (K form in *pa-ka-telej* ‘cause to be caught in trap’), corresponding with *telej-en* ‘get caught’, UVP form of an agentive verb ‘catch in a trap’.

Class K/Z also mainly encodes statives, particularly inherent properties.

(80) Pazih K/Z

{8} mental	<i>ka-hapet/haapet</i> ‘like, love, want to’
{10} state	<i>pa-ka-tatih/tatih</i> ‘little’, <i>pa-ka-baged/baged</i> ‘fat, obese’, <i>pa-ka-eder/eder</i> ‘true’, <i>pa-ka-haidej/haidej</i> ‘heavy’, <i>pa-ka-lamik/lamik</i> ‘cool, cold’, <i>pa-ka-sem/sem</i> ‘dark’

The data and analysis above is summarized in Table 16. Pazih evidently had a boundary between A- and P-verbs between {7} and {8}.

### 5.13 Teruku Seediq

Sources of data are Li (1978), Pecoraro (1979) and, with regard to verb class assignment, Tsukida (2005, 2009:193–237). The data include no inanimate activity verbs, and some other categories of the hierarchy are only scantily represented, but Teruku clearly has classes  $\emptyset/U$ ,  $\emptyset/MA$ ,  $\emptyset/Z$ , K/MA and K/Z.

**Table 16:** Semantic distribution of Pazih verb classes

		$\emptyset/U$	$\emptyset/Z$	K/MA	K/Z
{1}	agentive	x			
	{1a} collective			(x)	
{2}	weather	x			
{3}	perception	x	x (1)		
{4}	movement	x			
{5}	position	x			
{6}	involuntary activity	x		x (1)	
{7}	inanimate activity	x	x (1)		
{8}	mental event			x	x (1)
{9}	resultative			x (1)	
{10}	state			x	x

Class  $\emptyset/U$  encodes agentive verbs but also stretches down the hierarchy as far as {8}, mental activities.

- (81) Teruku  $\emptyset/U$
- |                |   |
|----------------|---|
| {1} agentive   | <i>bahu/b&lt;en&gt;ahu</i> ‘wash (clothes)’, <i>quti/q&lt;em&gt;uti</i> ‘excrete’, <i>kerut/k&lt;em&gt;erut</i> ‘cut’, <i>adas/m-adas</i> ‘carry’, <i>reñaw/r&lt;em&gt;eñaw</i> ‘talk’, <i>kiyux/k&lt;em&gt;iyux</i> ‘push’ |
| {2} weather    | <i>quyux/q&lt;em&gt;uyux</i> ‘rain’   |
| {3} perception | <i>qita/q&lt;em&gt;ita</i> ‘look, see’, <i>qbahan/[q-e]m-bahan</i> ‘listen, hear’   |
| {4} movement   | <i>ʔusa/m-usa</i> ‘go’, <i>ʔiyah/m-iyah</i> ‘come’  |
| {8} mental     | <i>lenlun/l&lt;em&gt;enlun</i> ‘think’, <i>sekuxul/s&lt;em&gt;ekuxul</i> ‘like’   |

Class  $\emptyset/MA$  is exceptional among F languages in covering much of the hierarchy from agentive verbs to states. Even its close relative, Mayrinax Atayal, does not extend the functions of  $\emptyset/MA$  in this way.

- (82) Teruku  $\emptyset/MA$
- |                 |   |
|-----------------|---|
| {1} agentive    | <i>diras/me-diras</i> ‘shout’, <i>karaw/me-karaw</i> ‘climb (tree)’   |
| {5} position    | <i>sañay/me-sañay</i> ‘rest’, <i>hiyug/me-hiyug</i> ‘stand’   |
| {6} invol. act. | <i>huqil/me-huqil</i> ‘die, be dead’, <i>tutiñ/me-tutiñ</i> ‘fall, stumble’, <i>takur/me-takur</i> ‘stumble’, <i>taqi/me-taqi</i> ‘sleep, be asleep’, <i>hulis/me-hulis</i> ‘laugh’ |
| {8} mental      | <i>kela/me-kela</i> ‘know’  |
| {10} state      | <i>tutuy/me-tutuy</i> ‘be awake’, <i>rudu/me-rudu</i> ‘be dishevelled’, <i>rutiq/me-rutiq</i> ‘be dirty’, <i>huriq/me-huriq</i> ‘be wet’, <i>dawi/me-dawi</i> ‘be lazy’             |

Class  $\emptyset/Z$  is a much smaller class but covers much of the same ground as  $\emptyset/U$ .

- (83) Teruku  $\emptyset/Z$
- |              |  |
|--------------|--|
| {1} agentive | <i>deñday</i> ‘boil water’, <i>terima</i> ‘bathe’, <i>tegeliqin</i> ‘hide oneself’, <i>sekiya</i> ‘fly’, <i>seteqtu</i> ‘go up, climb’ |
| {4} movement | <i>dehuq</i> ‘arrive’, <i>qeduriq</i> ‘flee’   |
| {5} position | <i>teluʔun</i> ‘sit’   |

Class  $K/MA$  includes resultatives and states. It appears to include nothing higher in the hierarchy, and in this it contrasts with the  $K/MA$  class in other F languages.

- (84) Teruku  $K/MA$
- |                 |   |
|-----------------|---|
| {8} resultative | <i>ke-hurah/me-hurah</i> ‘be broken’/(cf. <i>h&lt;em&gt;urah</i> ‘destroy (s.t.)’), <i>ke-gerun/me-gerun</i> ‘be broken’/(cf. <i>g&lt;em&gt;erun</i> ‘destroy (s.t.)’), <i>ke-deñul/me-deñu</i> ‘be dry’ (cf. <i>d&lt;em&gt;eñu</i> ‘dry (s.t.)’) |
| {10} state      | <i>ke-qudus/me-qudus</i> ‘be alive’, <i>ke-rutiq/me-rutiq</i> ‘be dirty’, <i>ke-keray/me-keray</i> ‘be hard’, <i>ke-nihur/me-nihur</i> ‘be bitter’, <i>ke-sekuy/me-sekuy</i> ‘be cold’  |

**Table 17:** Semantic distribution of Teruku Seediq verb classes

		Ø/U	Ø/MA	Ø/Z	K/MA	K/Z
{1}	agentive	x	x	x		
{2}	weather	x (1)				
{3}	perception	x				
{4}	movement	x		x		
{5}	position		x	x		
{6}	involuntary activity		x			
{8}	mental event	x	x			
{9}	resultative				x	
{10}	state		x		x	x

Class K/Z is apparently devoted to inherent qualities.

(85) Teruku K/Z

{10} state      *ke-banah/banah* ‘be red’, *ke-paru/paru* ‘be big’, *ke-bilaq/bilaq* ‘be small’, *ke-malu/malu* ‘be good’, *ke-behegay/behegay* ‘be white’

The form–meaning correspondences of Teruku Seediq verbs are shown in Table 17. Along with its close relative, Mayrinax Atayal, Teruku Seediq draws the line between A- and P-verbs further down the hierarchy than other F languages, namely between {8} and {9}.

## 5.14 Mayrinax Atayal

Sources are Li (1978), Elizabeth Zeitoun (personal communication 2003) and especially Huang (2000), who provides a semantic analysis of Mayrinax verb classes. Huang sees Mayrinax verb classes as forming a cline from dynamic to stative. Ø/U verbs are the most dynamic, followed by Ø/MA and Ø/Z verbs which are dynamic but intransitive or less agentive, followed in turn by K/MA verbs that encode a state resulting from an action, and finally K/Z verbs encoding simple states. There are, she says, a few exceptions in each class.

A sample of agentive verbs from Huang’s dynamic classes is given in (86)–(89). The overall distribution of the Mayrinax classes is quite similar to that of Teruku Seediq, its close relative, except that the latter has extended the functions of Ø/MA where Mayrinax hasn’t.

Mayrinax shows the results of a process similar to the split of Saisiyat U into <um> and *m-*, and as in Saisiyat the Mayrinax outcomes are unpredictable. Thus, in *βahiy/m-ahiy* ‘beat, hit’, *pasiaq/m-asiaq* ‘laugh’, *kitaal/m-itaal* ‘see’ and *qaniq/m-anig* ‘eat’ the root-initial obstruent is replaced by *av m-*. As a result, some *p*-initial roots end up in class P/M, as they do in Saisiyat. In the case of *kitaal* both *k<um>itaal* and *m-itaal* occur.

- (86) Mayrinax Ø/U
- {1} agentive *hakay/h<um>akay* ‘walk’, *aras/m-aras* ‘bring, take’, *cuβuq/c<um>uβuq* ‘roast’, *βahiy/m-ahiy* ‘beat, hit’, *qaniq/m-anig* ‘eat’, *kaluap/k<um>aluap* ‘hunt’
  - {2} weather *quwalax/q<um>uwalax* ‘rain’, *m-aβicu* ‘thunder’
  - {3} perception *kitaal/k<um>itaal*, *m-itaal* ‘see’, *m-uŋ* ‘listen, hear’, *s<um>auq* ‘smell’
  - {4} movement *usaq/m-usaq* ‘go’, *uwah/m-uwah* ‘come’
  - {5} position *s<um>iahay* ‘lie (down)’, *m-anthawnaq* ‘sit’, *m-atyayaay* ‘lie (on bed)’
  - {6} invol. act. *s<um>apiyal* ‘dream’, *s<um>ijuhuy* ‘snore’, *m-asuway* ‘yawn’, *m-itutul* ‘wake up’, *m-βilβil* ‘tremble’, *m-inuqil* ‘die’
  - {7} inan. act. *h<um>βiŋ* ‘leak’
  - {8} mental *lanluŋ/l<um>aŋluŋ* ‘think’, *siuwal/s<um>iuwal* ‘like’
- (87) Mayrinax P/M
- {1} agentive *pasiaq/m-asiaq* ‘laugh’
- (88) Mayrinax Ø/MA
- {1} agentive *qagal/ma-qagal* ‘take’, *qtuβainay/ma-qtuβainay* ‘sell’, *βahuq/ma-βahuq* ‘wash (clothes)’, *qaqaquh/ma-qaqaquh* ‘be busy’
  - {6} invol. act. *qilaap/ma-qilaap* ‘sleep’, *qianux/ma-qianux* ‘live (not die)’
- (89) Mayrinax Ø/Z
- {1} agentive *panaip* ‘fish’ ‘be busy’, *paqati* ‘throw away’, *kaβaux* ‘borrow’, *tulqin* ‘hide (ITR)’, *psihow* ‘suck’
  - {4} movement *awuni* ‘come’, *aruwa* ‘leave’, *tayhok* ‘arrive’
  - {6} invol. act. *psacoŋ* ‘cough’
  - {7} inan. act. *psiβqaq* ‘explode’
- Huang’s suggestion that K/MA verbs encode a state resulting from an action is true of many members of the class, but not of all.
- (90) Mayrinax Atayal state verbs ({10})
- {7} inan. act. *ka-taquq/ma-taquq* ‘fall down’
  - {8} resultative *ka-βuqaq/ma-βuqaq* ‘broken’ (cf. *βuqaq/m-qaq* ‘break (a bowl)’), *ka-yaqaŋ/ma-yaqaŋ* ‘missing, lost’ (cf. *yaqaŋ/y<in>aqaŋ* ‘lose’)
  - {10} state *ka-qicuq/ma-qicuq* ‘be afraid’, *ka-qoβaq/ma-qoβaq* ‘white; clean’, *ka-qoway/ma-qoway* ‘tired’, *ka-hnuq/ma-hnuq* ‘cheap; soft’, *ka-saβiq/ma-saβiq* ‘lazy’, *ka-tanah/ma-tanah* ‘red, ripe’, *k-uhaal/m-uhaal* ‘sick’

K/Z verbs are at the stative extreme of Huang’s cline. They usually encode states that are not the result of an action.

**Table 18:** Semantic distribution of Mayrinax Atayal verb classes

		∅/U	P/M	∅/MA	∅/Z	K/MA	K/Z
{1}	agentive	x	x	x	x		
{2}	weather	x					
{3}	perception	x					
{4}	movement	x			x		
{5}	position	x					
{6}	involuntary activity	x		x	x (1)		
{7}	inanimate activity	x (1)			x (1)	x (1)	
{8}	mental event	x					
{9}	resultative					x	
{10}	state					x	x

## (91) Mayrinax K/Z ({10})

{10} states *ka-capay/capay* ‘old (not new)’, *ka-isitiŋ/isitiŋ* ‘short’, *ka-qihma/qihma* ‘thick’, *ka-kithuq/kithuq* ‘fat (by nature)’, *ka-lawkah/lawkah* ‘strong’, *ka-rahuwal/rahuwal* ‘big’

Table 18 summarizes this section. Like Teruku Seediq (§5.13), Mayrinax Atayal draws the line between A- and P-verbs low in the hierarchy, between {8} and {9}.

## 6. Reconstructing PAn verb classes: combining two approaches

This section draws together the evidence from the subsections of §5 and proposes a construction-based reconstruction of PAn verb classes, summarized in Tables 19 and 20. This is collated with lexical reconstructions of the Kstem and Mstem of verbs in four of the six verb classes. No lexical reconstructions are made in classes ∅/MA or K/Z.

**Table 19:** The distribution of verb classes across Formosan languages

x = the class is present; – = the class is absent

Language key:									Puy BunI	Tso Tha	RukTa Paz	RukM Sai	Saa SedT	Kan AtaM	Sir Kav	Pai Ami									
	U								MA								Z								
∅	x	x	x	x	x	x	x	x	x	x	x	—	x	—	—	—	x	x	x	x	x	x	x	x	x
	—	x	x	x	x	x	x	x	x	x	x	—	x	x	x	x	—	x	x	x	x	x	x	x	—
K									x	—	x	x	x	x	x	x	x	x	—	—	—	—	—	—	x
									x	x	x	x	x	x	x	x	x	—	—	x	x	x	x	x	x

**Table 20:** Semantic distribution of verb classes in Formosan languages

Cell key:		the class is present in this semantic category										Note: P/M verbs are counted as $\emptyset$ /U verbs.											
		just one member of the class has been found in this semantic category																					
		the class is absent from this semantic category																					
		relevant evidence is missing																					
–		the class is not present in this language																					
				$\emptyset$ /U										$\emptyset$ /MA									



**Table 21:** The number of verbs in each PAn verb class reconstructed by the author

	$\emptyset/U + P/M$	$\emptyset/MA$	$\emptyset/Z$	K/MA	K/Z	Total
number	57 + 1	0	2	8	0	68
percent	83.8 + 1.5%	—	2.9%	11.8%	—	100%

Table 19 shows the distribution of verb classes across F languages. The immediate impression created by the table is that classes  $\emptyset/U$ ,  $\emptyset/Z$  and K/MA are certainly reconstructable, while  $\emptyset/MA$  and K/Z are reasonably well supported. However, when we look at Table 20, showing the distribution of the verb classes by semantic category, a more nuanced picture emerges.

Table 20 shows a number of facts. An obvious one is that in {2} to {8} a good many cells contain ‘1’, that is, only one member of this verb class has been found in this semantic category. This reflects the fact that in any F language (and perhaps many languages elsewhere) most verbs fall into {1} or {10}. Table 21 gives the number of verbs that I have reconstructed in each PAn verb class. The distributional imbalance is obvious. The categories between {1} and {10} are minority categories where A- and P-verbs tend to overlap, resulting in messy-looking distributions in {2} to {8}.

The P/M class is omitted from this and subsequent tables, since it appears to have been a spin-off from class  $\emptyset/U$  and had only one monomorphemic member, *\*paCay/\*m-aCay* (§6.3).

## 6.1 The distribution of verb classes

A salient fact about Table 20 is the complementary distribution of classes  $\emptyset/U$  and K/MA. Class  $\emptyset/U$  is reflected in {1}, agentive, in all languages except Isbukun Bunun, where it has merged with  $\emptyset/MA$ , transferring all its members to  $\emptyset/MA_1$  (§5.11). It is also reflected in {2}, weather, and {3}, perception, in most languages except Amis, where MA has spread up the hierarchy as far as {2} (§5.4). Class  $\emptyset/U$  is *not* reflected in {9}, resultatives, or {10}, states. Conversely class K/MA is reflected in {9} and {10} (§5.3), except in Thao, where it has merged with  $\emptyset/MA$ , apparently transferring all its members to  $\emptyset/MA$ . It is *not* reflected in {1}, {2} or {3} except in Amis.

This suggests strongly that PAn *\* $\emptyset/U$*  should be reconstructed in categories {1}, {2} and {3} and PAn *\*K/MA* in categories {9} and {10}. Table 20 also suggests that PAn *\* $\emptyset/U$*  occurred in categories {4}, movement, and {5}, position, with greater frequency than PAn *\*K/MA*.

However, such a statement cannot be made about either *\* $\emptyset/U$*  or *\*K/MA* with regard to categories {6}, {7} and {8}. Here Table 20 shows an overlap in modern languages, as there probably was in PAn. This is no surprise. Because {6}, involuntary activity, and {7}, inanimate activity, both entail activity but are non-agentive, they may be construed either as agentive-like activities or as non-agentive processes. Category {8}, mental event, entails an experiencer, which across the world’s languages is variously construed as agent-like or patient-like.

Lexical reconstructions of PAn verbs belonging to the various classes are presented in §6.2–§6.7. In each cognate set those cognates that reflect the reconstructed Mstem are listed first, followed by cognates that appear to have shifted class in parentheses. The Kstem is not given where it is predictable from the Mstem and the verb class. Where the form of the Kstem is not predictable, the

Mstem is preceded by a form which allows the Kstem form to be isolated. For example, in the Tsou example in (92a) the Mstem form *m-onʒi* is preceded by the location-subject undergoer-voice (UVL) form *tnʒis-i*, from which the Kstem form *tnʒis-* can be isolated.

## 6.2 Class Ø/U

### 6.2.1 Lexical reconstructions

Table 21 shows that Ø/UM was far and away the largest verb class, and it is no surprise that a number of lexical reconstructions are possible in categories {1} agentive, {2} weather, {3} perception and {4} movement.<sup>25</sup>

(92) PAn Ø/U agentive ({1})

- a. **PAn** \*C<um>anʒis ‘weep’ > **Tso** *tnʒis-i* (UVL)/*m-onʒi* **Kan** t<um>anʒi **Saa** t<um>anʒi **Tha** θ<um>anit **Bab** s<um>ani **Sai** h<em>anʒih **Sir** t<um>anʒi **Ami** t<um>anʒic (but **Puy** *ma-tanʒis*)
- b. **PAn** \*C<um>aqis ‘sew’ > **PuyT** t<em>a-taʒiq (IRR) **RukB** *ua-caisi* **Tso** tʔes-a (UVP)/t<um>eʔsi **Kan** t<um>a-taʔisi (IPFV) **Pai** ts<em>aqis **BunTt** *ma-taqais* **Tha** ʃ<um>aqif **Fav** s<umm>is **Paz** *mu-sais* **Sai** ʃ<om>æʔif **AtaM** c<um>aqis **SedT** s<um>ais **Sir** t<um>ahy **Kav** t<um>ais
- c. **PAn** \*C<um>uLuh ‘roast in fire’ > **RukB** *ua-culu* **Tso** *chu-a/c<um>uhu* **Kan** c<um>á-cunu (IPFV) **Saa** c<um>uʔu **Pai** ts<em>ulu **BunTt** *ma-tunuh* **Sai** θ<om>ulcəh **AtaM** c<um>uluh (but **Ami** *mi-tuluh* ‘burn’)
- d. **PAn** \*k<um>aen ‘eat’ > **RukB** *ua-kane* **Tso** ʔan-a (UVP)/b-oni **Kan** k<um>aén **Pai** k<em>an **BunI** *kaʔun-av* (IMP)/maʔun **Tha** k<um>an **Fav** *m-an* **Paz** *me-ken* **AtaM** *qaniq/m-anig* **Sir** k<um>an **Ami** k<um>aʔen **Kav** *qan/q<um>an* (but **Puy** *m-ekan*, **SedT** ʔekan/m-eʔkan<sup>26</sup>)
- e. **PAn** \*k<um>iRim ‘seek’ > **Tso** *iim-a* (UVP)/bi-bimi **Kan** k<um>írimi **Pai** k<em>im **Tha** k<um>iʔim **Sai** *kiʔim/k<om>iʔim* **AtaM** *k-um-iyum* **Sir** *kiim/k<um>iim* **Kav** *kirim-i/k<um>irim* (but **Saa** *kirimi/u-kirimi* **Ami** *mi-kilim*)
- f. **PAn** \*t<um>enuʔun ‘weave’ > **RukM** *o-tiono* **Tso** *m-invini* **Kan** t<um>a-tinun **Saa** t<um>ineene **Pai** c<em>enun **Tha** *tnun-in/t<u>nun* **Fav** t<um>enon **Paz** *mu-tuʔul* **Sai** t<om>non **AtaM** *cinun/t<um>inun* **SedT** t<um>inun **Kav** *tnun/t<um>nun* (but **Ami** *mi-tenuʔun*)

<sup>25</sup> Abbreviations of language names used in the examples are the first three letters of the language name, except that Seediq is abbreviated **Sed**. Languages not previously mentioned are the extinct Plains languages **Babuza**, **Favorlang** and **Taokas**. Dialects are specified by additional letters: **AtaM** *Mayrinax Atayal*; **BunI** *Isbukun Bunun*, **BunTt** *Takitudah Bunun*; **PuyK** *Katripul Puyuma*, **PuyT** *Tamalakaw Puyuma* (**Puy** is *Nanwang Puyuma*); **RukB** *Budai Rukai*, **RukM** *Mantauran Rukai*; **SedT** *Teruku Seediq*. Additional lexical data are drawn from Kuo 1985 (RukB), Li 1973, 1977 (Ruk), 1981 (Ata and Sed), 1988 (Bun); 2003 (Fav); Li & Tsuchida 2002 (Paz); Sung & Shen 2006 (Kav); Tsuchida 1969 (Bun), 1982 (Bab, Fav, Tao), 1983 (Puy); Zeitoun & Teng 2014 (Kan, Saa).

<sup>26</sup> Addition of initial *e-* to *-kan* may reflect the restoration of disyllabicity to a root that had been reduced to a monosyllable.

- g. **PAn** \*p(⟨um⟩)anaq<sup>27</sup> ‘shoot (arrow)’ > **PuyT** p⟨en⟩a-panaq (IRR) **RukM** o-pana **Tso** pn-aa/pono **Saa** u-pana **Kan** mu-paná? **Pai** p⟨en⟩anaq **BunTv** panaq-an/m-anaq **Tha** panaq ‘shoot, throw’ **Paz** mu-pana ‘throw at’ **Sai** panæ?/m-anæ? **Kav** p⟨m⟩ani (but **Ami** mi-panaq)

The Bunun and Saisiyat reflexes in (92g) look deceptively as if they reflect a P/M verb. The Saisiyat reflex, however, is regular (§5.5), and the Bunun reflex may be too, but Bunun morphophonemics are not well enough understood to be sure.

Weather verbs are formed by verbalising a root that also occurs alone as a noun.

(93) PAn ∅/U weather ({2})

- a. **PAn** \*q⟨um⟩uzaL ‘rain’, \*quzaL ‘rain (N)’ > **RukB** ua-udal **Tso** m-ichi **Kan** ?⟨um⟩ucan **Saa** um-usaŋe **Pai** q⟨em⟩ujal **Tha** q⟨m⟩usað **Fav** m-utas **Paz** m-udal **Sai** ?⟨æm⟩oral **AtaM** q⟨um⟩uwal-ax **Sir** udal **Kav** m-uzan (but **Puy** ma-?udal **Ami** ma-urad)
- b. **PAn** \*b⟨un⟩ali ‘(wind) blow’, \*bali ‘wind (N)’ > **Pai** v⟨en⟩alɿ **Paz** ma-bari **Kav** u-bari (**Ami** ma-bali ‘be windy’)
- c. **PAn** \*x⟨um⟩uReLa ‘snow’, \*xuReLa ‘snow (N)’ > **Pai** s⟨em⟩ula **Kav** s⟨m⟩uRna **Sai** hœla?/h⟨æm⟩ola? **Paz** ma-hahela

(94) PAn ∅/U perception ({3})

- a. **PAn** \*k⟨um⟩ita ‘see’ > **Tso** a-it-i (UvL)/b-a-ito **Saa** k⟨um⟩ita **Sirk**⟨m⟩ita **Paz** mi-kita **Sai** k⟨om⟩ita? **SedT** q⟨em⟩ita **AtaM** k⟨um⟩itaal, m-itaal **Kav** m-qita
- b. **PAn** \*t⟨um⟩aLam ‘taste (s.t.)’ > **Tha** t⟨m⟩aðam **Paz** mu-talam **Kav** t⟨m⟩alam (**Ami** mi-tanam)
- c. **PAn** \*S⟨um⟩ajek ‘smell (s.t.)’ > **Tha** ʃ⟨m⟩aðik **Paz** mu-sazek (**Ami** mi-sanek)

Movement and position verbs in F languages often have derived stems. In Puyuma, Thao, Saisiyat and Isbukun movement verbs are formed by prefixing a reflex of the PAn prefix \*u-/m-u- ‘go’ to a destination nominal. The source of this prefix is quite probably PAn \*uSa ‘go (to)’, the one reconstructable movement verb.

(95) PAn ∅/U movement ({3})

**PAn** \*uSa/\*(u)m-uSa ‘go (to)’ > **RukM** oa/om-oa ‘go’ **Tha** uʃa/m-uʃa ‘go’ **Paz** usa/m-usa ‘go (away)’ **Sai** m⟨in⟩oʃa? ‘went’ **AtaM** usaq/m-usaq ‘go’ **SedT** ?usa/m-usa (**Tso** us-a/uso ‘go toward’)

## 6.2.2 Reconstructing the morphology of U

The reflexes of \*⟨um⟩ in all F languages have phonologically conditioned allomorphs, shown in (96) and (97). The ‘default’ column shows the allomorph that occurs when none of the conditions

<sup>27</sup> The presence or absence of \*⟨um⟩ in labial-initial stems is discussed below.

in the four rightmost columns applies. The notation  $x/y$  means that the allomorph of the  $*\langle um \rangle$  reflex is  $x$  if the root-initial segment is  $y$ . The notation  $x//y$  means that the allomorph of the  $*\langle um \rangle$  reflex is  $x$ , replacing the root-initial segment if it is  $y$ . For example, Tsou  $b-//\varnothing$  applies in (92d–e).

(96)	default	/V	/labials	/sonorants	/other
PAn	$*\langle um \rangle$	$*(u)m-$	$*\varnothing/p, b (?)$	—	—
Puyuma	$\langle em \rangle$	$m-$	$\langle en \rangle/p, b$	$me-/l, \ell, r, n, \eta$	—
Tanan Rukai	$wa-$	$am-/disyll$	—	—	—
Man Rukai	$o-$	$om-/disyll$	—	—	—
Tsou <sup>28</sup>	$\langle m \rangle/c, s, t$	$m-$	$\varnothing/p$	—	$m-/s, k; m-//t, r; b-//\varnothing$
Saaroa	$\langle um \rangle$	$um-, m-$	$u/p, v, m$	—	$u-$ or $\langle um \rangle/k, t$ <sup>29</sup>
Kanakanavu	$\langle um \rangle$	$um-$	$mu-/p, v, m$	$mu-/ \eta$	$m-/V$ of prefix
Paiwan	$\langle em \rangle$	$[e]m-$	$\langle en \rangle/p, b, v, m$	$me-/n, \eta$	—
Thao	$\langle uN \rangle/C\_C; \langle m \rangle$	$m-$	$\varnothing/f, b, m, w$	$m-/ \_V, l, r, \delta$	$\varnothing/d, n, y$
Pazih	$mu$ <sup>30</sup>	$m-$	—	—	$me-/Ce$
Saisiyat	$\langle om \rangle$	$m-$	$\varnothing/p, \beta, m$ <sup>31</sup>	—	$m-//k, q$
Atayal	$\langle um \rangle$	$m-$	$m-//p, b$	—	$m-//k, y, q$
Seediq	$\langle m \rangle$	$m-$	$m-//p, b$	—	$m-//\varnothing$
Siraya	$\langle m \rangle$	$m-$	$\varnothing/p, v, m$	—	$m-//h$
Kavalan	$\langle m \rangle$	$m-$	$u-/b; m-/p$	—	see (97)

(97) Additional conditioned reflexes of  $\langle m \rangle$  in Kavalan:

$m-$  /V, wV, rV, RV, IV, zV, nV

$mu-$  /rC, RC, IC, zC, nC

$\langle um \rangle$  /p\_C, t\_C, k\_C, q\_C, s\_C

Default reflexes of  $*\langle um \rangle$  are mostly infixal, and we can assume this to be true of the proto-language. The exceptions are Rukai and Pazih. Mantauran  $o-$  and  $om-$  ‘AV realis’ reflect PAn  $*\langle um \rangle$  before a Kstem-initial consonant and before a vowel-initial disyllabic Kstem. Tanan (and Budai, Tona and Labuan) Rukai  $wa-$  reflects  $*um-$  plus  $*-a-$  ‘realis’ (Zeitoun 2003), perhaps a reflex of PAn  $*\alpha-$ . Its allomorph  $am-$  occurs with vowel-initial disyllabic stems and evidently reflects  $*\alpha-$  and  $*(u)m-$ . However, the Tanan AV subjunctive reflects  $*\langle um \rangle$  alone, so that alongside, for example, AV realis  $wa-kane$  ‘eat’ and  $am-a[a]$  ‘take’, we find AV subjunctive  $k\langle u \rangle ane$ , reflecting the infixation in PAn  $*k\langle um \rangle aen$ , and  $m-a[a]$ , reflecting PAn  $*(u)m-ala$ .

Pazih  $mu-$  and its  $mV-$  alternants and the first two rows of Kavalan forms in (97) reflect replacement of infix  $*\langle um \rangle$  by its prefixed alternant  $m(V)-$ .

<sup>28</sup> No source gives a full set of allomorphs, and we have no record of what happens when the root has initial  $b-, f-, v-, d-, m-, n-$  or  $\eta-$ . The  $\varnothing/p$  allomorph is reflected in the verb  $pn-aalpono$  ‘shoot’.

<sup>29</sup> The choice between  $u-$  and  $\langle um \rangle$  is lexically determined (Pan 2012:208).

<sup>30</sup> There is sporadic vowel harmony:  $mi-/Ci$  and  $ma-/Ca, Cu$ .

<sup>31</sup> With the exception of irregular  $\beta oay/mo-\beta ay$  ‘give’.

The form of the PAn infix is debatable. I take it to have been \*⟨um⟩ in conformity with the CV(C) syllable structure reconstructed for PAn. The languages in which it is reduced to ⟨m⟩ are, except for Siraya, languages in which the vowels of unstressed syllables are deleted under various conditions. Puyuma and Paiwan ⟨em⟩ [ə̃m] presumably reflect unstressed-vowel reduction.

Of the other allomorphs of \*⟨um⟩ shown in (96), those in the ‘sonorants’ and ‘other’ columns appear to be local developments. The ‘V’ column, however, suggests strongly that a vowel-initial root was preceded either by \*um- or \*m-.

The ‘labials’ column displays enough consistency to suggest that occurrence of \*⟨um⟩ with labials was restricted in PAn. This coincides with the observation that early Austronesian had a place harmony constraint whereby if, in a C<sub>1</sub>VC<sub>2</sub>V sequence, C<sub>1</sub> and C<sub>2</sub> had the same place of articulation, they also had the same manner of articulation. Under such a constraint *p-p*, *b-b* and *m-m* occur, but *p-b*, *p-m*, *b-m*, *b-p*, *m-p* and *m-b* normally do not (Uhlenbeck 1949, 1950). Clynes (1990) points out that such a constraint applies to Paiwan and probably applied to PAn. The restrictions in the ‘labials’ column of (96) reflect avoidance of \*p⟨um⟩... and \*b⟨um⟩..., later extended in some languages to reflexes of \*m⟨um⟩... and in Thao also of \*w⟨um⟩.... Avoidance strategies are:

- (98) a. replacing \*⟨um⟩ with zero (Thao, Saisiyat, Siraya)
- b. replacing \*⟨um⟩ with ⟨en⟩ (Puyuma, Paiwan)
- c. replacing \*⟨um⟩ with *m(V)*- (Kanakanavu, Kavalan /*p*)
- d. replacing the root-initial labial with \*m (Atayal, Seediq)
- e. deleting the -*m*- of \*⟨um⟩ (Kavalan /*b*)

Of these, (98c) does not honor the place harmony constraint and appears to be a more recent analogical development, (98d) reflects part of a wider set of phonological changes in Atayalic<sup>32</sup> and (98e) is restricted to Kavalan. This leaves (98a) and (98b) as possible PAn strategies. Of the two, (98a), the zero allomorph, is the more probable PAn form. It is a little more widely distributed, whilst (98b) occurs only in Puyuma and Paiwan, between which there is known to have been borrowing (Blust 1999b:47–51). This inference receives a little support from the reflexes of \*beRay, the putative AV form of ‘give’ in (99a), where Puyuma has *beray* rather than expected *b⟨en⟩eray*, (further discussion in §6.6.1). The items in (99), on the other hand, reflect putative PAn \*b⟨um⟩eRay, but could have arisen by simple analogy with other Ø/U verbs.

- (99) a. **PAn** \*beRay ‘AV.give’ > **Puy** beray **Tha** ʔay<sup>33</sup> **Fav** peʔi/pea
- b. **PAn** \*b⟨um⟩eRay ‘⟨AV⟩give’ > **RukM** o-vaʔay **Kav** [u-]bura **Saa** u-vuru **Pai** v⟨en⟩ay  
**Paz** baxa-i/mu-baxa **Sai** ʔoay/mo-ʔay **AtaM** baiq/m-aiq **SedT** begay/m-egay (but  
**Ami** pa-bri?)

If PAn class Ø/U did indeed have a zero allomorph with \*p- and \*b-initial roots, then these verbs would have displayed no formal difference from Ø/Z verbs, a matter further discussed in §6.6.1.

<sup>32</sup> Shown in the column labelled ‘other’ in (96).

<sup>33</sup> Presumably from earlier \*ʔay.

### 6.3 Class P/M

Most F representatives of PAn class P/M are derived forms, in other words it is the derivational prefix that carries the P/M alternation. However, one monomorphemic verb is reconstructable, namely PAn \*paCay/\*m-aCay ‘die’. Its reflexes, in (100), display considerable disagreement with regard to verb class, but the forms in (100a) almost certainly reflect the PAn forms, because (a) they are well distributed and (b) they display a shared uniqueness. The forms in (100b) reflect a reanalysis of the root, accommodating the verb to class  $\emptyset$ /MA. The Puyuma forms in (100c) reflect reanalysis of a nominalization and accommodation to class K/MA. Those in (100d), (100e) and (100f) retain the root \*patay and turn the verb into a member of class K/MA,  $\emptyset$ /U or  $\emptyset$ /MA respectively.

- (100) a. (P/M) **PAn** \*paCay/\*m-aCay ‘die’ > **Tso** mcoi (o-pcoi ‘kill’) **Kan** macai (pai-pacái ‘kill’) **Saa** maci?i (m-[p]a(-)ta-paci’i ‘become dead by being pounded’)<sup>34</sup> **Pai** patsay/m-atsay **BunTv** patað/matað (pa-patað ‘kill’) **Fav** machá **Sai** paθay/maθay  
 b. ( $\emptyset$ /MA) **Tha** aθay/m-aθay **SedT** acay/m-acay  
 c. (K/MA) **Puy** k-/m-inaŋay  
 d. (K/MA) **RukM** ka-/ma-pacay ‘dry’ **Sir** ma-patey (pa-ka-patey ‘kill’) **Ami** ka-patay-an/ma-patay (UVL)  
 e. ( $\emptyset$ /U) **RukB** ua-pácai  
 f. ( $\emptyset$ /MA) **Kav** patay/m-patay

It is possible that when PAn broke up, \*paCay/\*m-aCay ‘die’ was the only verb in class P/M and that it had survived as a unicum because of its high token frequency (§6.8). However, precisely because it was unique and thus had minimal type frequency, there was a countervailing tendency to accommodate it to other verb classes, as (100b–e) show.

It is fairly clear that \*p-replacement was derived from earlier \*⟨um⟩ and that the PAn P/M verb \*maCay ‘die’ reflects earlier but unattested \*p⟨um⟩aCay. The verbs in (100a–c) reflect further strategies to avoid \*p⟨um⟩... (§6.2.2).

### 6.4 Class K/MA

PAn evidently had a large number of K/MA resultative ({9}) and state ({10}) verbs.

K/MA and K/Z both encode states. However, K/MA also encodes resultatives, whereas K/Z never does. That is, K/MA encodes patient-subject processes, for example ‘the pot broke’, and the resultant states, e.g. ‘the pot is broken’, whereas K/MA and K/Z both encode states where the process leading to the state is not an intrinsic part of the meaning, for example ‘the pot is red’. Huang (2000), writing about Atayal, distinguishes between resultant states, encoded by K/MA, and inherent states, encoded by K/Z.

No attempt is made here to reconstruct PAn resultatives. It is clear from the languages in which they occur—Paiwan, Kavalan, Rukai, Seediq and Atayal—that there is at least a semiproductive derivational relationship between two-argument agentive  $\emptyset$ /UM verbs and patient-argument resultative K/MA verbs, and there must also have been a degree of productivity in PAn.

<sup>34</sup> Li (2009:206).

Tsou lacks both the K/MA and K/Z classes (§5.7). Class K/MA verbs are reflected in Tsou without a reflex of \*ma- (*m-* in *m-aski* ‘salty’ arguably reflects PAn \**um*). Bunun has effectively lost its K/MA class, whilst Paiwan, Kavalan and Amis have extended it.

(101) PAn K/MA states ({10})

- a. **PAn** \*ka-/ma-buSuk ‘drunk’ > **RukM** ma-voʔoko **Paz** ka-/ma-busuk **AtaM** ma-busuk **Ami** ma-busuk **Kav** qa/m-busuq (but **Tso** fsi **Sai** -k βofok/βofok **Sed** bsuk-an)
- b. **PAn** \*ka-/ma-Najam ‘be accustomed to, tame’ > **PuyT** ma-nadam **Saa** ma-lalame **Bab** ma-darram **Paz** ma-daxam **Ami** ma-nanam **Kav** q<n>a-nanam /m-nanam ‘be used to’ (but **Tso** a-hmo-hmo ‘unaccustomed’)
- c. **PAn** \*ka-/ma-qasiL ‘salty’ > **Kan** ma-ʔaini **Saa** m-aili **Fav** m-atshis **Ami** ka-hcil **Kav** ma-ʔsim (but **Tso** m-aski)
- d. **PAn** \*ka-/ma-zaLih ‘near’ > **RukM** maʔa-ðiiðali **Kan** ará-cani **Saa** ma-sali (but **Paz** alih **Sed** ke-dalih/dalih)
- e. **PAn** \*ka-/ma-zawiL ‘far’ > **RukM** maʔa-ðaili **Kan** ara-caini **Saa** ma-saila **Kav** m-laul (cf. q<n>a-laul-an ‘place far away’) (but **Puy** a-dawil, **Tso** covh-a/covhi **Ami** raay)

The third category regularly encoded by K/MA verbs consists of mental events. However, only one verb in this category is reconstructable.

(102) PAn K/MA mental events ({8})

**PAn** \*ka-/ma-bajaq ‘know’ > **Tha** ka-/ma-faðaq ‘know, understand, be able’ **Paz** ka-/ma-baza **Sir** ka-vana-n/ma-vana **Ami** ma-banaq (but **Sai** -k βaðæq/βaðæq ‘listen to’ **AtaM** baq ‘know, understand, be able’)

## 6.5 Class K/Z

The distribution of K/Z in Table 20 is straightforward: it encodes {10}, states, and is reflected almost nowhere else. The exception is Saisiyat, where some verbs seem to have shifted from class K/MA to class K/Z (§5.5). It is thus reasonably clear that PAn \*K/Z is reconstructable for {10}, states, and does not encode resultatives (§6.4).

Because Table 20 shows that class K/Z is missing from half the F languages (Tsou, Rukai, Saaroa, Kanakanavu, Siraya, Bunun, Thao) and Table 21 indicates that no PAn K/Z verb has been reconstructed, it is tempting to infer that it should not be reconstructed for PAn. However, this would be a mistake. The languages that reflect it agree rather strongly that it encodes pure states, and it is well distributed from south to north. Both facts support its reconstruction, and point to its loss from the languages that do not reflect it.

In most cases K/Z verbs are stative and denote those semantic classes that Dixon (1977) associates with adjective classes in languages that have them, as the listings in (103) to (108) show. Paiwan and Pazih each have a class of apparently invariable words that appear to be adjectives rather than verbs, and these encode a good many of the concepts encoded by K/Z verbs in other F languages.



- (103) Dimension: **Puy** *asat* ‘giant, high’ (but *ma-tina* K/MA ‘big’, *ma-kiteŋ* K/MA ‘small’), **Paz** *tatiŋ/tatih* ‘little, small’, **Sai** *θoβææh* ‘big’, *θopaloy* ‘big’, *ʔolʔolaʔan* ‘small’, **AtaM** *rahuwal* ‘big’, *tikay* ‘small’, *qanaruquh* ‘long’, *qihma* ‘thick’, *irarauq* ‘short (not tall)’, *isitiŋ* ‘short’, *kithuq* ‘fat’, *iqhiq* ‘slender’, *lihbiq* ‘thin (not thick)’, **SedT** *paru* ‘big’, *bilag* ‘small’, *bebaraw* ‘long’ **Kav** *raya* ‘big’, *kitut* ‘small’
- (104) Quantity: **Puy** *saqu* ‘many’ **Pai** *liaw* ‘many (things)’, *tjuruvu* ADJ ‘many (people)’ **Paz** *dahu* ‘much, many’ **Sai** *ʔakoy* ‘many’ **AtaM** *habaay* ‘many’ **SedT** *payux* ‘much, plenty’
- (105) Age: **Puy** *bekal* ‘new’ (but *ma-luman* K/MA ‘old (of people)’) **Sai** *tatiniʔ* ‘old’ **AtaM** *qiqas* ‘new’, *capaŋ* ‘old (not new)’, *qulaqiq* ‘young’, *nabakis* ‘old’ **Kav** *tasu* ‘new’, *zan* ‘old’
- (106) Evaluation: **Puy** *inaba* ‘good’ (*i-na-naba* ‘become better’ *pa-ka-inaba* ‘make better’), *buɭay* ‘beautiful’ (*pa-ka-buɭay* ‘make (s.t.) beautiful’), *baysar* ‘handsome’ **Pai** *na-ŋuaʔ* ‘good’ **Paz** *eder* ‘true’ **Sai** *kayðæh* ‘good’, *ʔæwhay* ‘bad’ **AtaM** *balaiq* ‘good, clever’, *qaqih* ‘bad’, *qayihun* ‘difficult’ **SedT** *malu* ‘good’, *naqih* ‘bad’ **Kav** *nŋi* ‘good’, *sukaw* ‘bad’
- (107) Color: **Puy** *aremeŋ* ‘dark’ (*a-ra-remeŋ* ‘become darker’) **Paz** *risilaw* ‘white’, *sem* ‘dark’, *tereheŋ* ‘black’ **SedT** *behegay* ‘white’, *banah* ‘red’ **Kav** *busaR* ‘white’, *tjan* ‘black’ (but *tbaRi* Ø/Z ‘red’)
- (108) Physical properties of objects and of human beings: **Puy** *litek* ‘cold’ (*la-litek* ‘become colder’ *pa-litek* (sic) ‘make cool’), *riksis* ‘dirty’, *bias* ‘fever, warm’ (cf. *b<en>ias* ‘heat (s.t.)’), *baɭaw* ‘life, alive’ (cf. *b<en>aɭaw* ‘save’), *sabeɭaw* ‘hungry’, *ŋakowaq* ‘noisy’, *tratoq* ‘cold (water)’, *mamatiq* ‘expensive, hard (not soft), capable’, **Pai** *laleqeɭ* ‘be cold; feel cold’, *valut* ‘be alive’, *saɭetju* ‘be painful’,<sup>35</sup> **Paz** *asikis* ‘painful’, *baged* ‘fat, obese’, *haiden* ‘heavy’, *lamik* ‘cool, cold’, *manex* ‘(rice) burnt’, **Sai** *ʔæ:ʔæiw* ‘dry’, *ʔæyæh* ‘ill’, *ayæh* ‘sick’, *qasqasay* ‘cooked/ripe’, *βa:βaok* ‘full’, *βaliq* ‘lean’, *hopay* ‘tired’, *kerpe:* ‘fat’, *konkononan* ‘round’, *eklaʔ* ‘awake’, *βofok* ‘drunk’, *rikrika:* ‘hot’, *fiɭʔi:* ‘heavy’, **AtaM** *baq* ‘able’, *bosinuq* ‘drunk’, *yihaaq* ‘cold (weather)’, *imuliq* ‘heavy’, *lihaqaq* ‘light (not heavy)’, *maskaiyun* ‘hungry’, *lawkah* ‘strong’

As one might expect, there are signs of movement between the stative classes K/MA and K/Z. Thus Puyuma *ma-tina* ‘big’, *ma-kiteŋ* ‘small’ in (103) and *ma-luman* ‘old (of people)’ in (105) are K/MA verbs which belong to typically K/Z semantic categories, and Saisiyat *bofok* and Atayal *bosinuq* ‘drunk’ in (108) seem to reflect a transfer from PAn K/MA verb \*ka-/\*ma-buSuk ‘drunk’ in (101).

In certain languages other semantic categories occur in class K/Z: there is no easy way of knowing whether these reflect retention or innovation. In the cases of (109), (110) and (111) this is unsurprising, as these are mainly stative verbs.

<sup>35</sup> Examples from Wei-chen Huang (personal communication 2013).

- (109) Taste and smell: **AtaM** *lalbiŋ* ‘sweet’, *sawkan* ‘stinking’, *shahiaq* ‘delicious’, *tatimoq* ‘salty’, **SedT** *sesibus* ‘sweet’ **Kav** *qarim* ‘sour’, *baŋtu* ‘sour’ (but *la-tbus* ‘sweet’ cf *tbus* ‘sugarcane’)
- (110) Speed: **Pai** *djalaw* ‘do quickly’ **Sai** *βalβalay* ‘slow’, *maʔŋel* ‘slow’, *qamoeh* ‘fast’, *qalikæh* ‘fast’ **AtaM** *lihkaq* ‘fast’
- (111) Emotional states: **Puy** *sagar* ‘like’ (*pa-ka-sagar* ‘bring pleasure to (s.o.)’), *igela* ‘ashamed’, *ilemes* ‘angry’ (*i-la-lemes* ‘become angry’) **Pai** *tjeŋelay* ‘love/like’, *saŋiŋa* ‘want’, *limuceŋ* ‘be sulky (about s.o./s.t.)’, *siŋ(e)lit* ‘to miss (s.t./s.o.)’ (but *ma-tjaŋaw* K/MA ‘angry’, *ma-siaʔ* K/MA ‘ashamed’) **Paz** *zemet* ‘annoyed’ **Sai** *fararaʔ* ‘like’, *beʔe:* ‘angry’

In Puyuma, Paiwan and Saisiyat class K/Z also contains verbs from the ‘in-between’ semantic categories. Whether these represent local innovations or reflect the situation in PAn it is impossible to be sure, in the absence of reconstructions.

- (112) {3} Perception: **Sai** *βaðæʔ* ‘hear’, *raam* ‘know’, *θaðek* ‘smell’, *shkelaʔ* ‘understand’  
 {4} Movement: **Puy** *bekas* ‘run’ **Sai** *loβih* ‘return’, *ræiw* ‘depart’  
 {6} Involuntary activity: **Puy** *alupe* ‘sleep’, *saeru* ‘laugh’, *a-remeŋ* ‘laugh’ **Pai** *me-ʔaCa-* ‘become tall’  
 {7} Inanimate activity: **Pai** *ka-b<a>ecuʔ* ‘explode’  
 {8} Mental events: **Sai** *raam* ‘know’, *sazek* ‘smell’, *fahæro:* ‘see’, *sekelaʔ* ‘understand’

From the agreement across languages that certain stative semantic categories fall into class K/Z, one can reasonably conclude that these categories were encoded by PAn K/Z verbs. Of the languages that lack class K/Z, we have sufficient data to infer that in Rukai and Thao class K/Z has merged with class K/MA. Thus the Mantauran Rukai and Thao K/MA verbs in (113) and (114) belong semantically to the K/Z categories reflected in (103) to (108).

- (113) **Mantauran Rukai** *ma-ðaw* ‘big, many’, *ma-votoŋoʔo* ‘fat’, *ma-ŋiʔemeðe* ‘thick’, *ma-ʔopiŋi* ‘old (of things)’, *ma-roðaje* ‘old (of people)’, *ma-taaðiʔi* ‘good’, *ma-takoŋa* ‘bad’, *ma-poli* ‘white’, *ma-ðiliaŋe* ‘red’, *ma-teleke* ‘cold (of water)’, *ma-keceŋe* ‘cold (of weather)’, *ma-teke* ‘hard’, *ma-meŋeʔe* ‘soft, weak, cheap’, *ma-keʔeceŋe* ‘strong’, *ma-caʔeme* ‘sick’, *ma-pacai* ‘dry’, *ma-ðalame* ‘like’, *ma-ʔamaðe* ‘dislike’
- (114) **Thao** *ma-raʔin* ‘big’, *ma-ramu* ‘fat’, *ma-naʔa* ‘many, plentiful’, *ma-qitan* ‘good, beautiful’, *ma-qarman* ‘bad, ugly’, *ma-humhum* ‘dark’, *ma-qusum* ‘black’, *ma-qula-qula* ‘very red’ (*ag-qula* ‘red’), *ma-puði* ‘white’, *ma-haðiwhiw* ‘cool, cold’, *ma-hnar* ‘hot (of weather)’, *ma-qaltaf* ‘alive’, *ma-kuʔraθ* ‘hard (of rice)’, *ma-dumuθ* ‘soft, tender’, *ma-dumur* ‘damp’, *ma-funuq* ‘soft (of soil)’

Amis class Z (§5.4) includes stative verbs that may have once belonged to class K/Z.

- (115) **Amis** *kuhtij* ‘black’ (cf. *ma-kuhtij* ‘become black’), *faqedet* ‘hot’ (cf. *mi-faqedet* ‘heat (s.t.)’), *puti* ‘white’ (cf. *ma-puti* ‘clean’), *lipahak* ‘happy’ (cf. *ma-lipahak* ‘make happy’), *tataqan* ‘big’, *adidiq* ‘small’, *signaw* ‘cold’, *faŋcal* ‘beautiful’

## 6.6 Class Ø/Z

The distribution of Ø/Z in Table 20 is quite different from the distributions of the classes discussed above. It is well represented among {1}, agentive verbs, but also reflected in six languages among {10}, state verbs. In other words, unlike Ø/U, K/MA and K/Z it is present at both ends of the hierarchy. Curiously, it is absent from the categories that border these extremes, namely from {2}, weather verbs, and {9}, resultatives, but sporadically reflected in categories {3} to {8} (its absence from {5}, position, may be fortuitous, as there are many gaps here). What are we to make of this peculiar distribution?

Reconstruction of individual verbs suggests that class Ø/Z is diachronically not one class but two: one agentive, the other stative. The agentive class consists of verbs that for a couple of reasons have migrated from Ø/U to Ø/Z, that is, their exponent of U has been lost (§6.6.1). The stative Ø/Z class resembles class K/Z in its distribution: it encodes ‘pure’ states, not resultatives, as shown in §6.6.2. The sporadic presence of Ø/Z verbs in {3} to {8} resembles the distribution of class Ø/U and probably reflects the agentive Ø/Z class.

### 6.6.1 Agentive

Verbs of the Ø/Z agentive subclass seem to have emerged from class Ø/U in two ways. The verbs in (116) have an initial labial stop, suggesting that they reflect avoidance of \*p<um>... and \*b<um>... , in other words they reflect the zero allomorph of \*<um> discussed in §6.2.2. By the rules in (96), however, they are irregular. The Puyuma verbs should have <en> in this environment and Atayal should have *p*-replacement. The best explanation of these forms is that the PAn allomorph of \*<um> with \*p- and \*b-initial roots was indeed zero (§5). This is reflected in the verbs in (116), whereas the synchronically regular reflexes referred to in (96) are the result of later innovations which failed to affect the verbs in (116).

- (116) a. Puyuma *beray* ‘give’, *pulan* ‘help’  
 b. Atayal *pawuni* ‘come’, *buina* ‘graduate’, *palawatux* ‘bark’, *panaip* ‘fish’, *panaturij* ‘point at’, *paqati* ‘throw away’, *pasilay* ‘thresh (rice)’, *psacoy* ‘cough’, *psibaq* ‘teach’, *psibqaq* ‘explode’, *psihow* ‘suck’

Only one verb stem is reconstructable as monomorphemic and trisyllabic, PAn \*timaLa ‘hear’, a Ø/Z verb except in Kanakanavu and Saaroa, where it belongs to class Ø/U.

- (117) **PAn** \*timaLa ‘hear’ > **Tha** tmaða-n (UVP)/tmaða ‘hear, listen’ **Paz** tumala ‘hear, listen’  
**Tao** temada, timaza, timala (**Kan** t<um>a-timana **Saa** t<um>a-timaŋa)

The common feature of the  $\emptyset/Z$  verbs listed in (118) is that, like \*timaLa, their stems have or once had three or more syllables, but unlike \*timaLa they appear to be or to have been polymorphemic. Polymorphemic verbs belong to the class of their first morpheme. The distribution of polymorphemic verbs across  $\emptyset$  verb classes is different from the distribution of monomorphemic verbs. The order of apparent frequency among monomorphemic verbs is  $\emptyset/U > \emptyset/MA > \emptyset/Z > P/M$ . The order of apparent frequency among verbs with derivational prefixes is  $\emptyset/Z > P/M > \emptyset/U$ , with the vast majority of derivational-prefix verbs belonging to classes  $\emptyset/Z$  and  $P/M$ . It thus seems likely that PAn had a further constraint on allomorphs of \* $\langle um \rangle$ , such that stems of three or more syllables, either of class  $\emptyset/U$  or polymorphemic, had by default the zero allomorph of \* $\langle um \rangle$ , and this is reflected in the verbs in (118).

- (118) a. Mantauran Rukai *omikace* ‘open one’s eyes’, *paða[o]o* ‘at once’  
 b. Tanan Rukai *arakai* ‘use’, *ʔaswala* ‘do first’  
 c. Tsou *eobak-a* (UVP)/*eobako* ‘beat’, *etamaku-a* (UVP)/*etamaku* ‘smoke’  
 d. Paiwan *petsugu* ‘to make a fire’, *kenama* ‘to breakfast’  
 e. Atayal *aruwa* ‘leave’, *alaluq* ‘swing’, *kabaux* ‘borrow’, *qabalay* ‘build’, *tayhok* ‘arrive’, *toraqis* ‘wash face’, *tulqin* ‘hide (intr.)’, *tusasiq* ‘rest’  
 f. Siraya *irua* ‘arrive’, *thabul* ‘leave, travel’, *tawrùma* ‘get lost’  
 g. Kavalan *satzay* ‘sing’, *saqmay* ‘cook’

Finally, (119) lists  $\emptyset/Z$  verbs that do not fit into any of the categories above. However, there is some evidence that at least some of them are migrants from other classes, as Tanan Rukai *davac* ‘walk’ reflects the PAn  $\emptyset/U$  verb \*zawaC ‘walk’ (Paiwan *j⟨em⟩avats* ‘walk’, Hoanya *dz⟨am⟩awat* ‘walk’) whilst Saisiyat *θawaʔ* ‘laugh at’ reflects the PAn  $\emptyset/MA$  verb \*ma-Cawa ‘laugh’ in (122).

- (119) a. Mantauran Rukai *aviʔi* ‘give’, *irovo* ‘pile up’  
 b. Tanan Rukai *nyaman* ‘come’, *rivo* ‘scold’, *davac* ‘walk’  
 c. Paiwan *aya* ‘speak to (s.o.)’, *vaik* ‘go, leave’, *kian* ‘follow (s.o.)’, *kaiv* ‘eat dinner’, *tsalum* ‘fetch water’  
 d. Saisiyat *θawaʔ* ‘laugh at’, *rima* ‘go’, *raawak* ‘dance’, *mariʔ* ‘take’  
 e. Siraya *irua* ‘arrive’, *lupux* ‘kill’  
 f. Kavalan *tmeq* ‘bake’, *wi* ‘go’

## 6.6.2 States

Two members of the  $\emptyset/Z$  stative subclass are reconstructable:

- (120) a. **PAn** \*baqeRuh ‘new’ > **RukM** *vaoʔae* **Tso** *faeva* **Saa** *varuʔu* **Kan** *vaʔurua* **Pai** *vaʔu-an* **BunI** *bahluʔ* **Tha** *faqʔu* **Sir** *vaho* **Ami** *vaʔluh*  
 b. **PAn** \*ZaZaL ‘old’ > **PuyT** *zazal-an* ‘old (seeds that no longer germinate)’ **Pai** *zazal-an* ‘last year’s millet harvest’ **Tha** *sasað* **Bab** *raras* **AtaM** *raral* **Kav** *ʔzann*

These two fall into class  $\emptyset/Z$  on the negative evidence that they are never reflected with \*ma- or \*ka-. Li & Tsuchida (2006) record a causative *pa-ʔzan* ‘cause to become old’ which attests to the absence of \*ka-. There are signs that this stative category had a few other members, but in the absence of evidence of a Kstem it is difficult to be certain that they did not belong to class K/Z:

- (121) a. Puyuma *palu* ‘demarcate, be until’  
 b. Tanan Rukai *bekelay* ‘little’, *edekay* ‘short’, *tikia* ‘small’  
 c. Kavalan *Riju* ‘be unable to’, *supaR* ‘know’, *sizɲes* ‘think of’

## 6.7 Class $\emptyset/MA$

The verbs of class  $\emptyset/MA$  pose a greater reconstructive challenge than any of the other four classes. Like class K/Z, they are present in only about half the F languages. But whereas class K/Z has a pronounced distributional profile (only {10}), class  $\emptyset/MA$  occurs sporadically across the semantic categories from {1} to {10}. This raises the question: Was it perhaps absent from PAn? Does it reflect verbs that have shifted from other classes? We know that in some cases it does. Isbukun Bunun has two  $\emptyset/MA$  classes. Bunun  $\emptyset/MA_1$  (mostly?) reflects  $\emptyset/U$  verbs in which U has been replaced wholesale by MA. Bunun  $\emptyset/MA_2$  (mostly?) reflects K/MA verbs from which K has been lost (§5.11). Thao K/MA resultative and state verbs have lost *ka-*, giving  $\emptyset/MA$  verbs in these categories (§5.3).<sup>36</sup> This means that Bunun  $\emptyset/MA$  verbs should perhaps be discounted in reconstructing putative \* $\emptyset/MA$ ,<sup>37</sup> and so should Thao \* $\emptyset/MA$  resultatives and states. But we must still account for Puyuma, Tsou, Tanan Rukai, Seediq and Atayal reflexes of  $\emptyset/MA$  in {1}, agentive verbs, and Kavalan reflexes in {3}, perception verbs, as well as a variety of reflexes in {5}, {6}, {7} and {8}. Writing about Atayal, Huang (2000) suggests that  $\emptyset/MA$  verbs occupy a semantic position somewhere between dynamic  $\emptyset/U$  verbs and stative  $\emptyset/Z$ . This accords with the scatter of  $\emptyset/MA$  verbs across the categories from {5} to {8}. Huang’s (2000:370) count of Mayrinax Atayal verbs shown in Table 22 is also instructive, as  $\emptyset/MA$  is the smallest category. Perhaps the nearest we can get to characterising PAn \* $\emptyset/MA$  is to say there is evidence (a) that it occurred, (b) that it was the smallest PAn class and (c) that its core area meaning lay somewhere between agentive and resultative verbs.

**Table 22:** The number of verbs in each verb class in Mayrinax Atayal (after Huang 2000:370)

	$\emptyset/U + P/M$	$\emptyset/MA$	$\emptyset/Z$	K/MA	K/Z	Total
number	148	9	20	23	37	237
percent	62.5%	3.8%	8.4%	9.7%	15.6%	100%

<sup>36</sup> Saisiyat, on the other hand, seems to have lost any agentive  $\emptyset/MA$  verbs it might once have had (§5.5). In Kavalan certain allomorphs of U and MA are ambiguous, and this may have led to a reanalysis of some  $\emptyset/MA$  verbs as  $\emptyset/U$ , the default agentive class (§5.2). In Amis the spread of *ka-* to the Kstem of agentive verbs has emptied class  $\emptyset/MA$  and transferred agentive verbs to class K/MA (§5.4). A similar change has occurred in Paiwan (§5.1), though there has been no Amis-like system-wide change.

<sup>37</sup> In Table 20 Bunun  $\emptyset/MA_2$  has in any case been treated as K/MA, not as  $\emptyset/MA$ .

No  $\emptyset$ /MA verb is reconstructable, and it is perhaps the location of its core meaning that is responsible for this. The categories from {3} to {8} are those with the fewest member verbs. This means that the type frequency of the  $\emptyset$ /MA class is low, and its members are therefore likely to have been transferred to larger classes (to be ‘regularized’: see §6.8). It is tempting to attribute the two involuntary activity verbs in (122) to class  $\emptyset$ /MA, but they may have been K/MA verbs. Reflexes of their Kstems are insufficient to decide. In (122a) Thao reflects K, but Kavalan reflects  $\emptyset$ . In (122b) there is evidence only for the Kavalan Kstem.

- (122) a. **PAn** \*ma-Cawa ‘laugh’ > **Tso** cocv-i (UVL)/cocvo **Saa** ma-caacaa **Sir** ma-tawa **Tha** ka- $\theta$ awa-i/ma- $\theta$ a- $\theta$ awa ‘laugh at’, **Sai**  $\theta$ awa? **Kav** tawa/m-tawa **Ami** ma-tawa  
 b. **PAn** \*ma-tu(dzD)uR ‘feel sleepy’ > **Saa** ma-ta-tusoro (IPFV) **BunTt** ma-tu?dul **Kav** m-tuRuz (pa-tuRuz ‘make s.o. sleepy’)

A number of  $\emptyset$ /MA verbs in various languages belong semantically to categories {3} to {8}:

- (123) a. Thao *ma-damdam* ‘taste, eat’, *ma-tunaw* ‘try one’s hardest’, *ma-duhray* ‘feel rested’  
 b. Pazih *ma-bunux* ‘swell’, *ma-hurik* ‘be sleepy’, *ma-ledalet* ‘shiver’, *ma-irat* ‘be alive’, *ma-iray* ‘be lame’, *ma-irix* ‘slip, slide’, *ma-anat* ‘burn (ITR)’ (cf. *m-anat* ‘catch fire’, *pa-anat* ‘burn (ITR)’), *ma-lalej* ‘dwell’  
 c. Atayal *ma-qaqa?uh* ‘be busy’, *ma-qilaap* ‘sleep’, *ma-qiyanax* ‘live (not die)’

Huang’s (2000) suggestion that  $\emptyset$ /MA verbs occupy a semantic position somewhere between dynamic  $\emptyset$ /U verbs and stative  $\emptyset$ /Z is true of Atayal and of some other F languages if ‘less dynamic’ means that the verb lacks agentivity but is not stative.

- (124) a. Puyuma *ma-ragan* ‘get up’, *ma-berek* ‘depart’, *ma-redek* ‘arrive’, *ma-alep* ‘crawl’  
 b. Thao *ma-ifif* ‘rub against s.t.’ (cf. *pa-ifif* ‘rub s.t.’), *ma-la $\theta$ aq* ‘fall from a height’ (cf. *pa-la $\theta$ aq* ‘lower’)  
 c. Pazih *ma-xutaxa* ‘wait’, *ma-urit* ‘collapse’, *ma-kulalah* ‘fall down’  
 d. Kavalan *m-suRaw* ‘fall over’, *m- $\eta$ aqsan* ‘do slowly’, *m-qasiR* ‘do quickly’, *m-buRaw* ‘move away’

## 6.8 An overview

Table 23 approximates a summary of the findings above, where the number of ‘x’s is a rough representation of the proportion of members of that class with meanings in each semantic category. For clarity’s sake, class  $\emptyset$ /Z is divided into  $\emptyset$ /Z<sub>A</sub> (prototypically agentive) and  $\emptyset$ /Z<sub>S</sub> (states).

If we look at these findings from the perspective of the hierarchy of semantic categories, the seeming messiness of {3} to {8} has a principled explanation. Two factors conspire here.

First,  $\emptyset$ /MA,  $\emptyset$ /Z, K/MA and K/Z are distinguished by a very restricted set of morphemes: the KStem by zero and K, the MStem by zero and MA. The possibility of a verb slipping sideways from, say, K/MA to K/Z is high.

**Table 23:** Semantic distribution of PAn verb classes

	PAn	Ø/U	Ø/MA	Ø/Z <sub>A</sub>	Ø/Z <sub>S</sub>	K/MA	K/Z
{1}	agentive	xxx	x	xx			
{2}	weather	xx					
{3}	perception	xx	x	x			
{4}	movement	xx	x	x		x	
{5}	position	x	x			x	
{6}	involuntary activity	x	x	x		x	
{7}	inanimate activity	x	x	x		x	
{8}	mental event	x	x	x		x	
{9}	resultative					xx	
{10}	state				xx	xx	xx

This possibility is magnified by the second factor, discourse frequency. The lexical reconstructions presented above are based on my compilation of forms. The numbers of PAn verbs in each class for which both stems could be reconstructed is given in Table 21.<sup>38</sup> The vast majority of these, 84%, are in class Ø/U, 12% in class K/MA, 3% in class Ø/Z and none in classes Ø/MA or K/Z. The probability of a morphological construction surviving over time is related to its frequency in discourse, namely its type frequency (Bybee 2007). Where the membership of a verb class is small, its type frequency is low, and its members will tend to shift to a larger class, that is to regularize, leaving the old class with vanishingly few members. This is a reason why members of the cognate sets in (122) belong to different verb classes.

One may appropriately ask in the light of this why \*paCay/\*m-aCay ‘die’ has survived as perhaps the sole monomorphemic member of the P/M class (§6.3). The answer is that the counter-vailing force to low type frequency is high token frequency (Bybee 1995:428–431). That is, a single item that occurs with high discourse frequency resists change, becoming increasingly irregular as less token-frequent members of its class move sideways into another class.

An effect of the class-to-class shifts of individual verbs is that the boundary between A- and P- verbs (i.e. between Ø and K class verbs) now falls at different points in different F languages. This is shown graphically in Table 24. From Table 23 and Table 24 it is clear that the PAn boundary cannot be exactly placed. What emerges from Table 24 is that, ignoring Amis, languages towards the south of Taiwan assign more semantic categories to P-verbs and languages towards the north more categories to A-verbs. However, this could be a contact phenomenon, and does not cast light on the PAn situation.

<sup>38</sup> Ten more reconstructed verbs have an Mstem in \*ma- but an unknown Kstem. Some of these verbs are evidently intransitive, and forms that reveal the Kstem therefore occur more rarely.



**Table 24:** A graphic representation of the distributions of A- and P-verbs in Formosan languages

		Puy	RukM	Saa	Sir	Pai	Tha	Paz	Sai	SedT	AtaM	Kav	Ami
		<div> <div></div> <div>A-verbs (<math>\emptyset</math>/U, <math>\emptyset</math>/MA, <math>\emptyset</math>/Z<sub>A</sub>)</div> </div>											
		<div> <div></div> <div>P-verbs (<math>\emptyset</math>/Z<sub>S</sub>, K/MA, K/Z)</div> </div>											
		<div> <div></div> <div>A- and P-verbs</div> </div>											
{1}	agentive												
{2}	weather												
{3}	perception												
{4}	movement												
{5}	position												
{6}	involuntary activity												
{7}	inanimate activity												
{8}	mental event												
{9}	resultative												
{10}	state												

## 6.9 Collective agentive verbs

Collective agentive verbs pose a different challenge. They appear to be a semantic category in at least seven languages, but their morphology differs: Puyuma  $\emptyset$ /MA, Pazih, Siraya, Paiwan, Kavalan and Amis K/MA and Saisiyat P/M (*pa-/ma-*). The common thread is an Mstem in *ma-*. Circumstantial evidence points to PAN \*pa-/\*ma-, since PAN reciprocals are reconstructed by Zeitoun (2002) with \*pa-Ca-/\*ma-Ca- with a  $\emptyset$  verb, and \*pa-R(a)-/\*ma-R(a)- with a K verb. Lichtenberk (2000) points out that there is considerable overlap morphologically in Austronesian languages between various concepts under the rubric ‘plurality of relations’, and it seems likely that the collective prefix formed part of the reciprocal prefixes. But this leaves us with an unresolved question:

Why did the Kstem shift from *pa-* to  $\emptyset$  in Puyuma and from *pa-* to K in Pazih, Siraya, Paiwan, Kavalan and Amis? If the latter shift did occur, it would be the nearest this paper comes to identifying a shared innovation diagnostic of a F subgroup—but it would be a subgroup defined by no other shared innovation.

## 7. Verb classes and Austronesian subgrouping: Malayo-Polynesian

One reason for reconstructing PAn verb classes is that an understanding of early Austronesian morphology may enable us to identify innovations that will improve our analysis of early Austronesian history and subgrouping. This short discussion of Malayo-Polynesian (MP) does not present detailed arguments, but indicates where this research might lead.

That all Austronesian languages outside Taiwan form a single subgroup, labelled MP by Blust (1977), has long been agreed by most scholars involved in Austronesian historical linguistics. Blust (2009:736–737) cites Dahl (1973) and Mills (1975) as originators of the hypothesis and summarizes the shared innovations that are evidence for it. Verbal morphology does not figure among them, partly because no detailed reconstruction of Proto Malayo-Polynesian (PMP) verbal morphology has ever been attempted. But as soon as one compares the morphological classes of verbs in conservative MP languages (i.e. Philippine languages) with those of F languages and those reconstructed here for PAn, it becomes evident that the morphology that characterizes PAn verb classes had undergone radical changes in PMP. Ivatan (Reid 1966; Tsuchida et al. 1989), a MP language spoken in the Batanes Islands between Taiwan and the Philippines and quite possibly an early offshoot of MP (Ross 2005), has an orthodox  $\emptyset$ /U class:

- (125)  $\emptyset$ /U *kum>an* ‘eat’, *tum>nek* ‘stand’, *tum>añis* ‘weep’, *m-ayam* ‘walk’, *tum>uwaw* ‘go out, take out’, *qum>ulib* ‘hide (INTR)’

Its K/MA class is formally orthodox, but, unlike F languages, includes weather verbs, the involuntary activity verb ‘die’ and movement verbs as well as statives.

- (126) K/MA *ka-/ma-timuy* ‘rain’, *ka-/ma-yayu* ‘flood’, *ka-/ma-diman* ‘die’, *ka-/ma-ŋay* ‘go, depart’, *ka-/ma-wara* ‘arrive’

Ivatan also has what appear to be  $\emptyset$ /MA verbs, such as *ma-pia* ‘good’ and *ma-rahēt* ‘shallow’, but these are arguably adjectives, not verbs, as the only morphological change they undergo is replacement of imperfective *ma-* by perfective *na-*.

There is a possible K/Z class, but it lacks an AV form: Reid (1966:24–25) gives *ka-anaruh-en* ‘lengthen’ (root *qanaru* ‘long’) and *ka-yñer-en* ‘shorten’ (root *qiñed* ‘short’) as examples.

But the most characteristically MP feature of Ivatan verbs is the massive expansion of AV forms with the two PMP P/M prefixes \*paN-/m-aN- and \*paR-/m-aR- at the expense of other verb classes.

- (127) PMP \*m-aN-/paN-: Ivatan *paŋ-/maŋ-amuqmu* ‘frighten’, *pan-/man-adiu* ‘buy’, *paŋ-/maŋ-ulib* ‘hide (ITR)’, *pan-/man-arip* ‘peel’, *pan-/man-adiu* ‘buy’, *paŋ-/maŋ-lup* ‘blow’, *pan-/man-laveŋ* ‘bury’, *paŋ-/maŋ-yaw* ‘shout’, *paŋ-/maŋ-aamuŋ* ‘fish’

The \*-N- symbol here indicates a process that places \*-ŋ- before a vowel and replaces the root-initial consonant with a homorganic nasal, a widespread MP process with reflexes of this prefix (Blust 2004). The origin of PMP \*paN-/m-aN- is far from clear: homorganic nasal substitution occurs infrequently and unpredictably in F languages (Starosta 2002:194–195; Tsuchida 1976:257–258; Wolff 2007). Kaufman (2009) derives \*paN- from earlier \*p<aŋ>- and glosses \*<aŋ> as denoting plurality of object or action.<sup>39</sup>

The original PMP use of \*paN-/m-aN- was perhaps distributive, as it retains this function in the Philippine language Binukid (Post 1992). Verbs of this new PMP class often replaced PAn Ø/U verbs, and Kaufman (2009) suggests that the trigger for the proliferation of the ‘new’ prefixes was the ambiguity of \*ma- as a marker of both dynamic (Ø/MA, D/P) and stative (K/MA) verbs.

PMP \*paR-/m-aR- apparently reflects the PAn stative reciprocal prefix \*pa-R(a)-/ma-R(a)- (Zeitoun 2002). Its function evidently expanded to include agentive reciprocals like ‘fight (each other)’ in (128), but then became bleached to mark any number of agentive verbs.

- (128) PMP \*paR-/m-aR-: Ivatan *pay-/may-diman* ‘fight (each other)’, *pay-/may-weswes* ‘turn back’, *pay-/may-weswes* ‘turn back’, *pay-/may-liliak* ‘speak’, *pay-/may-hojihoji* ‘move’, *pay-/may-piqsa* ‘hit’

Liao (2011:857–862) draws attention to the PMP ‘abilitative’ or ‘potentive’ AV prefix \*maka- which in Philippine languages is one of a set of voice markers that ‘deletes’ the agentive feature from agentive verbs. She points out that there are F cognates—in Puyuma, Paiwan and Amis—but none has the range of functions that its reflexes have assumed in Philippine languages.

This subsection has offered only an informal characterization of the changes in morphological verb classes that occurred in PMP, but hopefully suggests future directions of research.

## 8. Conclusion

This paper has provided a first reconstruction of the morphological classes of PAn verbs with monomorphemic stems. The data on which it is based are neither as extensive nor as consistent as one would desire, and this has probably resulted in errors. One finding that emerges from the analysis is that there is an acute need for finer morphological descriptions of F verbal systems and for the application by lexicographers of the categories that emerge from these descriptions. If the analysis of principal parts (Kstem and Mstem) here is correct, then it needs to be applied in dictionaries.

Another finding is that, although the correspondence between verb classes and the semantic categories that fall between {4} and {8}, that is between the extremes of agentive activity and state, is messy, there are principled diachronic reasons why this is so (§6.8).

A goal of the paper was to provide a reconstruction on the basis of which shared innovations can be identified in order to augment the foundations for subgrouping and the understanding of

<sup>39</sup> The infix \*<(a)ŋ> was presumably also present in Thao *kin-* ‘get obtain’ (<PAn \*ki- + \*<(a)ŋ>) and in Kavalan *sim- ~ sin-* ‘reciprocal’ in, as Kaufman notes. If so, then the infix was simply \*<ŋ>.

early Austronesian linguistic history. Section 7 is a brief attempt to show how this can be done, by showing that PMP underwent striking innovations in its verbal morphology.

## References

- Adelaar, K. Alexander. 2011. *Siraya: Retrieving the Phonology, Grammar and Lexicon of a Dormant Formosan Language*. Berlin & Boston: De Gruyter Mouton.
- Aronoff, Mark. 1994. *Morphology by Itself: Stems and Inflectional Classes*. Cambridge: MIT Press.
- Blevins, James P. 2006. Word-based morphology. *Journal of Linguistics* 42.3:531–573.
- Blust, Robert A. 1977. The Proto-Austronesian pronouns and Austronesian subgrouping: a preliminary report. *University of Hawai'i Working Papers in Linguistics* 9.2:1–15.
- Blust, Robert A. 1999a. Notes on Pazeh phonology and morphology. *Oceanic Linguistics* 38.2: 321–365.
- Blust, Robert A. 1999b. Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. *Selected Papers from the Eighth International Conference on Austronesian Linguistics*, ed. by Elizabeth Zeitung & Paul Jen-kuei Li, 31–94. Taipei: Institute of Linguistics, Academia Sinica.
- Blust, Robert A. 2003. *Thao Dictionary*. Taipei: Institute of Linguistics, Academia Sinica.
- Blust, Robert A. 2004. Austronesian nasal substitution: a survey. *Oceanic Linguistics* 43.1:73–148.
- Blust, Robert A. 2009. *The Austronesian Languages*. Canberra: Pacific Linguistics.
- Booij, Geert. 2010. *Construction Morphology*. Oxford & New York: Oxford University Press.
- Booij, Geert. 2012. Construction morphology, a brief introduction. *Morphology* 22.3:343–346.
- Bybee, Joan L. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10.5: 425–455.
- Bybee, Joan L. 2007. *Frequency of Use and the Organization of Language*. Oxford & New York: Oxford University Press.
- Cauquelin, Josiane. 1991. *Dictionnaire puyuma-français*. Paris: École française d'Extrême-Orient.
- Chafe, Wallace L. 1970. *Meaning and the Structure of Language*. Chicago: University of Chicago Press.
- Chang, Anna Hsiou-chuan. 2006. *A Reference Grammar of Paiwan*. Canberra: The Australian National University dissertation.
- Chang, Henry Y. 1997. *Voice, Case and Agreement in Seediq and Kavalan*. Hsinchu: National Tsing Hua University dissertation.
- Chang, Henry Y. 2009. Adverbial verbs and adverbial compounds in Tsou: a syntactic analysis. *Oceanic Linguistics* 48.2:439–476.
- Chen, Teresa M. 1987. *Verbal Constructions and Verbal Classification in Nataoran Amis*. Canberra: Pacific Linguistics.
- Clynes, Adrian J. 1990. Active articulator phonological patterning in Austronesian languages. Manuscript. Canberra: The Australian National University.
- Corbett, Greville G., & Norman M. Fraser. 1993. Network morphology: a DATR account of Russian nominal inflection. *Journal of Linguistics* 29.1:113–142.
- Croft, William. 2012. *Verbs: Aspect and Causal Structure*. Oxford & New York: Oxford University Press.

- Dahl, Otto Christian. 1973. *Proto-Austronesian*. Lund: Studentlitteratur.
- De Busser, Rik. 2009. *Towards a Grammar of Takivatan Bunun: Selected Topics*. Melbourne: La Trobe University dissertation.
- Dixon, R. M. W. 1977. Where have all the adjectives gone? *Studies in Language* 1.1:19–80.
- Dowty, David R. 1991. Thematic proto-roles and argument selection. *Language* 67.3:547–619.
- Egerod, Søren. 1999. *Atayal–English Dictionary* (2<sup>nd</sup> edition). Copenhagen: The Royal Danish Academy of Sciences and Letters.
- Egli, Hans. 1990. *Paiwangrammatik*. Wiesbaden: Otto Harrassowitz.
- Egli, Hans. 2002. *Paiwan Wörterbuch: Paiwan–Deutsch, Deutsch–Paiwan*. Wiesbaden: Otto Harrassowitz.
- Ferrell, Raleigh. 1982. *Paiwan Dictionary*. Canberra: Pacific Linguistics.
- Fey, Virginia. 1986. *Amis Dictionary*. Taipei: The Bible Society in the Republic of China.
- Finkel, Raphael, & Gregory Stump. 2007. Principal parts and morphological typology. *Morphology* 17.1:39–75.
- Foley, William A. 1976. *Comparative Syntax in Austronesian*. Berkeley: University of California at Berkeley dissertation.
- Foley, William A. 2005. Semantic parameters and the unaccusative split in the Austronesian language family. *Studies in Language* 29.2:385–430.
- Foley, William A., & Robert D. Van Valin. 1984. *Functional Syntax and Universal Grammar*. Cambridge & New York: Cambridge University Press.
- Himmelman, Nikolaus P. 2002. Voice in western Austronesian: an update. *The History and Typology of Western Austronesian Voice Systems*, ed. by Fay Wouk & Malcolm Ross, 7–16. Canberra: Pacific Linguistics.
- Himmelman, Nikolaus P. 2005. The Austronesian languages of Asia and Madagascar: typological characteristics. *The Austronesian Languages of Asia and Madagascar*, ed. by K. Alexander Adelaar & Nikolaus P. Himmelman, 110–181. London & New York: Routledge.
- Himmelman, Nikolaus P. 2006. How to miss a paradigm or two: multifunctional *ma-* in Tagalog. *Catching Language: The Standing Challenge of Grammar Writing*, ed. by Felix Ameka, Alan Dench & Nicholas Evans, 487–526. Berlin & New York: Mouton de Gruyter.
- Himmelman, Nikolaus P. 2008. Lexical categories and voice in Tagalog. *Voice and Grammatical Relations in Austronesian Languages*, ed. by Peter Austin & Simon Musgrave, 247–293. Stanford: CSLI.
- Ho, Dah-an. 1978. Wuzhong Paiwan fangyan de chubu bijiao [A preliminary comparative study of five Paiwan dialects]. *Bulletin of the Institute of History and Philology Academia Sinica* 49.4:565–681.
- Ho, Dah-an. 1997. Kanakanafuyu [The Kanakanavu language]. *The Formosan Languages of Kaohsiung*, ed. by Paul Jen-kuei Li, 228–271. Kaohsiung: Kaohsiung County Government.
- Huang, Lillian M. 2000. Verb classification in Mayrinax Atayal. *Oceanic Linguistics* 39.2:364–390.
- Huang, Shuanfan. 2002. Tsou is different: a cognitive perspective on language, emotion, and body. *Cognitive Linguistics* 13.2:167–186.
- Huang, Shuping, & Li-May Sung. 2008. The undergoer focus *ma-* in Kavalan. *Oceanic Linguistics* 47.1:159–184.
- Huang, Wei-chen. 2012. *A Study of Verbal Morphology in Puljetji Paiwan*. Hsinchu: National Tsing Hua University MA thesis.

- Huang, Yajun. 1988. *Amis Verb Classification*. Taipei: Fu-jen Catholic University MA thesis.
- Jiang, Yijuan. 2012. *A Sketch Grammar of Takibakha Bunun*. Hsinchu: National Tsing Hua University MA thesis.
- Kaufman, Daniel. 2009. On \*pa-, \*pa<R> and \*pa<ŋ>. Paper presented at the Eleventh International Conference on Austronesian Linguistics (11-ICAL), June 22–26, 2009. Aussois, France.
- Kuo, Ching-hwa. 1985. Budai vocabulary. Unpublished typescript.
- Lee, Amy Pei-jung. 1997. *The Case-marking and Focus Systems in Kavalan*. Hsinchu: National Tsing Hua University MA thesis.
- Li, Chao-Lin. 2009. The syntax of prefix concord in Saaroa: restructuring and multiple agreement. *Oceanic Linguistics* 48.1:172–212.
- Li, Paul Jen-kuei. 1973. *Rukai Structure*. Taipei: Institute of History and Philology, Academia Sinica.
- Li, Paul Jen-kuei. 1977. The internal relationships of Rukai. *Bulletin of the Institute of History and Philology Academia Sinica* 48.1:1–92.
- Li, Paul Jen-kuei. 1978. A comparative vocabulary of Saisiyat dialects. *Bulletin of the Institute of History and Philology Academia Sinica* 49.2:133–199.
- Li, Paul Jen-kuei. 1981. Reconstruction of Proto-Atayalic phonology. *Bulletin of the Institute of History and Philology Academia Sinica* 52.2:235–301.
- Li, Paul Jen-kuei. 1988. A comparative study of Bunun dialects. *Bulletin of the Institute of History and Philology Academia Sinica* 59.2:479–508.
- Li, Paul Jen-kuei. (ed.) 2003. *English–Favorlang Vocabulary by Naoyoshi Ogawa*. Tokyo: Research Institute for Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.
- Li, Paul Jen-kuei, & Shigeru Tsuchida. 2001. *Pazih Dictionary*. Taipei: Institute of Linguistics, Academia Sinica.
- Li, Paul Jen-kuei, & Shigeru Tsuchida. 2002. *Pazih Texts and Songs*. Taipei: Institute of Linguistics, Academia Sinica.
- Li, Paul Jen-kuei, & Shigeru Tsuchida. 2006. *Kavalan Dictionary*. Taipei: Institute of Linguistics, Academia Sinica.
- Liao, Hsiu-chuan. 2011. Some morphosyntactic differences between Formosan and Philippine languages. *Language and Linguistics* 12.4:845–876.
- Lichtenberk, Frantisek. 2000. Reciprocals without reflexives. *Reciprocals: Forms and Functions*, ed. by Zygmunt Frajzyngier, 31–62. Amsterdam & Philadelphia: John Benjamins.
- Liu, Dorinda Tsai-hsiu. 1999. *Cleft Constructions in Amis*. Taipei: National Taiwan University MA thesis.
- Liu, Dorinda Tsai-hsiu. 2014. Neutral and imperfective forms in Kanakanavu. *Argument Realisations and Related Constructions in Austronesian Languages: Papers from 12-ICAL*, Vol. 2, ed. by I Wayan Arka & N. L. K. Mas Indrawati, 175–195. Canberra: Asia-Pacific Linguistics Open Access College of Asia and Pacific, The Australian National University.
- Mei, Kuang. 1982. Pronouns and verb inflection in Kanakanavu. *Tsing Hua Journal of Chinese Studies*, New Series 14.1–2:207–232.
- Mills, Roger F. 1975. *Proto South Sulawesi and Proto Austronesian Phonology*. Ann Arbor: University of Michigan dissertation.
- Nichols, Johanna, David A. Peterson, & Jonathan Barnes. 2004. Transitivity and detransitivizing languages. *Linguistic Typology* 8.2:149–211.



- Nojima, Motoyasu. 1996. Lexical prefixes of Bunun verbs. *Gengo Kenkyu* [Journal of the Linguistic Society of Japan] 110:1–27.
- Pan, Chia-jung. 2012. *A Grammar of Lha'alua, an Austronesian Language of Taiwan*. Cairns: James Cook University dissertation.
- Pecoraro, Ferdinando. 1979. *Éléments de grammaire taroko, précédés de la présentation de la vie et de la culture des taroko*. Paris: Association Archipel.
- Post, Ursula. 1992. *Binukid Dictionary*. Manila: Linguistic Society of the Philippines and Summer Institute of Linguistics.
- Reid, Lawrence A. 1966. *An Ivatan Syntax*. Honolulu: University of Hawai'i Press.
- Ross, Malcolm. 2005. The Batanic languages in relation to the early history of the Malayo-Polynesian subgroup of Austronesian. *Journal of Austronesian Studies* 1.2:1–24.
- Ross, Malcolm. 2009. Proto Austronesian verbal morphology: a reappraisal. *Austronesian Historical Linguistics and Culture History: A Festschrift for Robert Blust*, ed. by K. Alexander Adelaar & Andrew Pawley, 295–326. Canberra: Pacific Linguistics.
- Ross, Malcolm. 2012a. Some Proto Austronesian coronals reexamined. Manuscript. Canberra: The Australian National University.
- Ross, Malcolm. 2012b. In defense of nuclear Austronesian (and against Tsouic). *Language and Linguistics* 13.6:1253–1330.
- Starosta, Stanley. 2002. The rise and fall and rise and fall of Proto Malayo-Polynesian. *Collected Papers on Southeast Asian and Pacific Languages*, ed. by Robert S. Bauer, 185–203. Canberra: Pacific Linguistics.
- Sung, Li-May, & Chia-chi Shen. 2006. Reciprocals in Kavalan and a typological comparison. *Streams Converging Into an Ocean: Festschrift in Honor of Professor Paul Jen-kuei Li on his 70<sup>th</sup> Birthday*, ed. by Henry Y. Chang, Lillian M. Huang & Dah-an Ho, 239–277. Taipei: Institute of Linguistics, Academia Sinica.
- Szakos, József. 1994. *Die Sprache der Cou: Untersuchungen zur Synchronie einer austronesischen Sprache auf Taiwan*. Bonn: Rheinische Friedrich-Wilhelms-Universität dissertation.
- Teng, Stacy F. 2008. *A Reference Grammar of Puyuma, an Austronesian Language of Taiwan*. Canberra: Pacific Linguistics.
- Tsuchida, Shigeru. 1969. Word list of three Bunun dialects. Unpublished mimeograph.
- Tsuchida, Shigeru. 1976. *Reconstruction of Proto-Tsouic Phonology*. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.
- Tsuchida, Shigeru. 1980. Puyuma-go (Tamarakao hōgen) goi: fu gohōgaisetsu oyobi tekisuto [Puyuma (Tamalakaw dialect) vocabulary: with grammatical notes and texts]. *Kuroshio no Minzoku, Bunka, Gengo* [Ethnology, Cultures and Languages along the Black Current], ed. by Kuroshio Bunka no Kai [Black Current Cultures Committee], 183–307. Tokyo: Kadokawa Shoten.
- Tsuchida, Shigeru. 1981. Unpublished wordlists of Amis dialects collected in 1980–81. Tokyo: University of Tokyo.
- Tsuchida, Shigeru. 1982. *A Comparative Vocabulary of Austronesian Languages of Sinicized Ethnic Groups in Taiwan*, Part I: *West Taiwan*. Tokyo: Faculty of Letters, University of Tokyo.
- Tsuchida, Shigeru. 1983. Puyuma-English index. *Working Papers in Linguistics, University of Tokyo* 83:10–63.

- Tsuchida, Shigeru. 1985. 'Classificatory prefixes' in the Tsouic languages (Formosa). Unpublished seminar handout, March 14, 1985. Canberra: Department of Linguistics, Research School of Pacific Studies, The Australian National University.
- Tsuchida, Shigeru. 1990. Classificatory prefixes of Tsou verbs. *Tokyo University Linguistics Papers* 89:17–52.
- Tsuchida, Shigeru. 1993. Short sketch of Kavalan grammar. Manuscript.
- Tsuchida, Shigeru. 2000. Lexical prefixes and prefix harmony in Siraya. *Grammatical Analysis: Morphology, Syntax and Semantics: Studies in Honor of Stanley Starosta*, ed. by Videia P. De Guzman & Byron W. Bender, 109–128. Honolulu: University of Hawai'i Press.
- Tsuchida, Shigeru. 2003. *Kanakanavu Texts (Austronesian Formosan)*. Osaka: Endangered Languages of the Pacific Rim, Osaka Gakuin University.
- Tsuchida, Shigeru, Ernesto Constantino, Yukihiko Yamada, & Tsunekazu Moriguchi. 1989. *Batanic Languages: Lists of Sentences for Grammatical Features*. Tokyo: Department of Linguistics, Faculty of Letters, University of Tokyo.
- Tsukida, Naomi. 2005. Seediq. *The Austronesian Languages of Asia and Madagascar*, ed. by K. Alexander Adelaar & Nikolaus P. Himmelmann, 291–325. London & New York: Routledge.
- Tsukida, Naomi. 2008. Verb classification in Amis. *The Typology of Semantic Alignment*, ed. by Mark Donohue & Søren Wichmann, 277–293. Oxford & New York: Oxford University Press.
- Tsukida, Naomi. 2009. *Sedekku-go no Bunpou [A Grammar of Seediq]*. Tokyo: University of Tokyo dissertation.
- Uhlenbeck, E. M. 1949. *De structuur van het Javaanse morpheem*. Bandoeng: Nix.
- Uhlenbeck, E. M. 1950. The structure of the Javanese morpheme. *Lingua* 2:239–270.
- Vendler, Zeno. 1967. *Linguistics in Philosophy*. Ithaca: Cornell University Press.
- Wolff, John U. 2007. Do the Malayo-Polynesian languages constitute a subgroup of the Austronesian language family? *Studies in Philippine Languages and Cultures* 15:13–24.
- Wu, Chunming. 2014. Focus system in Kanakanavu. Manuscript. Taipei: Institute of Linguistics, Academia Sinica.
- Wu, Joy. 2003. Voice affixes and verb classification in Amis. Paper presented at the Second Workshop on Formosan Languages, November 1–2, 2003. Taipei: Academia Sinica.
- Wu, Joy. 2006. *Verb Classification, Case Marking, and Grammatical Relations in Amis*. Buffalo: State University of New York at Buffalo dissertation.
- Yeh, Marie M. 1991. *Saisiyat Structure*. Hsinchu: National Tsing Hua University MA thesis.
- Yeh, Marie M. 2003. *A Syntactic and Semantic Study of Saisiyat Verbs*. Taipei: National Taiwan Normal University dissertation.
- Yeh, Marie M. (n.d.a). Bunun reference grammar. Manuscript.
- Yeh, Marie M. (n.d.b). Morphological causativization in Bunun. Manuscript.
- Zeitoun, Elizabeth. 2000a. Dynamic vs. stative verbs in Mantauran (Rukai). *Oceanic Linguistics* 39.2:415–427.
- Zeitoun, Elizabeth. 2000b. *Bunongyu Cankao Yufa [Bunun Reference Grammar]*. Taipei: Yuan-liou.
- Zeitoun, Elizabeth. 2001. Squib: Negation in Saisiyat: another perspective. *Oceanic Linguistics* 40.1:125–134.
- Zeitoun, Elizabeth. 2002. Reciprocals in the Formosan languages: a preliminary study. Paper presented at the Ninth International Conference on Austronesian Linguistics (ICAL-9), January 8–11, 2002. Canberra: The Australian National University.



- Zeitoun, Elizabeth. 2003. Toward a reconstruction of Proto-Rukai morpho-syntax. Keynote address at the 10<sup>th</sup> Annual Meeting of the Austronesian Formal Linguistics Association (AFLA-10), March 28–30. Honolulu: University of Hawai‘i at Mānoa.
- Zeitoun, Elizabeth. 2005. Tsou. *The Austronesian Languages of Asia and Madagascar*, ed. by K. Alexander Adelaar & Nikolaus P. Himmelmann, 259–290. London & New York: Routledge.
- Zeitoun, Elizabeth. 2007. *A Grammar of Mantauran (Rukai)*. Taipei: Institute of Linguistics, Academia Sinica.
- Zeitoun, Elizabeth, Chu Tai-hwa, & Lalo a tahesh kaybaybaw. 2015. *A Study of Saisiyat Morphology*. Honolulu: University of Hawai‘i Press.
- Zeitoun, Elizabeth, & Lillian M. Huang. 2000. Concerning *ka-*, an overlooked marker of verbal derivation in Formosan languages. *Oceanic Linguistics* 39.2:391–414.
- Zeitoun, Elizabeth, & Stacy F. Teng. 2014. The position of Kanakanavu and Saaroa within the Formosan languages revisited. Paper presented at the 14<sup>th</sup> International Symposium on Chinese Languages and Linguistics (IsCLL-14), June 4–6, 2014. Taipei: Academia Sinica.

[Received 25 August 2014; revised 1 October 2014; accepted 2 October 2014]

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

# 原始南島語動詞類別之重建

Malcolm Ross

澳洲國立大學

雖然構詞創新有助於語族歷史重建，但原始南島語動詞性構詞的重建卻相當少。本文嘗試做這一方面的努力，以彌補此一缺憾。之前的研究已經針對多個台灣南島語言進行了動詞類別的重建，本文則針對原始南島語的動詞類別展開初步的重建，並且檢視該項重建所面臨的挑戰、提供解決之道。本文最後一節簡單扼要地探討原始南島語動詞類別的重建如何帶動構詞創新的建立以及如何增進我們對南島語言分群的瞭解。

關鍵詞：台灣南島語言，動詞性構詞，歷史重建