Reconstruction of '2' in PAN and Related Issues[*]

Elizabeth Zeitoun[1], Stacy Fang-ching Teng[1], and Raleigh Ferrell
Academia Sinica[1]

Blust (1998, 2003, 2009) proposes to reconstruct two sets of numerals in PAN: the first is said to be underived and used for serial counting; the second is derived from the first through Ca- reduplication and is used to count humans. The aim of the present paper is to reassess the reconstruction of numerals from 20-90 in PAN and discuss the use of Ca- reduplication as a primary device to indicate a human distinction at the PAN level.

We argue that in the Formosan languages, there is a dichotomy between two sets of numeral forms, i.e. free vs. bound numerals; both sets must be reconstructed in PAN. Under 10, all numerals are free forms; bound and free numerals may occur together to form higher numerals. This hypothesis leads us to reassess the formation of numerals from 20-90 in PAN. While Blust (1998, 2003, 2009) proposes to reconstruct *ma-puSaN as '20', we show that *puSa- should actually be reconstructed as a bound form meaning '2' equivalent to the free form *duSa. It combines with PAN *ma- 'multiple of tens' and *-N 'recurrence' as *ma-puSa-N '20'. A similar pattern applies to all numerals up to 90.

We further show that the core meaning of Ca- reduplication is iterativity but has extended into marking plurality. Ca- reduplication can be used to indicate [+human] in numerals from 1-9 but this feature might be treated as an epiphenomenon of plural marking.

Key words: numerals, reconstruction, Formosan languages, Proto-Austronesian

1. Introduction

Sagart (2004) argues that the PAN numeral system was quinary. This inference follows from his subgrouping hypothesis: he claims that Pazeh and Saisiyat, which

exhibit quinary numeral systems, are first-order offshoots of PAN and thus reflect the PAN numeral system. Until there is strong evidence that Sagart's subgrouping hypothesis is independent of the reconstruction of a PAN quinary system,[1] we prefer to assume, as others have done (see Dempwolff 1924-1925, 1934-1938, Dyen 1947, 1975, Dahl 1981, Blust 1998, 2003, 2009), that PAN exhibited a decimal system and that numerals formed a closed word class.

In this paper, we propose a reassessment of the reconstruction of the numeral system for multiples of tens in PAN based on a description of numerals from 20-90 in the Formosan languages (see §3). In the second part of the paper, we show that while the core meaning of Ca- reduplication is iterativity, it has further extended into marking plurality. Ca- reduplication can be used to indicate [+human] in numerals from 1-9 but this feature might be treated as an epiphenomenon of plural marking (see §4). Our discussion is based on previous work reviewed in the following section (see §2).

2. Literature review

Li (2006) has provided a typological study on the counting system from 1 to 10 in Formosan languages (see §2.1). Reconstructed forms have also been given for these numerals by Blust (1998, 2003, 2009) (see §2.2). In §2.3, we reassess Blust's (1998, 2003, 2009) reconstruction and present the main hypotheses of this paper.

2.1 A bird eye's view of the numeral system in the Formosan languages

In this section, we provide a bird eye's view of the numeral systems of Formosan languages based on Li's data (2006:135ff). While the analysis that follows is much more complex than the one given in Li (2006), it follows his description in showing that the numeral systems of the Formosan languages are apparently based on a human/non-human distinction in enumeration as opposed to serial counting.

As a first approximation, the Formosan languages can be divided into three main groups as shown in Table 1. One group of languages does not make any distinction between serial counting and non-human/human enumeration. A second group exhibits a dichotomy between human enumeration as opposed to serial counting and non-human enumeration. A third group displays a tripartite distinction; i.e. it distinguishes serial counting, non-human counting and human enumeration. Three languages, Tsou, Atayal, and Thao cross over these three categories, as further explained below.

---

[1] See Winter (2010) for a critique of Sagart's subgrouping.

Table 1: Classification of the Formosan languages based on their (non-)distinction between serial counting and (non-)human enumeration

No distinction	2-way distinction	3-way distinction
Seediq	Bunun	Kavalan
Saisiyat	Paiwan	Kanakanavu
Pazeh	Rukai	Saaroa
Tsou (> 2)	Amis	Puyuma
	Atayal (1, 7-9)	Atayal (2-6)
	Thao (1, 6-9)	Thao (2-5)
	Tsou (< 2)	

The contrast between these three types of languages is illustrated with the numeral '3' below:

(1)   Based on Li (2006:148-149)

		Language	Dialect	serial counting	non-human enumeration	human enumeration
I.	a.	Seediq	Paran	*teru*[2]	*teru*	*teru*
	b.	Saisiyat	Tungho	*too'*	*too'*	*too'*
	c.	Pazeh	–	*turu*	*turu*	*turu*
	d.	Tsou	Duhtu	*turu*	*turu*	*turu*
II.	a.	Bunun	Takbanuaz	*tau*	*tau*	*ta-tau*
	b.	Paiwan	Stimur	*telru*	*telru*	*ma-telru*
	c.	Rukai	Mantauran	*tolro'*	*ma-tolro'*	*ma-tolro'*
	d.	Amis	Farangaw	*tolo'*	*tolo'*	*ta-tolo'*
	e.	Tsou	Duhtu	*turu*	*turu*	*turu*
III.	a.	Kavalan	Hsinshe	*turu*	*u-turu*	*kin-turu*
	b.	Kanakanavu	–	*tuulu*	*u-tulu*	*ta-tulu*
	c.	Saaroa	–	*tuulu*	*u-tulu*	*ta-tulu*
	d.	Puyuma	Katipul	*telru*	*ta-telru[w]-a*	*ta-telru*
	e.	Atayal	Mayrinax	*tuu'*	*tugal*	*ta-tuu'*
	f.	Thao	–	*turu*	*la-turu*	*ta-turu*

---

[2]   The orthographic system adopted in this paper generally follows that of the Ministry of Education and the Council of Indigenous Peoples that was decreed on December 15, 2005. Abbreviations used in this paper include: AV: Agent Voice, Red: Ca- Reduplication, CS: Change of State, Dyn: Dynamic, E: Exclusive, Exp: Experiencer, Fin: Finite, Fut: Future, Gen: Genitive, hum: human, I: Inclusive, Loc: Locative, LocNmz: Locative nominalization, mvt: movement, NFin: Non-Finite, Nom: Nominative, P/plur: plural, PF: Patient Focus, Real: Realis, Rec: Reciprocal, Red: Reduplication, S: Singular, Stat: Stative, Subj: Subjunctive, SubjNmz: Subjective nominalization.

Elizabeth Zeitoun, Stacy Fang-ching Teng, and Raleigh Ferrell

The data in (1) show that the human/non-human distinction can be expressed either through Ca- reduplication, as in Bunun for instance, or through prefixation, as in Kavalan.

Out of 14 Formosan languages, only seven (Bunun, Amis, Kanakanavu, Saaroa, Puyuma, Atayal, and Thao) use, to some extent, Ca- reduplication to mark the human/non-human distinction.

Note that in Amis Ca- reduplication is optional for the numeral '1' (*ca*)*cecay* 'one person' (Li 2006:148). In Katipul Puyuma, Ca- reduplication is used *both* for human and non-human enumeration, except for the numerals from 7 to 9, as shown in (2). The distinction between human and non-human enumeration lies in the suffixation of -*a* in numerals used for non-human enumeration.

(2) Katipul Puyuma (Based on Li 2006:148-150)

	serial counting	non-human enumeration	human enumeration
1.	*isa*	*sa-sa*[*y*]-*a*	*sa-sa*
2.	*zua*	*za-zua*[*y*]-*a*	*za-zua*
3.	*telru*	*ta-telru*[*w*]-*a*	*ta-telru*
4.	*pat*	*pa-pat-a*	*a-'apat*
5.	*lrima*	*la-lu*[*w*]*atr-a*	*la-lu*[*w*]*atr*
6.	*unem*	*na-nem-a*	*a-anem*
7.	*pitu*	*pitu-a*	*pa-pitu*
8.	*walru*	*walru-walru-a*	*wa-walru*
9.	*iwa*	*iwa-iwa*[*y*]-*a*	*a-iwa*

Mayrinax Atayal numerals from 2 to 6 are different when used in serial counting and in the enumeration of (non)humans. This is exemplified in (3a) with the numeral '2'. On the other hand, neutralization occurs in the enumeration of non-humans and humans for the numeral '1' and for the numerals from '7' to '9' in contradistinction with the numerals used in serial counting. This contrast is illustrated for the numeral '8' in (3b).

(3) Mayrinax Atayal (Li 2006:148-149)

	Gloss	serial counting	non-human enumeration	human enumeration
a.	'2'	*rusa'*	*'usaying*	*rarusa'*
b.	'8'	*maqisu'*	*mamaqisu'*	*mamaqisu'*

In Kanakanavu, the prefix *ta*- is used for '1' and '2' in human enumeration; Ca-reduplication is used for numerals above '3'. This is shown in (4).

(4) Kanakanavu (Based on Li 2006:148-149)

	serial counting	non-human enumeration	human enumeration
1.	*cáni*	*u-cáni*	*ta-cíni*
2.	*cuúsa*	*u-rúcini*	*ta-súsa*
3.	*tuulu*	*u-túlu*	*ta-túlu*
4.	*suúpate*	*u-súpate*	*sa-súpate*
5.	*liíma*	*u-líma*	*la-líma*

In Thao, there is a distinction between serial counting and human/non-human enumeration for the numerals 2-5, as shown in (5). This distinction is expressed through *la-* (non-human enumeration) and Ca- reduplication (human enumeration). Note, however, that the prefix *la-* can be used with the same numerals to refer to the enumeration of (non-)humans, as shown in (6b-c).

(5) Thao (Li 2006:148-150)

	serial counting	non-human enumeration	human enumeration
2.	*tusha*	*la-tusha*	*ta-tusha*
3.	*turu*	*la-turu*	*ta-turu*
4.	*shpat*	*la-shpat*	*sha-shpat*
5.	*rima*	*la-rima*	*ra-rima*

(6) Thao (Based on Blust 2003:934)

a. *kan*   *sha-shpat*   *yamin.*
    go   Red-four   1PE.Nom
    'Four of us are going.'

b. *yaku*   *q<m>alush*   *madahun*   *la-shpat*   *iza*
    1S.Nom   <AV>distribute   candy   four   already
    sa   madahun.
        candy
    'I distributed four candies to everyone.'

c. *la-shpat*   *ita*   *inay*   *ilhugqu.*
    four   1PI.Gen   here   sit/visit sb. to socialize
    'Four of us are here.'

Five languages, Nanwang Puyuma, Kavalan, Mantauran Rukai, Paiwan, and Tsou make use of affixation to distinguish between human and non-human enumeration. Three types of affixes are used: verbal affixes, sortal affixes, and the plural prefix *a-* (see Zeitoun 2009a for a discussion on this plural prefix).

Nanwang Puyuma makes use of a verbal affix, *mi-*, which co-occurs with a prefix *a-* tentatively treated as plural for numerals above '2', e.g. *mi-sasa* 'one person' vs. *mi-a-dru[w]a/mi-a-dra-dru[w]a* 'two persons'.

Kavalan, Paiwan, and Tsou make use of sortal affixes: Kavalan *kin-* (Li 2006), Paiwan *ma-* and/or *mane* (Ferrell 1982, Li 2006, Tang 2004). In Tsou, sortal affixes are used in specific contexts. Tung et al. (1964) mention *to-…hu/hʉ* 'person' as opposed to *zo-…hu/hʉ* 'family'; cf. *to-m-tue-hu* 'thirty persons' vs. *zo-m-tue-hu* 'thirty families'. Note in passing that Tsou makes no distinction between serial counting and human/non-human enumeration for numerals above '3'. It makes such a distinction for '1' and '2' through lexical replacement, cf. *coni* '1' vs. *cihi* '1 (person)' and *ruso* '2' vs. *roso* '2 (persons)'.

In Mantauran Rukai, there is no distinction between (non-)human enumeration, as shown in (7a-b), i.e. the same verbal prefix, *ma-* 'stative', is used in both contexts.

(7) Mantauran Rukai (Zeitoun 2007:254)

    a. <u>*ma-eaea*</u>   *ta'olro*   *ta-iki*   *dhona.*
        Stat.Fin-one:Red   dog   SubjNmz-be at   that
        'There is one dog over there.'

    b. <u>*ma-eaea*</u>   *ocao*   *ta-iki*   *dhona.*
        Stat.Fin-one:Red   person   SubjNmz-be at   that
        'There is one person over there.'

Mantauran Rukai exhibits, however, a series of sortal affixes, some of which are specifically used when the referent is [+human]. The prefix *ta'a-* refers to a group of persons (8a); *'ano-Ca-* and *taro-/tao-* both refer to a number of persons or animate referents in movement but occur in complementary distribution: *'ano-Ca-* with bound forms below four (8b) and *taro-/tao-* with bound forms above four, as shown in (8c). The circumfix *maa-…-e* (~ *paa-…-e*) tentatively glossed as 'together' (< *maa-* 'dual reciprocal' prefix for dynamic verbs) is used exclusively with *-dho'a* 'two' as in *maa-dho'a-e* to refer to two persons working or owning a field in common (8d).

(8) Mantauran Rukai (Zeitoun 2007)

    a. <u>*ta'a-tolro*</u>   *ta-'ongolo*   *vavaa.*
        hum not in mvt-three   with-Dyn.NFin:drink   wine
        'Three persons drink/drank wine.' (p.271)

    b. <u>*'ano-ta-tolro*</u>   *dhoace*   *la-ma'a-ali-ali.*
        hum in mvt-Red-three   Dyn.Subj:leave   plur-Rec-Red-(female) friend
        'The three friends left together.' (p.272)

c. <u>*taro-lrima*</u>   *maava'i/*kone.*
   hum in mvt-five   Dyn.Subj:come/*Dyn.Subj:eat
   'Five persons came/*ate.' (p.272)

d. <u>*maa-dhoa'-e-mita*</u>   *ooma.*
   together-two-together-1PI.Nom   field
   'The two of us own/work (the) field together.' (p.274)

We have shown above that some (but not all the) Formosan languages make a distinction between human and non-human enumeration as opposed to serial counting. Such a distinction is marked through two different devices: Ca- reduplication and/or affixation.

This discussion is summarized in a tabular form below (Parentheses indicate low productivity):

Table 2: Devices co-occurring with numerals to mark a human/non-human distinction in Formosan languages

Languages	Ca-reduplication	Lexical/phonological devices	Sortal affix		Verbal affix	Plural affix
			–human	+human		
Seediq	–	–	–	–	–	–
Saisiyat	–	–	–	–	–	–
Pazeh	–	–	–	–	–	–
Tf Tsou	–	+ (1-2)	–	(+)	–	–
Mt Rukai	–	–	–	+	–	–
Paiwan	–	–	–	+	–	–
Kavalan	–	–	+	+	–	–
Ng Puyuma	+ (optional for non-human)	–	–	–	+	+ (*a-* above 2)
Kp Puyuma	+ (also for non-human from 1-6)	–	–	–	–	–
Fa Amis	+ (optional for 1)	–	–	–	–	–
Mx Atayal	+ (2-5)	–	–	–	–	–
Thao	+ (2-5, 7)	–	–	–	–	+ (*la-*)
Tkz Bunun	+	+	–	–	–	–
Saaroa	+	+	–	–	–	–
Kanakanavu	+ (above 3)	+	–	+ (1-2)	–	–

2.2 Two reconstructed sets in PAN

Blust (1998, 2003) makes the following two claims regarding PAN numerals:

(1) He suggests that PAN had "two morphologically related sets of numerals, an unaffixed set [Set A] used in serial counting and in the enumeration of non-human referents, and a second set [Set B] derived from the first by Ca- reduplication and used in counting human referents" (Blust 2003:205). A Thao example is given as an illustration in (9):

(9) Thao (Blust 2003)
    a. <u>*tusha*</u>   *wa*   *qali*
        two   Lig   day
        'two days' (p.1026)

    b. *a*   *kan*   <u>*ta-tusha*</u>   *yamin*   *mu-qariwan.*
        Fut   go   Red-two   1PE.Nom   go-Puli
        'We two (excl.) are going to Pu-li.' (p.1027)

These two sets are reconstructed as follows:

Table 3: Simple and reduplicated PAN numeral forms (after Blust 1998:31)

Set A	Set B	
*pija	*pa-pija	'how much, how many?'
*esa/*isa	*a-esa	'one'
*duSa	*da-duSa	'two'
*telu	*ta-telu	'three'
*Sepat	*Sa-Sepat	'four'
*lima	*la-lima	'five'
*enem	*a-enem	'six'
*pitu	*pa-pitu	'seven'
*walu	*wa-walu	'eight'
*Siwa	*Sa-Siwa	'nine'
*puluq	*pa-puluq	'ten'

Blust argues that while the evidence for reconstructing Set B numerals is not abundant, it is found in languages distributed over a fairly wide geographical area that includes Taiwan (Thao, Bunun, Kanakanavu, Amis, Puyuma), the Philippines, and Indonesia and concludes that "although reflexes of Set B numerals have been reported from fewer than

twenty of the more than 1,000 AN languages, the antiquity of this derivational pattern is beyond dispute" (Blust 1998:34).

(2) He reconstructs the word for '20' as *ma-puSaN and provides as evidence the following data: Proto-Atayal *ma-pusal, Bunun /*ma-pusan*/, Tsou /*mpusku*/, Rukai (Budai) /*ma-posale*/ '20', Kavalan /*-pusan*/ 'twice, do something twice'. (Blust 2003: 206)

2.3 Assessment of Blust's analysis and goals of the paper

Blust's (2003) PAN reconstructions raise a number of questions.

The reconstruction of *ma-puSaN as '20' is problematic when the semantic contrast between Proto-Atayal *ma-pusal 'twenty' as opposed to Kavalan /[*qi-*]*pusan*/ 'twice, do something twice' is taken into account. It thus calls for a reassessment. Based on a description of the formation of numerals from 20 to 90 in the Formosan languages, we shall show in §3 that *ma-puSaN should actually be reconstructed as *ma-puSa-N (see Dahl 1981:52-53, Ossart 2004:112, and Li 2006:140): *puSa- represents the bound counterpart form of *duSa and means '2'; *ma-…-N is composed of two affixes, i.e. *ma- 'multiple of tens' and *-N[3] tentatively glossed as 'recurrence'. We shall try, at the same time, to pin down the use and function of the reflexes of *-N in the Formosan languages and show the validity for its reconstruction at the PAN level.

The reconstruction of two sets of numerals in PAN and in particular the implication that the human/non-human dichotomy already existed at the PAN level is questionable. Ca- reduplication is used for many purposes in Austronesian languages, and it is not surprising that it came to be used as a device to mark the opposition between human vs. non-human. In Table 3, the reconstruction of *a-esa in set B (as opposed to numerals from 2 to 10) is not supported by any reflexes in the Formosan languages where the formation of set B numerals is best represented. Blust (1998:33) mentions that in Muna (Southeast Sulawesi), a "reduplicated" set of numerals contains the forms: *se-ise* 'one', *ru-dua* 'two', *to-tolu* 'three', and so on, and points out that while it is possible that "CV-reduplication in Muna has no historical connection with Ca- reduplication in Formosan and Philippine languages […] the functional agreement in distinguishing a numeral set used for counting humans, and the partial formal agreement in both cases, make convergence a rather unlikely explanation." Reid (2009:240), however, warns against the methodological risk of reconstructing reduplicative forms based on data found in widely separated though genetically related languages as opposed to reconstructing lexical items for which reflexes are consistently found in daughter languages. Finally,

---

[3] As mentioned by Teng & Ross (in press), *-N has an allomorph *-eN: *-N occurs with words ending with an open syllable and *-eN with words ending with a closed syllable.

Blust does not comment on whether the forms for higher numerals (i.e. numerals above 10) also exhibited the same human/non-human distinction in PAN, and whether the word for '20', *ma-puSaN, for instance, had a counterpart used for humans.

We shall show in §4 that while Ca- reduplication has been reconstructed with at least two distinct functions in PAN (① human counting in numerals, ② instrumental nouns) (Blust 2009), its core meaning actually is "iteration" and the [+human] feature that Ca- reduplication marks might be treated as an epiphenomenon of plural marking.

3. Numerals from 10 through 90: evidence for bound numeral forms in Formosan languages

Most Formosan languages[4] fall into two groups. The first group includes languages where tens are expressed by a numeral followed by a ligature and the word 'ten' (see §3.1). In the second group, tens are formed by a bound numeral form to which is attached a complex affix which can be reconstructed as *ma-…-N (see §3.2). While we shall focus on synchronic data, we shall show that these two competing word formations co-existed in PAN.

3.1 Numeral + Ligature + 10

Languages where the numerals 20-90 are made up of a primary cardinal followed by a ligature and the word 'ten' include Amis (10), Paiwan (11), and Pazeh (12). While Amis and Paiwan both exhibit a decimal system, Pazeh displays a quinary system (see Dahl 1981, Sagart 2004, Li 2006:142).

(10)   Amis (Ossart 2004:111)

1	*cecay*		10	*cecay*	*a*	*poloq*
2	*tosa*		20	*tosa*	*a*	*poloq*
3	*tolo*		30	*tolo*	*a*	*poloq*
4	*sepat*		40	*sepat*	*a*	*poloq*
5	*lima*		50	*lima*	*a*	*poloq*
6	*'enem*		60	*'enem*	*a*	*poloq*
7	*pito*		70	*pito*	*a*	*poloq*
8	*falo*		80	*falo*	*a*	*poloq*
9	*siwa*		90	*siwa*	*a*	*poloq*
10	*poloq, mo'tep*					

---

[4]   Data on Kavalan are not included here, as this language falls into none of these categories and thus does not fit into our discussion (see Li 2006).

(11) Paiwan (Ferrell 1982:41)

1	*ita*	10	*pulruq, ta-pulruq*		
2	*drusa*	20	*drusa*	*a*	*pulruq*
3	*tjelu*	30	*tjelu*	*a*	*pulruq*
4	*spatj*	40	*spatj*	*a*	*pulruq*
5	*lrima*	50	*lrima*	*a*	*pulruq*
6	*unem*	60	*unem*	*a*	*pulruq*
7	*pitju*	70	*pitju*	*a*	*pulruq*
8	*alu*	80	*alu*	*a*	*pulruq*
9	*siva*	90	*siva*	*a*	*pulruq*

In either Amis or Paiwan, there are two different forms for the numeral '10', cf. Amis *poloq* vs. *mo'tep* and Paiwan *pulruq* vs. *ta-pulruq*. The first of these forms, *poloq*/*pulruq*, is only used in serial counting and in the formation of tens. The second, *mo'tep*/*ta-pulruq*, is used in non-serial counting.

In Pazeh, the form for 'ten' is *isit*. Multiples of ten are made up the same way as in Paiwan and Amis: Num *a isit*.

(12) Pazeh (Li & Tsuchida 2001)

1	*ida*	10	*isit*		
2	*dusa*	20	*dusa*		*isit*
3	*turu*	30	*turu*	*a*	*isit*
4	*supat*	40	*supat*	*a*	*isit*
5	*xasep*	50	*lima*	*a*	*isit*
6	*xaseb-uza*	60	*xaseb-uza*		*isit*
7	*xaseb-i-dusa*	70	*xaseb-i-dusa*		*isit*
8	*xaseb-i-turu*	80	*xaseb-i-turu*		*isit*
9	*xaseb-i-supat*	90	*xaseb-i-supat*		*isit*

Dahl (1981:52) points out that Western plains languages share the same etymon as in Pazeh:[5] "Taok[as] (*ta*)*isid*, Luilang *isit*, Fav[orlang] *zchiett*, H[o]ny[a] (*miata*)*isi*, perhaps also Pap[ora] (*me*)*tsi* 'ten'" which he tentatively reconstructs as *it₂id. Interestingly, he notes that (i) outside Taiwan, "the numerals from 20 to 90 are generally formed with *puluq preceded by the numerals from 2 to 9" (1981:53), with a nasal (most often *m*) as a linker in between and that (ii) within Taiwan, Paiwan and Amis are the only two languages to have this form. We can therefore tentatively hypothesize that in PAN, multiples of ten were formed *Num + Lig + puluq, as shown in (13):

---

5 Such a remark is not intended to imply the genetic relationship between Pazeh and the Western plains languages.

(13) 20   *duSa   a   puluq
      30   *telu   a   puluq
      40   *Sepat   a   puluq
      50   *lima   a   puluq
      60   *enem   a   puluq
      70   *pitu   a   puluq
      80   *walu   a   puluq
      90   *Siwa   a   puluq

3.2 *ma- + Numeral + -N

As shown in Amis, Paiwan and Pazeh (10)-(12), the lower numerals 2-9 are used in the formation of higher numerals from 20 to 90. These numerals do not undergo any phonological, morphological, or lexical changes. Such changes are observed, on the other hand, in languages where numerals from 20-90 are formed through the affixation of *ma-…-N as shown in (14)-(23) below.

3.2.1 The data

Languages where the multiples of ten for 20-90 are composed of a bound numeral form in co-occurrence with *ma-…-N include (among others) Mantauran Rukai (14), Isbukun Bunun (15), Mayrinax Atayal (16), Truku Seediq (17), Tsou (18), Saaroa (19), Kanakanavu (20), Thao (21), Tungho Saisiyat (22), and Nanwang Puyuma (23), each language presenting characteristics of its own, and some having already innovated new forms for decades.

(14)   Mantauran Rukai (Zeitoun 2007:254, 256)
       1   *neka*   10   *polroko*, *ma-ma-nge-le*
       2   *nosa*   20   *ma-ma-po'a-le*
       3   *tolro*   30   *ma-ma-tolro-lo*
       4   *pate*   40   *ma-ma-'epate-le*
       5   *lrima*   50   *ma-ma-lrima-le*
       6   *neme*   60   *ma-ma-neme-le*
       7   *pito*   70   *ma-ma-pito-lo*
       8   *valro*   80   *ma-ma-valro-lo*
       9   *vangate*   90   *ma-ma-vangate*

Mantauran Rukai has two forms for '10', viz. *polroko*, *ma-ma-nge-le*. The word *polroko* is an irregular reflex of PAN *sa-puluq: PAN *q was lost in PR so the expected

form is **pulru (Li 2006:140, Zeitoun 2007:253). *polroko* should thus be treated as a loan because it is phonologically irregular.

(15) Isbukun Bunun (Li 1997:551-554)

1	*tasa'*	10	*ma-s'a-n*
2	*dusa'*	20	*ma-pusa-n*
3	*tau*	30	*ma-tiu-n*
4	*paat*	40	*ma-sipat-un*
5	*'ima'*	50	*ma-ima-un*
6	*nuum*	60	*ma-num-un*
7	*pitu'*	70	*ma-pitu-un*
8	*vau'*	80	*ma-vau'-un*
9	*siva'*	90	*ma-siva-un*

(16) Mayrinax Atayal (Li 2006:141 & 148-150, C.-M. Wu p.c.)

1	*qun*	10	*pug, ma-gal-pug*
2	*rusa'*	20	*ma-pusa-l*
3	*tuu'*	30	*ma-tuu-l*
4	*sepat*	40	*ma-spat-ul*
5	*t-ima'*	50	*ma-ima-l*
6	*ma-tuu'*	60	*ma-tuu' p-ga-l*
7	*pitu'*	70	*ma-pitu' p-ga-l*
8	*ma-spat*	80	*ma-spat p-ga-l*
9	*ma-qisu'*	90	*ma-qisu' p-ga-l*

(17) Truku Seediq (Tsukida 2005:297)

1	*'uin, kingal*	10	*ma-xa-l*
2	*deha*	20	*ma-pusa-l*
3	*teru*	30	*me-teru-l*
4	*sepat*	40	*ma-spat-ul*
5	*rima*	50	*m-rima-l*
6	*ma-taru*	60	*ma-taru ke-mexa-l-an*
7	*em-pitu*	70	*em-pitu ke-mexa-l-an*
8	*ma-spat*	80	*ma-spat ke-mexa-l-an*
9	*mengari*	90	*mengari ke-mexa-l-an*

There are two forms for one in Truku Seediq, *'uin* and *kingal*. Tsukida (2005:297) argues that *'u[w]in* is used when counting things while *kingal* is used in enumeration.[6]

---

[6] Pecoraro (1977:324) mentions, on the other hand, that *'uin* is only used in serial counting while *kingal* is used to count [±human] counting.

(18) Tfuya Tsou (Zeitoun 2005:265)

1	*coni*	10	*ma-s-kʉ*
2	*yuso*	20	*ma-pus-ku*
3	*tuyu*	30	*m-tuyu-hu*
4	*sʉptʉ*	40	*m-sʉptʉ-hʉ*
5	*imo*	50	*m-eimo-hʉ*
6	*nomʉ*	60	*m-onmʉ-hʉ*
7	*pitu*	70	*m-pʉtvʉ-hʉ*
8	*voyu*	80	*m-voyvʉ-hʉ*
9	*sio*	90	*m-sio-hʉ*

(19) Saaroa (Li 1997:551-554 and Pan forthcoming)

1	*caani*	10	*ku-ma-a-lhe ~ ma-a-lhe ~ 'umara-ma-a-lhe*
2	*suusa*	20	*ma-pua-lhe*
3	*tuulu*	30	*ma-tulu-lhu*
4	*paate*	40	*ma-upate-lhe*
5	*ku-lima*	50	*ma-lima-lhe*
6	*ke-neme*	60	*ma-eneme-lhe*
7	*ku-pitu*	70	*ma-pitu-lhe*
8	*ku-alu*	80	*ma-ale-lhe*
9	*ku-sia*	90	*ma-sia-lhe*

Pan (forthcoming) suggests that there is a distinction between serial counting (*ku-ma-a-lhe* '10') and (non-)human counting (cf. *ma-a-lhe* '10 non-human enumeration' vs. *'umara-ma-a-lhe* '10 non-human enumeration') for numerals above 10 (and up to 1000). It is not clear at this stage, whether this distinction is kept for numerals from 20 through 90.

(20) Kanakanavu (Li 1997:551-554)

1	*caani*	10	*ma-a-ne*
2	*cuusa*	20	*ma-pusa-ne*
3	*tuulu*	30	*ma-tuu-nu*
4	*suupate*	40	*ma-spatu-nu*
5	*liima*	50	*ma-ima-en*
6	*neeme*	60	*ma-enem-en*
7	*piito*	70	*ma-pitu-un*
8	*aalu*	80	*ma-alu-n*
9	*siia*	90	*ma-sia-en*

(21) Thao (Blust 2003:204-205)

1	*tata*	10	*maqcin*
2	*tusha*	20	*ma-pusha-z*
3	*turu*	30	*ma-turu-z*
4	*pat*	40	*ma-shpat-iz*
5	*rima*	50	*ma-rima-z*
6	*makalh-turu-turu*	60	*makalh-turu-turu a maqcin*
7	*pitu*	70	*ma-pitu-'un*
8	*makalh-shpat-shpat*	80	*makalh-shpat-shapt a maqcin*
9	*tanacu*	90	*makalhtanacu a maqcin*

Thao exhibits a mixed system with both the occurrence of the circumfix *ma-…-z* for 20-50, *ma-…'un* for 70 (a loan from Bunun, according to Blust 2003:206) and the occurrence of the primary numerals 6, 8, and 9 followed by the ligature *a* and the word *maqcin* 'ten' for 60, 80, and 90.

(22) Tungho Saisiyat (Zeitoun et al. forthcoming)

1	*'aehae'*	10	*langpez*
2	*rosha'*	20	*sha-m'iaeh*
3	*too'*	30	*ma-too-l*
4	*shopat~shepat*	40	*mon-oshpat*
5	*(h)aseb*	50	*mon-aseb*
6	*shayboshii*	60	*mon-shayboshii*
7	*shayboshii=o 'aehae', yoehae'*	70	*mon-shayboshii=o 'aehae'*
8	*kashpat*	80	*mon-may-kashpat*
9	*(h)ae'hae', tae'hae'*	90	*mon-ae'hae'*

Saisiyat has undergone a series of lexical replacements: *langpez* '10' is an innovated form; *sham'iaeh* '20' is composed of *sha-* which means '1' and *m[a]'iaeh* 'person' (with further resyllabification: *sham.'iaeh*, see Li 2006:140); *matool* '30' is composed of the complex affix *ma-…-l* 'decade' (< PAN *ma-…-N); from 40 on, the prefix is *mon-*. The reason why *-l* does not appear on numerals above 30 in Saisiyat is that they all end with a consonant,[7] and consonant clusters in final position are disallowed in that language.

---

[7] In Tungho Saisiyat, the long vowel *ii* is *shayboshii* '6' results from the deletion of the flap (see Li 1978).

(23)   Nanwang Puyuma

1	*sa*	10	*pulru'*
2	*drua*	20	*maka-betra'an*
3	*telru*	30	*maka-telru-n*
4	*pat*	40	*maka-pet-el*
5	*lrima*	50	*maka-luwatr*
6	*enem*	60	*maka-nem-en*
7	*pitu*	70	*maka-pitu*
8	*walru*	80	*maka-walru*
9	*iwa*	90	*maka-iwa*

Teng & Ross (in press) note that Puyuma has two reflexes of *N, namely *n* and *l*, *l* being more frequent than *n* in Puyuma reflexes. They propose that final PAN *N has become *n* when "the onset of the final syllable is the retroflex lateral -*lr*- (< PAN *l): *ma-ka-telru-n* 'thirty' (< PAN *ma-telu-N), *par-telru-n* 'do three times' (< PAN *paR-telu-N) and *ma-ka-walru-n* 'eighty' (< PAN *ma-walu-N)." In Nanwang, the -*n* in *maka-walru-n* has been dropped.

3.2.2 Bound numeral forms in Formosan languages

Based on the above data and discussion, we suggest positing for Mantauran Rukai, Isbukun Bunun, Tsou, Saaroa, Kanakanavu, Thao, Tungho Saisiyat, Paiwan, and Nanwang Puyuma the following bound numeral forms (24)-(32) as opposed to the free numeral forms found in serial counting.

Bound forms are used productively in all the Formosan languages. Besides being used to form higher/complex numbers (e.g. 10 to 90, 100 to 900, and above), they may co-occur with different verbal prefixes, time/frequency and ordinal affixes to form composite numerals which usually function as predicates (see Zeitoun 2009b, Zeitoun et al. in preparation). In cases where more than one bound form has been reported for a specific language, we provide an example to show the distribution of such bound forms.

(24)   Mantauran Rukai (Zeitoun 2007:254)

1	*neka*	-*nge*, as in *ma-ma-**nge**-le* '10' (*Lit.*: 'one ten')
		-*ngea*, as in *ta'a-**ngea***
		-*eaea*, as in *ma-**eaea*** '(there is) one…'
		-*ea*, as in *'aka-**ea*** 'the first'
		-*a*, as in *'a-paka-**a**-le* 'the first day/month/year'
2	*nosa*	-*dho'a*, as in *otara-**dho'a*** 'two months/years old'
		po'a-, as in (*o*-)***po'a**-le* 'two bowls'

3	*tolro*	*-tolro*
4	*pate*	*-'epate*
5	*lrima*	*-lrima*
6	*neme*	*-(e)neme*
7	*pito*	*-pito*
8	*valro*	*-valro*
9	*vangate*	*-vangate*

(25) Isbukun Bunun

1	*tasa'*	*s'a-*
2	*dusa'*	*pusa-*
3	*tau*	*tiu-*
4	*paat*	*sipat-*
5	*'ima'*	*ima-*
6	*nuum*	*num-*
7	*pitu'*	*pitu-*
8	*vau'*	*vau'-*
9	*siva'*	*siva-*

(26) Tfuya Tsou (Tung et al. 1964, unless noted otherwise)

1 *coni* (V)*s-*, as in *ma-**s**-kʉ* '10' (*Lit.*: 'one ten')
cni-, as in *zo-**cni*** 'one family'
nʉs-, as in *moh-**nʉs**-kʉ* 'go once'
nso-, as in *to-**nso**-ha* 'one year'

2 *yuso* *pus-*, as in *o-m-**pus**-ku* 'go twenty times'
pso-, as in *o-**pso**-hʉ* 'go twice'
po-, as in *ae-**po**-hʉ* 'two at a time'
-eso, as in *meha-**eso*** 'two pieces of trees'
yus-, as in *ma-yu-**yus**-no-hie* 'every two days' (Pan 2007:120)

3 *tuyu* *tuy-*, as in *o-m-**tuy**-hu* 'go thirty times'
-teu, as in *mo-**teu*** 'get three prey'

4 *sʉptʉ* *sʉptʉ-*, as in *no-**spʉtʉ**-hʉ* 'stay for four days'
spotʉ-, as in *o-**spotʉ**-hʉ* 'go four times'

5 *imo* *iemo-*, as in *to-**iemo**-ha* 'five years'
emo-, as in *o-**emo**-hʉ* 'go five times'

6 *nomʉ* *onmʉ-*, as in *m-**onmʉ**-hʉ* 'sixty'

7 *pitu* *pʉtvʉ-*, as in *moh-**pʉtvʉ**-kʉ* 'go seven times'
ptu-, as in *o-**ptu**-hu* 'go seven times'

	8	*voyu*	*voyvɨ-*, as in *moh-**voyvɨ**-kɨ* 'go eight times'
			-veo, as in *meha-**veo*** 'eight pieces of trees'
	9	*sio*	*sio-*

(27) Saaroa

	1	*caani*	*a-*
	2	*suusa*	*pua-*, as in *ma-**pua**-lhe* 'twenty'
			-sua, *u-**sua*** '2 (things)', *sa-**sua*** 'two (persons)'
	3	*tuulu*	*tulu-*
	4	*paate*	*upate-*
	5	*ku-lima*	*lima-*
	6	*ke-neme*	*eneme-*
	7	*ku-pitu*	*pitu-*
	8	*ku-alu*	*ale-*
	9	*ku-sia*	*sia-*

(28) Kanakanavu

	1	*caani*	*a-*
	2	*cuusa*	*pusa-*
	3	*tuulu*	*tuu-*
	4	*suupate*	*-supate*, as in *u-**supate*** 'four (things)'
			spatu-, as in *ma-**spatu**-nu* 'forty'
	5	*liima*	*enem-*
	7	*piito*	*pitu-*
	8	*aalu*	*alu-*
	9	*siia*	*sia-*

(29) Thao

	1	*tata*	*ta-*[8]
	2	*tusha*	*pusha-*
	3	*turu*	*turu-*
	4	*pat*	*shpat-*
	5	*rima*	*rima-*
	6	*makalhturuturu*	--

---

[8] Blust (2003) considers *-tal* '1' to be a single (bound) word, but the occurrence of forms such as *mu-shpat-iz* 'do/go four times', *mu-rima-z* 'do/go five times' shows that *-l* is part of the complex affix *mu-…*(C), where C = *-l*, *-(i)z* or ∅. Paul Jen-kuei Li (p.c.) also confirmed that *-(i)z* and *-l* often occur in free variation, e.g. *fural ~ furaz* 'moon'.

7	*pitu*	*pitu-*
8	*makalhshpatshpat*	--
9	*tanacu*	--

In Thao, no bound forms can be given for the numerals '6', '8', or '9', since these numerals represent innovated forms.

(30) Tungho Saisiyat (Zeitoun et al. forthcoming)

1	*'aehae'*	*sha-*, as in ***sha**-mi'aeh* 'twenty'
		ha-, as in (*h*)*in-**ha**-l* 'beat once'
2	*rosha'*	*posha-*
3	*too'*	*too-*
4	*shopat~shepat*	*shpat-*
5	(*h*)*aseb*	*aseb-*
6	*shayboshii*	*shayboshii-*
7	*shayboshii=o 'aehae', yoehae'*	*shayboshii=o 'aehae'-*
8	*kashpat*	*may-kashpat-*
9	*hae'hae', tae'hae'*	*ae'hae'-*

In Saisiyat, *sha-* '1' is found only in *sham'iaeh* '20'; in all other composite numerals, where a bound numeral form co-occurs with an affix to form a verb, *ha-* is used instead.

(31) Paiwan

1	*ita*	*ta-*
2	*drusa*	*pusa-*
3	*tjelru*	*tjelru-*
4	*spatj*	*si-matj-*
5	*lrima*	*lrima-*
6	*unem*	*nem-*
7	*pitju*	*pitju-*
8	*alu*	*valu-*
9	*siva*	*siva-*

(32) Nanwang Puyuma

1	*sa*	*sa-*
2	*drua*	*betra'an-*
3	*telru*	*telru-*

4	*pat*	*pet-*
5	*lrima*	*luwatr-*
6	*enem*	*nem-*
7	*pitu*	*pitu-*
8	*walru*	*walru-*
9	*iwa*	*iwa-*

The paradigms given in (24)-(32) show that some numeral forms exhibit two distinct forms; others do not. Table 4 tabulates the differences (marked by ✓) perceived between bound and free numeral forms.

Table 4: Differences between bound and numeral forms in Formosan languages

	1	2	3	4	5	6	7	8	9
Mt Rukai	✓	✓		✓		✓			
Is Bunun	✓	✓	✓	✓		✓			
Tf Tsou	✓	✓	✓	✓	✓	✓	✓	✓	
Saaroa	✓	✓	✓	✓		✓		✓	
Kanakanavu	✓	✓	✓	✓	✓	✓	✓	✓	
Thao	✓	✓		✓					
Tg Saisiyat	✓	✓		✓					
Paiwan	✓	✓		✓		✓		✓	
Ng Puyuma		✓		✓	✓	✓			

The recognition of bound numeral forms in the Formosan languages might not represent a breakthrough in itself, as other languages in the world are known to have free and bound numerals as well (e.g. *four ~ for-ty*, *five ~ fif-ty*, B. Bender p.c). However, it has some consequences on the reconstruction of the PAN numeral system (see §3.2.3).

3.2.3 Evidence for the reconstruction of *puSa- as '2'

In this section we reassess the reconstruction of the word '20' in PAN and show that the formation of this numeral tells something about the formation of the multiples of ten.

As mentioned above, the word for '20' has been reconstructed as PAN *ma-puSaN by Blust (1998, 2003). Its occurrence in the Formosan languages is somehow intriguing since in languages outside Taiwan, the formation of multiples of ten is generally the same: *Num (+ Lig) puluq. Dahl (1981:55) wonders whether *puSaN could be treated as "an older word for ten, having acquired the meaning 'twenty' from the doubling of

*ma-." This is improbable, though, as in many Formosan languages, *ma-…-N forms multiples of ten and constitutes a whole paradigm for numerals from 10 to 90.

We note, on the other hand, that all the above-mentioned languages, with the exception of Saisiyat (22) and Puyuma (23), exhibit a parallel between *duSa* and *puSa-*, e.g. Isbukun Bunun *dusa'* '2' vs. *ma-pusa-n* '20'.[9] This observation corroborates earlier findings by Ferrell (1982:43), Ossart (2004:112), and Li (2006:140). Comparative data further show that in the Formosan languages, reflexes of PAN *duSa and *puSa- cooccur in complementary distribution: reflexes of *duSa appear alone, as in (33), or preceded/followed by prefixes and/or suffixes (with the exception of *-N), reflexes of *puSa- always co-occur with a *-N 'recurrence' (see §3.3 below), as in (34). We thus posit *puSa- '2' as the bound counterpart form of the free numeral form *duSa.

In (33), we show the contrast between the PAN forms '2', *duSa used in serial counting and *puSa- used in the formation of the numeral '20', and their reflexes in the Formosan languages.

(33) Reflexes of:

	*duSa '2'	*puSa- '2'	Notes
Mantauran Rukai	*nosa*	*ma-po'a-le*	(PAN *S > PR *s > Mt ', PAN *N > PR *l)
Isbukun Bunun	*dusa'*	*ma-pusa-n*	(PAN *N > PB *n*)
Tfuya Tsou	*ruso*	*ma-pus-ku*	(PAN *N > Tsou *h*, *h* → *k/s___*)
Saaroa	*suusa*	*ma-pua-lhe*	(PAN *S > Saaroa Ø, *N > Saaroa *lh*)
Kanakanavu	*cuusa*	*ma-pusa-ne*	(PAN *N > Kan *n*)
Mayrinax Atayal	*rusa'*	*ma-pusa-l*	(PAN *N > PA *l*)
Truku Seediq	*daha*	*m-pusa-l*	(PAN *N > PA *l*)
Thao	*tusha*	*ma-pusha-z*	(PAN *S > Thao *sh*, *N > Thao *z*)[10]

Further examples with reflexes of *duSa and *puSa- in co-occurrence with other types of affixes are evidenced in five languages (Rukai, Saisiyat, Thao, Paiwan, and Puyuma) and show the same distribution between these two reflexes (Zeitoun 2007:255). Note that in Saisiyat and Puyuma, while the change from *duSa to *puSa- is not apparent in the word for 'twenty', for which these two languages have suppletive forms, it becomes obvious in other related forms meaning 'two N …' and/or '… two times'.

---

[9] Similar phenomena are found in other affixed forms in the daughter languages, Paiwan *maka-si-matj-el* 'for four days' (< *sepatj* 'four') and *-ngida* 'when?' (< *pida* 'how many?')

[10] Blust (2003:752) glosses Thao *pushaz* as "two, in certain collocations" but does not account for the complementary distribution between *pushaz* and *tusha* '2'.

(34) Mantauran Rukai (Zeitoun 2007)
    a. *o-tara-dho'a*   'two (months/years)' (p.263)   **o-tara-po'a*
    b. *o-ka'a-po'a-le*   '(drink/eat) two bowls' (p.275)   **o-ka'a-dho'a-le*

    Tungho Saisiyat (Zeitoun et al. forthcoming)
    c. *kina-rosha'-an*   'two glasses'   **kina-posha-an*
~   c′. *kina-posha-l-an*   'two glasses'   **kina-rosha-l-an*
    d. *sha-posha-l*   'see twice'   **sha-rosha'*

  Thao (After Blust 2003)
    e. *min-tusha*   'become two' (p.1027)   **min-push*a
    f. *lhin-tusha-in*   'be divided in two' (PF) (p.1027)   **lhin-pusha*
    g. *makim-pusha-z*[11]   'two floors' (p.752)   **makim-tusha-z*

  Paiwan (After Ferrell 1982:89)
    h. *si-ka-drusa*   'second'   **si-ka-pusa*
    i. *maka-pusa-l*   '(for) two days, on two occasions'   **maka-drusa-l*

  Nanwang Puyuma
    j. *puka-dra-drua*   'second'   **puka-pa-pua*
    k. *par-pua-n*   'twice'   **par-dua-n*

The next conclusion that imposes itself is that multiples of ten in PAN were also formed by *ma-…-N.

(35) 20   *ma-puSa-N
    30   *ma-telu-N
    40   *ma-Sepat-eN
    50   *ma-lima-N
    60   *ma-enem-eN
    70   *ma-pitu-N
    80   *ma-walu-N
    90   *ma-Siwa-N

---

[11] Blust (2003:100) defines *-iz* (~ *-z*) as "a suffix used with numerals, often in combination with other affixes […]. *-iz* appears to have a variety of functions, although this is difficult to determine when it combines with other affixes, since the individual contributions of each affix cannot always be disentangled unambiguously."

3.3 Reconstruction of *-N 'recurrence'

The widely distributed *-N in the word for 'twenty' has been shown in (33) above and its occurrence across such disparate languages as Atayalic, Tsouic, Saisiyat, Bunun, Thao, and Rukai (which form primary subgroups according to Blust 1999) strongly suggests that *-N must have been a PAN suffix, retained unevenly but in sufficiently different uses in each language so that later diffusion appears an unlikely source.

Aside from the examples given in (34), this suffix can appear with different prefixes in occurrence with bound numeral forms, to indicate, for instance, periods of time, as in Paiwan *paka-...-l*/*maka-...-l* and Mantauran Rukai *paka-...-le*/*-lo*/*maka-...-le*/*lo* as in (36) (see Dahl's 1981:56 reconstruction *(ma)ka-…-(a)ɬ) or frequency as in (37).

(36)　a.　Paiwan
　　　　　paka-...-l ~ maka-...-l 'for....N days'
　　e.g.　*paka-ta-l ~ maka-ta-l* 'for one day'
　　　　　paka-pusa-l ~ maka-pusa-l 'for two days'

　　　　b.　Mantauran Rukai
　　　　　paka-...-le/*-lo ~ maka-...-le*/*-lo* 'for....N days'
　　e.g.　*paka-a-le ~ maka-a-le* 'for two days'
　　　　　paka-po'a-le ~ maka-po'a-le 'for two days'

(37)　a.　Tsou (Tung et al. 1964:589, 615)
　　　　　i'-...-hu/*-hɨ* 'N times'
　　e.g.　*i'-nɨs-kɨ* 'once'
　　　　　i'-po-pso-hɨ 'twice'

　　　　b.　Paiwan
　　　　　kin-...-l 'N times'
　　e.g.　*kin-ta-l* 'once'
　　　　　kin-musa-l 'twice' (where *pusa- > musa-*)

　　　　c.　Mantauran Rukai (Zeitoun 2007:265)
　　　　　'o-...-le 'N times'
　　e.g.　*'o-a-le* 'once'
　　　　　'o-po'a-le 'twice'

　　　　d.　Nanwang Puyuma
　　　　　par-ia-...-l~par-... (-*n*) 'N times'
　　e.g.　*paria-sa-l* 'once'
　　　　　par-pua-n 'twice'

e.   Tungho Saisiyat (Zeitoun et al. forthcoming)
      'on-...(-l) ~ mon-...(-l) 'N times'
e.g. *'on-ha-l ~ mon-ha-l* 'once'
      'on-posha-l ~ mon-posha-l 'twice'

The reflex of *-N can combine with other affixes, e.g. lexical affixes, sortal affixes, ordinal affixes, etc. (see Zeitoun et al. in preparation). In Thao (Blust 2003), Bunun (Lin et al. 1998), Saisiyat (Zeitoun et al. forthcoming), Paiwan (Ferrell 1982), Puyuma (Cauquelin 1991), and Tsou (Tung et al. 1964, Pan 2007, Chang 2009), bound numeral forms attach to different verbal affixes, made up of two parts: the lexical prefix gives the semantic meaning; e.g. Saisiyat *(h)in-* 'beat', *kal-* 'go by foot', *min-* 'blossom' etc., while the reflex of *-N refers to 'a number of times'. The derived numeral (referred to as a composite numeral) thus functions as a verb whose meaning depends on the prefix and refers to that action made 'a number of times'. Examples (38) are drawn from Saisiyat and Thao and show the contrast between the lexical prefix used with a verb or a noun as opposed to when it co-occurs with a numeral. Other examples are given in (39). Note that in Paiwan, a lexical prefix can attach to the sequence Num- + -(*e*)*l* or a numeral can attach to Verb + -(*e*)*l*.

(38) Tungho Saisiyat (Zeitoun et al. forthcoming)
      a.   *yako*   <u>*kal-'aish*</u>   *kala   koko'   rima'*
           1S.Nom   pass through-in passing Loc   grandmother go
           ka-kish-kaat-an.
           Real-read aloud-write-LocNmz
           'I went by grandmother's home on my way to school.'

      a′. *yao*   *'ina=*<u>*kal-ha-l*</u>*=ila*   *rima'   ray'in.*
           1S.Nom   Exp=go by foot-one-N times=CS   go   Penglai
           'I went to Penglai once (implied: by foot).'

      Thao (Based on Blust 2003)
      b.   *yamin*   <u>*k<un>saqazi*</u>   *k<m>an.*
           1PE.Nom   eat<AV>noon   <AV>eat
           'We are eating lunch.' (p.864)

      b′. <u>*k<un>turu-z*</u>   *yaku*   *k<m>an.*
           eat<AV>three-N times   1S.Nom   <AV>eat
           'I ate three times.' (p.1021)

(39) a. Paiwan
        pa-kin-ta-l   'to do with one action/movement'
        tja-uma-l   'do again differently'
        ta-meqe-l   'to swallow in one gulp' (< *ri-mequ* 'to swallow')

    b. Isbukun Bunun
        ma-pit-s'a-n   'to cook once' (< *ma-pit-* 'to cook')
        tal-pusa-n   'to stay two nights' (< *tal-* 'to stay overnight')

4. Another look at the use of Ca- reduplication in numerals

We are now left with a number of questions and hypotheses regarding the PAN numeral system that need to be addressed and carefully considered. Was the PAN numeral system based on a dichotomy between 'bound' and 'free forms'? Was the human/non-human distinction a primary device in PAN? Was it exclusively marked by Ca- reduplication?

Evidence in support of the reconstruction of a distinction between bound and free forms is not strong. The only bound numeral form that is different from the free numeral forms is PAN *puSa-. However, in modern languages (see (14)-(23)), it is the bound numeral forms that typically reflect more closely PAN forms, e.g. Mantauran Rukai -(*ea*)*ea* '1' (as opposed to *neka*), -*dho'a* '2' (as opposed to *nosa*), Isbukun Bunun *s'a-* '1' (as opposed to *tasa'*) and in many languages the bound numeral form for '4' contains a reflex of *S, e.g. Mantauran Rukai *'epate-* '4' as opposed to *pate*, Isbukun Bunun *sipat-* '4' as opposed to *paat*, Thao *shpat-* '4' as opposed to *pat*, while it is deleted in serial counting. Based on comparative evidence, it is thus highly plausible that the PAN numeral system was based on a dichotomy between 'bound' and 'free forms'.

The reconstruction of a human/non-human distinction as a primary device at the PAN level is questionable. A close inspection of the numerals from 1 to 9 in the Formosan languages (see §2.1) raises doubts as to what the original device to mark the human/non-human distinction was. The gaps regarding the use of Ca- reduplication for certain numerals from 1 to 9 in Thao and Mayrinax Atayal can easily be explained because the numerals that do not undergo Ca- reduplication represent forms derived from '3' and '4', e.g. Thao *makalh-turu-turu* 'six', *makalh-shpat-shpat* 'eight', and Mayrinax Atayal *ma-tuu'* 'six', *ma-shpat* 'eight'. We note, on the other hand, that the same form in Tsou, Kanakanavu, and Saaroa is expressed by three different processes, i.e. lexical replacement, affixation, and Ca- reduplication. Besides, in the same language family (e.g. Atayalic), some languages/dialects make a distinction while others do not. Finally Ca- reduplication usually occurs for numerals greater than '1'.

While Blust (2009) posits at least two functions for Ca- reduplication in PAN (① human counting in numerals, ② instrumental nouns), Yeh (2009) and Reid (2009) have attempted to unify the different functions of Ca- reduplication in the Formosan and Philippine languages respectively. Though the development paths they propose for Ca- reduplication differ, we are inclined to follow their analysis. We agree with Reid (2009:247) that "once divorced from their original verbal context, the reduplications could then be extended to derive a third person plural demonstrative pronoun […], quantifiers and numerals referring specifically to humans […] and eventually to general (that is non-human) plural nouns, as shown in (40).

(40) Paths of development of Ca- reduplication according to Reid (2009:247)

> PL HUMAN DEMONSTRATIVE

ITERATIVE > DISTRIBUTIVE > PL HUMAN AGENT > PL HUMAN NOUN > HUMAN QUANTIFIERS

> GENERAL PLURAL

Following Reid (2009), we posit that the marking of plurality is an extension of iterativity and that the marking of [humanness] is an epiphenomenon of plurality.

We have questioned Blust's (1998) reconstruction of *a-esa '1 (set B)' (see Table 3). If Ca- reduplication is understood as conveying a plural meaning, then it becomes clear (i) why it is not found in Atayal and Thao for the numeral '1' and (ii) why it is optional in Amis while in the use of a suppletive form undergoing Ca- reduplication in both Saaroa and Bunun can be further understood as a result of analogy (see Table 5). Its associated meaning [+human] also becomes obvious, since plurality is usually associated with the notion of [+humanness].

Table 5: Marking of the numeral form '1' used for human counting

Languages	Ca- reduplication	Suppletive form	Affixation with the plural *a-*	Affixed form	
				Sortal affix	Verbal affix
Mx Atayal	–	–	–	–	
Thao	–	–	–	–	
Amis	optional	–	–	–	
Ng Puyuma	+ (optional for non-human)	–	– (as opposed to forms above 2)	–	+
Kp Puyuma	+ (also for non-human)	–	–	–	
Tf Tsou	–	+	–	–	
Is Bunun	+	+	–	–	
Saaroa	+	+	–	–	
Kanakanavu	–	+	–	+	–

If Ca- reduplication is treated as a secondary development, it also explains why in many languages, it is the bound numeral form that undergoes Ca- reduplication, and not the free form. Bound forms must have existed prior to their reduplication with Ca-. Consider the following paradigms:

(41)

	Gloss	Serial counting	Bound numeral form	Ca- reduplicated form
Saaroa	2	*suua*	*sua-*	*sa-sua*
Bunun	4	*paat*	*sipat-*	*sa-spat*
Thao	4	*pat*	*shpat-*	*sha-shpat*
Kanakanavu	4	*suupate*	*-supate*	*sa-supate*
Saaroa	4	*paate*	*upate-*	*a-upate*
Katipul Puyuma	5	*lrima*	*luwatr-*	*la-luwatr*

5. Conclusion

We have shown that while earlier studies have concentrated exclusively on a distinction between "human" and "non-human" in primary cardinal numbers, there exists another more basic distinction that has been overlooked in the past, i.e. a distinction between free vs. bound numeral forms. The recognition of the existence of bound numeral forms has led us to question the validity of the reconstruction of PAN *ma-puSaN as '20' and posit instead PAN *ma-puSa-N as a complex numeral form composed of *puSa-* '2' the bound counterpart of PAN *duSa '2', *ma- 'multiple of ten' and *-N 'recurrence' while showing that the reconstruction of PAN multiples of 10 through 90 follows the same pattern, i.e. *ma- + primary cardinal + *-N. Another method of expressing multiples of '10', lower numeral + Lig *a* + 'ten', also needs to be reconstructed as it is found in two languages which do not form a subgroup (Dahl 1981). We have argued that Ca- reduplication is not the only device to mark [+human] distinction. Following Reid (2009), we have posited that the "core" meaning of Ca- reduplication is iterativity. Its meaning further extended to mark plurality. The marking of [humanness] is treated as an epiphenomenon of plurality.[12]

---

[12] Reid (2009) argues that Ca- reduplication must have originally been CV- due to the strong analogical pressure from the preponderance of *a*-vowels in the first syllable of reduplicated bases, and *a*-vowel prefixes.

References

Aikhenvald, Alexandra Y. 2000. *Classifiers: A Typology of Noun Categorization Devices*. Oxford & New York: Oxford University Press.

Bender, Andrea, and Sieghard Beller. 2006. Numeral classifiers and counting systems in Polynesian and Micronesian languages: common roots and cultural adaptations. *Oceanic Linguistics* 45.2:380-403.

Blust, Robert A. 1998. Ca- reduplication and Proto-Austronesian grammar. *Oceanic Linguistics* 37.1:29-64.

Blust, Robert A. 1999. Subgrouping, circularity and extinction: some issues in Austrone-sian comparative linguistics. *Selected Papers from the Eighth International Conference on Austronesian Linguistics*, ed. by Elizabeth Zeitoun & Paul Jen-kuei Li, 31-94. Taipei: Institute of Linguistics, Academia Sinica.

Blust, Robert A. 2003. *Thao Dictionary*. Taipei: Institute of Linguistics, Academia Sinica.

Blust, Robert A. 2009. *The Austronesian Languages*. Pacific Linguistics 602. Canberra: The Australian National University.

Cauquelin, Josiane. 1991. *Dictionnaire puyuma-français*. Paris & Jakarta: École Française d'Extrême-Orient.

Chang, Henry Y. 2009. Adverbial verb constructions in Tsou. Paper presented at the Workshop on Formosan Languages, March 27, 2009. Kaohsiung: National Kao-hsiung Normal University.

Corbett, Greville G. 2000. *Number*. Cambridge & New York: Cambridge University Press.

Dahl, Otto Chr. 1981. Austronesian numerals. *Historical Linguistics in Indonesia*, Part I, ed. by Robert A. Blust, 46-58. Jakarta: Universitas Atma Jaya.

Dempwolff, Otto. 1924-1925. Die l-, r- und d- Laute in austronesischen Sprachen. *Zeitschrift für Eingeborenen-Sprachen* 15:19-50, 116-138, 223-238, 273-319.

Dempwolff, Otto. 1934-1938. 3 Vols. *Vergleichende Lautlehre des austronesischen Wortschatzes*. Zeitschrift für Eingeborenen-Sprachen, Supplement 1. *Induktiver Aufbau einer indonesischen Ursprache* (1934), Supplement 2. *Deduktive Anwendung des Urindonesischen auf austronesische Einzelsprachen* (1937), Supplement 3. *Austronesisches Wörterverzeichnis* (1938). Berlin: Dietrich Reimer.

Dyen, Isidore. 1947. The Malayo-Polynesian word for 'two'. *Language* 23.1:50-55.

Dyen, Isidore. 1975. A reconstructional confirmation: the Proto-Austronesian word for "two". *Oceanic Linguistics* 14.1:1-11.

Ferrell, Raleigh. 1982. *Paiwan Dictionary*. Pacific Linguistics C 73. Canberra: The Australian National University.

Ferrell, Raleigh. 2009. Notes on PAN suffix *N recurrence. Manuscript.

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. (ed.) 1997. *The Formosan Languages of Kaohsiung*. Kaohsiung: Kaohsiung County Government. (In Chinese)

Li, Paul Jen-kuei. 2006. Numerals in Formosan languages. *Oceanic Linguistics* 45.1: 133-152.

Li, Paul Jen-kuei, and Shigeru Tsuchida. 2001. *Pazih Dictionary*. Taipei: Institute of Linguistics, Academia Sinica.

Li, Paul Jen-kuei, and Shigeru Tsuchida. 2006. *Kavalan Dictionary*. Taipei: Institute of Linguistics, Academia Sinica.

Lin, Tai, Wensu Li, and Sheng-hsien Lin. 1998. *The Moon Passes the Time*. Taichung: Morning Star Press. (In Chinese)

Ossart, Nicolas. 2004. Les systèmes de numération dans les langues austronésiennes et leur fonctionnement. *Faits de Langues* 23-24:107-121.

Pan, Chia-jung. 2007. *The Grammatical Realization of Temporal Expressions in Tsou*. Chiayi: National Chung Cheng University MA thesis.

Pan, Chia-jung. (forthcoming). *A Grammar of Lha'alua, An Austronesian Language of Taiwan*. Cairns: The Cairns Institute, James Cook University PhD dissertation.

Pecoraro, Ferdinando. 1977. *Essai de dictionnaire Taroko-Français*. Paris: SECMI.

Reid, Lawrence A. 2009. On the diachronic development of C_1V_1- reduplication in some Austronesian languages. *Morphology* 19.2:239-261.

Ross, Malcolm. 2009. Proto Austronesian verbal morphology: a reappraisal. *Austronesian Historical Linguistics and Culture History: A Festschrift for Robert Blus*t, ed. by K. Alexander Adelaar & Andrew K. Pawley, 295-326. Pacific Linguistics 601. Canberra: The Australian National University.

Sagart, Laurent. 2004. The higher phylogeny of Austronesian and the position of Tai-Kadai. *Oceanic Linguistics* 43.2:411-444.

Sung, Margaret Mian Yan. 1969. Morphology of the Kanakanavu language of Formosa. *Bibliography Quaterly* 14.2:89-112.

Szakos, József. 1994. *Die Sprache der Cou: Untersuchungen zur Synchronie einer Austronesischen Sprache auf Taiwan*. Bonn: University of Bonn dissertation.

Tang, Chih-Chen Jane. 2004. Two types of classifier languages: a typological study of classification markers in Paiwan noun phrases. *Language and Linguistics* 5.2: 377-407.

Tang, Chih-Chen Jane. 2005. Nouns or classifiers: a non-movement analysis of classifiers in Chinese. *Language and Linguistics* 6.3:431-472.

Teng, Stacy Fang-ching. 2008. *A Reference Grammar of Puyuma, an Austronesian Language of Taiwan*. Pacific Linguistics 595. Canberra: The Australian National University.

Teng, Stacy Fang-ching, and Malcolm Ross. (in press). Is Puyuma a primary branch of Austronesian? A reply to Sagart. *Oceanic Linguistics* 49.2.

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.

Tung, T'ung-ho, Sung-hsing Wang, Tung-kuei Kuan, Tsai-fa Cheng, and Margaret Yan. 1964. *A Descriptive Study of the Tsou Language, Formosa*. Taipei: Institute of History and Philology, Academia Sinica.

Winter, Bodo. 2010. A note on the higher phylogeny of Austronesian. *Oceanic Linguistics* 49.1:282-287.

Wu, Chun-ming. 2009. The development of the path morpheme *maka-* in Paiwan, Mayrinax Atayal and other Formosan languages. Paper presented at the Workshop on Austronesian Subgrouping and Dispersal, January 7, 2009. Taipei: Academia Sinica.

Yeh, Marie M. 2009. Ca- reduplication in Formosan Languages. *Grazer Linguistische Studien* 71:135-156.

Zeitoun, Elizabeth. 2000. *A Reference Grammar of Tsou*. Taipei: Yuanliou. (In Chinese)

Zeitoun, Elizabeth. 2002. Reciprocals in Formosan languages. Paper presented at the 9th International Conference on Austronesian Linguistics (9-ICAL). Canberra: The National Australian University.

Zeitoun, Elizabeth. 2005. Tsou. *The Austronesian Languages of Asia and Madagascar*, ed. by K. Alexander Adelaar & Nikolaus P. Himmelmann, 259-290. London & New York: Routledge.

Zeitoun, Elizabeth. 2007. *A Grammar of Mantauran* (*Rukai*). Taipei: Institute of Linguistics, Academia Sinica.

Zeitoun, Elizabeth. 2009a. Reassessing the reconstruction of plural affixes in PAN: evidence from the Formosan languages. *Austronesian Historical Linguistics and Culture History: A Festschrift for Robert Blust*, ed. by K. Alexander Adelaar & Andrew K. Pawley, 348-361. Pacific Linguistics 601. Canberra: The Australian National University.

Zeitoun, Elizabeth. 2009b. Bound and free numeral forms in Formosan languages. Paper presented at the bi-monthly colloquium of the Institute of Linguistics at Academia Sinica, May 18, 2009. Taipei: Academia Sinica. Also presented at the 11th International Conference on Austronesian Linguistics (11-ICAL), June 22-26, 2009. Aussois, France.

Zeitoun, Elizabeth et al. (in preparation). Composite numerals in Formosan languages. Manuscript. Taipei: Institute of Linguistics, Academia Sinica.

Zeitoun, Elizabeth, Tai-hwa Chu, and Lalo' a Tahesh Kaybaybaw. (forthcoming). *A Study of Saisiyat Morphology*. Manuscript. Taipei: Institute of Linguistics, Academia Sinica.

Elizabeth Zeitoun
Institute of Linguistics
Academia Sinica
130, Sec. 2, Academia Road
Nankang, Taipei 115, Taiwan
hsez@gate.sinica.edu.tw

Stacy Fang-ching Teng
Institute of Linguistics
Academia Sinica
130, Sec. 2, Academia Road
Nankang, Taipei 115, Taiwan
stacyt@gate.sinica.edu.tw

Raleigh Ferrell
raleighferrell@yahoo.com

再論古南島語「二」以及相關議題

齊莉莎[1]　　鄧芳青[1]　　費羅禮

中央研究院[1]

　　Blust (1998, 2003, 2009) 認為應該在古南島語構擬兩套數詞系統：一套為未附加詞綴形式，專門用於數數；另一套則衍生於第一套，透過 Ca- 重疊，用於數人。本文旨在重新探討古南島語 20-90 的構擬，並重新思考在古南島語數詞系統中，Ca- 重疊之核心語意為區分計算人或非人之可行性。

　　我們主張，台灣南島語有兩套不同且須被構擬於古南島語的數詞系統：自由式和附著式的數詞。10 之下的所有數詞為自由式；10 以上的數詞則可同時由自由式和附著式的數詞構成。這個假設讓我們重新分析古南島語 20-90 之構詞方式。Blust (1998, 2003, 2009) 構擬 *ma-puSaN 為古南島語的 '20'，而我們認為 *puSa- 應分析為 '2'，是 *duSa 的附著式對應；puSa- 和古南島語 *ma-「十倍數」和 *-N「反覆」構成 *ma-puSa-N '20'。同樣的模式可說明其他數詞（從 30 至 90）的衍生方式。

　　我們進一步說明 Ca- 重疊的核心語意為「反覆」，後被延伸用於標註「複數」。雖然 Ca- 重疊可以用來構成計算人的 2-9 數詞，但是這個特徵應該來自於「複數」的延伸。

關鍵詞：數詞，構擬，台灣南島語，古南島語