

# Some Remarks on Sagart’s New Evidence for a Numeral-Based Phylogeny of Austronesian

Alexander D. Smith

NATIONAL UNIVERSITY OF SINGAPORE

This paper presents a critical evaluation of a recent update to Sagart’s “numeral-based phylogeny” of Austronesian languages. The update takes the form of new evidence, including new etymologies and reconstructions of words meaning ‘six’ and ‘ten’ which differ from conventional reconstructions, and updated and expanded evidence for “Southern Austronesian,” a subgroup that contains Kra-Dai and Malayo-Polynesian. This paper argues that Sagart’s new evidence is unconvincing and does not provide additional support for the numeral-based phylogeny. Rather, this paper details shortcomings in new etymologies for ‘six’ and ‘ten’, as well as issues in the comparisons made between Kra-Dai and Malayo-Polynesian. It is concluded that conventional subgrouping proposals remain superior to the numeral-based phylogeny despite recent updates.

**Keywords:** Subgrouping; Historical Linguistics; Numeral-Based Phylogeny; Comparative Method

**1. INTRODUCTION.**<sup>1</sup> There are many competing proposals on the higher order subgrouping of Austronesian (AN) languages, many of which are based on familiar approaches to subgrouping and reconstruction. For example, Blust (1999) has proposed a subgrouping with ten primary branches, based solely on phonological evidence. Alternatively, Ross (2009) offers a subgrouping with four primary divisions and evidence coming mostly from morphosyntax. Additional proposals include Li (2008), Aldridge (2016), and others, all of which rely on the evaluation of phonological or morphosyntactic evidence. But perhaps the most novel, and controversial, approach to higher order subgrouping in the last twenty years is provided by Laurent Sagart, who, in a series of publications, has proposed a “numeral-based phylogeny” of AN (Sagart 2004, 2008, 2014, 2021a, 2022). The core of Sagart’s proposal is that the distribution of the conventional Proto-Austronesian (PAN) numerals 1 through 10 (table 1) are implicational. That is, certain numerals only exist in a subset of

1. Special thanks to Robert Blust and Daniel Kaufman, who both added valuable input to this document, as well as to the anonymous referees. Remaining mistakes are my responsibility.

TABLE 1. CONVENTIONAL PAN NUMERALS.

*isa, *əsa, *asa	‘one’
*duSa	‘two’
*təlu	‘three’
*Səpat	‘four’
*lima	‘five’
*ənəm	‘six’
*pitu	‘seven’
*walu	‘eight’
*siwa	‘nine’
*sa-puluq	‘ten’

TABLE 2. SAGART’S NUMERAL IMPLICATION.

10 << 8, 9 << 6 << 5 << 7 << 1–4

34 languages that contain other numerals on an implicational hierarchy. A sche-  
 35 matic of that implication is given in table 2, from Blust (2014).

36 According to this implicational distribution, if a reflex of the conventional  
 37 reconstruction of 10, \*puluq, is found in a language, then reflexes of conven-  
 38 tional 1–9 are implied. If conventional 8 and 9 are found in a language, then  
 39 reflexes of conventional 1–7 are implied, but a reflex of conventional 10 is not.  
 40 If conventional 6 is found in a language, then reflexes of conventional 1–5 and  
 41 7 are implied, but reflexes of conventional 8–10 are not. Sagart uses this impli-  
 42 cational distribution to construct a totally novel family tree, where major divi-  
 43 sions are based on the implicational distribution of numerals rather than on  
 44 shared phonological or morphological innovations. His most recent proposal  
 45 is given in table 3 (from Sagart 2021a, 2022).

46 The proposal is notable for several reasons. First, and most obviously, is the  
 47 fact that it is fundamentally based on the distribution of numerals, whereas  
 48 other popular proposals are based on more conventional methods such as pho-  
 49 nological or morphosyntactic innovation. Second is the inclusion of Kra-Dai  
 50 (KD) as a daughter of PAN and sister of Malayo-Polynesian (MP). Although  
 51 some have suggested that KD and AN may be distantly related as sister-families  
 52 (Benedict 1942; Ostapirat 2005; Smith 2022), Sagart is forced to include KD in  
 53 the same daughter node as MP, “Southern Austronesian,” since KD languages  
 54 reflect the entire series of conventional numerals 1–10.

55 An additional novelty of Sagart’s approach is how he derives the conven-  
 56 tional numerals from a new set of reconstructed PAN numerals based largely  
 57 on the numeral system of Pazeh. Sagart’s PAN numerals resemble the conven-  
 58 tional reconstruction in the first four numerals only. The numerals five through  
 59 ten are new. Table 4 compares Sagart’s reconstruction with the Pazeh numeral  
 60 system on which they are based.

61 Several techniques are devised by Sagart to derive the conventional numerals  
 62 from his new reconstruction. For the numeral 5, Sagart states that \*lima was

**TABLE 3. SAGART'S NUMERAL-BASED PHYLOGENY  
(SAGART 2021a, 2022).**

Proto-Austronesian
Luilang
Saisiyat
Pazeh
Pituish
Favorlang-Taokas
Limaish
Atayalic-Thao
Enemish
Siraya
Walu-Siwaish
West coast Walu-Siwaish
Papora, Hoanya
Central Walu-Siwaish
Rukai-Tsouic
Bunun
Eastern Walu-Siwaish
North Formosan
Kavalan, Basay, Trobiawan
Puluqish
Northern Puluqish
Amis, Puyuma
Paiwan
Southern Austronesian
Kra-Dai
Malayo-Polynesian

---

**TABLE 4. SAGART'S PROPOSED PAN NUMERALS.**

AQ2	PAN (Sagart 2021a,b)	Pazeh	
	*əsa/*isa	ida	one
	*duSa/*tuSa	dusa	two
	*təlu	turu	three
	*Səpat	supat	four
	*RaCəp	xasəp	five
	*NəmNəm	xasəb uza	six
	*RaCəp-i-tuSa	xasəb-i-dusa	seven
	*RaCəp-a-təlu	xasəb-a-turu ~ xasəb-i-turu	eight
	*RaCəp-i-Səpat	xasəb-i-supat	nine
	*sa-iCit	isit	ten

---

63 innovated through analogy with the word \*qalima ‘hand’ in his Limaish group.  
 64 The numerals 7–9 are derived with a series of lenitions, deletions, and prunings  
 65 that affect his new numerals. The ... p-i-tu ... sequence from \*RaCəp-i-tuSa  
 66 yields the more familiar \*pitu ‘seven’. The sequence ... p-a-təlu yields \*walu  
 67 ‘eight’ after \*p lenites to w, schwa deletes, and \*-tl- reduces to l. The sequence  
 68 ... -i-Səpa ... yields \*Siwa ‘nine’ after schwa assimilates to \*i, \*p becomes w,  
 69 and everything else deletes. Sagart’s ‘six’ and ‘ten’ cannot be derived in a similar  
 70 fashion. They were originally considered innovations without historical explana-  
 71 tion, but recently, Sagart offers \*NəmNəm as a possible reconstruction of ‘six’ and

72 derives ‘ten’ from a semantic extension dealing with the word for ‘to separate’ in  
73 Amis. These two cases are discussed more in the following sections of this paper.

74 The numeral-based phylogeny has faced tough criticism, notably from Blust  
75 & Aq3 (2014), Ross (2012), and Winter (2010). Common criticisms are that Sagart  
76 relies on ad hoc explanations and irregular sound changes, that the numerals  
77 ‘six’ and ‘ten’ are exceptional in his proposal since they are not derived from  
78 earlier numerals as 5, 7, 8, and 9 are, and that his proposal creates numerous  
79 conflicts with relatively well-established subgroups. Those interested in these  
80 criticisms of the proposal in its earlier stages will reference the citations above.  
81 The criticisms will not be restated at length here. The ultimate conclusion  
82 drawn from those who criticize the numeral-based phylogeny is that the impli-  
83 cational distribution of numerals is a product of chance, and does not imply  
84 anything about higher order subgrouping.

85 Sagart has responded to some of these criticisms, particularly to the criticism  
86 of ‘six’ and ‘ten’ being exceptional. Recently, Sagart (2021a) has given a com-  
87 prehensive update to his numeral-based phylogeny at the *International*  
88 *Conference of Austronesian Linguistics (ICAL)*, where he presented “new evi-  
89 dence” and etymologies that he claims support the tree in table 3. His new evi-  
90 dence is listed on his personal website (Sagart 2022) and, at the time of this  
91 writing, has not otherwise appeared in publication. Some of the new evidence  
92 is shown in the following non-comprehensive list:

- 93 • An etymology for ‘six’ based on a newly reconstructed root \*Nəm ‘three’,  
94 which Sagart claims is reduplicated in his PAN \*NəmNəm ‘six’.
- 95 • Etymology for ‘ten’ based on a verb, Amis *polo* ‘which meant ‘to sepa-  
96 rate’. Sagart claims that \*puluq had an original meaning of separate and  
97 was used to separate numerals of ten from singular numerals in numbers  
98 greater than ten.
- 99 • Several additional lexical, phonological, and morphological innovations  
100 at various intermediate nodes, including for Southern Austronesian, a  
101 node which contains MP and KD as sister groups.

102  
103 This paper is a critical response to Sagart’s proposal. An analysis of his new  
104 evidence is given, with special attention to his updated etymologies of ‘six’,  
105 ‘ten’, in sections 3 and 4, respectively, as well as his new evidence for  
106 Southern Austronesian, in section 5. This paper argues that Sagart’s new  
107 numeral-based phylogeny continues to be plagued by the same issues that  
108 Blust, Ross, and Winter have pointed out in older versions. That is, a reliance  
109 on ad hoc and irregular sound changes, inaccuracies in his reporting of PAN  
110 reconstructions, and irregular correspondences between his MP–KD shared  
111 innovations. It is argued that his new evidence is unconvincing and that it does  
112 not improve upon the original numeral-based proposal and its various updated  
113 proposals. Rather, more conventional models of AN higher order subgrouping  
114 remain better references for the ongoing comparative study of AN languages.

115 **2. EVALUATING THE EVIDENCE WITH THE COMPARATIVE**  
 116 **METHOD.** There is not much need to give an exhaustive review of the compar-  
 117 ative method and why it is such a powerful tool for linguistic subgrouping  
 118 and reconstruction. Such reviews are readily found in most introductory books  
 119 on historical linguistics (Campbell [2021], and Crowley and Bowerman [2010] to  
 120 name only two). It is necessary, however, to give a brief reminder of what the  
 121 comparative method is, and how it is utilized in comparative and historical anal-  
 122 ysis, since the criticisms within this paper are made from a perspective that is  
 123 heavily influenced by the comparative method.

124 In short, the comparative method is a strict set of procedures by which his-  
 125 torical linguists test and justify claims that similarities between two or more  
 126 languages are inherited from a common ancestor, and are not the product of  
 127 chance, diffusion, linguistic universals, or any other nongenetic explanation  
 128 whereby two unrelated languages may seem similar. Critical to any compara-  
 129 tive endeavor is (i) the recognition that similarity is not always inherited, (ii) the  
 130 recognition that inheritance is only a reasonable explanation for similarity after  
 131 other explanations are ruled out, and (iii) that sound change is for the most part  
 132 regular. In addition, following the logic of the comparative method and the  
 133 regularity of sound change, we can state that determining if two languages  
 134 are similar through inheritance relies on identifying regular, recurring sound  
 135 correspondences between suspected cognate words.

136 When reconstructing a proto-lexeme, these principles remain important. If  
 137 one suspects that similar words in related languages share a common origin,  
 138 then one must show that (i) the sound correspondences between phonemes  
 139 in those words are regular when compared with other words of similar shape,  
 140 and (ii) that the words appear in a sufficiently diverse group of languages to  
 141 warrant their reconstruction. If one wishes to propose a new proto-lexeme that  
 142 fits into an already well-defined proto-language, then one must also show that  
 143 the sounds of the proposed proto-lexeme have undergone the expected regular  
 144 sound changes in the languages where reflexes are found. This is all to show  
 145 that the similarity in the suspected cognates is not just chance, but rather the  
 146 result of inheritance.

147 Say, for example, that one was to propose a new PAN reconstruction \*Caka  
 148 (a nonce-word). The evidence for \*Caka must ideally consist of words from  
 149 various AN languages in more than one primary branch, and the words must  
 150 have regular reflexes of \*C, \*a, and \*k in their respective positions. A  
 151 Pazeh reflex of \*Caka should begin with *s*, since *s* is the regular reflex of  
 152 \*C, an Amis reflex should begin with *t*, a Thao reflex with *c*, and so on. It  
 153 is normal, of course, to find irregular reflexes here and there. One may still  
 154 propose \*Caka even if a small percentage of the languages with reflexes of  
 155 \*Caka contain an irregular reflex of \*C. The allowance of minimal irregularity  
 156 is necessary, since language evolution is almost never as clean as originally  
 157 envisioned by the neogrammarian school of early historical linguistics. It would  
 158 not be acceptable, however, to propose a new word \*Caka, if the majority of

159 languages, where reflexes of \*Caka are found, have irregular reflexes of \*C.  
 160 That is, although irregularity exists, most sound change is still overwhelmingly  
 161 regular.

162 When evaluating the new proposals from Sagart, these principles will be  
 163 strictly adhered to. Each proposal will be evaluated based on the recurrence  
 164 of sound correspondences in correspondence sets, the regularity of sound  
 165 change in each proposal, the existence of sufficiently diverse evidence for  
 166 new reconstructions, and the absence of reasonable alternative explanations  
 167 for similarity.

168 **3. EVALUATING \*NəmNəm ‘SIX’.** Some critics of the numeral-based phylogeny  
 169 have pointed out that ‘six’ was the only numeral below ‘ten’ that was not  
 170 derived from some already existent compound numeral. Blust (2014:364) goes so  
 171 far as to cite ‘six’ as “an embarrassment for the numeral-based phylogeny” and  
 172 goes on to note that “If PAN was like Pazeḥ, the numerals 6–9 should have had  
 173 the structure 5 + 1, 5 + 2, 5 + 3, and 5 + 4.” Sagart has therefore gone to great  
 174 lengths to construct an etymology for ‘six’ that solves this apparent shortcoming.  
 175 To that end, Sagart proposes that the word for ‘six’, commonly reconstructed as  
 176 PAN \*ənəm, is instead derived from a reduplication of a newly proposed root  
 177 \*Nəm ‘three’ (Sagart [2022] credits Benedict [1995] as first proposing a redupli-  
 178 cated ‘six’ with his \*ʔəmləm reconstruction). The root \*Nəm is claimed to  
 179 have been reduplicated as \*NəmNəm ‘six’, which is then analyzed as an additive:  
 180 ‘3 + 3’ (Sagart). Several problems with Sagart’s proposed reconstruction for ‘six’  
 181 are organized below. These include (i) 3 + 3 rather than 5 + 1 addition, (ii) prob-  
 182 lems with the source data cited, (iii) irregularity in reflexes of \*N if ‘six’ was  
 183 indeed \*NəmNəm, (iv) alternative explanations for some of his observations,  
 184 and (v) issues with Sagart’s use of a root \*Nəm ‘three’ which he uses to derive  
 185 \*NəmNəm.

186 **3.1. THE ADDITION METHOD.** From the start, there is a significant break  
 187 between the new proposal and past proposals within the numeral-based phylog-  
 188 eny. In Sagart’s earlier proposals, all numerals higher than four, other than six,  
 189 are reconstructed as 5 +  $x$ , where  $x$  may be the numerals ‘two’, ‘three’, and  
 190 ‘four’ according to Sagart’s proposal. If a new word for ‘six’ is to be proposed,  
 191 and if that word was to fit into the existing 5 +  $x$  formula, then it is unexpected  
 192 that ‘six’ would be uniquely composed of a reduplicated root with the inferred  
 193 meaning of 3 + 3, not the expected 5 + 1. Pazeḥ, which neatly reflects all other  
 194 reconstructions of Sagart’s PAN numerals, has *xasəb uza*, or ‘5 + 1’. In fact, the  
 195 reduplication method of 3 + 3 for ‘six’ is rare in the AN family (but Sagart cites  
 196 Blust [2003] for some similar examples from Thao). Somewhat similar systems  
 197 of addition are only found in AN languages that are in close contact with Papuan  
 198 languages, such as Gapapaiwa of Milne Bay, New Guinea, which has *rua ma*  
 199 *rua* (two plus two) for ‘four’ (McGuckin 2002). Such examples are rare, clearly  
 200 the product of contact, and still structurally distinct from Sagart’s \*NəmNəm.

201 **3.2. THE SOURCE DATA.** The unlikelihood of a 3 + 3 method for deriving  
 202 'six' is not the only problem with this proposal, however. Sagart reconstructs  
 203 \*N, rather than \*n as the onset of the ultima on data from notes recently pub-  
 204 lished in Li and Toyoshima (2006), a collection of centuries-old field notes  
 205 gathered by different linguists and anthropologists with varying levels of lin-  
 206 guistic training.<sup>2</sup> Sagart claims that unexplained *l* and *r* reflexes of \*n suggest  
 207 that the reconstructed phoneme should be \*N. The logic of his argument is that  
 208 since *l* and *r* are liquids, and \*N was a liquid, that these must reflect \*N, not \*n.  
 209 The data which he cites are given below:

- (1) Tsou #38a *lomu*  
 Saaroa #47a *rumu, urumu,*  
 #48 *urumu*  
 Rukai #63 *ulum*  
 Paiwan #59a, 62a *urum*  
 Kavalan #108, 108 *arum*  
 Basay #106-107-108 *arum, rarum*  
 Papora #129b *rum*  
 Hoanya #138 *alim, ilim,*  
 #142b.2 *rom.*

211 There are problems with Sagart's use of these data. First, nearly every lan-  
 212 guage on the list, where an unexpected *l* or *r* is transcribed, is contradicted by  
 213 more recent and linguistically reliable material, where *n* is transcribed. These  
 214 are listed below with relevant citations. Note that in addition to the consonants  
 215 being more regularly transcribed as *n* and *m*, that the vowels are also more con-  
 216 sistent transcribed as *ə*, which more accurately matches the known historical  
 217 phonologies of these languages.

- (2) Tsou *nomə* (Tsuchida 1975)  
 Saaroa *ənəmə* (Tsuchida 1975)  
 Budai Rukai *ənəmə* (Chen 2006)  
 Paiwan *unəmə* (Chang 2006; Chen 2006)  
 Kavalan *ʔnəm* (Li and Tsuchida 2006)  
 Basay *anəmə* (Li 2004)  
 Papora *nə-nom* (Blust and Trussel 2020)

218 Second, even in Li and Toyoshima there are numerous alternative transcrip-  
 220 tions given for these numerals. For example, Paiwan has ten listed entries for  
 221 'six'. Sagart lists #59a, 62a *urum*, but fails to list 59b, 62c *onum*, 60, 61, 62d,  
 222 62e, 62f, *unum*, all gathered by various individuals from various locations.<sup>3</sup>  
 223 The individuals gathering the information had different levels of linguistic train-  
 224 ing. Most of the words for 'six' listed by Sagart which have an *l* or *r* reflex were  
 225 gathered by Ito, an anthropologist, and are contradicted by other words with more

2. PAN \*n and \*N are usually reconstructed as a typical alveolar or dental nasal, \*n [n], and a lateral \*N ([H]?) distinct from \*l [l]. The exact phonetic shape of \*N is debatable, but it was probably a voiceless lateral or lateral fricative.

3. On page ix, it is noted that words were "[...] compiled from 163-plus sources, collected by various people at various times."

226 familiar \*n reflexes (Li and Toyoshima 2006:xi–xv). This is true both within the  
 227 transcripts of Li and Toyoshima and in more recent materials listed above in (2).  
 228 That is, it is almost always Ito who records *l* and *r* where other researchers record  
 229 *n*. The data given by Sagart in his proposal are therefore limited only to data which  
 230 may be used to further his argument, regardless of the reliability of the source or  
 231 the presence of contradictory data. Scholars evaluating the proposal should be  
 232 well aware of these potential issues with the data being used and should not rush  
 233 to make conclusions based on material whose reliability may be in question.<sup>4</sup>

234 **3.3. IRREGULAR REFLEXES OF \*N.** Another shortcoming is Sagart’s  
 235 reliance on superficial similarity, rather than inheritance, in his claim.  
 236 Superficial, as I use it here, means that the similarities are surface-level, not  
 237 the product of inheritance from a common ancestor. Sagart’s argument is that  
 238 because \*N was a liquid and some words for ‘six’ contain liquids in Li and  
 239 Toyoshima (2006) then the correct reconstruction for ‘six’ should contain  
 240 \*N rather than \*n. This similarity is superficial, and the similarity of *l* and *r*  
 241 to \*N is only relevant if the correspondences between these consonants are reg-  
 242 ular. Regular reflexes of \*N in many of the Formosan languages cited by Sagart  
 243 are, however, not *l* or *r*, which indicates that any similarity between *l*, *r*, and \*N  
 244 in these languages is not inherited. In table 5, Sagart’s reflexes of \*N in words  
 245 for ‘six’ are compared with other, actual reflexes of \*N in multiple examples  
 246 from Tsou, Saaroa, Paiwan, Kavalan, Basay, and Papora. As the table demon-  
 247 strates, possible reflexes of \*N in Sagart’s words for ‘six’ are irregular.

248 Not on this list are Rukai and Hoanya. Rukai reflexes of \*N are typically *l* or  
 249 *ʃ*, whereas Hoanya evidence is mixed but also points to *l*. Both might support an  
 250 \*N reconstruction for six under Sagart’s proposal, although this is still unlikely,  
 251 because of the numerous recorded counter examples where ‘six’ is reflected  
 252 with an *n*, not *l* or *r*, in Rukai and Hoanya.

253 In defense of Sagart’s proposal, one may suggest that the cluster \*-mN-  
 254 results in a different reflex than normal. This might explain *l* and *r* reflexes  
 255 in Tsou, Saaroa, Paiwan, Kavalan, Basay, and Papora. Such an explanation  
 256 is not satisfactory, however, since the Paiwan reflex of \*NəŋNəŋ ‘to stare, look  
 257 fixedly’ is *ʃ-em-əŋʃəŋ*, with the regular *ʃ* reflex of \*N despite a cluster envi-  
 258 ronment quite similar to the one proposed in Sagart’s \*NəmNəm. One cannot,  
 259 then, appeal to ad hoc explanation to get rid of the fact that the majority of  
 260 Sagart’s evidence for \*NəmNəm is irregular.

261 Yet another issue with the evidence is that the *l* and *r* phonemes in these lan-  
 262 guages belong to correspondence sets regularly assigned to other proto-phonemes.  
 263 Examples from four of these languages, Saaroa, Paiwan, Kavalan, and Basai

4. In the introduction to Li and Toyoshima (2006:x), the compilers are careful to state that they “[...] do not know what some of the special symbols and diacritic marks stand for [...],” that “[...] some forms are inaccurate, as based on our knowledge of the languages [...],” and that they “[...] may have been misled by stains of ink, spots in paper, or accidental marks in the original manuscript.” Based on this introduction, one expects researchers to exercise reasonable caution when using these data.



**TABLE 5. REGULAR REFLEXES OF \*N COMPARED WITH REFLEXES OF \*NəmNəm.**

Language	Sagart's *N reflex	Regular *N reflex	Supporting evidence		
Tsou	l	h	*aNak >	me-ahʔo	'give birth'
			*lukəNaw >	eʔuho	'clouded leopard'
			*qaNiCu >	hicu	'spirit, god'
Saaroa	r	lh	*waNan >	alhanə	'right'
			*daNum >	salhumu	'water'
			*CaNəm >	Cuma-calhəmə	'to cover with earth'
Paiwan	r	lʷ	*aNak >	alʷak	'child'
			*CuNuh >	tsulyu	'roast something'
			*daNum >	zalyum	'water'
Kavalan	r	n	*daNum >	zanum	'water'
			*qaNiŋu >	niŋu	'shadow'
			*CaNəm >	tanəm	'grave'
Basay	r	n	*daNum >	ranum	'water'
			*quzaN >	uran	'rain'
			*Naŋuy >	naŋuy	'to swim'
Papora	r	l	*qaNuaŋ >	loan	'buffalo'
			*waNiS >	walis	'tusk of wild boar'

264 should be enough to prove this point. First, Saaroa *r*, which Sagart claims supports  
 265 an \*N reconstruction for 'six', belongs to a correspondence set that regularly  
 266 reflects PAN \*R, not \*N. Evidence includes Saaroa *a-vəraə* 'husked rice', *laarə*  
 267 'flying squirrel', and *caraʔə* 'blood', from PAN \*bəRas, \*lawar, and \*daRaŋ,  
 268 respectively. Second, Paiwan *r* regularly reflects PAN \*r and \*R, not \*N.  
 AQ6 Evidence from Paiwan includes *garuts* 'comb', *ka-viri* 'left (hand)', and *virivir*  
 270 'lips', from PAN \*garuC, \*ka-wiRi, and \*biRbiR, respectively. Kavalan has  
 AQ7 two possible phonemes which may correspond to the orthographic *r* from Li  
 272 and Toyoshima (2008) that Sagart assigns to \*N, *r* and *R* (a uvular). Neither  
 273 of these reflects \*N. Rather, they both reflect PAN \*l and \*R, with a split giving  
 274 *r* in some reflexes and *R* in others. Evidence includes PAN \*ala > Kavalan *ara*  
 275 'take it', \*bulaN > *buran* 'moon', \*ka-wiRi > *kawirian* 'left', \*luSeq > *Rusi*  
 276 'tears', \*qulu > *uRu* 'head', \*kaRaC > *q<m>aRat* 'bite', and \*beRaS >  
 277 *bRas* 'husked rice'. Finally, Basai *r* seems to reflect a merger between PAN  
 278 \*d, \*z, and \*R, although Basai evidence is inconsistent, some evidence includes  
 279 PAN \*quzaN > Basai *uran* 'rain', \*daNum > *ranum* 'water'. It can therefore be  
 280 concluded that the *r* and *l* consonants that Sagart assigns to \*N has no historical  
 281 phonological grounding.

282 **3.4. GEMINATION FROM SCHWA, NOT FROM CLUSTER ASSIMI-**  
 283 **LATION.** Sagart also points to apparent gemination of the medial consonant -  
 284 *n-* in some recordings of 'six' as evidence for a past cluster. He lists the  
 285 following examples and explains the geminates, *nn*, as arising from the assimila-  
 286 tion of \*-mN-, first via the nasalization of \*N, producing \*-mn-, then via  
 287 place assimilation to *-nn-*:

- (3) Siraya #143 *annim, annum, nnum, ninnum, ninnim*  
 Paiwan #71, 72 *unnum*  
 Amis #82c1, 82d, 83cd, 83k.2 *num*  
 Kavalan #88b, 92a, 92b *num*;  
           #92c *unnum, innim*;  
           #101 *innoom*  
 Basay #102b *num*  
 Siraya #143 *annim, annum, nnum, ninnum, ninnim*  
 Paiwan #71, 72 *unnum*

289

288

290

291

2: AQ8

2: AQ9

294

295

296

297

298

299

300

301

302

303

These cases of geminates are unsurprising, however, when one considers the shape of the conventional reconstruction for ‘six’, \**ənəm*. In a process that has been extensively documented, schwa, when it appears in an open penultimate syllable, often triggers the automatic lengthening of the following consonant (Blust 2018; Smith, [to appear](#)). This automatic lengthening of post-schwa consonants is prevalent in MP, but is also recorded in some Formosan languages. For example, Blust (2014b) also recorded a geminate in ‘six’ for Central Amis, *nəm* < \**ənəm*. Importantly, Blust also recorded geminate consonants in several other words, all of which have a schwa-trigger in the penult: *səm.aʔ* < \**Səma* ‘tongue’, *həm:ot* ‘spine’, *təfos* < \**təbuS* ‘sugarcane’, *həc:iʔ* < \**Səsi* ‘muscle; flesh’. There are many more examples, but enumerating each here is not necessary. Gemination in reflexes of six is therefore best explained as being triggered by the schwa penult, not by cluster assimilation. Sagart’s cluster-reduction proposal ignores the better supported explanation for gemination in words for ‘six’ in preference for a less-supported explanation that better suits his \**NəmNəm* reconstruction.

304

305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

### 3.5. AN ROOTS AND THE UNLIKELYHOOD OF \**Nəm* ‘THREE’.

Finally, there are issues with the methods that Sagart uses in his argument for the putative \**Nəm* root meaning ‘three’. First, in AN linguistics “root” is a term often used in reference to semantically similar submorphemic -CVC sequences similar to phonestemes. These roots are found in word-final position. They do not form active morphology, but rather are inseparable parts of base words which consist of a meaningless and inseparable CV- segment followed by the -CVC root. An example of such a root is \*-pit, found in Malay (*h*) *apit* ‘pressure between two disconnected surfaces’, *capit* ‘pincers’, *mən-capit* ‘to nip’, *dəmpit* ‘pressed together, in contact’, (*h*) *impit* ‘squeezing pressure’, or *-ket* in Kankanaey *busiket* ‘clay; the thick, gluey black earth of rice fields’, *niket* ‘resin’, and *paŋkét* ‘sticking, adhering, cleaving’ (Blust 1988). Some roots do include forms that are found in reduplicated monosyllables; however, these types of roots also combine with meaningless CV- segments just as any other root does. CVC roots do not, however, combine with CV or V segments in productive ways, nor do they combine with segments in other-than-final position. Sagart claims that independent evidence for the meaning ‘three’ attached to a root \**Nəm* can be found in the following “roots”<sup>5</sup>:

5. Hyphens in these words are original. It is not clear what they are indicating.

(4) Basai	#112, 113	<i>pinum, peinum</i>	‘four’
Siraya		<i>kulom</i>	‘three’
		<i>kulom-taʔ</i>	‘four’
Makatao	M3	<i>ra-rum-a</i>	‘three’
Makatao	M7	<i>nunta</i>	‘four’
Makatao	M8	<i>lum-ta</i>	‘four’

323 All of the suggested forms for Makatao involve uses of “roots” that do not  
 324 match the typical position of roots in AN words. That is, they are CVC bases to  
 325 which morphology is attached, rather than an inseparable word-final -CVC ele-  
 326 ment that combines with an unanalyzable CV- initial segment to form a base.  
 327 This is also true for the Basai words, which are obviously multimorphemic. The  
 328 initial element that appears in Basai is a \*pa- numeral prefix, which denotes  
 329 cardinal numerals (Blust and Trussel 2020).

330 There are other issues with the comparisons in (4). For one, PAN bases are  
 331 largely disyllabic, and monosyllabic elements are limited only to grammatical  
 332 words and onomatopoeic words (Chrétien 1965; Blust 2013). An independent  
 333 monosyllabic base word, \*Nəm ‘three’, in the numerals is highly idiosyncratic  
 334 as a PAN reconstruction. Finally, the words in (4) have a mixture of meanings,  
 335 both ‘three’ and ‘four’. It is possible that Sagart interprets the -ta element as an  
 336 additive formed from PAN \*isa ‘one’, but Siraya *asu* ‘dog’, from PAN \*asu,  
 337 contradicts the \*s > t change required for such an interpretation.

338 It should be clear by now that there is little hope of salvaging the \*NəmNəm  
 339 reconstruction for ‘six’. First, Sagart has based the reconstruction on datasets  
 340 gathered by an anthropologist which are contradicted by data gathered by  
 341 trained linguists. Second, even if the data cited are accepted as legitimate,  
 342 the majority of example languages have regular reflexes of \*N that differ from  
 343 those found in the reflexes of ‘six’ that Sagart cites to further his hypothesis.  
 344 Third, the *r* and *l* phonemes in the various languages may be assigned to cog-  
 345 nate sets which do not reflect \*N. Fourth, because the proposed root \*Nəm does  
 346 not operate in a way that is consistent with roots in AN and even if \*Nəm was an  
 347 active base, then we still cannot reconcile the idiosyncrasy of a monosyllabic  
 348 numeral in a language that was otherwise disyllabic as well as the differences in  
 349 meaning. Fifth, if \*Nəm ‘three’ is the basis for \*NəmNəm ‘six’, then it would  
 350 have been necessary for \*Nəm to have replaced \*təlu, which it did not. Within  
 351 the scope of Sagart’s numeral-based phylogeny, then, six remains a numeral  
 352 that cannot be explained as occurring through the reanalysis of an already exist-  
 353 ing numeral. The proposal that ‘six’ was \*NəmNəm and that this follows from a  
 354 root \*Nəm ‘three’ is therefore not convincing.

355 **4. EVALUATING \*puluq ‘TO SEPARATE FROM, AND LEAVE**  
 356 **ALONE’.** Sagart uses \*puluq ‘ten’, an innovation in his model, to define  
 357 the Puluqish node on his tree, which includes Amis, Puyuma, Paiwan, MP,  
 358 and KD. Like ‘six’, the innovation of \*puluq was problematic for Sagart’s origi-  
 359 nal proposal, since it innovates “out of nowhere,” while other innovative

numerals are based on changes to existing words. Sagart attempts to right this wrong by proposing a source for \*puluq from an already existing word which he reconstructs as Proto-Eastern Walu-Siwaish \*puluq ‘to separate from, and leave alone’, which was reinterpreted as ‘ten’ through a series of semantic changes.

Sagart’s etymology for \*puluq is given as follows. In the Eastern Walu-Siwaish branch a word \*puluq ‘to separate from, and leave alone’ was used in numerals greater than ten. He gives a hypothetical example with Proto-Eastern Walu-Siwaish ‘35’ which he reconstructs as \*təlu-baCaqan-puluq-lima ‘35’ which is literally translated as ‘three-ten-leave those alone-five’. The role of \*puluq in numerals greater than ten is to separate the numerals of place value ten, in this case \*təlu-baCaqan ‘30’, from the numerals of the singular place value. The logic of this role is that, according to Sagart, ‘to separate from, and leave alone’ matches “. . . the mental operation of isolating *n* sets of ten objects and putting them aside before expressing the remainder” (Sagart 2022). According to Sagart’s theory, \*baCaqan ‘ten’ and \*puluq ‘to separate from, and leave alone’ are redundant in the fully articulated form \*təlu-baCaqan-puluq-lima, since \*puluq will always separate the word ‘ten’ from the ones-place numeral. Therefore, an alternative, truncated numeral of \*təlu-puluq-lima appeared and eventually replaced \*təlu-baCaqan-puluq-lima in Proto-Puluqish. Some criticisms of this proposal are organized below. These are, (i) the word \*puluq ‘to separate from, and leave alone’ cannot be reconstructed outside of Amis, (ii) the semantic leap from ‘separate’ to ‘ten’ is too large and not supported by parallel developments elsewhere, and (iii) an alternative hypothesis, that Amis *puluq* ‘to separate’ and \*puluq ‘ten’ are unrelated, is not given adequate consideration in Sagart’s proposal.

**4.1. THE RECONSTRUCTABILITY OF \*puluq ‘TO SEPARATE FROM, AND LEAVE ALONE’.** Sagart’s proposal requires a word of the meaning ‘to separate from, and leave alone’ of the shape \*puluq in his Proto-Eastern Walu-Siwaish. A requirement of reconstruction is that evidence for a proposed reconstruction must exist in at least two primary branches in order for a word to be confidently reconstructed to a proto-language. But the reconstruction \*puluq, be it ‘to separate, and leave alone’ or ‘ten’ is only available in Sagart’s Puluqish branch, with evidence from Amis, Puyuma, Paiwan, and the non-Formosan languages. Based on this alone, \*puluq cannot be further reconstructed to the level of Proto-Eastern Walu-Siwaish unless evidence from outside of Puluqish is given.

But what about the reconstruction of \*puluq ‘to separate from, and leave alone’ to Proto-Puluqish? Here, too the evidence does not support Sagart’s conjecture. To make the jump from ‘separate, leave alone’ to ‘ten’, one must have evidence of \*puluq being used in such an environment. The only Pituish language which has a form that may reflect \*puluq with a meaning ‘separate, leave alone’ is in fact Amis. The Amis word *puluq* ‘separate, leave alone’, however,

403 is not found in numerals as Sagart reconstructs. It is a stand-alone verb. In Amis  
 404 the word 35 is *tolo' polo' ira ko lima*. In Amis *ira ko* is not a marker of separa-  
 405 tion but rather one of inclusion (translatable as 'there is', or 'has'). Puyuma,  
 406 Paiwan, KD, and MP all fail to provide evidence that \*puluq, meaning 'sepa-  
 407 rate, leave alone', is reconstructable past Proto-Amis. Of course, all of these  
 408 languages provide evidence for \*puluq as 'ten', but this alone cannot justify  
 409 the reconstruction of \*puluq 'to separate, and leave alone', since the semantic  
 410 gap between 'ten' and 'separate' is so large.

411 **4.2. THE SEMANTIC LEAP FROM 'SEPARATE' TO 'TEN'.** This  
 412 brings forward another issue with Sagart's proposed 'separate and leave alone'  
 413 → 'ten' history of \*puluq, namely, the likelihood of such a semantic shift. As far  
 414 as I am aware, there are no AN languages, other than Sagart's proto-Puluqish, that  
 415 utilize a word meaning 'separate' or 'leave alone' to remove the lower numeral  
 416 from the tens numeral in words like '35'. There are, of course, AN languages with  
 417 a ligature separating reflexes of \*puluq from the following numeral. Examples  
 418 include Ilokano *sa-ŋa-púlo ket duá* or Tondano *ma-pulu? wo-rua*, both '12'  
 419 (Blust 2013). These ligatures are not, however, derived from words with a dis-  
 420 creet meaning. Amis, as demonstrated earlier, uses a word that means 'have'.  
 421 Puyuma separates the ten and singular numerals with *misama* 'remaining' and  
 422 Tagalog uses *labi* 'the remains' in *labing dalawa* 'twelve'. Some languages  
 423 use a word to add to a ten base, such as Woq Helaq Modang *siḡh hin səway*  
 424 '11', literally, *siḡh* 'one' *hin* 'from' *səway* 'ten', but the use of *hin* 'from' does  
 425 not extend to separating the ten numerals from the singular numerals in words  
 426 like '35'. In all, there is little reason to accept a history whereby 'separate from,  
 427 leave alone' was used to separate numerals, both because of the semantic leap  
 428 that one must take, and because of a lack of evidence for any sort of similar pro-  
 429 cess in other AN languages.

430 **4.3. AN ALTERNATIVE EXPLANATION: CHANCE SIMILARITY.**  
 431 The most likely explanation for Amis *polo'* in its usage as a verb meaning  
 432 'to separate' is that it is a homophone with the ten-multiplier *polo'*. That is *polo'*  
 433 'to separate' and *polo'* 'ten' have nothing to do with one another. Homophones  
 434 are by no means rare in languages. Examples of numerals with non-numeral  
 435 homophones are abundant: English *two* : to/too, Mei Lan Modang *maw?* 'five'  
 436 : *maw?* 'farm; field', Kanowit *tuju?* 'seven', *tuju?* 'to push', and many others.  
 437 Why must the verb for 'separate' be related to the numeral 'ten' in Amis if not  
 438 just to motivate the replacement of Sagart's Proto-Eastern Walu-Siwaish  
 439 \*baCaqaŋ? Amis *polo'* 'to separate' appears in no other AN language with  
 440 the same or similar meaning. No languages, other than Sagart's proposed  
 441 Puluqish languages, are known to utilize a word of the meaning 'separate'  
 442 or 'leave alone' in the formation of numerals above ten. The entire premise  
 443 is based on a semantic leap whose only purpose serves to link a functionally  
 444 distinct verb to the numerals. Sagart appears unwilling to accept chance as a  
 445 probable explanation for similarity in this case.

446 **4.4. \*baCaqan ‘TEN’ AND ITS RELATIONSHIP TO \*puluq.** Sagart also  
 447 proposes an intermediate word for ‘ten’, \*baCaqan, which replaced his PAN  
 448 \*iCit, at the Proto-Eastern Walu-Siwaish level, between PAN and Proto-  
 449 Puluqish. His evidence is from Kavalan, Basay, and Ketagalan (closely related  
 450 languages in the Kavalanic group, Sagart’s “North Formosan” group), as well  
 451 as Sakizaya (Amis) and Puyuma. There are some issues with this reconstruction  
 452 as well, but first, the implication that \*baCaqan has for Sagart’s history of  
 453 \*puluq should be discussed.

454 A potential issue with the proposal that \*puluq originates from a verb mean-  
 455 ing ‘to separate’ is that if \*puluq was utilized only in forms with an other-than-  
 456 zero final digit, then there is no motivation for Sagart’s \*duSa-baCaqan ‘20’ or  
 457 \*təlu-baCaqan ‘30’ to ever become \*duSa-puluq or \*təlu-puluq, since \*puluq  
 458 only appears in 21–29, 31–39, 41–49, and so forth. How does Amis innovate  
 459 *tosa polo* ‘20’ when \*puluq never appeared in the history of that word? One  
 460 can imagine a conjecture that \*duSa baCaqan ‘20’ becomes \*duSa puluq  
 461 through analogy with \*duSa puluq ita, from earlier \*duSa baCaqan puluq  
 462 ita ‘21’. Sagart (2022) recognizes this and states that in Sakizaya and  
 463 Puyuma \*baCaqan was “predictably retained where there was no remainder  
 464 to ‘leave alone’.” Examples from Sakizaya include *cəcay a bataʔan* ‘10’  
 465 and *tusa bataʔan* ‘20’. This is misleading, however. Sakizaya *bataʔan* is not  
 466 only restricted to cases where there was no remainder to leave alone. It is also  
 467 utilized in words with a remainder, like *tusa bataʔan izau ku səpat* ‘24’. In  
 468 Sakizaya, therefore, the \*təlu-baCaqan-puluq-lima > \*təlu-puluq-lima simpli-  
 469 fication assigned to Proto-Puluqish appears to have never taken place. The  
 470 internal chronology of Sagart’s proposal and its relationship to the tree structure  
 471 that he proposes is therefore inconsistent.

472 It is not clear from where Sakizaya *bataʔan* originates. Since it is found  
 473 nowhere else in Amis, it cannot be reconstructed without external evidence.  
 474 Sagart gives Puyuma as an external witness. Puyuma has a special numeral for  
 475 multiples of ten from 20 and above formed with a prefix *maka*. Sagart cites  
 476 *maka-bəʔaʔan* ‘20’, *maka-telun*, and so on from Cauquelin (2015). Note that  
 477 Teng (2008) reports different numerals for twenty and thirty when used in count-  
 478 ing, citing instead *qua pulu* ‘twenty’ and *telu pulu* ‘thirty’. Forms with *maka*-  
 479 therefore indicate cardinal numerals spoken in isolation. There is an obvious issue  
 480 within the Puyuma evidence. The proposed reconstruction for ‘ten’, \*baCaqan,  
 481 appears only in the numeral ‘twenty’. The word for ‘ten’ is *mukəʔəp* and ‘thirty’  
 482 is *makətəʔun*. Why would the word for ‘ten’ appear only in the word for ‘twenty’,  
 483 and nowhere else? Sagart’s explanation is that *maka-bəʔaʔan* is derived from ear-  
 484 lier \*maka-qua-bəʔaʔan, where ‘dua’ was deleted. But this is not a satisfactory  
 485 answer. Why was \*qua deleted but not \*təlu from hypothetical \*maka-təlu-  
 486 bəʔaʔan? Furthermore, what explanation can be given for *mukəʔəp* ‘ten’ if we  
 487 assume a similar reconstruction \*maka-əsa-bəʔaʔan or \*muka-əsa-bəʔaʔan?

488 The other external witness that Sagart proposes is from Kavalan, Basay, and  
 489 Ketagalan. Here, Sagart’s historical explanation is less problematic. He cites Li

490 and Tsuchida (2006) for Kavalan *Rabtin* ‘10’ and *zusabtin* ‘20’. Note that PAN \*a  
 491 regularly becomes \*i when adjacent to \*q in Kavalan. Basay and Ketagalan have  
 492 similar words. The problem that now arises with Sagart’s \*baCaqaŋ proposal is its  
 493 distribution. The Puyuma evidence does not seem to reflect \*baCaqaŋ ‘ten’ con-  
 494 sidering its restriction to ‘twenty’. What remains are languages that all belong to  
 495 the East Formosan subgroup under more conventional classifications. Is \*baCaqaŋ  
 496 ‘ten’ really an Eastern Walu-Siwaish innovation or might it be better explained as  
 497 an East Formosan innovation? Considering the issues with Sagart’s numeral-based  
 498 subgrouping in general, there appears to be no good reason to prefer the former  
 499 over the latter.

500 Finally, Sagart attempts again to derive \*baCaqaŋ ‘ten’ from a preexisting  
 501 non-numeral. This time, he looks to Bunun, where he cites Tsuchida (1988)  
 502 *bataq-an* from the Qato and Idhokan dialects as an “L-shaped carrying device.”  
 503 His logic here is that Bunun, which lies outside Eastern Walu-Siwaish, retains  
 504 the original meaning ‘carrying rack’ from Proto-Walu-Siwaish \*baCaqaŋ,  
 505 which replaced \*iCit as ‘ten’ in Proto-Eastern Walu-Siwaish. The argument  
 506 here is similarly flawed to that which Sagart used to reconstruct \*puluq ‘sepa-  
 507 rate from, and set aside’ to Proto-Eastern Walu-Siwaish. That is, the evidence  
 508 for a meaning of ‘carrying rack’ is found only in Bunun, a Central Walu-  
 509 Siwaish language. There is not evidence from other Walu-Siwaish languages  
 510 to support reconstructing \*baCaqaŋ ‘carrying rack’ to Proto-Central Walu-  
 511 Siwaish. All potential cognates outside of Central Walu-Siwaish have the  
 512 meaning ‘ten’, and again, the semantic leap necessary to get from ‘carrying  
 513 rack’ to ‘ten’ is far too great to take the Bunun evidence as is. Without any  
 514 corroborating evidence from another Central Walu-Siwaish language, and with-  
 515 out any evidence from Eastern Walu-Siwaish that \*baCaqaŋ ever meant any-  
 516 thing other than ‘ten’, the etymology that he proposes is difficult to accept.

517 **5. SOUTHERN AUSTRONESIAN.** KD was first included within Puluqışh  
 518 in Sagart’s original 2004 proposal, but was further refined in Sagart (2014), where  
 519 it was placed alongside MP in what he calls “Southern Austronesian.” The pro-  
 520 posal was not originally defended with evidence, but Sagart (2021a,b) has since  
 521 provided some evidence for the subgroup, largely in the form of exclusively  
 522 shared lexical innovations and semantic shifts in KD and MP. The innovations  
 523 are listed below, with Sagart’s Proto-Southern Austronesian reconstruction:

- |                                 |               |                  |
|---------------------------------|---------------|------------------|
| (5) Proto-Southern Austronesian | *baqbaq       | ‘mouth’          |
| Proto-Southern Austronesian     | *biRáq        | ‘kind of taro’   |
| Proto-Southern Austronesian     | *-ŋəl         | ‘deaf’           |
| Proto-Southern Austronesian     | *píntu        | ‘door’           |
| Proto-Southern Austronesian     | *qi(d)zúR     | ‘sputum, saliva’ |
| Proto-Southern Austronesian     | *qa-sáuŋ      | ‘canine tooth’   |
| Proto-Southern Austronesian     | *(ma)-sapəléd | ‘astringent’     |
| Proto-Southern Austronesian     | *buŋaHl       | ‘flower’         |
| Proto-Southern Austronesian     | *bujak        | ‘flower’         |

525  
524  
526  
527  
528  
529  
530  
531  
532  
533

Upon close inspection, much of these lexical data suffer from inconsistencies, inaccuracies, and irregularities which bring into question their status as true lexical innovations. In the sections below, each proposed innovation is inspected. Special attention is given to the correspondences between PAN and Proto-Malayo-Polynesian (PMP) reconstructions and proposed cognates in KD. Although the study of sound correspondences between AN and KD is still in its early stages and the relationship between AN and KD is a matter of ongoing debate, enough is known about AN–KD sound correspondences that it is possible to identify several irregularities in Sagart’s proposals.<sup>6</sup>

534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545

**5.1. PROTO-SOUTHERN AUSTRONESIAN \*baqbaq ‘MOUTH’.** Sagart claims that \*ŋusuq is the earliest reconstruction available for ‘mouth’, which is reconstructed to PAN in Blust and Trussel (2020), and to Proto-Walu-Siwaish in his subgrouping. He argues that \*baqbaq originally meant ‘jaw’ or ‘chin’ but shifted to ‘mouth’ in Proto-Southern Austronesian, where it was in competition with \*ŋusuq. According to Sagart \*baqbaq eventually won out in many MP languages, as well as KD (Proto-Tai [PT] \*pa:k<sup>D</sup>, POB \*ʔa:k<sup>D1</sup>, for example). Problems with Sagart’s claims about ‘mouth’ are as follows: (i) the reconstruction of \*ŋusuq is misrepresented; (ii) supposed reflexes of \*baqbaq in Formosan languages are irregular; (iii) Sagart does not consider the semantic gap for ‘mouth’ at PAN significant, although the absence of a basic vocabulary item suggests that the gap may be artificial.

546  
547  
548  
549  
550  
551  
552  
553  
554  
555

First, Sagart’s claim that \*ŋusuq meant ‘mouth’ is a misrepresentation, since \*ŋusuq meant ‘nasal area, snout’ not ‘mouth’, with evidence from Amis *ŋoso?* ‘nose’, Kulisusu *me-ŋusu* ‘to sniff’, Motu *udu* ‘mouth; nose; beak’, Kapampangan *ŋusu?* ‘upper lip’, Tagalog *ŋuso?* ‘snout; muzzle; nose; mouth; jaw; trunk’,<sup>7</sup> Mussau *ŋusu* ‘nose’, and many others with a meaning that includes mouth, nose, snout, and beak. The interesting thing about PAN ‘mouth’, then, is that PAN, as conventionally reconstructed, contained a semantic gap for this basic vocabulary item. Sagart’s argument, that \*baqbaq competed with and eventually replaced \*ŋusuq, is flawed in this respect, since \*ŋusuq meant ‘snout’ and the two were never in competition.

556  
557  
558  
559  
560  
561

Another aspect of Sagart’s theory is that a Formosan word \*baqbaq existed, but must have meant ‘(lower) jaw’ or ‘chin’ and shifted to ‘mouth’ in Proto-Southern Austronesian. For example, Sagart points to Amis *babaq* ‘jaw’ (Sagart 2022, citing Brill, p.c.) and Bunun *vaqvaq* ‘chin’ (Wolff 2010) as evidence. It is not clear that these two words are related to \*baqbaq, however, since \*b normally becomes *f* in Amis and is normally retained unchanged in Bunun.

6. PT reconstructions in this section are from Pittayaporn (2009). Kra data, including Proto-Kra reconstructions, Paha, Pubiao, Qabiao, Gelao, and Buyang, are from Ostapirat (2000).

7. This gloss is from the *Austronesian Comparative Dictionary (ACD)*, but an anonymous reviewer points out that Tagalog *ŋuso?* is more restricted, meaning only ‘snout’ or ‘upper lip’ and suggests that this may be true for other cognates with ‘mouth’ listed as a possible meaning. This only adds to the issues with Sagart’s reconstruction.



562 There are thus no regular Formosan reflexes of \*baqbaq and no recon-  
 563 struction for the semantic category ‘mouth’ to PAN. It is expected that such a basic  
 564 vocabulary item should exist, and according to the hypothesis that KD and AN  
 565 are sisters, the KD reflexes may help explain this gap. If the KD reflexes are  
 566 truly cognate with \*baqbaq then this suggests that \*baqbaq was the word for  
 567 ‘mouth’ in an ancestor to both AN and KD, which is retained in KD and MP, but  
 568 lost in Formosan languages. Since there are so few Formosan languages when  
 569 compared with MP, it is not surprising that some PAN words may be retained  
 570 only in MP.<sup>8</sup> In cases where there is an empty category for a basic vocabulary  
 571 item at the PAN level and a corresponding MP-only reconstruction for the same  
 572 category, it is hard to ignore the value that KD languages may have in filling in  
 573 these PAN gaps. Retention from PAN is therefore a viable explanation for the  
 574 suspected KD cognates.

575 **5.2. PROTO-SOUTHERN AUSTRONESIAN \*-ŋəl ‘DEAF’.** The root  
 576 \*-ŋəl ‘deaf’ and apparent Kra cognates reflecting Proto-Kra \*ŋəl seem to have  
 577 regular correspondences. Although the word is rather restricted KD, appearing  
 578 only in the Kra branch, the necessary regularity is there, and the comparison  
 579 appears legitimate.

580 **5.3. PROTO-SOUTHERN AUSTRONESIAN \*biRaq ‘WILD TARO’.**  
 581 Sagart proposes that PAN \*biRaq originally meant ‘leaf’ and narrowed to ‘taro’  
 582 in Proto-Southern Austronesian, an innovation that groups KD and MP together.  
 583 To support this, he claims that Formosan witnesses of \*biRaq mean ‘leaf’ not  
 584 ‘taro’. This is not true, however, as reflexes of \*biRaq mean ‘wild taro’ in  
 585 Rukai (Li 1977). The proposed semantic shift ‘leaf’ > ‘taro’ in Southern  
 586 Austronesian is therefore invalid, since the meaning ‘taro’ is a retention.

587 **5.4. PROTO-SOUTHERN AUSTRONESIAN \*pintu ‘DOOR’.** Sagart  
 588 assigns \*pintu to Proto-Southern Austronesian after first assigning it to PMP  
 589 with Philippine and Malay evidence. This approach is flawed, however, since  
 590 words for ‘door’ in Philippine languages with shapes similar to Malay *pintu* are  
 591 phonologically irregular and are likely borrowed from Malay (Blust and Trussel  
 592 2020). To claim a relationship between Malay *pintu* and KD words of similar  
 593 shape is therefore methodologically flawed because KD words are being com-  
 594 pared with a word from a single AN lower level subgroup which cannot inde-  
 595 pendently be reconstructed to a first order AN node.

596 The KD comparisons also do not have the expected correspondences with  
 597 Sagart’s PMP \*pintu. PT, for example, typically has diphthongs where PAN has  
 598 word-final high vowels: PT \*kraw<sup>C</sup> : PAN \*qulu ‘head’, PT \*traw<sup>A</sup> : PAN

8. A reviewer points out that there are more primary branches in Taiwan, so there should be more opportunities for retention of vocabulary in Taiwan rather than outside of Taiwan. There may be more branches, but there are far fewer languages. The implication is that with so few languages in Taiwan, that fewer lexical replacement events are necessary in Taiwan to obscure a reconstruction than are necessary outside of Taiwan.

599 \*kuCu ‘louse’, PT \*nay<sup>C</sup> : PAN \*ni ‘this’. Tai reflexes do not show expected  
 600 diphthongization in suspected cognates of *pintu*: PT \*tu:<sup>A</sup>. So, in addition to the  
 601 questionable PMP reconstruction \*pintu, the proposed KD cognates suffer  
 602 from their own irregularities, which calls into question the comparison.

603 **5.5. PROTO-SOUTHERN AUSTRONESIAN \*qi(d)zúR ‘SPUTUM,**  
 604 **SALIVA’.** Sagart suggests that \*qi(d)zúR ‘sputum, saliva’ was innovated at  
 605 the Proto-Southern Austronesian level and existed along reflexes of earlier  
 606 Proto-Walu-Siwaish \*ŋalay (PMP \*qizuR and PAN \*ŋajay in Blust and  
 607 Trussel [2020]). Reflexes of \*ŋalay and \*qi(d)zúR are found in KD and MP,  
 608 but only \*ŋalay reflexes are found in Formosan languages. However, proposed  
 609 KD cognates of \*qi(d)zúR suffer from irregularities in sound correspondences  
 610 and may not be actual cognates, as explained below.

611 First, the reflex of \*z in KD is not clear, since there are no widely accepted  
 612 potential cognates. Smith (2022) suggests that \*Səzam ‘to lend; borrow’ may  
 613 have cognates, with the correspondence \*z : y in Tai (PT \*<sup>2</sup>yw:m<sup>a</sup>), and  
 614 Southern Kam *yam*<sup>55</sup>. However, without multiple witnesses there is little other  
 615 evidence supporting either of these comparisons and it is not clear if \*z > y is  
 616 regular. Second, Sagart suggests (d) may have been a medial consonant in his  
 617 reconstruction, but \*d (from earlier \*d and \*j) belongs to a different correspon-  
 618 dence set than that of Sagart’s suggested reflex. So the logical question to ask is  
 619 to which, if any, correspondence set do KD words for ‘sputum, saliva’ belong?  
 620 In evaluating the suspected reflexes of \*qi(d)zúR, one finds the following cor-  
 621 respondence set for the medial consonant(s): Paha *ðu*<sup>45</sup> : Pubiao (also Qabiao)  
 622 *tau*<sup>B1</sup> : Buyang *tuu*<sup>B2</sup>. A *ð* : *t* : *t* correspondence between Paha, Pubiao/Qabiao,  
 623 and Buyang typically corresponds to PAN \*t and \*C, not \*z or \*d. For example,  
 624 table 6 demonstrates this correspondence with PAN, Paha, Qabiao, and Buyang  
 625 ‘seven’ and ‘louse’.

626 It is therefore unlikely that the Kra words for ‘saliva’ correspond to PMP  
 627 \*qi(d)zúR because the medial consonant cannot reflect \*-(d)z- but would regu-  
 628 larly reflect \*C/\*t.

629 **5.6. PROTO-SOUTHERN AUSTRONESIAN \*qa-sáuŋ ‘CANINE TOOTH’.**  
 630 Sagart objects to the reconstruction \*baŋəliS on the grounds that it is specifi-  
 631 cally a word for ‘tusk’ and suggests that the only PAN word for ‘canine tooth’ is  
 632 a Formosan only word, \*waqit. According to his history of the word, an inno-  
 633 vation, \*qa-sáuŋ ‘canine tooth’ was innovated at Proto-Southern Austronesian  
 634 with reflexes in Philippine languages and KD. Issues with this comparison

TABLE 6. REGULAR CORRESPONDENCES BETWEEN KRA LANGUAGES  
 AND PAN \*t AND \*C.

AQ11

PAN	Paha	Qabiao	Buyang	
*pitu	ðu <sup>33</sup>	mə0 tu53	tuuA2	seven
*kuCu	ðu <sup>33</sup>	qa0 tau53	(qa0)tu54	louse

635 include (i) an inaccurate representation of the reconstruction *\*baŋəliS*, (ii) the  
 636 unreconstructability of *\*qa-sáuŋ* ‘canine tooth’ to PMP, and (iii) irregularities in  
 637 the proposed KD cognates of *\*qa-sáuŋ*.

638 First, Sagart’s representation of the reconstruction *\*baŋəliS* as ‘specifically a  
 639 word for ‘tusk’” (Sagart 2022) is inaccurate; *\*baŋəliS* is reflected as both ‘tusk’  
 640 and ‘canine tooth’, not only ‘tusk’. But even so, Sagart’s objection to its recon-  
 641 struction as ‘tusk; canine tooth’ is odd, since this same semantic crossover is  
 642 also found in his proposed Proto-Southern Austronesian reconstruction *\*qa-  
 643 săuŋ* ‘canine tooth’, which has reflexes only in the northern Philippines and  
 644 means ‘tusk’ as well as ‘canine tooth’ in every language recorded, Yami,  
 645 Itbayaten, Ivatan, Ibatan, Ilokano, Bontok, and Kankanaey. The same semantic  
 646 crossover is again found in the Formosan only word *\*waqit*, with reflexes like  
 647 Atayal ‘canine tooth; fang; tusk; tooth or tool of an instrument’. Sagart’s initial  
 648 objection to the *\*baŋəliS* reconstruction therefore is unfounded. *\*qa-sáuŋ* is not  
 649 a PMP replacement innovation, but rather a Philippine-only innovation, while  
 650 *\*baŋəliS* continued into Kelabit *bəŋəlih*. The fact that ‘tusk’ is a possible use of  
 651 the word is irrelevant, since that usage appears consistently across different  
 652 words meaning ‘canine tooth’.

653 An additional issue with this comparison is that it is not clear that the  
 654 reflexes in Kra are regular, since *\*s* becomes Paha *t*, Pubiao *ʔ*, and Gelao  
 655 *ts*, in the word for ‘one’, from PAN *\*isa*. Examples are shown below:

- (6) PAN *\*isa*  
 Paha *ti*<sup>55</sup>  
 Pubiao *ʔeja*<sup>C1</sup>  
 Gelao *tsi*<sup>55</sup>

657 Supposed Kra reflexes of *\*qa-sáuŋ*, on the other hand, appear to belong to a  
 658 different correspondence set:

- AQ12 (7) Paha *jəŋ*<sup>322</sup>  
 Pubiao *θuaŋ*<sup>A1</sup>  
 Gelao *pləŋ*<sup>31</sup>

660 Although it is true that *\*s* may be reflected differently in the onset of the  
 661 penult, the fact that Gelao has a *pl* cluster in initial position casts doubt on a  
 662 reconstruction with *\*s-*. It is therefore unlikely that Kra words for ‘tooth’ are  
 663 related to Northern Philippine words for ‘tusk; canine tooth’, due both to the  
 664 unreconstructability of *\*qa-sáuŋ* and irregularities in the proposed KD cognates.

665 A final note about the reconstruction *\*qa-sauŋ* is that the Philippine evidence  
 666 does not clearly support an initial *\*qa-*, which appears in none of the examples  
 667 listed in the *ACD*. A more appropriate reconstruction is *\*sauŋ*, but again, this  
 668 word does not appear to be related to any KD word.

669 **5.7. PROTO-SOUTHERN AUSTRONESIAN *\*(ma)-sapələd* ‘ASTRI-**  
 670 **NGENT’.** Sagart notes that *\*sapələd* replaced PAN *\*qasəpa* ‘astringent’ in  
 671 Proto-Southern Austronesian, while PAN *\*qasəpa* shifted to ‘tasteless’ in

672 Proto-Southern Austronesian. He then compares \*sapələd to similar reflexes in  
 673 KD, including PT \*<sup>h</sup>wuət<sup>D</sup> and Buyang *phat*<sup>54</sup>. He proposes that the innovation  
 674 of \*sapələd is shared between KD and MP, providing more evidence for Proto-  
 675 Southern Austronesian. The first issue with this comparison lies in the vowel  
 676 correspondences. PAN and PMP \*ə in final syllables typically corresponds to a  
 677 short \*a in PT reconstructions, as shown below:

- (8) PT \*t.nam<sup>A</sup> : PAN \*tanəm ‘to plant’  
 PT \*kam<sup>A</sup> : PAN \*kəmkəm ‘to hold in the fist’  
 PT \*C.dam<sup>A</sup> : PAN \*dəmdəm ‘dark; black’/  
 \*tidəm ‘black’

679 Exceptions to this include cases where PAN/PMP \*ə corresponds a short  
 680 back vowel, such as PT \*C.tɔŋ : PMP \*datəŋ ‘arrive’. Cases where PT words  
 681 have either \*u: or uə in the final syllable tend to correspond to PAN \*a, and are  
 682 further conditioned by the presence of a high vowel in the reconstructed penult,  
 683 as shown below:

- (9) PT \*ɬluən<sup>A</sup> : PAN \*bulaN ‘moon’  
 PT \*C.buə<sup>A</sup> : PAN \*tubah ‘derris root’  
 PT \*d̥wu:.<sup>A</sup> : PAN \*puja ‘navel’

685 The given cognate of \*(ma)-səpələd in PT is surprising in that the schwa in  
 686 PMP corresponds to \*uə, rather than to \*a (and sometimes ɤ). PT \*uə suggests  
 687 an \*a in the PMP final syllable, not schwa.

688 Another issue with this proposed cognate concerns the initial consonants.  
 689 \*sapələd is a three syllable word which would have been reduced to a mono-  
 690 syllable in KD. The typical path of such reduction is pələd > pləd, ignoring for  
 691 now the issue of the antepenultimate syllable. An initial \*pl- cluster would  
 692 remain as \*pl- in PT, whereas \*<sup>h</sup>w follows from an earlier \*C.p-, for example  
 693 PAN \*qaCipa ‘soft shelled turtle’ : PT \*<sup>h</sup>wuə ‘soft shelled turtle’. There is no  
 694 explanation for why \*l deleted and no viable path from Proto-Southern  
 695 Austronesian \*sapələd to PT \*<sup>h</sup>wuət<sup>D</sup>.

696 **5.8. PROTO-SOUTHERN AUSTRONESIAN \*bujak AND \*buŋa**  
 697 **‘FLOWER’.** Sagart argues that the PAN word for flower was \*buRay, a  
 698 Formosan-only word found in Saisiyat and Paiwan which Blust and Trussel  
 699 reconstruct alongside PAN \*buŋa. Sagart rejects PAN \*buŋa as a reconstruction  
 700 with the meaning ‘flower’, arguing that because it has meanings such as blos-  
 701 soms, fruit buds, fruits, saplings, and many other much looser referents than the  
 702 strict meaning ‘flower’, another reconstruction for ‘flower’ is warranted. Sagart  
 703 concludes that \*buRay is the only PAN word for flower. He then proposes that  
 704 \*buRay was replaced by \*buŋa and \*bujak as ‘flower’ in Proto-Southern  
 705 Austronesian. Both \*buŋa and \*bujak are then inherited in KD and MP, respec-  
 706 tively, with no trace of \*buRay.

707 There are many issues with these reconstructions. First, \*bujak cannot be  
 708 reconstructed to PMP. Its evidence is restricted to Philippine languages, except

709 in Kadazan Dusun and Kelabit, both languages of Borneo, where the suspected  
 710 reflexes listed on the *ACD* are *vusak* and *busak*, respectively. The reflex of \*j is  
 711 irregular in these languages, since \*j typically merges with \*d the common  
 712 ancestor of Kelabit and Kadazan Dusun (Smith 2017). There is thus no basis  
 713 for reconstructing \*bujak outside of the Philippines. Like \*pintu, this word can-  
 714 not be reconstructed to PMP with internal evidence, and therefore the same  
 715 criticisms apply to it as apply to *pintu*.

716 Second, Sagart compares his PMP \*bujak to PT \*6lo:k ‘flower’. The issues  
 717 with this comparison are twofold. First, PAN and PMP \*a corresponds to PT \*a:  
 718 or \*u:/uə/wu if conditioned by a high vowel in the penult, as already discussed.  
 719 PT \*6lo:k would therefore be the only case where PMP \*a corresponds to \*o:.  
 720 Where an \*o(:) is found in a PT syllable closed with a voiceless stop, the most  
 721 likely PAN correspondence is with \*-uC, where C is any voiceless consonant. For  
 722 example, Zhuang *tok*<sup>DS1</sup> : PAN \*-tuq ‘to fall’ and PT \*C.nok<sup>D</sup> : PAN \*manuk  
 723 ‘bird’. It is therefore unlikely that \*bujak is a shared cognate in KD and MP.

724 The second issue is with his explanation of the / reflex of \*j in PT \*6lo:k. Sagart  
 725 explains / as follows: “When forming a cluster with a preceding labial stop (as  
 726 inside a trisyllable, where syncope of the second vowel occurs), the stop is retained  
 727 and \*d (like \*t-) lenites to l.” To support this interpretation he points to PT \*6li:  
 728 ‘gall’, which is said to reflect \*qapəju. This comparison is also problematic, since  
 729 it contains a sound change of \*u > i: in the final syllable. Sagart has an explanation  
 730 for this, too, and claims that a change of \*u to i: after \*j in PT \*6li: is regular, being  
 731 conditioned by the preceding \*j. The issue here is with the chronology. It is estab-  
 732 lished that \*j and \*d merged not only in Tai, but in all other KD languages where  
 733 they can be compared (Ostapirat 2005). PT \*6li: therefore does not provide evi-  
 734 dence that \*d became l next to a labial, since the vowel irregularity can no longer  
 735 be explained away as a consequence of its adjacency to \*j, which had already  
 736 merged with \*d in Proto-Kra-Dai. PT \*6li: probably has no relation to any AN  
 737 word. There is therefore no confirming evidence that \*bVd- became PT \*6l-.

738 Regarding Sagart’s objections to reconstructing \*buŋa as ‘flower’, it is not  
 739 clear that these objections mean that the narrowing of \*buŋa to ‘flower’ in sus-  
 740 pected KD cognates is best described as a shared innovation. PAN \*buŋa likely  
 741 meant ‘flower, fruit bud, fruit, sprout, fruiting part of a plant’ and other mean-  
 742 ings derived from a core meaning of ‘early stage visible extrusion from a plant’.  
 743 Formosan comparisons like Saaroa *vuŋa-vuŋa* ‘ear of foxtail millet’, *vuu-vuŋa*  
 744 ‘flower’, or Kananabau *buŋa-buŋa* ‘flower’ (Sagart 2022) follow naturally  
 745 from this reconstruction. Importantly, the direction of semantic change in this  
 746 term is clearly one of narrowing, and there are multiple independent instances  
 AQ13 of narrowing toward a meaning of ‘flower’, as well as narrowing’s toward other  
 748 meanings in both Formosan and MP languages. A parallel development in sus-  
 749 pected KD cognates cannot therefore be ruled out. If Proto-Kra \*hŋa ‘flower’  
 750 truly meant only ‘flower’ and nothing else, and if it is truly cognate with \*buŋa,  
 751 then parallel semantic narrowing applies to this word just as it does to

752 Kanakanabu *buŋa-buŋa* or Malay *buŋa*. One cannot rule out drift as an expla-  
753 nation in this case.

754 **5.9. PROTO-SOUTHERN AUSTRONESIAN \*báNaS ‘HUSBAND’.** PAN  
755 \*baNaS meant ‘male, of animals’ but shifted to ‘husband’ in PMP. Sagart points  
756 to Proto-Hlai \*ali<sup>A</sup> as a potential cognate, also with the meaning to ‘husband’.  
757 There is no immediate reason to discard this comparison. Hlai tends to delete the  
758 initial consonant in words with AN cognates, for example, Proto-Hlai \*aRi:u<sup>C</sup> ‘to  
759 sell’ : PAN \*saliw/baliw ‘to sell’, Proto-Hlai \*ava<sup>C</sup> ‘shoulder’ : PAN \*qabaRa  
760 ‘shoulder’, and Proto-Hlai \*uRa<sup>A</sup> ‘to plant’ : PAN \*mula ‘to plant’. The \*l :  
761 N correspondence appears in Proto-Hlai \*ali:k ‘child’ : PAN \*aNak ‘child’, as  
762 does the \*i : \*a correspondence. It is not known if the deletion of \*S in final  
763 position is regular, and of course, with the deletion of the initial consonant, it  
764 is not clear if the original initial was in fact \*b-, since there are no KD cognates  
765 in other branches to which \*ali<sup>A</sup> may be compared. In all, this is one of the stron-  
766 ger comparisons that Sagart proposes, but with its restriction to Hlai and the fact  
767 that it relies on semantic rather than phonological evidence, it alone does not  
768 inform a compelling argument for Southern Austronesian.

769 **5.10. KD WORDS FOR ‘TEN’ THAT END IN *t*.** It has been noted that sus-  
770 pected KD cognates of PAN \*puluq ‘ten’ end with a *t*, which is irregular since \*-  
771 AQ14 q usually corresponds to *-k* in KD (Ostapirat 2005; Sagart 2021a,b; Smith  
772 2022). Explanations for this irregularity vary. Ostapirat proposes that \*-q  
773 becomes *-t* when it appears after *u*. Smith questions Ostapirat’s explanation,  
774 citing other words that appear to have ended in \*-uq that did not undergo simi-  
775 lar fronting. Sagart proposes that *-t* is from a linker, \*at, which appears between  
776 \*puluq and following numerals in some Philippine languages. Sagart’s original  
777 proposal is problematic because the linker, *at*, is only found in northern  
778 Philippine languages, and cannot be reconstructed to a higher node.  
779 Subsequent revisions of that initial proposal has led Sagart to propose that  
780 an original linker, \*atu ‘and’, was reduced to just \*at in Proto-Southern  
781 Austronesian, and that \*at was further reduced to *t* as a linker between \*puluq  
782 and other numerals in the Philippines and KD. Evidence for \*atu is from Amis  
783 *atu* ‘and’ and Paiwan *katu* ‘and’. Sagart notes that both *atu* and *katu* are ana-  
784 lyzed as *a-tu* and *ka-tu*, respectively, with *a* and *ka* originally meaning ‘and’,  
785 and an addition of the *tu* oblique common noun marker. There are two problems  
786 with Sagart’s \*atu reconstruction. First, Amis and Paiwan *atu* and *katu* are  
787 formed with different bases, *a* and *ka* from PAN \*a and \*ka ‘and’. They do  
788 not agree on the shape of a putative ‘and’ reconstruction so it is not clear that  
789 \*atu > \*at is a legitimate Southern Austronesian innovation since \*atu would  
790 be an amalgamation of different root words. Second, Paiwan *t* reflects PAN \*s,  
791 not \*t. PAN \*t, on the other hand, regularly becomes Paiwan *tj*.<sup>9</sup> The compari-  
792 son is therefore also invalid on phonological grounds.

9. Thanks to Robert Blust for pointing this out.

793 In summary, Southern Austronesian lexical innovations do not withstand  
 794 close scrutiny and do not strengthen the argument that KD and MP are sister  
 795 subgroups within AN. The major piece of evidence remains the fact that both  
 796 reflect \*puluq, but this evidence is only important if one accepts Sagart's  
 797 numeral-based phylogeny. Under a more conventional model of AN higher  
 798 order subgrouping, \*puluq is reconstructable all the way to PAN, so its presence  
 799 in KD does not suggest that KD belongs within the AN family tree.

## 800 6. CONCLUSION: NO MORE CONVINCING NOW THAN BEFORE.

801 This paper, a critical review of Sagart's new evidence for the numeral-based phy-  
 802 logeny, analyzed the \*NəmNəm 'six', \*puluq 'to separate from, and alone', and  
 AQ15 Southern Austronesian portions of Sagart (2021a,b). Overall, Sagart's new evi-  
 804 dence does not form a convincing argument for the numeral-based phylogeny.  
 805 Rather, it was shown that etymologies for 'six' and 'ten' remain problematic,  
 806 and that the new proposals suffer from the same kinds of problems that Blust,  
 807 Ross, and Winter have pointed out for older versions of the numeral-based phy-  
 808 logeny. Lexical evidence for an Southern Austronesian subgroup does not fare  
 809 any better. Inaccuracies in his reporting of PAN reconstructions, the semantics of  
 810 reflexes, and a lack of regularity in proposed KD cognates plague Sagart's list of  
 811 lexical innovations. In the end, there is little evidence to support a subgroup that  
 812 contains KD and MP, other than the Puluqish node proposal itself, which is not  
 813 considered viable by this study.

814 The numeral-based phylogeny, while novel, does not adhere to the principles  
 815 of the comparative method. Irregularities in sound correspondences are ignored in  
 816 favor of interpretations that favor the numeral-based phylogeny hypothesis. Great  
 817 semantic differences are presented as non-problematic. Reconstructions are made  
 818 to Proto-languages without evidence from multiple primary branches. Although  
 819 this paper only looked at a portion of the newly proposed evidence, one is left  
 820 with little confidence that other evidence will fare any better. What the numeral-  
 821 based phylogeny does show us, however, is the value of the comparative method  
 822 not only as a means to accurately and confidently uncover the historical devel-  
 823 opment of languages, but as a tool to determine the strength of new proposals. In  
 824 the case of the numeral-based phylogeny, the comparative method shows once  
 825 again that Sagart's novel approach to AN higher order subgrouping leaves much  
 826 to be desired. This paper therefore must conclude that the numeral-based phylog-  
 827 eny, even with this new evidence, is unable to challenge more conventional pro-  
 828 posals on AN subgrouping which are grounded in more traditional methods.

829

## REFERENCES

- 830 Aldridge, Edith. 2016. Ergativity from subjunctive in Austronesian languages.  
 831 *Language and Linguistics* 17(1): 27–62.  
 832 Benedict, Paul K. 1942. Thai, Kadai, and Indonesian: A new alignment in Southeastern  
 833 Asia. *American Anthropologist* 44:576–601.

- 834 ———. 1995. Extra-austronesian evidence for Formosan etyma. In *Austronesian studies relating to Taiwan*, ed. by Paul J. K. Li, C. H. Tsang, Y. K. Huang, D. A. Ho, and  
835 C. Y. Tseng, 399–454. Symposium series of the Institute of History and Philology,  
836 Academia Sinica 3. Taipei: Academia Sinica.
- 837 Blust, Robert. 1988. *Austronesian root theory: An essay on the limits of morphology*.  
838 Amsterdam/Philadelphia: John Benjamins.
- 839 ———. 1999. Subgrouping, circularity and extinction: Some issues in Austronesian  
840 comparative linguistics. In *Selected Papers from the Eighth International*  
841 *Conference on Austronesian Linguistics, Number 1 in Symposium series of the*  
842 *Institute of Linguistics, Academia Sinica*, ed. by Elizabeth Zeitoun and Paul J.  
843 K. Li, 31–94. Taipei: Academia Sinica.
- 844 ———. 2003. *Thao dictionary. Language and linguistics monograph series*. Taipei:  
845 Institute of Linguistics, Academia Sinica.
- 846 ———. 2013. *The Austronesian languages*. 2nd edition. Canberra: Pacific Linguistics.
- 847 ———. 2014a. Some recent proposals concerning the classification of the Austronesian  
848 languages. *Oceanic Linguistics* 53(2): 300–91.
- 849 ———. 2014b. *Blust field notes*. Kaipuleohone Digital Language Archive. [https://hdl.  
850 handle.net/10125/33115](https://hdl.handle.net/10125/33115).
- 851 ———. Two birds with one stone: The aerodynamic voicing constraint and the lan-  
852 guages of Borneo. *Journal of the Southeast Asian Linguistics Society* 11(2): 1–18.
- 853 **AQ16** Blust, Robert, and Stephen Trussel. 2020. *Austronesian Comparative Dictionary*.  
854 <https://www.trussel2.com/ACD/>.
- 855 Campbell, Lyle. 2021. *Historical linguistics: An introduction*. 4th edition. Cambridge:  
856 MIT Press.
- 857 Cauquelin, Josiane. 2015. *Nanwang Puyuma-English dictionary*. Taipei: Academia  
858 Sinica.
- 859 Chang, Anna Hsiou-chuan. 2006. A reference grammar of Paiwan. Doctoral thesis, The  
860 Australian National University.
- 861 Chen, Chun-Mei. 2006. *A comparative study on Formosan phonology: Paiwan and*  
862 *Budai Rukai*. Austin: The University of Texas at Austin.
- 863 Chretien, C. Douglas. 1965. The statistical structure of the Proto-Austronesian morph.  
864 *Lingua* 14:243–70.
- 865 Crowley, Terry, and Claire Bower. 2010. *An introduction to historical linguistics*.  
866 Oxford: Oxford University Press.
- 867 Li, Paul Jen-kuei. 1977. The internal relationships of Rukai. *Bulletin of the Institute of*  
868 *History and Philology, Academia Sinica* 48: 1–92.
- 869 ———. 2004. Origins of the East Formosans: Basay, Kavalan, Amis, and Siraya.  
870 *Language and Linguistics* 5:363–76.
- 871 ———. 2008. Time perspective of Formosan Aborigines. In *Past human migrations in*  
872 *East Asia: Matching archaeology, linguistics and genetics*, ed. by Alicia Sanchez-  
873 Mazas, Roger Blench, Malcolm Ross, Ilia Peiros, and Marie Lin, 211–18. London:  
874 Routledge.
- 875 Li, Paul Jen-kuei, and Masayuki Toyoshima, eds. 2006. *Comparative vocabulary of*  
876 *Formosan languages and dialects, by Naoyoshi Ogawa*. Tokyo: Institute for  
877 Languages and cultures of Asia and Africa, Tokyo University of Foreign Studies.
- 878 Li, Paul Jen-kuei, and Shigeru Tsuchida. 2006. *Kavalan dictionary*. Taipei: Academia  
879 Sinica.
- 880 McGuckin, Catherine. 2002. Gapapaiwa. In *The Oceanic languages*, ed. by John Lynch,  
881 Malcolm Ross, and Terry Crowley, 297–321. Curzon Language Family Series.  
882 Richmond, Surrey: Curzon Press.
- 883 Ostapirat, Weera. 2000. Proto-Kra. *Linguistics of the Tibeto-Burman Area* 23(1): 1–251.
- 884 ———. 2005. Kra–Dai and Austronesian: Notes on phonological correspondences  
885 and vocabulary distribution. In *The peopling of East Asia: Putting together*  
886



- 887 *archaeology, linguistics and genetics*, ed. by Laurent Sagart, Roger Blench, and  
 888 Alicia Sanchez-Mazas, 107–31. London: Routledge.
- 889 Pittayaporn, Pittayawat. 2009. The phonology of Proto-Tai. Doctoral diss., Department  
 890 of Linguistics, Cornell University.
- 891 Ross, Malcolm. 2009. Proto Austronesian verbal morphology: A reappraisal. In  
 892 *Austronesian historical linguistics and culture history: A festschrift for Robert*  
 893 *Blust*, ed. by K. Alexander Adelaar and Andrew Pawley, 295–326. Canberra:  
 894 Pacific Linguistics.
- 895 ———. 2012. In defense of Nuclear Austronesian (and against Tsouic). *Language and*  
 896 *Linguistics* 13:1253–330.
- 897 Sagart, Laurent. 2004. The higher phylogeny of Austronesian and the position of Tai-  
**AQ17** Kadai. *Oceanic Linguistics* 43(2).
- 899 ———. 2008. The expansion of Setaria farmers in East Asia: A linguistic and archaeo-  
 900 logical model. In *Past human migrations in East Asia: Matching archaeology, lin-*  
 901 *guistics and genetics*, ed. by Alicia Sanchez-Mazas, Roger Blench, Malcolm Ross,  
 902 Ilia Peiros, and Marie Lin, 133–58. London: Routledge.
- 903 ———. 2014. In defense of the numeral-based model of Austronesian phylogeny, and  
 904 of Tsouic. *Language and Linguistics* 15:859–82.
- 905 Sagart, Laurent. 2021a. A more detailed early Austronesian phylogeny. Plenary presen-  
 906 tation at the *15th International Conference on Austronesian Linguistics (15 ICAL)*,  
 907 Palacký University Olomouc, Czech Republic, June 28–July 1, 2021.
- 908 ———. 2022. Sino-Tibetan-Austronesian. <https://stan.hypotheses.org/> (Accessed July  
 909 24, 2022).
- 910 Smith, Alexander D. 2017. The Western Malayo-Polynesian problem. *Oceanic*  
 911 *Linguistics* 56(2): 435–90.
- 912 ———. 2022. More Austro-Tai comparisons and observations on vowel correspond-  
 913 ences. In *Papers from the 30th Conference of the Southeast Asian Linguistics*  
 914 *Society (2021)*, ed. by Mark Alves and Paul Sidwell, 112–34. JSEALS Special  
 915 Publication No. 8. Honolulu: University of Hawai‘i Press.
- 916 ———. To appear. Reconstructing non-contrastive stress in Austronesian and the role of  
**AQ18** the mora in stress shift, gemination, and vowel shift. *Diachronica*.
- 918 Teng, Stacy Fang-ching. 2008. *A reference grammar of Puyuma, an Austronesian lan-*  
 919 *guage of Taiwan*. Canberra: Pacific Linguistics.
- 920 Tsuchida, Shigeru. 1975. Reconstruction of Proto-Tsouic phonology. Doctoral diss.,  
 921 Yale University.
- 922 ———. 1988. Comparative word lists of Bunun dialects. Report of the research carried  
 923 out in 1983. Unpublished manuscript.
- 924 Winter, Bodo. 2010. A note on the higher phylogeny of Austronesian. *Oceanic*  
 925 *Linguistics* 49(1): 282–87.
- 926 Wolff, John. 2010. *Proto-Austronesian phonology with glossary*, vol. 2. Ithaca: Cornell  
 927 University Press.

## QUERIES

- AQ1: There are two “Blust (2014)” references in the Reference section (denoted as 2014a and 2014b). Please specify which is being cited here.
- AQ2: Although Sagart (2021b) is cited, there is no corresponding Reference entry. Please provide a Reference entry for this citation, or indicate the citations should be deleted throughout the article.
- AQ3: There are two “Blust (2014)” references in the Reference section (denoted as 2014a and 2014b). Please specify which is being cited here.
- AQ4: There are two “Blust (2014)” references in the Reference section (denoted as 2014a and 2014b). Please specify which is being cited here.
- AQ5: Reference year for Sagart and the closing parenthesis are missing here. Please check and provide the same.
- AQ6: Closing single quote seems missing. Please check and provide the same.
- AQ7: Although Li and Toyoshima (2008) is cited, there is no corresponding Reference entry. Please provide a Reference entry for this citation, or indicate the citations should be deleted throughout the article.
- AQ8: Although Blust (2018) is cited, there is no corresponding Reference entry. Please provide a Reference entry for this citation, or indicate the citations should be deleted throughout the article.
- AQ9: Please update publishing year. If possible.
- AQ10: Although Sagart (2021b) is cited, there is no corresponding Reference entry. Please provide a Reference entry for this citation, or indicate the citations should be deleted throughout the article.
- AQ11: Please make sure whether the "0", "53", "A2", and "54" in table 4 should be in superscript.
- AQ12: Is this "32" or "322"? Please clarify.
- AQ13: Is this narrowing? Please clarify.
- AQ14: Although Sagart (2021b) is cited, there is no corresponding Reference entry. Please provide a Reference entry for this citation, or indicate the citations should be deleted throughout the article.
- AQ15: Although Sagart (2021b) is cited, there is no corresponding Reference entry. Please provide a Reference entry for this citation, or indicate the citations should be deleted throughout the article.
- AQ16: Please provide publishing year for this reference.
- AQ17: For Sagart (2004), please provide page range.
- AQ18: For Smith (to appear), please update publishing year, volume number, and page range.