



**HAL**  
open science

## Is Puyuma a primary branch of Austronesian ?

Laurent Sagart

► **To cite this version:**

Laurent Sagart. Is Puyuma a primary branch of Austronesian?. *Oceanic Linguistics*, 2010, 49 (1), pp.194-204. hal-00781049

**HAL Id: hal-00781049**

**<https://hal.science/hal-00781049>**

Submitted on 25 Jan 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**IS PUYUMA A PRIMARY BRANCH OF AUSTRONESIAN ?**

Laurent Sagart  
EHESS and INaLCO, Paris

**Abstract:** Malcolm Ross's new theory of early Austronesian phylogeny is examined. The author describes evidence that \*-en served to mark verbs in undergoer voice, patient subject in a language ancestral to Puyuma, as well as evidence that \*<in> marks undergoer voice, patient subject perfective in one sociolect of Nanwang Puyuma. This evidence falsifies the claim that Puyuma reflects an early Austronesian stage at which \*-en and \*<in> had not yet been reinterpreted from nominalizers into voice markers. It also falsifies the phylogeny which takes that putative innovation as its central event. A hypothetical scenario is offered to account for the replacement of the \*-en, \*-an and \*Si- (or \*Sa-) series of voice markers by the series now found in Puyuma independent verbs.

In a recent paper, Malcolm Ross (2009) presents a new account of early Austronesian phylogeny, based in large part on new data from Teng's useful Puyuma grammar (Teng 2008). My intention here is to point out some problems with Ross's theory.<sup>1</sup> In discussing it, I will for convenience follow Ross in designating the four traditional 'focus' categories as: AV (Actor Voice), UVP (Undergoer Voice, patient subject), UVL (Undergoer Voice, location subject) and UVC (Undergoer Voice, circumstance subject).

Ross's starting point is that the familiar Austronesian neutral undergoer-voice markers UVP \*-en, UVL \*-an and UVC \*Si-, as well as the perfective aspect marker \*<in>, regarded as part of PAn verbal morphology since Wolff (1973), never occur in Puyuma, Rukai and Tsou as voice or aspect markers on verbs, but only as nominalizers on deverbal nouns. From this he infers (somewhat tentatively) that the reinterpretation of verb nominalizers into voice markers had not yet taken place in PAn. This development was assumed to be the mechanism behind the formation of the voice system of Philippine and Formosan languages in the theory of Starosta, Pawley and Reid (1982; known as 'SPQR'<sup>2</sup>). Instead, Ross proposes (2009:304, 306) that this reinterpretation occurred at a later stage, and that it defines one primary branch of PAn ("Nuclear An"), which includes all An languages except Puyuma, Rukai and Tsou; each of these three representing a primary An branch (FIGURE 1). In this theory, the neutral set of UV markers found in Puyuma: UVP \*-aw, UVL \*-ay and UVC \*-anay, occurred in that function in PAn. Ross calls them 'first-generation' affixes. The set consisting of \*-en, \*-an, \*Si- and \*<in> also existed in PAn, but according to him, only as patient, location, conveyance and perfective *nominalizers* respectively: he claims that they acquired voice-marking functions only in proto-Nuclear Austronesian, displacing the first-generation affixes in all but their irrealis functions. He calls them 'second-generation' affixes.

Ross gives two reasons for thinking that the state of affairs represented by Teng's Puyuma must be original: first, if it were not, "we would expect Puyuma to preserve some reflex of the alleged intervening PAn stage, but it doesn't". This means that if Puyuma was descended from a language where \*-en, \*-an, \*Si- and \*<in> had served as UV markers and not just as nominalizers, Puyuma should show traces of that usage; and second, one would have to assume that "PAn (first-generation) undergoer-voice optative/hortative forms have extended

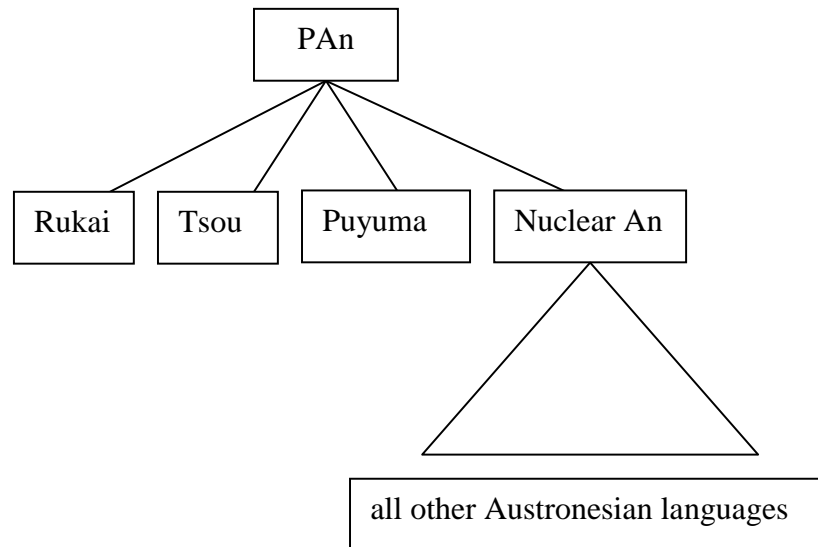
---

<sup>1</sup> Special thanks go to Sander Adelaar, Thomas Pellard, Lawrence Reid, Malcolm Ross, John Wolff and Elizabeth Zeitoun for useful discussion and information; and above all to Josiane Cauquelin for sharing with me her knowledge of Puyuma and making her notes available to me. I retain full responsibility for the final product.

<sup>2</sup> In punning reference to the Roman senate.

their function in Puyuma to include the realis, displacing the PAn second-generation forms—a step which seems quite implausible".

**FIGURE 1. AUSTRONESIAN PHYLOGENY IN ROSS (2009)**



Ross's reason for thinking Rukai and Tsou are primary branches of An is that, like Puyuma, their verbal morphology does not make use of the undergoer-voice markers \*-en, \*-an and \*Si- (or \*Sa-), or of the perfective marker \*<in>. However, alternative explanations for the absence of these affixes are available. In Tsou, auxiliary verbs expressing focus and aspect have become obligatory. Virtually every sentence has them: the lexical verbs in each sentence are the auxiliaries' dependents. Their voice markers: UVP -a, UVL -i, UVC -(n)eni probably belong to early Tsou *dependent*, rather than independent, verb morphology. Independent verb morphology, including the neutral UV markers, has disappeared outside of the auxiliaries. Even the auxiliaries do not take overt UV markers: an auxiliary verb belongs to undergoer voice by default, and to actor voice when it carries the AV prefix *m-*. As to the perfective marker \*<in>, it has been made redundant by the fact that the auxiliaries have taken over the function of indicating aspect.

For Rukai, Zeitoun and Teng (in press) demonstrate that passive constructions using verbs with *ki-* have taken over the functions of the old undergoer-voice constructions: competition from, and later replacement by, *ki-*passives, provides sufficient explanation for the disappearance of verb forms marked by \*-en, \*-an and \*Si- (or \*Sa-). The reasons for the disappearance of \*<in> as perfective marker may be independent. Mantauran Rukai marks perfective aspect by a suffix: *-nga* (Zeitoun 2007:157). Ross's theory does not throw any

supplementary light on the history of voice marking in Tsou or Rukai: it simply assigns to PAn certain characters which have ready explanations as post-PAn innovations. The evidence from Puyuma is therefore central to Ross's theory.

Before examining whether Puyuma conforms to Ross's account, a couple of preliminary observations are in order. First, Ross's phylogeny contains a single internal node which he regards as the site of a complex innovative grammatical event. Successful phylogenies combine a number of *independent* innovations into a single tree. Here it must be pointed out that although the defining innovation of Ross's Nuclear Austronesian branch is complex, its components —reinterpretation of cleft sentences as verb-initial; of genitive pronouns as agentive; of nominalizers as voice markers— are crucially dependent upon one another: the SPQR mechanism can only have taken effect as a single reanalysis event, which must have occurred only once, among one group of speakers. In no way can the SPQR mechanism be regarded as a collection of independent innovations. Ross points out that his phylogeny is by and large compatible with the phonological innovations in Blust (1999); yet, because Blust's phylogeny is a ten-branch star with little higher-level structure, this point remains on the whole rather trivial. Apparently missing are compatible higher-level innovations in diffusion-resistant areas like morphology and the basic vocabulary. Indeed, Ross's phylogeny clashes with several independently established innovations in these two areas: it cannot accommodate the nodes which are needed to express Tsuchida's Tsouic innovations \*ramuCu 'hand' and \*cani 'one' (Tsuchida 1976),<sup>3</sup> or the innovations proposed in Sagart (2004) for the numerals \*pitu 'seven', \*walu 'eight' and \*Siwa 'nine'; nor can it accommodate nodes for morphological innovations like the elimination of \*-en in UVP perfective verb forms, or the extension to verb roots of \*ki- prefixation (both innovations are discussed in Sagart, 2009) or, as we shall see below, for verbs of the shape \*paR-Numeral-en.

Even more problematic are facts coming from Puyuma itself: I will show in section 1 that \*<in> is still alive as a perfective aspect marker on verbs in another sociolect of Nanwang Puyuma, and that certain Puyuma verbs contain fossilized traces of the UVP marker \*-en. In section 2, I will outline a hypothetical scenario for the replacement of \*-en, \*-an and \*Si- (or \*Sa-) as independent UV markers by a series of markers which had until then served for

---

<sup>3</sup> Ross suggests that Tsouic-only forms such as these could be retentions: if so, the widely represented \*lima 'hand' and \*isa 'one' (reflected in Puyuma by *Lima* 'hand' and *sa* 'one') must be post-PAn innovations displacing \*ramuCu and \*cani. Since however Ross's tree has Puyuma branching off directly from the root node, Puyuma should reflect \*ramuCu and \*cani, not \*lima and \*isa.

future or irrealis: \*-aw, \*-ay and \*-anay.

## 1. THE FATE OF THE OLD UNDERGOER VOICE MARKERS IN PUYUMA.

1.1. **The survival of \*<in>.** Since Wolff (1973) it has been assumed that \*<in> was a PAn perfective marker which could co-occur on verbs simultaneously with the markers of at least three of the four voice categories, as well as derive nouns out of verbs. Ross (2009) restricts PAn \*<in> to the derivation of nominals, and perhaps of verbs in Actor Voice. The basis for Ross's claim about \*<in> can be found in Teng (2008:130). She states that Puyuma words with <in> are nominals because they "cannot take a nominative pronoun enclitic like verbs do". Thus she takes a sentence like *ku=tr<in>ima na tilril* ("this is the book I bought") to be a cleft sentence with two noun phrases: *ku=tr<in>ima* 'my buying' (where *ku=* 'my' is a genitive pronoun proclitic) and *na tilril* 'the book', literally "this book is my buying".

Josiane Cauquelin has worked on the same Nanwang dialect of Puyuma since the late seventies. She has these observations about the division of the village into two main social networks and the sociolects corresponding to these networks:

Christianisation of the Nanwang Puyuma village began soon after World War II. All ethnic Puyuma villages have a Catholic and a Protestant Church (...). What strikes one on arrival in Nanwang is the division of the village between the two churches, *which has followed the original dual structure of the village* (italics mine, LS). (...) It is not possible to speak of rivalry between the two parishes but rather between these and the shamans. (...) Minor differences in language usage occasionally distinguish Protestants and Catholics. At one point I questioned elderly Puyuma about the archaic, compound personal pronoun *kunu-* 'I (do to) you,' as in *ku-nu-beray-ay* 'I give you'; the usual form in Nanwang Puyuma is *ku-beray-ay kanu*. Protestants and unconverted Puyuma asserted unhesitatingly that this compound pronoun does not exist in Nanwang Puyuma, though they acknowledged its occurrence in other Puyuma dialects. Catholics, on the other hand, unanimously maintained that it exists in Nanwang Puyuma. (...) (Cauquelin, to appear).

Compound proclitics are certainly ancient in Puyuma, since Tsuchida (1980:200) finds them in Tamalakaw Puyuma, as Cauquelin has pointed out to me (p.c., Nov. 2009). Teng's grammar allows only noncompound proclitic pronouns. Her principal informants are Protestants —although she has been careful to collect texts from Catholics and unconverted persons as well as Protestants, and her grammar is based on generalizations extracted from those texts. Her interactions with Puyuma speakers are conducted in Chinese or through an interpreter. In contrast, Cauquelin's informants are non-Protestants and her interactions with





3P.GEN-do DEF.OBL head-hunting

"I have heard from the old Puyuma men their way of doing head-hunting"

(6)

*k<in>ibuLas-an=ku* *rumaq*

<PERF>borrow-UVC=1S.NOM house

"I have borrowed the house"

Similar forms occur in the ritual texts published by Cauquelin (2008): *p<in>utia=ku...* 'I have dreamt ...' (#15-13); *p<in>ar-ki-anu-anun=ku...* 'I have already said the prayers...' (#16-34). The words with <in> in these sentences take nominative pronoun enclitics: by Teng's criterion they must be verbs. Nanwang Puyuma thus has perfective verb forms marked with <in>, at least in UVP (sentences 1 through 3) and UVC (sentences 4-6). Table 1 shows the UV markers in Cauquelin's Nanwang.

	UVP	UVL	UVC
Neutral	V-aw	V-ay	V-anay
Perfective	<in>V	?	<in>V-an

Table 1: UV markers in the Nanwang Puyuma sociolect studied by J. Cauquelin (simplified)

In Sagart (2009) I argued that in PAN, UVP perfective verb forms had free variation between \*<in>V-en and \*<in>V. I showed that no later than proto-Walu-Siwaish, the simpler variant won out and \*<in>V-en was eliminated. According to my phylogeny (FIGURE 2), Puyuma, a Walu-Siwaish language, should have undergone loss of \*-en in perfective UVP verb forms. As Table 1 shows, the UVP perfective form of the verb in Puyuma: <in>V, is what my phylogeny leads me to expect. The other Puluqish languages (Paiwan, Amis, PMP) also show \*<in>V for UVP perfective, confirming that UVP perfective was \*<in>V in proto-Puluqish.

**1.2. Traces of \*-en.** The clearest traces of the UVP marker \*-en in Nanwang Puyuma are found in a set of verbs of the shape \*paR-Numeral-en, meaning 'do N times'. An example is *parpuan* 'do twice' (Cauquelin 1991) which reflects \*paR-puSa-en. Its stem is \*puSa, the bound morpheme for 'two' also found, often in collocation with suffixed \*-N 'times', in PAN

\*ma-puSa-N 'twenty' and other forms like Thao *pushaz* 'two, in certain collocations' (Blust 2003), Siraya *-pohal* 'double' (Adelaar, unpublished materials), Seediq *mutupussal* 'twice' (Asai 1953:53) and Mantauran Rukai *po'a* 'two, in circumfixes' (Zeitoun 2007:255).<sup>4</sup> The observation that \*puSa must have been the bound counterpart of \*duSa 'two' in PAn belongs to Zeitoun (2009). *Parpuan* is prefixed with *par-*, a verbalizer of nouns: compare *par-bua* 'to create' (*bua* 'fruit'), *par-isiq* 'be incontinent' (*isiq* 'urine'), *par-baqaw* 'have an erection' (*baqaw* 'life'). Puyuma *-n* reflects PAn \*-n unambiguously, so that *parpuan* cannot be from \*paR-puSa-N.<sup>5</sup> It cannot be from \*paR-puSa-an either: that would give \*parpuayan in Nanwang, with *-y-* insertion. Josiane Cauquelin (p.c., July 29-August 2, 2009) gives these examples:

(7)

<i>ku=parpuan-ay</i>	<i>ma-reṅay</i>
1S.GEN=do twice-UVL	AV-speak

"I speak twice."

(8)

<i>p&lt;en&gt;arpuan=ku</i>	<i>s&lt;em&gt;enay</i>
do twice-AV=1S.NOM	AV-sing

"I sing twice."

Although it contains a fossilized voice suffix, synchronically *parpuan* is an unanalyzeable verb root which can occur as a head verb in (at least) AV and UVL once it is made to carry the requisite markers.<sup>6</sup> The old UVP marker \*-en has become part of the root. Nanwang Puyuma also has *parteLun* 'do three times' (Cauquelin 1991:40 and p.c., Aug. 2, 2009), reflecting \*paR-telu-en. As with *parpuan*, in *parteLun* \*-en has become part of the root. Here is an example in UVC (Cauquelin, p.c. August 2, 2009):

(9)

<i>ku=parteLun-anay</i>	<i>s&lt;em&gt;ilud</i>	<i>Da</i>	<i>enay</i>	<i>na</i>	<i>laqub</i>
1S.GEN=do three times-UVC	<AV>draw	OBL.INDEF	water	NOM	ladle

<sup>4</sup> "The distinction between *-dho'a* and *-po'a* 'two' lies in the type of affixes they take: *-dho'a* is only preceded by prefixes, while *-po'a* co-occurs with circumfixes (...)"

<sup>5</sup> Nanwang Puyuma (Cauquelin) has an example of the \*-N suffix meaning 'times' discussed above, in *pariasal* 'do one time' < \*paR-i-asa-N. Puyuma verbs meaning 'do N times' are constructed with *-n* (< PAn \*-en) when the number is two or more. The verb for 'do once' varies between *-n* and *-N* across dialects. The *-n* forms are probably the result of leveling.

<sup>6</sup> Cauquelin (p.c., July 29 and Aug. 1, 2009) indicates that her informant rejects UVP \*parpuan-aw.

"I draw water three times with the ladle."

While Nanwang Puyuma has only *parpuan* and *parteLun*, more extensive paradigms of these verbs can be found in other Puyuma dialects: Katipul (Zeng 1997:154) has *par-asa-n* 'do once', *par-puwa-n* 'do twice', *par-tulu-n* 'do three times', *par-pat-en* 'do four times', *par-nem-en* 'do six times', *par-pitu-n* 'do seven times', *par-walu-n* 'do eight times'. Tamalakaw (Tsuchida 1980:287-288) has *paR-asan* 'do once', *paR-puwan* 'do twice', *paR-terun* 'do three times', *paR-epat-en* 'do four times', *paR-nem-en* 'do six times', *paR-warun-n* 'do eight times'. In the Nanwang sociolect investigated by Teng, only *partelrun* (in her notation) apparently occurs. Commenting on an example involving *partelrun* she wrote: "it is not clear why in this case *telru* becomes *telrun*" (Teng 2008:75 fn. 17).

Evidently the Puyuma facts by themselves are not sufficient to show that the *-(e)n* endings in these verbs reflect the old UVP marker *\*-en*. What demonstrate the UVP origin of these endings are facts from Philippine languages. The Puyuma verbs just cited have cognates in Cebuano, Samar-Leyte, Tagalog and other Philippine languages: Tagalog *pag-isah-in* 'combine into one', Bikol *pag-apa't-on* 'divide into four; send four at a time', *pag-ano'm-on* 'divide into six; send six at a time; go six by six', *pag-walo'-on* 'divide into eight; send eight at a time'. Based on the agreement between Tamalakaw Puyuma and Philippine languages, Blust (n.d.) reconstructed PAN forms with *\*-en* suffix for each of them. Semantic shift from 'do X times' to 'divide/combine into X' is straightforward ('do X times' = 'divide action into X segments'). Crucially, these Philippine verbs are *synchronically* UVP forms where Tagalog *-in* and Bikol *-on* are the expected outcomes of the UVP marker, PAN *\*-en*. Constructions involving them are normal UVP construction with a definite patient subject and a genitive agent, as in the following Tagalog example (English 1986:398):

(10)

<i>pag-dalawah-ín=mo</i>	<i>ng</i>	<i>grúpo</i>	<i>ang</i>	<i>mangá</i>	<i>báta?</i>
Verbalizer-two-UVP=2SG.GEN	OBL	groups	NOM	PL	child

"(You) divide the children into two groups."

It would be difficult to argue that these verb forms are really nominalizations, whether they are in Puyuma or in Philippine languages. It is clear, therefore, that the common ancestor of Puyuma, Tagalog, Bikol etc. had these verbs in UVP form. There can be no doubt that

Puyuma forms like *parpuan* contain a fossilized UVP suffix \*-en.

Less conspicuous traces of PAN \*-en as a voice marker are found incorporated as fossils in the final -VC of certain isolated Puyuma verb roots, where they are without any synchronic morphological function. Here is an example, first presented in my (2009). Several languages of the northern Philippines: Isneg, Agta, Casiguran Dumagat<sup>7</sup> etc. have a verb reflecting \*(q)unik 'to climb'. Nanwang Puyuma (Cauquelin 1991) has (*q*)*unkun* 'to jump over, to climb',<sup>8</sup> a verb eligible for (at least) agent and non-agent voice marking: *m-unkun* (AV), *unkun-ay* (UVP), *unkun-aw* (UVL).<sup>9</sup> Examples from Cauquelin's dictionary:

(11)

*ku=unkun-aw*                      *na*      *guŋ*

1S.GEN=jump over-UVP      NOM    ox

"I jump over the ox."

(12)

*tu=unkun-ay=ku*                      *Da*      *suan*

3S.GEN=jump over-UVL=1S.NOM AGT    dog

"The dog jumps over me."

Although Cauquelin's dictionary only glosses *unkun* as 'to jump over', it also means 'to climb' (up a ladder), like in the Philippine forms. Cauquelin (p.c., July 29, 2009) gives this example sentence:

(13)

*m-unkun=ku*                      *piaisaT*                      *kana*    *raripaqaŋ*

AV-climb=1S.NOM    upwards                      OBL    ladder

"I climb up the ladder."

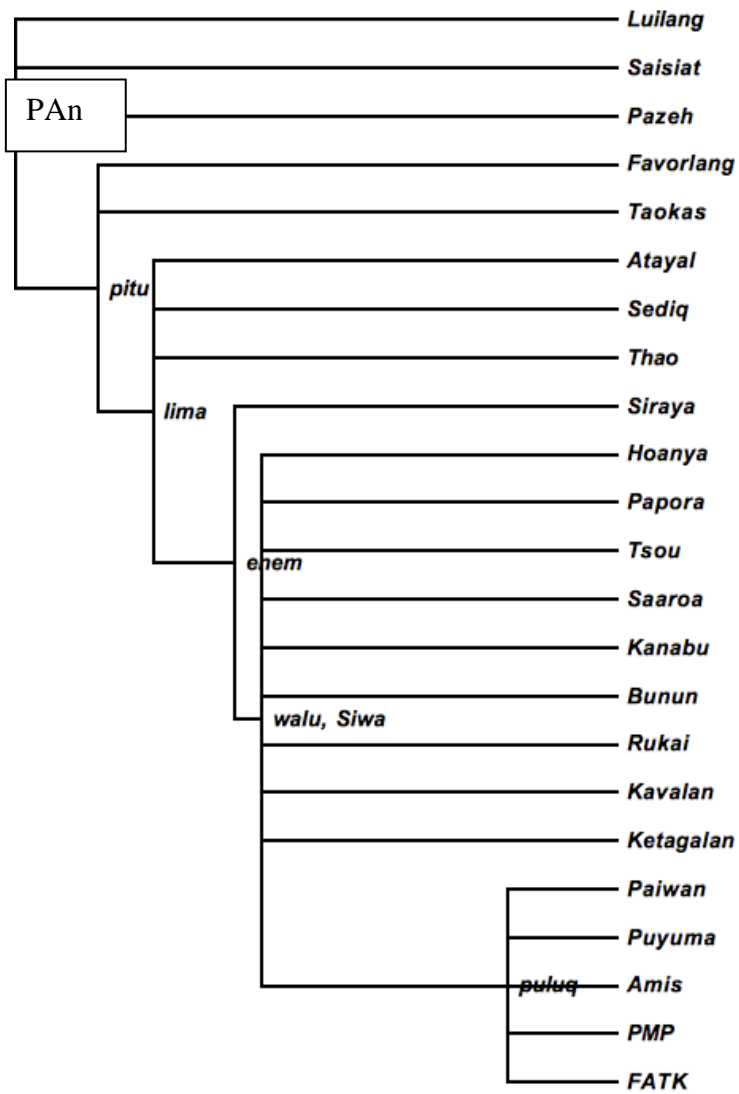
---

<sup>7</sup> Isneg *ʔ<um>uneʔ*, Casiguran Dumagat *ʔunek* 'climb up a tree', Agta *ʔ<im>unek* (Reid 1971). The last vowels in those forms reflect \*i (Lawrence Reid, p.c., June 2009).

<sup>8</sup> Listed by Cauquelin under *unkun*.

<sup>9</sup> Nanwang, the dialect of Puyuma investigated by Cauquelin, preserves \*q word-internally: \*q- has been preserved in *mu-qunkun* thanks to the *mu-* prefix.

**FIGURE 2. EARLY AN PHYLOGENY<sup>†</sup>**



<sup>†</sup> All nodes are based on innovations in the numerals 5-10: \*pitu 'seven', \*lima 'five', \*enem 'six', \*walu 'eight', \*Siwa 'nine', \*puluq 'ten'. Note: this tree is based on Sagart (2004) with the modification that the Muish node has been abandoned and replaced with a Puluqish node (for explanations why, see Sagart 2008).

Under the phylogeny in FIGURE 2, Philippine languages are within Puluqish, like Puyuma. Although they are not spoken in Taiwan, phylogenetically they are as closely related to Puyuma as are Amis and Paiwan, and more closely related to Puyuma than are Rukai, Tsou, Bunun or Kavalan. Puyuma *unkun* contains the string *-nk-*, two heterorganic consonants between which a vowel was evidently lost. *Unkun* can derive from \**qunik-en*, the UVP form of a hypothetical proto-Puluqish \**qunik*<sup>10</sup> 'to climb'. In section 2, I will suggest that after Puyuma had individualized out of proto-Puluqish, the inherited neutral Undergoer Voice markers \*-en, \*-an and \*Si- (or \*Sa-) were eliminated. When this happened certain UVP verbs with \*-en were reanalyzed so that \*-en became part of the root: thus \**qunik-en* became \**quniken*, a trisyllabic verb root compatible with all voice constructions and affixes. The presumably unstressed second syllable of \**quniken* underwent vowel syncope to \**qunken*, and eventually to (*q*)*unkun*, the modern form (schwas in Puyuma tend to be reflected as *u* if there is a /u/ earlier in the word).

Another likely example of fossilized \*-en is *mu-LuDun* 'to sink'. This can be compared to Blust's WMP \**qeled* 'sink' (Blust, n.d.), reconstructed on the basis of Philippine forms like Isneg *allad* 'sink', Dumagat *eled* 'sink', Maranao *led* 'drown', Tiruray *eled* 'sink' plus Old Javanese and Balinese words meaning 'to swallow'. Blust treats \**qeled* as containing a root \*/-led/ 'sink'. A hypothetical proto-Puluqish \**qeled* 'sink' would be \**qeled-en* in UVP form; incorporation of \*-en would yield a new verb root \**qeleden*; reduction to a disyllable, effected through loss of the first syllable rather than vowel syncope (perhaps because *-LD-*, the expected Puyuma reflex of *-ld-*, is not a possible cluster), would result in Puyuma *LeDen* 'to sink'; prefixation of *mu-* (characterized by Teng as anticausative) would trigger rounding of schwas to *mu-LuDun*, the modern form.

I have not identified secure traces of the UVL and UVC markers \*-an and \*Si- (or \*Sa-) in Puyuma. These would presumably consist of verb roots having incorporated these affixes (whose reflexes should be *-an*, and *i-* or *a-*) as fossils. In the case of \*-an, another voice marker — the UVC marker for negative and irrealis sentences (Tsuchida 1980, Teng 2008)— is synchronically alive and has exactly the desired reflex: *-an*. The difficulty of distinguishing

---

<sup>10</sup> This form may have been more complex, perhaps \**qunahik*, to take into account Tagalog *pa-nhik* 'climb' and Dempwolff's \**na'ik* 'go up, climb' (based on Toba Batak, Malay, Javanese and Ngadju Dayak). Puyuma, Casiguran Dumagat, Isneg, Toba Batak, Malay, Javanese and Ngadju Dayak all reflect PAN \**h* as zero, while Tagalog reflects it as /h/: a proto-Puluqish \**qunahik* would produce the right outcomes for this consonant. Explanations would have to be found for loss of \**qu-* and *-a-* in some languages.

UVL and UVC markers purely from lexical entries, without reference to the sentences in which they once occurred, has made it impossible to reach a conclusion regarding the presence or absence of the UVL marker \*-an at an earlier stage of Puyuma. A similar difficulty arises in UVC: Puyuma has several *i*- and *a*- prefixes which could potentially reflect \*Si- or \*Sa-.<sup>11</sup> What is certain is that a language ancestral to Puyuma had both UVP markers \*-en (neutral) and \*<in> (perfective): this is sufficient to falsify Ross's theory.

**2. WHAT HAPPENED TO PUYUMA ?** What, then, is the explanation for the absence of \*-en, \*-an, \*Si- (or \*Sa-) and \*<in> in Puyuma verbs ? why do we find instead a series of undergoer-voice markers: UVP \*-aw, UVL \*-ay, UVC \*-anay, which (at least for the first two of them) clearly denote the future or irrealis in other Formosan languages ? the only possible explanation is that one series of markers has displaced the other. The implausibility Ross speaks of comes from having a more marked set of independent verb forms (irrealis/future) replacing a less marked set (realis/neutral). Yet one admittedly hypothetical scenario leading to this very result does not look too improbable. It is commonplace for languages to replace simple verb forms with complex forms consisting of an auxiliary and the lexical verb as the auxiliary's dependent. In Taiwan, Tsou has followed that path, as we have seen. Tsou auxiliaries are obligatory and sentence-initial; pronouns for subject of AV (nominative agent) and agent of UV ("agentive" in Tsuchida 1976) attach to them as enclitics. The lexical verb follows, and any other pronouns occur as free forms after the verb. Should the auxiliaries become lost, the clitic pronouns would presumably reattach on their right as *proclitics* to the lexical verb.

Unlike Tsou, Puyuma does not have obligatory auxiliaries, but it does have proclitic agent-of-UV pronouns (genitive, as in a majority of Formosan and Philippine languages, for reasons well explained by Starosta, Pawley and Reid 1982): this is an intriguing suggestion that Puyuma may have had obligatory auxiliary verbs, now lost, in UV sentences. If it did, the

---

<sup>11</sup> One candidate reflects \*Sa-: it occurs in forms like *a-iray-an* 'a mill' (*iray* 'to grind'), *a-imey-an* 'a cell, detention room' (*imey-u* 'to keep in custody') in Tamalakaw Puyuma (Tsuchida 1980:239, 251) and *a-kan-an* 'food' (*kan* 'eat') in Nanwang Puyuma (Cauquelin 1991), which I interpret as reflecting \*Sa-verb-an, where -an is the agent for nominalization. Confirmation that *a*- in these forms reflects \*Sa- comes from Amis and Kavalan nouns with the same structure. Amis has *sa-holo-an* 'kitchen' (*holo* 'to cook'), *sa-faco-an* 'design' (*faco* 'make plans'), *sa-ati-an* 'small wine cup' (*ati* 'to welcome') (Pourrias and Poinot 1996). In Kavalan we find *sa-rumuz-an* 'steamer' (*rumuz* 'to steam'), *sa-Riqutu-an* 'comb for catching lice' (*Riqutu* 'catch lice'), *sa-Ramaz-an* 'fuel' (*Ramaz* 'to cook rice'), *sa-puysiw-an* 'food taken with liquor' (*puysiw* 'eat a snack while drinking') (Li and Tsuchida 2006). Amis *sa*- and Kavalan *sa*- both reflect PAN \*Sa-, Amis unambiguously. These forms could have their origin in UVC verbs with \*Sa-, nominalized by \*-an. The unresolved duality of \*Si- and \*Sa- and remaining questions about the derivational history of these forms make it difficult to conclude with certainty that this \*Sa- was a voice marker, however.

modern Puyuma neutral UV forms are the former dependent verbs in these constructions, and the UV markers they carry: \*-aw, \*-ay and \*-anay, do not continue the PAn independent-verb UV markers \*-en, \*-an and \*Si- (or \*Sa-). Those were eliminated when auxiliaries became obligatory in UV constructions. Rather, they are special undergoer voice markers for dependent verbs of the auxiliaries.

To understand why a future/irrealis series of UV markers may have been recruited to mark undergoer voice in dependent verbs, we would need to know out of what verbs the auxiliaries were grammaticalized. Unfortunately the auxiliaries, if they existed, are lost and we may never know the answer to that question. At any rate, promotion of the future/irrealis voice marker series would have been facilitated thanks to an advantage they had over the neutral set: they were just UV markers, while those in the neutral set were both UV markers and nominalizers. By generalizing the future/irrealis set as dependent UV markers, and by reducing the old neutral set to its nominalizing function, it was possible to make the nature (nominal or verbal) of words containing verbal roots more transparent, this facilitating the parsing of UV sentences. One notes that by eliminating the neutral \*-en, \*-an and \*Si- (or \*Sa-) series from their verbal system, while retaining them as nominalizers, Tsou and Rukai have achieved the same result as Puyuma. This probably indicates that pressure to decouple voice marking and nominalization marking existed in proto-Walu-Siwaish, the common ancestor of these three languages. Puyuma, Rukai and Tsou are the three languages singled out by Ross to be primary An branches in his phylogeny: under the present view, Puyuma, Rukai and Tsou have independently implemented different solutions to the same functional problem.

**Conclusion.** Although the voice-marking affixes of Puyuma are atypical in a Formosan context, modern Puyuma is in most other respects a normal Puluqish language. Lexically, it has innovated all of \*lima 'five', \*enem 'six', \*pitu 'seven', \*walu 'eight', \*Siwa 'nine' and \*puluq 'ten'; it has escaped the Tsouic innovations \*ramuCu 'hand' and \*cani 'one'; morphologically it has lost \*-en in UVP perfective verbs, extended prefixation of \*ki- to verb roots and innovated verbs meaning 'do N times' from numerals and the \*paR...en circumfix, an innovation it shares exclusively with MP. In these respects, Puyuma aligns with its Puluqish sisters, including PMP, rather than with the languages of the west coast of Taiwan. Far from being one of the first Formosan languages to branch off, Puyuma was one of the last to do so.



Finally, although I disagree with Ross on the time at which the SPQR mechanism took effect, and on which languages were affected by it, I concur with him in thinking that it remains the most attractive explanation of the formation of the early Formosan-Philippine "focus" system. Under the phylogeny in FIGURE 2, the first An languages to branch off are Pazeh and Saisiat. Both possess a distinction between Agent and Undergoer voice constructions, and among the latter between the patient subject, location subject and conveyance subject subtypes. This shows that the full "focus" system was already in place in PAn: if the SPQR mechanism occurred at all, it must have done so before PAn.

### REFERENCES

- Asai, E. 1953. *The Sedik language of Formosa*. Kanazawa: Cercle Linguistique de Kanazawa.
- Baxter, W. H. III. 1996. *Old Chinese origins of Middle Chinese palatal sy-: evidence from Min dialects*. 29th Conference on Sino-Tibetan Languages and Linguistics, Leiden, 10-13 October 1996.
- Blust, Robert. 1999. Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. In: Elizabeth Zeitoun and Paul Jen-kuei Li (eds.) *Selected Papers from the Eighth International Conference on Austronesian linguistics*, 31-94. Taipei: Institute of Linguistics (preparatory office).
- 2003. *Thao dictionary*. Language and Linguistics Monograph Series, No. A5. Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.
- n.d. Austronesian Comparative dictionary. Manuscript in progress quoted with the author's permission.
- Cauquelin, J. 1991. *Dictionnaire puyuma-français*. Paris, Jakarta: École Française d'Extrême-Orient.
- 2008. *Ritual texts of the last traditional practitioners of Nanwang Puyuma*. Language and Linguistics monograph series A23. Nankang: Institute of Linguistics, Academia Sinica.
- to appear. From one religion, the other 'old wine into new wineskins'.
- English, Fr. Leo, 1986. *Tagalog-English Dictionary*. Manila: Congregation of the Most Holy Redeemer.
- Li, P. J-K, and S. Tsuchida. 2006. *Kavalan dictionary*. Language and Linguistics monograph series A19. Taipei: Institute of Linguistics (preparatory office), Academia Sinica.
- Pourrias, Louis and Maurice Poinsoot. 1996. *Dictionnaire Amis-français*. Privately published.

- Reid, Lawrence A. 1971. *Philippine minor languages: Word lists and phonologies*. Oceanic Linguistics Special Publication, 8. Honolulu: University of Hawaii Press.
- Ross, Malcolm D. 2009. Proto Austronesian verbal morphology: a reappraisal. In A. Adelaar and A.K. Pawley (eds): *Austronesian historical linguistics and culture history. A festschrift for Robert Blust*, 285-316. Canberra: Pacific Linguistics.
- Sagart, L. 2004. The higher phylogeny of Austronesian and the position of Tai-Kadai. *Oceanic Linguistics* 43, 2: 411-444.
- 2008. The expansion of Setaria farmers in East Asia: a linguistic and archaeological model. In Sanchez-Mazas A, Blench R, Ross M, Peiros I, Lin M, eds. *Past human migrations in East Asia: matching archaeology, linguistics and genetics*, 133-157. Routledge Studies in the Early History of Asia, London: Routledge.
- 2009. PAn morphology in phylogenetic perspective. Paper presented at the Panel on Reconstruction of PAn morpho- syntax and implications for the An settlement on Taiwan. *11th International Conference on Austronesian Languages*. June 22-25, Aussois, France.
- Starosta, S., A. Pawley and L. Reid. 1982. The evolution of Focus in Austronesian. In Halim, Carrington and Wurm (eds.) *Papers from the Third International Conference on Austronesian Linguistics, vol. 2*, 145-70. Pacific Linguistics, C-75 (1982).
- Teng, Stacy Fan-ching. 2008. *A reference grammar of Puyuma, an Austronesian language of Taiwan*. Pacific Linguistics 595. Canberra: the Australian National University.
- Tsuchida, S. 1976. *Reconstruction of Proto-Tsouic phonology*. Study of Languages of Asia and Africa monograph series N°5. Tokyo: Institute for the Study of Language of Asia and Africa.
- 1980. Puyuma (Tamalakaw dialect) vocabulary with grammatical notes and texts. *Institute for the study of languages and cultures of Asia and Africa*, Tokyo University of foreign studies.
- Wolff, John U. 1973. verbal inflection in Proto-Austronesian. in Andrew Gonzales. (ed.) *Parangay kay Cecilio Lopez*, 71-91. Quezon city: Linguistic Society of the Philippines.
- Zeitoun, E. 2007. *A grammar of Mantauran (Rukai)*. Language and Linguistics monograph series A4-2. Nankang: Institute of Linguistics, Academia Sinica.
- 2009. Bound and free numeral forms in Formosan languages. Paper presented at 11ICAL, June 22-25, Aussois, France.

Zeitoun, E. and S. Teng. in press. From ki-N ‘get N’ in Formosan languages to ki-V ‘get V-ed’ (passive) in Rukai, Paiwan and Puyuma.

Zeng Jianci 曾建次. 1997. *Beinan zu muyu huilu* 卑南族母語彙錄. [A record of the vocabulary of the Puyuma language]. Hualien: Caituan faren tianzhujiao hui Hualian jiaoqu 財團法人天主教會花蓮教區.