PAN MORPHOLOGY IN PHYLOGENETIC PERSPECTIVE

Laurent Sagart

▶ **To cite this version:**

Laurent Sagart. PAN MORPHOLOGY IN PHYLOGENETIC PERSPECTIVE. 11ICAL (11ème conférence internationale sur la linguistique austronésienne), Jun 2009, Aussois, France. hal-00409999

HAL Id: hal-00409999
https://hal.science/hal-00409999

Submitted on 14 Aug 2009

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.

PAN MORPHOLOGY IN PHYLOGENETIC PERSPECTIVE[1]

Laurent Sagart

EHESS/INALCO/CNRS

Paris

0. Background

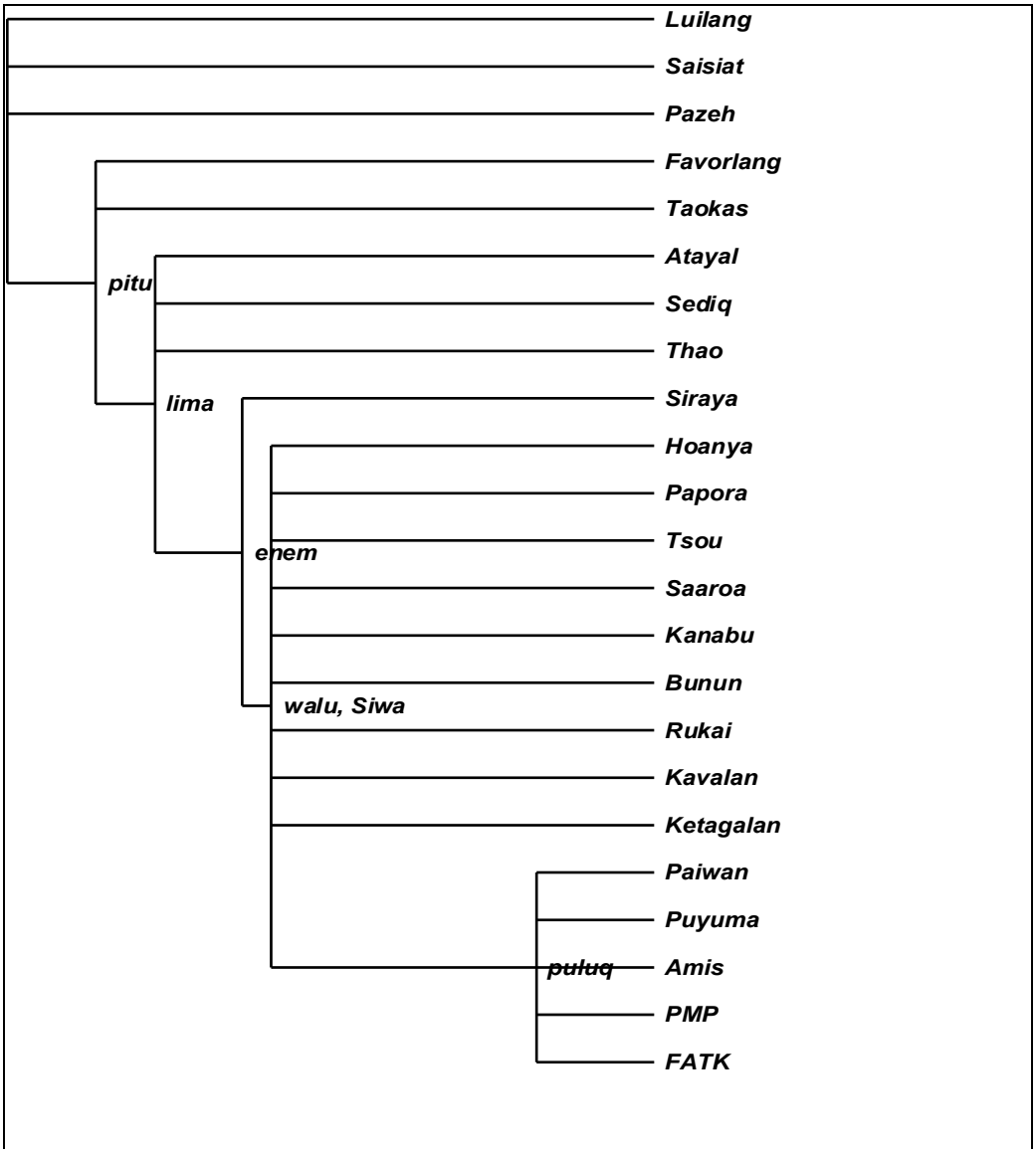
A recurrent problem when building phylogenetic trees from sound changes is that while they are relatively easy to identify[2] and their directionality is usually clear, the objects these changes take for their targets —sounds— are not highly language-specific: neighboring languages tend to have much overlap in their sound inventories, so that when a sound change reaches a language boundary, the risk is high that it will cross it, provided that there are bilinguals and that the target sound exists on the other side. For this reason, shared phonological innovations often do not reflect shared inheritance. Blust's sound-change-based phylogeny for Formosan languages (1999) assumes a ten-branch star shape because the phylogenetic signal is hard to extract from the phonological isoglosses of a group of related languages in contact.

Lexical innovations take words as their targets. As associations of a string of phonemes and a meaning, words are far more language-specific than sounds. A lexical innovation affects the meaning of a particular word. When a lexical innovation reaches a sharp language boundary, the target word will usually not be present on the other side: the change cannot cross the boundary, even if there are bilinguals. True, a word having undergone a semantic change can be borrowed by a neighboring language, giving the appearance of lexical change in the borrowing language. The confounding effect of borrowing can be minimized by excluding cultural vocabulary from consideration.

In my (2004), based on a pattern of phylogenetic compatibility among six innovative characters in the An numerals from 5 to 10, I presented a new higher phylogeny of An. This was slightly modified in my communication to 10ICAL and in my (2008). I reproduce the tree in Figure 1. The story told by this tree is that of a founder group crossing from northern Fujian, where the straits are narrowest and a mountain top in NW Taiwan can be seen with the naked eye: expanding from their first settlements in NW Taiwan southward along the west coast, reaching southern Taiwan and continuing their expansion along the east coast: from where a second founder group left to establish Austronesian-speaking colonies ancestral to PMP and Tai-Kadai in the Philippines and on the south China coast.

[1] Thanks go to John Wolff and Lawrence Reid for useful discussion and information.

[2] All phonological mergers are innovations.



In this paper I begin to examine the contribution of morphological innovations to early Austronesian (Formosan) phylogeny, using the same compatibility approach as in my 2004: I will discuss three nested characters, each of which is clearly innovative. The Austronesian languages which show these characters form are distributed in three nested isoglosses, forming a *clique* (sensu Meacham and Estabrook 1985), like the six numerals in Sagart 2004. When a node dominates another node in a phylogenetic tree, the innovations which define those nodes should show a nesting structure in geographical space.

1. Three early Austronesian morphological innovations.

In terms of their readiness to spread across language boundaries, morphological changes are probably intermediate between sound changes and lexical changes: on the one hand they are associations of a string of phonemes and a morphological meaning, making them more language-specific than sounds; on the other they are typically shorter than words, making them more like phonemes. Moreover, at least when closely related languages are in geographical contact, affixal inventories are likely to be shared to a significant extent: this

increases the risk that a morphological change spreading from across the boundary will find its target on this side, and cross. At the same time morphological changes are reputed to be relatively resistant to borrowing.[3] In this, morphological changes are like changes affecting the *basic* vocabulary.

Starosta (1994, 1995) accepted that phylogenies should be built from innovations. He had the insight that because of their resistance to borrowing, morphological innovations are choice phylogenetic material. He built a phylogeny for Formosan languages based exclusively on morphological arguments. He accepted that Taiwan is the Austronesian homeland: he also accepted that MP subgroups with Amis (Harvey; Reid). He regarded Philippine morphosyntax as highly elaborate and assumed the history of morphosyntax between PAn and the Philippine subgroup to have been one of gradual complexification. He insisted that if Rukai, the Formosan language with the simplest morphosyntax, had lost morphological processes found in other languages of Taiwan or the Philippines, lexicalized traces of these processes should be found in Rukai.

The directionality of morphological change is not easy to establish on general principles. Starosta took as his guiding idea the view that in the absence of lexicalized vestiges, the Formosan language with the simplest morphology —Rukai— must be the first to branch off. My approach is different. I accept that morphological complexity can decrease over time in a language, leaving only faint traces in the lexicon. Arguments from 'absence of vestiges' are a variant of *argumenta e silentio* (arguments from silence) and must be used with caution: absence of evidence is not evidence of absence. My guiding principle is the idea, established in my 2004, that the PAn homeland was in NW Taiwan and that the first languages to branch off were Pazeh and Saisiat (whether as a single subgroup or not). This is because they show none of the early An innovations in the numeral system, but have the long additive expressions that gave rise to the shortened numerals *pitu 'seven', *walu 'eight' and *siwa 'nine'. In general I regard morphological characters shared by Pazeh or Saisiat and at least one other Austronesian language not in contact with them as PAn. When applied to neutral focus markers, this yields the system of PAn focus markers in Table 1:

non-actor focus markers	patient	V-en
	location	V-an
	instrumental, beneficiary etc.	Si-V
actor focus marker		mu-V/<um>V

Table 1: PAn neutral focus markers

On the same grounds, PAn perfective aspect was marked by <in> infixed in the verb.

1.1. Third-person pronouns replaced

Ross (2006:536-537) described two sets of third-person pronouns in Formosan languages: one, reflected in Pazeh and Saisiat, is formed on a base *Cia. The nominative third-person pronouns in Pazeh and Saisiat, *sia* and *sia* (singular and plural not distinguished) reflect that base. Singular and plural forms are distinguished in the genitive: singular *ni-sia*, plural *n-a-sia* in both languages, reflecting *ni-Cia (singular) and *ni-a-Cia (plural). The second set is reflected in proto-Atayal and in PMP: only the singular forms are cognate. The base form is *sia, the genitive *nia. In terms of the tree in Figure 1, the first set can be regarded as PAn

[3] I make a distinction between transfer of a feature by spread and by borrowing: I speak of borrowing when a feature which is already fixed in a language —not in the process of spreading— is transferred from that language into another.

and the second set as an innovation at the Pituish node of the tree, as Ross notes. The path by which the second set arose out of the first can be described as follows: first, the genitive singular *ni-Cia was reduced to *nCia: *s and *C would be hard to distinguish following *n, and *nCia was reinterpreted as *nsia. In an analogically-motivated change, the base was *Cia was levelled to *sia. Finally *nsia was reduced to *nia. This sequence of changes was completed in proto-Pituish. In Pituish languages outside of Atayal and Malayo-Polynesian, the history of third-person pronouns is one of replacement of the new set by demonstratives. These replacements took place independently in each branch, as shown by the fact that the third-person pronouns in these languages are non-cognate.

1.2. Loss of -en in perfective patient focus forms.

There is no principle reason why we should expect an asymmetry in the way aspect and focus marking affixes combine in early An verb forms. Yet in the An world the perfective aspect marker (PERF) *<in> and patient focus marker (PF) *-en are found attached to the same verb stem only in four West coast Formosan languages: Saisiat, Pazeh, Thao and Siraya. Elsewhere such forms are not found.

Here is a Saisiat example (Zeitoun et al. 1996):

hiza	*ʔalaw*	*maʔan*	*minayʔangsow*	*s<in>iʔael-ən*
that	fish	1S.Gen	left	<PRF>-eat-PF

'That fish is what I have left of my eating'

Here are two more Saisiat examples, from Huang (2005):

noka	*kinaat*	*in-sawi-in,*	*ya'o*	*kayni'*	*ba:iw*
GEN	book	PERF-torn-PF	1sg	not.want	buy

'(because) the book is torn, I do not want to buy (it)'

ka	*kapapama'an*	*in-tani-in*	*ma'an*	*m-in-tani*
Nom	vehicle	PERF-stop-PF	1S.Gen	AF-PERF-stop

'(because) I stopped (in this way), the car stopped'

For Pazeh, Li and Tsuchida (2001: 28; 38 n. 17) state that "In most Formosan languages and western Austronesian languages, the perfective form of the patient-focus does not bear the focus affix *-en*, but the infix *<in>*. Pazih may bear only the the affix *<in>* (e.g. *b<in>aket* 'to have been beaten'), *-en* (e.g. 'to be beaten or to have been beaten'), or both affixes (e.g. *b<in>aked-en* 'to have been beaten') to indicate the perfective aspect of the Patient-focus." They do not provide sentence examples.

Verb forms with both <in> and -in (< *-en) are grammatical in Thao (Blust 2003:238). Examples: *in-dahip-in* 'was helped', *in-fari-n* 'was blown by the wind', *lh-in-irik-in* 'was poked or pierced', *sh-in-umshun-in* 'was worshipped by someone' etc.

aki	*sh-in-umshum-in*	*iza*	*ita*
God	worship-PERF-PF	already	1Pi-NOM

'we worshipped God'

nak	*a*	*kuskus*	*d-<in>uruk-in*	*sa*	*shkish*	*kahiwan*
1s-GENLIG		foot	pierce-PERF-PF	by	slender-bamboo	some-time-ago

'my foot was pierced by a slender bamboo some time ago'

i-nay	*a*	*hulus*	*f<in>ariw-in*	*suma*
this	LIG	shirt	buy-PERF-PF	someone

'Someone has bought this shirt'

Siraya, an extinct language of the SW coast of Taiwan known through 17th-century missionary materials, has been studied by Adelaar (1997). Adelaar gives the past tense of undergoer-oriented verbs ('PF') as either *ni*-V *or ni*-V-*ən*, where past tense *ni-* is clearly the reflex of the PA *<in> (although it is prefixed to the verb stem, not infixed in it) and *-ən* is the reflex of *-ən. Examples of *ni*-V-*ən* from Adelaar (1997):

ni-sulat-ən	*da*
PAST-write-PF	?

'It has been written...'

ni-patĭmxa-ən	*tĭn*	*ta*	*vare*	*vaung-appa*
PAST-punish-PF	(by) him	TM	wind	sea-also

'he rebuked the winds and the sea'

Are we dealing with a retention from PAn or with an innovation by four western Formosan languages ? Under Blust's ten-branch phylogeny, an innovation is strongly indicated, since the four languages fall into two primary branches of An: a retention would take eight independent losses in order to account for the form's absence outside of the Western plains and Northwest Formosan branches. Predictably, Blust (1998), to whom the Siraya, Pazeh and Saisiat facts were moreover not available at the time, came strongly on the side of the innovation view, arguing that in Thao PF -in (< *-en) could have been added to PF-PERF forms originally without it, in order to avoid homonymic clash with AF-PERF forms where the AF marker had a zero allomorph: if it was not for -*in*, he argued, nothing would distinguish AF-PERF forms like *d<in>uruk* 'stabbed' and *f<in>ariw* 'bought' from the corresponding PF-PERF verbs. He noted that among Thao verb stems in perfective patient focus, the PF marker -*in* (< *-en) is optional in some and obligatory in others; he suggested there is a tendence for those in which it is obligatory to have a zero allomorph of the AF marker *<um>* (like *d<in>uruk* 'stabbed' and *f<in>ariw* 'bought'), and to be optional in verbs where the allomorph of the AF marker is not zero (like *s-m-iraq* 'to kiss', *s-m-in-iraq* 'kissed' whose PF-PERF form is either *s-in-iraq* or *s-in-iraq-in*); he warned that the correlation is not perfect.

But the homonymic clash problem he outlined only exists under the view that -en is not original in Thao PERF-PF verbs. If *loss* of *-en is the innovation, Thao never lost it: it could have permitted zero allomorphs of *<um>* precisely because *-en was still there. Blust's discussion of this issue does not provide evidence that PF-PERF verb forms were without *-en in PAn; rather, it is a speculative account of why *-en could have been added to PF-PERF verbs in Thao assuming PF-PERF verbs were without that suffix in PAn.

In fact, consideration of the Saisiat, Pazeh and Siraya facts clearly shows that avoidance of homonymic clash is not what accounts for the presence of *-en in Pazeh *b<in>aked-en* 'to have been hit', Saisiat *in-tani-in* 'stopped' and Siraya *ni-kita-(ə)n* 'was seen', for these verbs require full allomorphs of the AF marker in their respective languages: Pazeh *m<in>u-baket* 'to have hit' (AF-PERF); Saisiat *m-in-tani* 'stopped' (AF-PERF), Siraya *ni-k<m>ita* 'saw' (AF-

PAST).

Because <in>V-en verbs are found in Saisiat and Pazeh as well as in two other languages — Thao and Siraya—, I regard them as part of PAn, and treat their absence in the rest of the An world as innovative. Judging from their low frequency of occurrence in Pazeh and Saisiat and their optionality in Siraya, it is likely that *-en was already optional in PAn when *<in> was present, in other words that loss of *-en in PF-PRF verbs was already in progress in PAn.

The motivation for this change appears to have been simplification. Removing *-en allowed to decrease morphological marking on verbs forms which were presumably very frequent in discourse. Since the *<in> perfective marker only occurred in verbs marked for verbal focus, removal of *-en could be achieved without losing a distinction, by taking advantage of a gap in the system.

1.3. Extension of ki- prefixation to verb stems.

In the following discussion, I rely in part on a draft paper by Stacy Teng and Elisabeth Zeitoun[4]. There attention is drawn to a prefix *ki- that attaches to noun stems, deriving verbs with the broad meaning of "get N" (or "collect N", "harvest N", "cut N" etc.) in Saisiat, Kavalan, Rukai, Kanakanabu, Saaroa, Paiwan and Puyuma (as well as in Bunun where it assumes the specialized meaning of 'take off, remove'). Thao has a prefix kin- which serves the same function; it is unclear how it relates to *ki-. No Amis examples can be found in Fey (1986), and Teng and Zeitoun accordingly do not mention Amis among the languages having such forms; yet Zeng (1991) and Pourrias and Poinsot (ms) list some.

Here are some illustrative word pairs: **Saisiat** *kaehoey* 'tree, wood, brushwood' vs. *ki-kaehoey* 'gather brushwood', **Kavalan** *tamun* 'vegetable' vs. *qi-tamun* 'pick vegetables', **Kanakanavu** *tamemi* 'sweet potato' vs. *ki-tamemi* 'gather sweet potatoes', **Rukai** (Mantauran) *paiso* 'money' vs. *i-paiso* 'earn money'[5] **Paiwan** *sudju* 'sweetheart', *ki-sudju* 'go courting, look for a sweetheart', **Puyuma** *daqiŋ* 'a share' vs. *ki-daqiŋ* 'claim one's share' **Amis**[6] *runaŋ* 'mud pool' vs. *ki-runaŋ* 'wallow in mud' (of water buffaloes).

As expected of verbs which incorporate their object, denominal *ki-* verbs are one-argument intransitives: their unique argument is a nominative agent. Most cannot cooccur with any voice affixes. Teng and Zeitoun argue that derivation of denominative verbs by *ki- is a good candidate for PAn ancestry. This is likely from my point of view because this process is seen in Saisiat. The affix may have grammaticalized out of a verb meaning 'get, collect, gather'. A denominal verb *ki-pañay 'harvest rice' was probably part of PAn (Saisiat *ki-pazay* 'harvest rice', Paiwan *ki-paday* 'harvest rice').[7]

Teng and Zeitoun describe a second construction involving the prefix *ki-, reflected in a more limited collection of languages. There the *ki- prefix attaches to a *verb* root to derive another verb which can be described as 'get V-ed': having reflexive or middle voice. This process is seen, with semantic variation, in Paiwan, Puyuma, Rukai, Bunun, Kavalan and, again, Amis.[8] It is not seen in Saisiat, Pazeh, Siraya, Thao, Tsou, Kanakanabu and Saaroa. In

[4] ***The passive ki- in Rukai, Paiwan and Puyuma: borrowing, shared innovation or parallel development?*** shown me by E. Zeitoun in August 2005.
[5] This example shows that the process was still productive during the Spanish occupation of Taiwan.
[6] This example is from Zeng (1991)
[7] See Sagart (2004) on the reconstruction of the PAn phoneme traditionally referred to as *j.
[8] Zeng (1991:29) cites the pair *ʔadiŋ* 'keep off, fend off, shelter from' (遮擋) vs. *ki-ʔadiŋ* 'to be protected from

Paiwan, Puyuma and Rukai one sees a further development to passive verbs derived from verb s by *ki-. Teng and Zeitoun warn that this development may involve contact or parallel innovation. The following examples are reproduced from Teng and Zeitoun:

Tona Rukai
...ky-a-avase *na* *sakabaane.*
...Pass-Real-Dyn.NFin:rob Nom village
'The village was robbed.' (Formosan Language Archive, DRUTo_11_004_b)

Nanwang Puyuma
m-uka-Ø *i* *trau-trau;* *m-uka-Ø* *ki-bekas-a.*
AV-go Loc Red-person AV-go Pass-interrogate-Proj
'He went to others; he went to be interrogated.'

Southern Paiwan
'i-pangulr *ti* *kivi.*
Pass-beat Nom Kivi
'Kivi got beaten.'

As Teng and Zeitoun noted, the nominative argument or subject in these constructions is the verb's patient. This is in contrast to verbs derived by ki- out of nouns (see above). Moreover Teng and Zeitoun contrasted ki-passives in Puyuma and Paiwan with patient-focus constructions in the same languages, observing that the ki-passives (where X gets V-ed according to his/her wishes) are more volitional than the Patient-focus construction. In a 'tentative conclusion' they further suggested that the rise of ki-passives is what triggered the realignment of verbal morphology around an active/passive distinction in Rukai:

"In Rukai, **ki-**V expanded so much that it came to replace the N[on-]A[ctor]V[oice] affixes while losing its volitional feature; the earlier NAV affixes were preserved in their nominalizing functions"

While this conclusion was offered tentatively, it is attractive. Several Walu-Siwaish languages aside from Rukai have undergone wide-ranging morphosyntactic realignments: competition between verbs with ki- and NAF (Non-Actor Focus) constructions has the potential to explain why Puyuma *ki-passives have replaced the old *neutral* NAF constructions with affixes *-en, *Si-, *-an;[9] while the NAF "projective" constructions with *-aw, *-ay and *-anay and with them, the entire NAF category were able to maintain themselves.

Ross (this panel) argues that Puyuma never had verbal uses of *-en, *Si- and *-an because no vestiges of verbs carrying these affixes can be found in that language. Here is a possible counter-example. Several languages of the northern Philippines: Isneg, Agta, Casiguran Dumagat etc. have a verb reflecting *[q?]unik 'to climb.[10] Puyuma (Cauquelin) has a stem *qunkun* 'to jump over',[11] eligible for agent and non-agent focus marking: *munkun* or *mu-*

the ouside by some object' ('中間有東西擋住')
[9] *-an was maintained in its nominalizing function.
[10] Isneg ʔ*umune*ʔ, Casiguran Dumagat ʔ*unek* 'climb up a tree', Agta ʔ*imunek* (Reid 1971). The last vowels in those forms reflect *i (Lawrence Reid, p.c., June 2009).
[11] Listed by Cauquelin under *unkun*.

qunkun (AF), *unkun-ai* (PF), *unkun-aw* (LF).[12] Examples from Cauquelin's dictionary:

ku	*unkun-aw*	*na*	*gung*
1S	jump-over-LF	the	ox

I jump over the ox

tu	*unkun-ai*	*ku*	*Da*	*suan*
3S	jump-over-PF	1S	the	dog

The dog jumps over me

If the semantics are not judged too divergent,[13] Puyuma *unkun* can derive from *qunik-en, the PF form of *qunik. I suggest that when neutral patient-focus constructions were abandoned for ki-verbs, the PF suffix of *qunik-en was incorporated into the stem, making it eligible for all focus constructions and moving stress to the right. *quniken then underwent unstressed vowel syncope to *qunken and eventually (*q)unkun*, the modern form. I speculate that vowel syncope helped the stem to survive in recognizable form. Verb stems having incorporated *-en normally had three syllables: this made them prone either to be discarded as too cumbersome, or to have their first syllable pruned off to make them disyllabic. Pruning would however have a side-effect: to make them hard to etymologize.[14] Here we see how morphological simplification can fail to leave conspicuous vestiges in the lexicon.

Like Puyuma, Paiwan developed a ki-passive construction but competition between NAF constructions and ki-passives did not lead to the loss of certain NAF markers as in Puyuma or of the entire NAF category as in Rukai. Likewise, in Amis and Bunun, the entire NAF category is maintained. Amis, Bunun and PMP settled the competition in favour of the NAF constructions by adding prefixes like ma-, pa-, pi- and mi- before ki- verbs, thereby giving those verbs the grammatical properties attached to these prefixes. A gradation can be seen among the prefix-adding languages: Bunun retains many unprefixed ki- verbs, Amis only a few: the MP languages keep none.

2. A morphology-based phylogeny

I present in Figure 2 a phylogenetic tree built from the three innovations discussed in the preceding paragraphs.

[12] The variety of Puyuma investigated by Cauquelin loses *q word-initially. Initial q- in this word is preserved in one of the AF forms thanks to the *mu-* prefix. The other AF form *munkun* is analogically motivated.

[13] Semantically the main difference between 'jump over' and 'climb' is the ballistic character of the former.

[14] Puyuma has a number of disyllabic verbs ending in -un and without cognates in other languages, which are candidates for incorporation of *-en and loss of first syllable: *repun* 'to assemble, get together', compare Puy. *reprep* 'to swarm with, be infested with' (insects); *kuLun* 'to roll', compare root -kul 'curl, bend'; *LuDun* 'to sink', compare Isneg *allad* 'sink', Tiruray *eled* 'sink' (forms from ACD, under *qeled 'sink'; Blust mentions a rot -led 'sink').

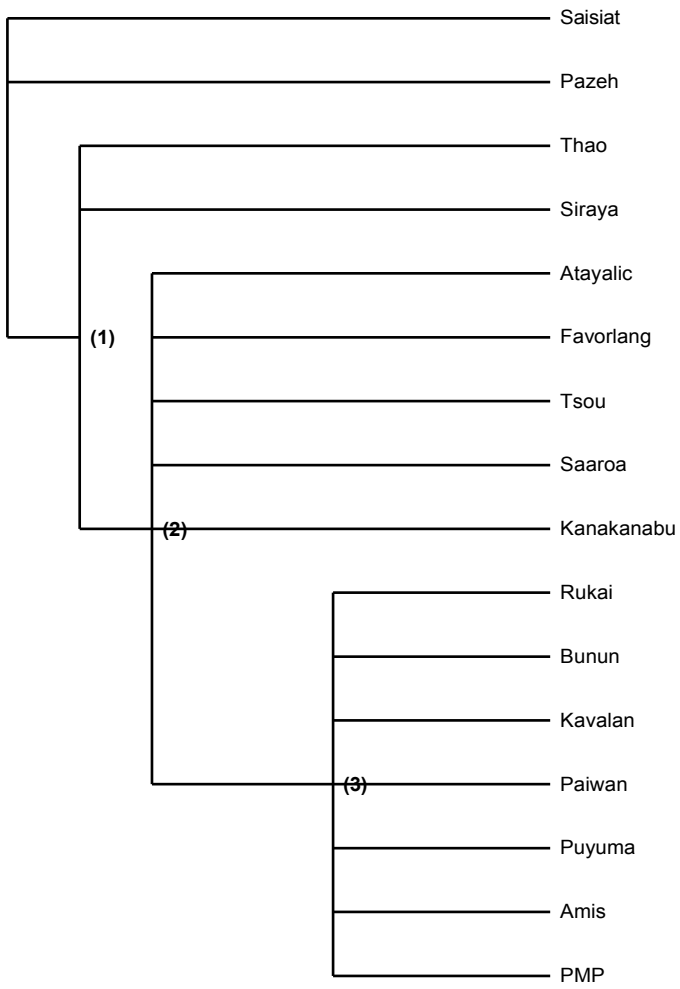


Figure 2: Higher An phylogeny based on three morphological characters. (1): replacement of 3rd-person pronouns; (2) loss of *-en in perfective PF; (3) extension of ki- prefixation to verb roots.

This tree conflicts with the tree in Figure 1 in that the ***enem** isogloss in the numerals-based tree overlaps with the **Loss of *-en** isogloss in the morphology-based tree: thus Atayalic is above Thao and Siraya in the lexical-based tree, but below them in the morphology-based tree.[15] The problem can be fixed by allowing *-en to be lost twice: once in Atayalic and another time in a node corresponding to Enemish in the numerals-based tree. This is not unrealistic for a character that was already optional in PAn.

It is not surprising, in any case, that the two trees are entirely compatible, given the likelihood of contact effects and of independent innovations. What matters is that they tell very similar stories: a PAn homeland on the northwest coast of Taiwan, full settlement of the west coast before southern Taiwan and finally the east coast are settled, and an origin of the MP migration in the East coast languages. This view, now supported by seven[16] lexical and three morphological innovations, is consistent with archaeological dates —earlier on the west coast than on the east coast of Taiwan—. Population genetics (Sanchez-Mazas, this panel) give a similar picture.

[15]It is impossible to be certain that extinct Favorlang did not allow *-en in at least some perfective patient-focus verb forms.

[16] The Enemish node is also supported by the displacement of *kawaS by *CawiN as 'year' (Sagart 2004).

Why, then, do Gray, Drummond and Greenhill (2009), who work with the basic vocabulary, find a star-like phylogeny —although unlike in Blust's, and like in mine, PMP appears as coordinate with Paiwan— ? this is undoubtedly because of their choice of *both* Old Chinese and Tai-Kadai as outgroups. Probably no Austronesianist thinks that both Tai-Kadai and Chinese are simultaneously related to An and exterior to it. Yet this is the basic assumption that underlies the analysis of Gray and his colleagues. By forcing Tai-Kadai, really a subgroup of An, to assume the role of an outgroup (that is, a language family outside of Austronesian but related to it), they are forcing their Bayesian statistics algorithm to treat every item shared by Tai-Kadai and *any* Austronesian language as a retention from PAn, thus preventing it from making use of post-PAn innovations like the 'standard' numerals between 5 and 10 or the shifts in 1st-⁻ and 2nd-person pronouns.[17]

I predict that if Gray and his colleagues give up Tai-Kadai as an outgroup and allow it to place itself in the tree *rooted in Old Chinese*, Tai-Kadai will branch off the Austronesian tree near where Malayo-Polynesian branches off, and the phylogeny they will find for Formosan languages will be very similar to those in Figure 1 and Figure 2.

This paper is dedicated to the memory of Stanley Starosta

References

Adelaar, K. Alexander. 1997. Grammar notes on Siraya, an extinct Formosan language. *Oceanic Linguistics* 36:2. 362-397.

Blust, R. (1998)A Note on the Thao Patient Focus Perfective. *Oceanic Linguistics*, Vol. 37, No. 2 (Dec., 1998), pp. 346-353.

Blust, R. (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).

Blust, R.A. (2003) *Thao Dictionary*. Nankang: Institute of Linguistics.

Cauquelin, J. (1991) *Dictionnaire puyuma-français*. Paris, Jakarta: Ecole Française d'Extrême-Orient.

Fey, V. (1986) *Amis dictionary*. The Bible Society in the Republic of China.

Gray, R.D., A.J. Drummond and S. Greenhill (2009) Language Phylogenies Reveal Expansion Pulses and Pauses in Pacific Settlement. *Nature*, 323, 479-483.

Huang, S.-F. (ms., 2005) *Topics in Saisiyat grammar.* Paper presented at the 'Conference on NSC Projects 2001-2004', Taiwan.

Li, P. J-K, and S. Tsuchida (2001) *Pazih dictionary*. Nankang: Institute of Linguistics (preparatory office), Academia Sinica

Meacham, C., and G. Estabrook. 1985. Compatibility methods in systematics. *Annual Review of Ecology and Systematics* 16:431-446.

Pourrias, L. and M. Poinsot (ms) *Dictionnaire amis-français.*

Reid, Lawrence A. 1971. *Philippine minor languages: Word lists and phonologies.* Oceanic Linguistics Special Publication, 8. Honolulu: University of Hawaii Press.

Ross, Malcolm D. (2006) Reconstructing the case-marking and personal pronoun systems of Proto-Austronesian. In: Henry Y. Chang, Lillian M. Huang and Dah-an Ho (eds.) *Streams converging into an ocean*, 521-563. Language and Linguistics monograph series number W-5. Taipei: Institue of Linguistics, Academia Sinica.

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

[17] Note that the choice of TK as an outgroup distorts the tree only above the point where TK *really* branches off: the non-Formosan part of the phylogeny in Gray et al. (2009) does not suffer from that problem.

Sagart, L. (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.

Starosta, Stanley (1994) Rukai-Tsouic: subgroup or treetop ? Paper presented at the Seventh International Conference on Austronesian Linguistics. Leiden: Leiden University.

Starosta, Stanley (1995) A grammatical subgrouping of Formosan languages. In: Paul Jen-kuei Li, Dah-an Ho, Ying-kuei Huang, Cheng-hwa Tsang, and Chiu-yu Tseng (eds), *Austronesian Studies Relating to Taiwan*. Symposium series of the Institute of History and Philology, Academia Sinica, no. 3. Taipei: Academia Sinica, pp. 683-726.

Teng, S. and E. Zeitoun (ms c. 2005) *The passive ki- in Rukai, Paiwan and Puyuma: borrowing, shared innovation or parallel development?*

Zeitoun, E., L. M. Huang, M. Yeh, A. Chang, J. Wu. 1996. The temporal/aspectual and modal systems of the Formosan languages: a typological perspective. *Oceanic Linguistics*, 35.1: 21-56.

Zeng Siqi (1991) *Taiwan Ameisi-yu yufa* [A Grammar of Amis of Taiwan]. Beijing: Zhongyang Minzu Xueyan. 283 p.