

John Benjamins Publishing Company



This is a contribution from *Language and Linguistics* 18:4
© 2017. John Benjamins Publishing Company

This electronic file may not be altered in any way.

The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.

Permission is granted by the publishers to post this file on a closed server which is accessible only to members (students and faculty) of the author's/s' institute. It is not permitted to post this PDF on the internet, or to share it on sites such as Mendeley, ResearchGate, Academia.edu.

Please see our rights policy on <https://benjamins.com/content/customers/rights>

For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com).

Please contact rights@benjamins.nl or consult our website: www.benjamins.com

The pitfalls of negative evidence

‘Nuclear Austronesian’, ‘Ergative Austronesian’, and their progeny

Robert Blust and Victoria Chen

University of Hawai‘i

Beginning with publications in the early 1980s there have been attempts to use syntactic data to determine the highest-order subgroups of Austronesian. These efforts fall into two categories: those which claim that the voice affixes of Philippine-type languages originally had exclusively nominalizing functions, and those which claim that the affixes themselves were innovated after the separation of Rukai from the ancestor of all other Austronesian languages. Although these ideas lay dormant for some years, recently both have been revived in renewed efforts to show that the Austronesian family tree is not ‘rake-like’ in its highest nodes, but shows extensive embedding of subgroups that can be justified by successive layers of syntactic innovations. This paper questions the methodological soundness of both types of arguments on the grounds that they appeal to negative evidence, and logically any such appeal can do no better than reach an inference of indeterminate status rather than the positive conclusions that have been proposed.

Keywords: subgrouping, negative evidence, Nuclear Austronesian hypothesis, Ergative Austronesian hypothesis

1. The nature of scientific arguments

Academic fields have their individualizing differences in terms of working methods, but the basic logic of scientific inference is trans-disciplinary. One feature of scientific arguments that is consistent in any academic field is the recognition that arguments based on negative evidence are castles built on sand. The old saw that ‘absence of evidence is not evidence of absence’ is not just a clever play on words – it illustrates a fundamental principle of inductive logic, namely that science is built on observational data, not on its absence. This can easily be misunderstood, and often is.

In arguments about prehistory it is all too easy to claim that, because some feature has not been found, it never existed. Superficially this even appears to meet the requirement that science be built on observational data. How can we argue for the past existence of some feature if no trace of it remains? The simple answer is: we cannot. However, the more complete answer is: we cannot, but neither can we argue the contrary case, since we cannot logically exclude the possibility that the feature was once present but disappeared before it could be observed. Where observational evidence is negative, then, the proper inference always is indeterminacy.

Cases of this kind are not uncommon in the literature of various academic disciplines. Since the focus of this paper is on the historical development of the Austronesian (AN) languages, a useful case study from another, empirically intersecting discipline, might be the archaeological arguments against the presence of rice during the initial AN settlement of Taiwan. An early discussion of this issue is found in Ferrell (1969: 10), who noted that “For a long time it was considered that rice cultivation was unknown to the Formosan aborigines until it was introduced by Chinese immigrants in modern times.” Ferrell himself was quick to point out that widespread cognate sets relating to rice are difficult to explain under this view, and he supported the contrary position that “Intensive grain cultivation (millet and probably rice) in Taiwan began about 2,500 BC with the arrival from mainland Asia of the Yüanshan Horizon in the Taipei Basin in the northern west coast region, and the Taiwan Lungshanoid Horizon in the south” (Ferrell 1969: 5).

However, Ferrell’s use of positive evidence for this inference was linguistic, not archaeological, and the usual barriers between academic disciplines meant that the archaeologists themselves were unwilling to accept arguments based on evidence from another field. The result was a further period of over three decades in which most archaeologists working in Taiwan were reluctant to believe that rice cultivation had been introduced to Taiwan by the founding Neolithic population, who in linguistic terms were speakers of Proto-Austronesian (PAn). Although pottery and other items of material culture had been found in the earliest archaeological levels, rice was absent, and there was thus no reason to believe that rice agriculture arrived with the first AN-speaking colonists. What is noteworthy is that the line between ‘absence of evidence’ and ‘evidence of absence’ blurred easily, with the result that some archaeologists believed the founding Neolithic communities in Taiwan must have lacked rice agriculture. Although it is hard to find explicit published statements to this effect, this view was often expressed in conversations about Taiwan prehistory in the early 1990s.

Then, in excavations conducted between September, 2002 and March, 2003 in advance of extensive government-sponsored construction, the whole archaeological edifice which saw the founding Dabenkeng archaeological culture as non-agricultural was demolished through the groundbreaking work of the Taiwanese

archaeologist Tsang Cheng-hwa. As told by Bellwood (2011: 98) "... remains of rice and millet were universally absent from sites of the Dabengkeng phase in Taiwan (3,500–2,500 BC) until both were found in unprecedented carbonized quantities dating to c. 2,800 BC, in hitherto unique waterlogged conditions, in the Nanguangli sites in the Tainan Science-based Industrial Park (Tsang 2005, Tsang (personal communication))." In just six months, the observational void that had been used as 'evidence' for the absence of rice among the earliest AN-speaking settlers of Taiwan was filled, and the whole theory of prehistory that depended on it collapsed.

More recently, and more concretely (because it is captured in print) the Australian archaeologist David Bulbeck (2008) has argued explicitly that, although speakers of PAN possessed rice agriculture in Taiwan, they abandoned it on leaving for insular Southeast Asia, becoming 'maritime foragers', hunting, fishing and gathering wild plants, with no further use of grain agriculture. Like its Formosan predecessor, this inference conflicts with a rich body of linguistic evidence for both rice and millet continuing as important cultigens among AN speakers as they expanded into the island world of the tropics. Bulbeck's position is a classic example of a positive argument built on negative evidence; he is quite certain that, once they had left Taiwan, AN speakers no longer had access to rice, as shown in this remarkable quotation, which pays lip service to interdisciplinary cooperation, yet ignores relevant linguistic evidence for prehistoric rice agriculture in the Philippines and the Malay archipelago (Bulbeck 2008: 48):

I hope it is clear from my contribution that I heartily endorse a multidisciplinary approach to ISEA [Island Southeast Asian] prehistory, and the 'triangulation' of the linguistic, archaeological and biological evidence relevant to the Neolithic (e.g. Sagart et al. 2005). However, I have little sympathy for taking a particular interpretation of the historical linguistics, one based on idealist culture history and the assumption of an expanding Malayo-Polynesian monoculture, and using it to overwrite the archaeological and biological evidence. In my view this reduces the number of disciplines that counts to one, which is the opposite of multidisciplinary research.

Bulbeck's position, like that of the skeptical Taiwan archaeologists before Tsang's work at Nanguangli, is based entirely on negative evidence (again, ignoring substantial countervailing linguistic data which served as *positive* evidence for a contrary view). In particular, his survey of the archaeology of Indonesia shows few prehistoric sites of any time-depth that include rice, and based on this observation he concludes that rice could not have been present in the founding AN-speaking cultures of this area. It is worth noting that the majority of these sites are rock shelters and caves, where one would not expect to find rice, but despite whatever

skewing factors might be present, the overarching error in this argument is the simple fact that it relies implicitly on the belief that ‘absence of evidence is evidence of absence.’

These examples are taken from an archaeological perspective on the AN homeland in Taiwan and the subsequent diaspora into insular Southeast Asia. Given the difference of discipline some linguists may be asking ‘What does this have to do with linguistics?’, but such a reaction would be short-sighted. As already noted, academic disciplines may differ in the details of how they organize or analyze data, but one feature that is consistent across all fields of science is that arguments based on negative evidence are inherently flawed because they attempt to draw positive conclusions from what, on methodological grounds, are inescapably indeterminate arguments.

The remainder of the paper is organized as follows. Section 2 discusses the methodological issues that arise in using negative evidence for diachronic inferences in linguistics. Section 3 evaluates two recent Austronesian higher-order subgrouping proposals in the light of comparative data from Malayo-Polynesian languages that allows an unambiguous determination of the direction of change, and hence a clear distinction between innovation and retention. Section 4 re-examines the two hypotheses with Formosan data, with a particular focus on the chronology of *-en under each hypothesis. Section 5 concludes.

2. The morphosyntax-based arguments for Austronesian higher-order subgrouping

The first serious attempt to reconstruct the verb system of PAN was that of Wolff (1973), which was founded on an application of the traditional Comparative Method of linguistics. Since the Comparative Method operates with substance, not with form, Wolff’s reconstruction aimed to find positive evidence for the morphemes which did the work of expressing the syntactic categories of verbal voice in a language ancestral to all members of the AN family. To do this he used comparative data from Atayal and Tsou in Taiwan, Samar-Leyte Bisayan in the Philippines, and Javanese in western Indonesia. The determination of function in the proto-language then followed from agreements between the functions of cognate affixes in the languages compared. The table below presents a somewhat scaled-down version of Wolff’s reconstruction, omitting details that have no role in the following argument (1):

(1) The morphological core of the PAn verb system (modified from Wolff 1973)

Voice	Nonpast	Past
Actor	*<um>	*<inum>
Direct passive	*-en	*<in>
Local passive	*-an	*<in>an
Instrumental passive	*i-	*i-...<in> (?)

Wolff designated these affixes ‘independent’ verbal voice markers, and in addition he posited a more fragmentary set of what he called ‘future-general action dependent subjunctive’ markers, of which only *-ay ‘local passive’ need be mentioned here. He could as easily have called the two sets ‘indicative’ and ‘non-indicative’, which would perhaps have provided a clearer indication of their syntactic distribution. In any case he evidently was the first to note that the direct passive (patient, undergoer) voice has a zero allomorph in the past (more commonly called perfective today). In addition, we now know that *i- should be *Si-. While Wolff’s analysis of this system as containing three passives is now out of favor, his use of ‘voice’ has come back in recent years, and is in competition with the problematic term ‘focus’; both terms will therefore be used here interchangeably.

Wolff’s reconstruction was not flawless, but it provided a solid beginning in our attempt to understand the nature of the PAn verb system and its transformations over time, because: (1) like all sound work in historical linguistics, it was rooted in the reconstruction of *morphemes*, not abstract schemata; (2) it showed that these morphemes exemplified a system of relationships that was internally consistent and capable of expressing a wide range of functions; and (3) in relation to the accepted guidelines of general scientific method, it was based on *positive* evidence. Most languages that have retained the system Wolff reconstructed are located in the Philippines, and for this reason such languages, regardless of where they are spoken (Taiwan, Sabah, Madagascar, Sulawesi, Marianas, etc.) have come to be known as ‘Philippine-type languages’.

The first challenge to Wolff’s reconstruction was that of Starosta et al. (1982), who argued that the affixes which Wolff saw as the core of the verb system had exclusively nominalizing functions in PAn, and only later acquired their verbal functions through a process of reanalysis. The fact that these affixes in the great majority of Philippine-type languages have both verbalizing and nominalizing functions dependent on syntactic context was passed over rather quickly, and although it was claimed that the reinterpretation of original nominalizers as voice markers should have important subgrouping implications, none were drawn (1982: 166).

Wolff (1973) was concerned only with morphosyntactic reconstruction, not with subgrouping, but following the publication of Starosta et al. (1982) attention turned to the use of syntactic reconstruction for purposes of phylogenetic classification. Nearly all of the methodological problems that plague current AN subgrouping proposals based on syntax trace their origin to Starosta (1995), who spoke confusingly of ‘Proto-Formosan’ as though it needed to be distinguished from PAN, and who argued that the first split in the AN family tree separated Rukai from a language, simply labelled F1, that was ancestral to all others (1995: 691).

As noted in Blust (1999: 62–66), Starosta’s argument is flawed in a number of respects, but the central problem that concerns us is the ‘evidence’ for Rukai vs. F1 at the top of the AN family tree. Starosta holds (1995: 689–690) that his subgrouping is founded on a clear recognition of “... the importance of establishing shared innovations. If a linguistic account of relatedness among a group of languages just counts similarities but makes no attempt to establish and justify shared innovations and distinguish them from similarities having other causes, it makes no contribution to the determination of the prehistory of a language family and of the people who spoke it [sic].” No one can quarrel with this general statement of method, since it is universally agreed that subgroups must be based on exclusively shared innovations that are not likely to be products of convergence. The problems appear when Starosta applies this acceptable model to actual language data. He begins by noting (1995: 692) that Rukai lacks “the focus morphology” that is common to most Formosan and Philippine languages, and he proposes to explain this difference by claiming that the widespread voice-marking or focus-marking affixes of Philippine-type languages are post-PAN innovations, since they do not appear in the verb system of Rukai. As noted in Blust (1999: 63) “The logic of this argument is reasonably clear: if two languages differ in that one has a feature that is absent in the other it follows *ipso facto* that the language which lacks the feature must be conservative. Stated differently, Starosta assumes (without argument) that *zero* : *nonzero* comparisons must have a zero starting point.”

Despite the difference of content, the form of this argument should look familiar, as it is precisely parallel to the arguments of the Taiwan archaeologists who insisted before the work of Tsang proved them wrong that speakers of PAN had no rice agriculture, because up to that point they had found no evidence for it. In short, Starosta’s position was that Rukai lacks a focus system, and therefore the focus system must have been innovated after the separation of Rukai from F1. Needless to say, this conclusion is no more justified than the view that speakers of PAN lacked grain agriculture when they settled Taiwan because evidence for it was unknown until the excavations at the Tainan Science-based Industrial Park. In the archaeological case we were fortunate in having an opportunity to see this erroneous view overturned by the subsequent discovery of conflicting evidence.

With linguistic data this is unlikely to happen, since it would require either written documents in an earlier form of Rukai, or Rukai dialects that have some or all of the affixes in question as key elements of the voice system. However, this does not change the fact that a subgrouping hypothesis which opposes Rukai to the rest of AN on the basis of exclusively shared ‘innovations’ in F1 has not ruled out the possibility that these proposed innovations could be retentions. Starosta (1995: fn. 3) had a retort for this: “Several colleagues have dismissed this whole hypothesis by saying that all the innovations I propose are actually just retentions, and all the focus morphology missing from Rukai for example has just been “lost”. In the absence of a detailed case-by-case accounting for each of the “lost” items, of course, this has no more explanatory force than a claim that, say, Lapita pottery actually originated in Taiwan, but that all the Lapita shards in Taiwan got lost.”

Starosta’s error here, and it is a fundamental one, was to assume that those who disagreed with his proposal were themselves arguing for a positive claim based on negative evidence. In other words, he characterized his opponents as claiming that the focus affixes must be retentions, rather than the actual claim that they *could* be retentions – and that as long as this possibility remains open the assertion that a language ancestral to Rukai never had them is logically indefensible. To the very end, then, Starosta continued to build a subgrouping hypothesis on the basis of negative evidence, with all the hazards that this type of argument entails.

To evaluate the possibility that Rukai could be descended from an ancestor with a full-blown Philippine-type voice system that underwent a radical transformation through innovations that are unique to it, we need only consider other languages for which the evidence of such a restructuring is unambiguous. The simple fact is we do not have to look far. English is descended from Proto-Indo-European, which had an extensive case-marking system that was still largely intact in Proto-Germanic, and that survives in a fairly robust form in both High German and Icelandic. However, in Dutch, which is a close relative of German, much of this system is gone. More to the point, in English almost nothing remains of the Proto-Germanic case system outside of the first and third-person pronouns. While current indications suggest that these remnants will resist further change long into the future, the accusative and dative form of the interrogative pronoun, which Sapir (1921: 156ff) discussed as still present but on its way out a century ago, is now virtually defunct (it is virtually unheard of today, at least in American English, for speakers to say ‘Whom did you see?’, or ‘To whom did you give the book?'). It is true that traces of a formerly much richer case-marking system remain in English, and so one might argue that the same should be true of a language like Rukai – if Rukai once had a Philippine-type voice system there should be at least some trace of it in the modern language, if only in relic forms. This was in fact a position that Starosta took in debates with colleagues: he simply could not believe that the

entire complex morphological system that supported voice-marking and a variety of other functions could disappear without a trace, and he could point to English as supporting evidence for his position.

The dramatic typological transformation of English from an inflectional language to a more isolating one is, of course, paralleled by most or all of the Romance languages in relation to Latin. These well-known examples alone make it clear that syntactic typology can change fundamentally over time, and it is useful to cite them since they are part of the common knowledge of nearly all linguists. However, there are many more cases that apply directly to the problem at hand.

What makes the syntactic history of typologically aberrant languages like Rukai, Tsou, or Puyuma especially tempting to those with a penchant for speculation is that they are near the ‘top’ of the Austronesian family tree – that is, each of these languages is either a primary branch of the family (Rukai, Puyuma), or part of a very small group of languages that itself is a primary branch (Tsou).¹ Given this relative independence, a subgrouping proposal that separates one of these languages from all others has dramatic implications for prehistory, as in Starosta’s proposal that all AN languages apart from Rukai share a corpus of exclusively shared innovations, meaning that Rukai split off from a proto-language (F1) that was immediately ancestral to every other language in the family. However, proposing subgrouping arguments near the ‘top’ of a family tree also encounters hazards that either do not exist or that can be controlled at lower nodes.

As noted in Blust (1999: 57), in order to convert exclusively shared linguistic properties into what we can confidently consider exclusively shared innovations, it is first necessary to solve the problem of directionality. In phonology this can be done in either of two ways. First, some sound changes are known to take place in one direction only, as $*s > h$, which is common in the world’s languages, while $*h > s$ is essentially unknown, or that take place only rarely in the opposite direction, as with $*p > f$, which is part of an erosion sequence $*p > *f > *h > \emptyset$ that is common in whole or in part to many language families, while $*f > p$, though known, is rare. The second way that the problem of directionality can be solved in phonology is through reference to the Regularity Hypothesis, and what might be called the ‘one-to-many vs. many-to-one problem.’ In general, if a language A has two phonemes corresponding to one phoneme in a related language B and they are not in complementary distribution, the most likely direction of change is from an ancestor that

1. The existence of a Tsouic subgroup that includes Tsou, Kakanavau, and Saaroa has provoked considerable controversy, with leading scholars coming down on both sides of the debate (Tsuchida 1976; Chang 2006; Ross 2012; Sagart 2014, Zeitoun & Teng 2016). We find the lexical evidence presented for Tsouic in Tsuchida (1976), and Sagart (2014) persuasive.

resembled Language A in this feature (many-to-one), since otherwise we would be compelled to acknowledge an unconditioned phonemic split, as shown in (2):

(2) The one-to-many vs. many-to-one problem

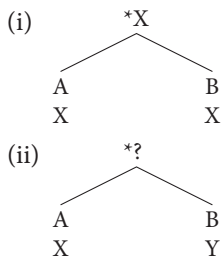
Lg. A	Lg. B
r	r
l	r

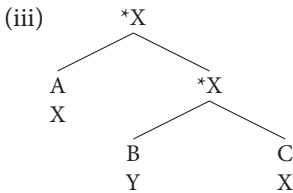
For at least the past century and a half these two considerations have served phonology well in justifying subgrouping arguments that the majority of scholars find plausible, and they remain today the safest basis for subgrouping precisely because they offer time-tested controls on the problem of directionality.

As also noted in this publication (Blust 1999: 60), the controls on directionality of change that are made possible by the nature of phonological correspondences do not exist with lexical correspondences. To choose a simple example, suppose we wish to reconstruct the ancestor of a language family that divides into two branches, A and B. The word for 'dog' in all A languages reflects *patu, while the word in all B languages reflects *kanit. What was the word for 'dog' in Proto-AB? A moment's reflection will show that it could have been *patu, with an innovative replacement in Proto-B, it could have been *kanit, with an innovative replacement in Proto-A, it could have been neither, with both *patu and *kanit being innovations that postdate the A-B split, or it could have been both, with some still undetected difference of meaning. Unless we have some indication of a semantic difference in the glosses of attested forms the last of these possibilities is perhaps the least likely, but without further information there is no obvious way to evaluate the relative likelihood of the other possibilities.

The logical problem of distinguishing innovations from retentions with lexical evidence was diagrammed in Blust (1999: 60) with a figure that is repeated in (3) to draw attention to the dangers of circular reasoning when subgrouping at the 'top' of a family tree based on the presence or absence of cognate lexical data:

(3) Inherent limitations on determining the directionality of change with lexical evidence





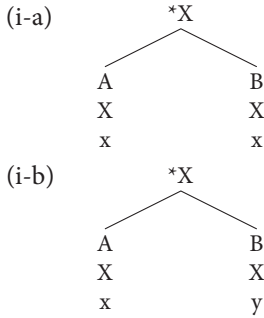
As noted in the earlier publication, the tree diagrams in (3) represent differing conditions that can be encountered in attempting to use lexical evidence for subgrouping. In each case a language family divides into two primary branches. Diagram (i) presents a case where confident lexical reconstruction is possible since cognates ‘X’ appear in both branches. In diagram (ii) the terms are not cognate, thus leaving a reconstruction for Proto-AB in doubt (the hypothetical case of ‘dog’ cited above). In diagram (iii) a similar situation holds for Proto-BC, but can be resolved through appeal to a more distant relative A, which shares a cognate with Language C.

As also noted in the earlier discussion of this diagram, Greenberg (1957: 49–50) drew attention to the problem of circularity in assuming that a lexical distribution is an innovation when the possibility that it is a retention cannot be excluded, and his advice was to resolve the indeterminacy by incorporating evidence from more distantly related languages. The discussion in Blust (1999) was directed at problems with what Isidore Dyen (1990) called the ‘homomeric method’ of subgrouping by lexical evidence, but the lesson that syntacticians should take away from it is that it applies equally to arguments in syntax. To see this, it is necessary to understand that morphosyntactic reconstruction, insofar as it must depend on cognate material to be possible at all, uses lexical data. This was certainly the case for Starosta (1995), who argued that Rukai lacks “the focus morphology” through which the Philippine-type voice system of most other Formosan languages is expressed, and all morphemes, whether free or bound, grammatical or not, are lexical items subject to precisely the constraints on historical inference diagrammed in (3).

With regard to the claim of Starosta et al. (1982) that the focus affixes arose from morphemes that had exclusively nominalizing functions, a moment’s reflection will show that it, too, fits naturally into the sets of logical relationships diagrammed in (3). In order to confidently reconstruct a morpheme in a given meaning or function, as in diagram (i) it is not enough to establish that the items compared are cognate – they must also agree semantically. If English and German had no linguistic relatives the comparison of German *Hund*, English *hound* would not suffice to allow a reconstructed form meaning ‘dog’, since this meaning would be limited to a single primary branch of the family, and from the standpoint of

semantics diagram (i) would be equivalent to diagram (ii) in relation to lexicon. To make this relationship clearer we can restrict upper case 'X' and 'Y' to relationships of cognation, and lower case 'x' and 'y' to relationships of meaning. Diagram (i) is then seen as having two subtypes, as in (4):

(4) Two subtypes of diagram (i)



Type (a) describes homosemantic cognates – words of the same origin and meaning, like German *Auge*, English *eye*. Type (b) describes heterosemantic cognates – words of the same origin but different meaning, like German *Hund*, English *hound*. The figure in (3) must be understood, then, as applying both to lexical distributions and to semantic or functional distributions. Applied to the cases at hand, if the voice morphology of most Philippine-type Austronesian languages is reflected only with nominalizing functions in one language, but with both verbalizing and nominalizing functions in another, how can we tell the directionality of change? As shown above, (i-b) is just as indeterminate in relation to meaning/function as diagram (ii) is in relation to cognation. To anticipate part of the later discussion, Pawley & Reid (1979: 111) assumed that the nominalizing functions of Wolff's indicative voice markers are historically secondary, while a scant three years later Starosta et al. (1982) reversed this position, totally rejecting the earlier interpretation of two of the three co-authors.

What makes all of this relevant to current debate about higher-level subgroups in AN is that other scholars have revived the Starosta et al. (1982) and Starosta (1995) proposals in new form with little or no discussion of the methodological issues that come with them. The arguments of these scholars will be discussed in § 3 and § 4, but before doing this it is important to point out that both of the claims initiated by the work of Starosta, and adopted in new form by younger scholars in recent times can be tested empirically. These claims as described in current debate are:

- (5) Two claims of recent morphosyntax-based arguments
- a. CLAIM A: Nominalization into verb (N-into-V), the claim that the focus affixes originally were nominalizers. This is the key element in what Ross (2009) calls the ‘Nuclear Austronesian’ (NAn) hypothesis, and in the further modification of this interpretation by Zeitoun & Teng (2016).
 - b. CLAIM B: Rukai as treetop, the claim that Rukai separated from the ancestor of all other AN languages before the Philippine-type voice system was innovated. This is the key element in what Aldridge (2014, 2016) calls the ‘Ergative Austronesian’ (EA) hypothesis.

3. Malayo-Polynesian as a test case

What has been strikingly absent from recent morphosyntax-based higher-level subgrouping proposals for AN languages is the use of controls on historical inference. This point may not be immediately obvious to everyone. What kinds of controls on historical inference are possible short of inventing a time machine that would allow us to revisit the past? The answer takes us back to the figure in (3), where tree diagram (iii) shows how an external witness can be used to guide the reconstruction of forms at lower levels of a phylogenetic tree. However, external controls not only provide a check on reconstruction, they also allow us to test hypotheses of change that cannot be tested at the highest node in a collection of related languages that divides into two primary branches.

Keeping this general conceptual framework in mind, we can now add that the consensus AN family tree recognizes more than one primary branch of the language family in Taiwan, and a single enormous branch called ‘Malayo-Polynesian’ (MP) that includes all others (Ross 2006). With more than 1,200 languages scattered over 206 degrees of longitude (Lewis et al. 2016), the MP branch of AN shows great typological diversity, ranging from languages that have a full-blown Philippine-type voice system, through others with a reduced system that clearly reflects the more elaborate one, to others that retain only vestiges of the original system, to others that retain nothing of it. Despite this variety, one thing is non-controversial, namely that PMP had a Philippine-type voice system, a fact established by Wolff (1973), and accepted in all subsequent work on the languages (Ross 2002: 48).

It should be apparent that in using PMP as the starting point for tracing the syntactic history of its descendants we are on much firmer ground than is possible when using PAN, since the reconstruction of PMP is essentially the lower node of diagram (iii) in (3), while the reconstruction of PAN grammatical morphemes, meanings or functions is essentially diagram (ii), in which ‘X’ or ‘Y’ is zero. Stated

differently, diagram (ii) covers cases in which the comparata are non-cognate, whether both contain positive but unrelated data, or whether one contains forms that the other lacks (hence negative evidence). Before pursuing the use of MP data further for testing claims about the highest-level subgroups of AN, it will be worthwhile to mention a few prominent features about the geographical distribution of voice systems in AN languages.

The syntactic typology of AN languages shows two striking geographical traits. The first is that languages that preserve the structure of a Philippine-type voice system tend to be concentrated in a compact area that corresponds to the AN homeland and those areas settled immediately after departure from the homeland. Thus, most Formosan aboriginal languages for which adequate descriptions are available (a number became extinct before good records were made), nearly all languages of the Philippines, and the languages of northern Borneo and northern Sulawesi have verb systems with at least two undergoer voices opposed to an actor voice. A few languages outside this compact geographical block have also preserved much of the original system for various reasons that remain obscure. Most prominent of these is Malagasy, which is geographically far-removed from the main body of AN languages, but which preserves the PMP voice system in greater detail than any other language outside the central block of Philippine-type languages. Other outliers that preserve more Philippine-type voice features than is typical of languages in their geographical area are the Batak languages of northern Sumatra, which have undergone substantial changes from a canonical Philippine-type ancestor, but which still preserve much more of both the verbalizing and the nominalizing morphology of the PMP system than is true of such neighboring languages as Malay, Acehnese, Nias, or Rejang. In addition, Old Javanese, preserved in texts on palm-leaf manuscripts dating from the 9th to the 15th centuries, contains several features of Philippine-type languages that have essentially disappeared in the modern language. Finally, Chamorro, the product of an early, isolated migration from the Philippines to the Mariana islands of western Micronesia, has a number of Philippine-type features, although these have been altered in various ways through distinctive innovations.

Other AN languages present a striking cline: from those with several remnant features of an earlier Philippine-type voice system (including Kelabit, Bintulu, and the Melanau languages of Borneo, Palauan, and, to some extent, some of the languages of New Ireland and the western Solomons); to those with the barest remaining traces; to those with no trace at all. Examples representing the more innovative of these languages will be given to show that what was uncontroversially a Philippine-type voice system in PMP has disappeared with few or no traces in some of its descendants.

3.1 The history of Malayo-Polynesian, 1: The Nuclear Austronesian hypothesis

It will be recalled from (1) and the discussion immediately following it that in addition to what he called the ‘independent’ verbal voice markers, Wolff (1973) posited a rather fragmentary set of ‘future-general action dependent subjunctive’ markers which included imperative and subjunctive morphemes for each of the voice categories established for the dominant independent set. In reconsidering the nature of the PAN voice system based on a much closer familiarity with Puyuma, Ross (2009: 303ff) took two important steps. First, based on Formosan data that was not available in the early 1970s, he filled out Wolff’s account of the non-indicative voice markers as shown in (6):²

(6) PAN optative/hortative morphology in Ross (2009)	
Actor voice	<i>um-√-a</i>
Patient subject	<i>√-aw</i>
Location subject	<i>√-ay</i>
Circumstance subject	<i>√-anay</i>

Second, and much more controversially, he labeled these affixes ‘first-generation forms’ and Wolff’s indicative voice markers ‘second-generation forms’, the terminology reflecting his claim that the traditional voice/focus markers of Philippine-type languages were originally nominalizers as claimed by Starosta et al. (1982), while the affixes in (6) exemplified the indicative mood in PAN verbs.

According to this interpretation AN has four primary branches: 1. Puyuma, 2. Tsou, 3. Rukai, and 4. Nuclear Austronesian (the rest). In Ross’s view PAN had a Philippine-type voice system, but one in which the verbal affixes were restricted to **<um>*, **-aw*, **-ay* and **-anay*, the more familiar voice markers of typical Philippine-type languages as reconstructed by Wolff (1973) (**<in>*, **-en*, **-an*, **Si-/Sa-*) functioning only as nominalizers. In his view, then, the complete absence of a Philippine-type verb system in Rukai must be due to loss, and the presence of the formally similar but morphologically distinct Philippine-type verb system in Proto-Nuclear Austronesian (PNAN) was due to the reinterpretation of original nominalizers as voice markers. Ross recognized that these claims raise crucial questions, and he tried to answer some of them as follows (2009: 303–304):

There are two alternative explanations of this state of affairs. Either (a) Puyuma has innovated by undoing the reanalysis of predicate nominalizations as verbs which had allegedly occurred by PAN times, or (b) Puyuma continues unchanged

2. For simplicity only the Realis forms of Ross’s (2009: 304) Table 4 are given here, and his *M-STEM* is altered to the more iconic *um-STEM*.

the state of affairs reconstructed for pre-PAN. If (b) is true, then the reanalysis of predicate nominalizations as verbs had not occurred in PAN, nor had it occurred in any interstage ancestral to Puyuma.

Answer (a) requires that precisely the verbal functions of second-generation forms which were gained in PAN were lost again in Puyuma. This is unlikely: we would expect Puyuma to preserve some reflex of the intervening PAN stage, but it doesn't. Answer (a) also requires that PAN (first-generation) undergoer-voice optative/hortative forms have extended their function in Puyuma to include the realis, displacing the PAN second-generation forms – a step which seems quite implausible. Answer (b) on the other hand requires no innovations.

In this statement Ross assumes that the voice markers as reconstructed by Wolff (1973) were verbal affixes that developed from pre-PAN nominalizers, and to account for the absence of verbal uses for their reflexes in Puyuma would violate Occam's razor, since this would entail a history of nominalizer to verbalizer to nominalizer rather than a simpler history of no change at all. He does not consider the possibility that the PAN affixes in question may have had *both* nominal and verbal functions depending on the larger syntactic context, in which case the verbal functions may well have been lost in Puyuma, as we *know* to be the case in various Malayo-Polynesian languages. His second assumption, that the forms in (6) could not plausibly have been non-indicative voice markers that replaced the original indicative voice markers in this language is not accompanied by any kind of argument showing why this change could not have occurred. All languages change over time, and given his own subgrouping assumptions Puyuma has been an independent branch of the AN language family for over 5,000 years – plenty of time for dramatic transformations of the morphosyntactic typology, as he himself implicitly assumes for Rukai, which has completely lost what he reconstructs as a Philippine-type voice system, along with the associated morphology.

Most critically, Ross's assumption that the verb system of Puyuma matches that of PAN is not based on comparative data of any kind. As he himself notes (2009: 305) "Puyuma, Rukai and Tsou do not reflect the PAN system and also have systems that have little in common with each other." This is a striking departure from normal application of the Comparative Method, and is completely at odds with the way the Philippine-type voice system of PMP has been reconstructed. Rather, his rejection of the possibility that PAN *-<in>, *-an, *-en, etc. could have had verbal uses is based entirely on negative evidence (2009: 304): "we would expect Puyuma to preserve some reflex of the intervening PAN stage, but it doesn't." Read: since Puyuma has no trace of verbal uses for its reflexes of these affixes the ancestral forms could never have been verbalizing. We must be reminded again that 'absence of evidence is not evidence of absence'.

To show how the Nuclear Austronesian hypothesis neglects relevant data that should force its adherents to reconsider the absolute position they take, we need only consider that the so-called ‘second-generation forms’ have both verbalizing and nominalizing properties in the majority of Philippine-type languages, both in Taiwan and in MP languages. This is illustrated in (7) with reflexes of *<in>, *-an, and *-en drawn from two languages that are only distantly related (Thao being the last surviving member of the Western Plains branch of AN languages in Taiwan, Tagalog an MP language):³

(7) Reflexes of *<in>, *-an and *-en in Thao and Tagalog

I. THAO

- (a) <in> as voice affix: *agqaruz* ‘current’ : *in-agqaruz* ‘was carried off by a current’
qtut ‘fart’ : *q<in>tut* ‘farted’
t<u>nun ‘to weave’ : *t<i>nun* ‘was woven’
- (b) <in> as nominalizer: *macay* ‘die, dead’ : *m<in>acay* ‘burial place, cemetery’
saran ‘path’ : *s<in>aran-an* ‘the place where someone walked’
pa-shizuq ‘put around the neck’ : *p<in>a-shizuq* ‘necklace’
- (a) -an as voice affix: *c<m>anit* ‘to weep’ : *canit-an* ‘be wept over, mourned’
iup ‘to blow’ : *iup-an* ‘to blow on something’
kuskus ‘foot/leg’ : *k<m>ay kuskus-an* ‘be hit on the foot or leg’
- (b) -an as nominalizer: *flhuq* ‘to wash, bathe’ : *flhu-flhuq-an* ‘washroom, bathing place’
kalhus ‘to sleep’ : *ka-kalhus-an* ‘sleeping place (not necessarily a bed)’
t<m>iktik ‘chop meat or vegetables with a cleaver’ :
ta-tiktik-an ‘cutting board’
- (a) -in as voice affix: *fariw-in* ‘be bought by someone’, *khlit-in* ‘be cut by someone’, *in-fari-n* ‘be blown by the wind’, *rinuz-in* ‘be shaken by an earthquake’
- (b) -in as nominalizer: *kan-in* ‘food’, *ushnaw-in* ‘thing liked or desired’

3. Thao data from Blust (2003), Tagalog data from several sources, including Schachter & Otones (1972) and English (1986).

II. TAGALOG

- (a) <in> as voice affix: *bilí* ‘purchasing price’ : *b<in>ilí* ‘bought’
l<um>agót ‘to snap or break’ : *l<in>agót* ~ *ni-lagót*
‘snapped or broken’
ubús-in ‘to devour, consume’ : *in-úbos* ‘used up’
- (b) <in> as nominalizer: *pukpók* ‘beating, pounding’ : *p<in>ukpók* ‘abaca
cloth made by pounding’
i-sigáy ‘put cooking utensils on a stove’ : *s<in>igáy*
‘stew of fish or meat’
tápay ‘dough’ : *t<in>ápay* ‘bread’
- (a) -an as voice affix: *bató* ‘stone’ : *bátúh-an* ‘throwing stones at one
another’
lakád ‘walking’ : *lakar-an* ‘to walk on something’
ñiwí? ‘wry face’ : *ñiwí-ñán* ‘make a bad face at
someone’
- (b) -an as nominalizer: *bató* ‘stone’ : *batuh-án* ‘stony place’
gizík ‘to thresh’ : *gizik-án* ‘threshing place’
gúlay ‘vegetables’ : *guláy-an* ‘vegetable garden’
lakád ‘walking’ : *lakar-án* ‘place to walk on’
i-sigáy ‘place cooking utensils on stove’ : *sigáy-an*
‘stove’
t<um>ágo? ‘to hide’ : *tagú?-an* ‘hiding place’
- (a) -in as voice affix: *dalá* ‘carried, brought’ : *dalh-in* ‘to carry, to bring
(emphasis on the thing that is brought)’
lagnát ‘fever’ : *lagnat-in* ‘to have a fever’
namnám ‘the taste of something’ : *namnam-in* ‘to
taste’
- (b) -in as nominalizer: *mag-áral* ‘to study’ : *arál-in* ‘something to study’
um-áwit ‘to sing’ : *awít-in* ‘song’
mag-lúto? ‘to cook’ : *lutú?-in* ‘the cooking’

Needless to say, one and the same affixed form can function as a verb or a noun depending on the larger syntactic context in which it is placed, as with *s<in>aran-an* in Thao *mun-saháy s<in>aran-an caycuy tu maku-na-nay uan* [go-there walked-on/across 3pl. TU come from still] ‘They took the path they had come by’ (verbal) vs. *m-ihu a s<in>aran-an yanan sapaz* [2sg gen. LIG place-of-walking have footprint] ‘the place where you walked has footprints’ (nominal).

Many other languages have bifunctional reflexes of these affixes like those of Thao and Tagalog. In some languages one affix may be monofunctional, but others bifunctional. What is especially noteworthy is that reflexes of these markers in a number of MP languages that have reduced or completely lost the Philippine-type

voice system now have only nominalizing functions. Here are a few examples to show that affixes that once functioned in both verbal and nominal contexts now function only in the latter:

1. Hoava and other languages of New Ireland and the Western Solomons

Hoava is an Oceanic language spoken near the western end of the island of New Georgia in the western Solomons (Davis 2003). Although Hoava has completely lost its inherited Philippine-type voice system, it retains reflexes of both *<in> and *-an, and the sole remaining function of both affixes is to create deverbal nominals. These are often abstract nouns with <in> and locative nouns with -an, as shown in (8) and (9); in Hoava, as in other AN languages, reflexes of *<in> are inserted before the first vowel of a base, surfacing as an infix in consonant-initial stems but a prefix in vowel-initial stems:

(8) Hoava deverbal nouns formed with a reflex of *<in>

<i>asa</i> 'to grate'	:	<i>in-asa</i> 'pudding of grated cassava'
<i>babana</i> 'to tow'	:	<i>b<in>abana</i> 'load, towed object'
<i>bukulu</i> 'defecate'	:	<i>b<in>ukulu</i> 'feces'
<i>haqala</i> 'to run'	:	<i>h<in>aqala</i> 'act of running'
<i>ropa</i> 'to clear land'	:	<i>r<in>opa</i> 'cleared land'
<i>to</i> 'be alive'	:	<i>t<in>o</i> 'life'

(9) Hoava deverbal nouns formed with a reflex of *-an⁴

<i>bagere</i> 'hunt possums'	:	<i>bagere-ana</i> 'possum hunting place'
<i>deo</i> 'clam'	:	<i>de-deo-na</i> 'clam bed'
<i>eko</i> 'lie down'	:	<i>ek-eko-ana</i> 'sleeping place, bed'
<i>igunu</i> 'to play'	:	<i>ig-igunu-ana</i> 'sportsground'
<i>qato</i> 'tree'	:	<i>qato-qato-ana</i> 'forest'

These nominalizing reflexes of *<in> and *-an are worth citing, because a similar situation is found in a number of the languages of New Ireland and the western Solomons, but in the same languages the PMP instrumental/benefactive prefix *Si-, the actor voice infix *<um> and the patient voice suffix *-en have vanished virtually without a trace, as have all verbal functions of *<in> and *-an, showing the clear evolution of earlier bifunctional morphemes into reflexes that are exclusively nominalizers.

4. PMP *-an appears as Hoava -ana (-na in some forms) through a regular sound change that added echo vowels to consonant-final bases (PMP *hikan > *igana* 'fish', *ma-nipis > *manivisi* 'thin, of materials', *ənəm > *onomo* 'six', etc.).

2. Kelabit

This non-Philippine-type language, spoken in northern Sarawak, retains reflexes of *<um>, *<in>, *-an and *-en. While reflexes of *<um>, and *-en are used only in verbal constructions, reflexes of *<in> are used both verbally and nominally, and the reflex of *-an is used only to form nouns of location, as shown in (10):

(10) Reflexes of *-an in Kelabit

<i>guta</i> ‘wade across a river’	:	<i>gəta-an</i> ‘fording place in a river’
<i>m-irup</i> ‘to drink’	:	<i>rup-an</i> ‘watering hole for jungle animals’
<i>nalan</i> ‘to walk’	:	<i>dələn-an</i> ‘path made by repeated walking over the same course’
<i>nələn</i> ‘to swallow’	:	<i>tələn-an</i> ‘throat, gullet’

3. Wolio

Anceaux (1952: 30) gives <in> as a noun formative only. He provides only two forms, and a search of his later dictionary (Anceaux 1987) shows no others:

(11) Reflexes of *<in> in Wolio

<i>kande</i> ‘to eat’	:	<i>k<in>ande</i> ‘food, dish, meal, boiled rice’
<i>tauraka</i> ‘to leave behind, to put down’	:	<i>t<in>auraka</i> ‘heritage’

This survey has been brief primarily because of the difficulty of obtaining relevant information from available sources, and because sources exist for only a small percentage of AN languages. Nonetheless it is sufficient to remind us of the truism that languages change over time, some more radically than others. In these cases the PMP focus affixes, which had both verbal and nominal uses, have lost their verbal functions, and now exist only as nominalizers. And once again, it is important to state that the proclivity of a language to change has no connection with its phylogenetic status. In other words, a ‘higher level’ language like Rukai, Tsou, or Puyuma is no less likely to transformative change than a ‘lower level’ language like any member of the MP group.

3.1.1 Implications for Puyuma

The demonstration that some MP languages have transformed the bifunctional voice markers of PMP into monofunctional nominalizers is sufficient in itself to show that a similar process could have happened in Formosan languages like Puyuma. However, the argument that such a functional reduction actually did take place in Puyuma is stronger than this. First, from a purely theoretical standpoint it would be strange for a language to innovate a perfective marker with nouns before it ever marked perfective in verbs. Puyuma has a reflex of PAN *<in> in forms such as the following (Cauquelin 2015):

(12) Reflexes of *<in> in Puyuma

<i>bəray</i> ‘to give’	:	<i>b<in>əray ~ ni-bəray</i> ‘a gift’
<i>dapal</i> ‘sole of foot’	:	<i>d<in>apal-an</i> ‘footprints’
<i>daway</i> ‘to create, produce’	:	<i>d<in>away</i> ‘creation’
<i>kasu</i> ‘to bring’	:	<i>k<in>asu-an</i> ‘things brought’
<i>ʔasi</i> ‘to melt in the fire’	:	<i>in-ʔasi</i> ‘small beads baked in the fire’
<i>mə-riap</i> ‘sow by broadcasting’	:	<i>nanku ni-riap-an</i> ‘my seeds’ (= ‘the seeds I have sown’)
<i>tənun</i> ‘weave’	:	<i>t<in>ənun-an</i> ‘woven material’
<i>tukuɟ</i> ‘to strengthen, buttress’	:	<i>t<in>ukuɟ-an</i> ‘a strengthened thing’
<i>ʔapa</i> ‘roast far from the fire’	:	<i>ʔ<in>apa</i> ‘roasted’; roasted millet’
<i>ʔufus</i> ‘darn roughly sewn clothes for mourners’	:	<i>ʔ<in>ufus</i> ‘mourning clothes’

Although most of these words are nouns that describe the product of some action, *ʔ<in>apa* is glossed as marking both an achieved state (‘roasted’) and a product of the action of roasting (‘roasted millet’). This is, of course, familiar from MP languages with a Philippine-type voice system, and is most simply explained as a residual verbal use of an affix that has largely lost its earlier verbal functions. Under the NAn hypothesis *<in> was a perfective marker only with nouns (as it is in Puyuma) that eventually evolved into a perfective marker for verbs. While the nominalizing functions of <in> emerge fairly clearly from its use as a perfective marker in verbs (achieved state >> product of an achieved state), it is far more difficult to find a plausible reason why a perfective marker would have been used in nouns that were eventually reinterpreted as verbs in PNAn. Moreover, some examples of Tamalakaw Puyuma *-an* are found in constructions that are unequivocally verbal, as in the following, all of which are given as imperatives (Tsuchida 1980, page numbers in parentheses):

(13) Instances of verbal *-an* in Tamalakaw Puyuma

<i>vuiH-an</i> ‘eject from the mouth’	(232)
<i>TimaH</i> ‘price’ : <i>TimaH-an</i> ‘to buy/sell’	(251)
<i>atez-an</i> ‘escort a drunk home’	(252)
<i>seHiz-an</i> ‘hang s.t. which has no hook’	(257)
<i>p-antip-an</i> ‘to insert’	(262)

Although none of these examples are given in sentences, Tsuchida (1980: 183) is explicit in stating that the above bases with *-an* are verbs: “Verbs in the vocabulary are given in an inflected form – whenever available in durative realis (approximately equivalent to imperfective or progressive) of actor focus (AV); in some cases in punctual realis (approximately equivalent to perfective). This inflected form is immediately followed in parentheses by the imperative form of object

focus (PV), or, if OF is not available, by that of AF.” All of the above examples from Tsuchida (1980) are found in parentheses immediately following what he calls ‘the inflected form,’ and there is no clear way to avoid interpreting them as verbal, since imperatives admit no other option.

The obvious objection to using these examples as evidence for *-an* as a verbalizer is that imperatives are not part of the indicative voice-marking system, and the NAn hypothesis is concerned only with changes in the morphological marking of voice distinctions in the latter (from **<um>*, **-aw*, **-ay*, **-anay* to **<um>*, **-en*, **-an*, **Si-/Sa-*). However, the morphology of the Tamakakaw *-an* imperatives shows that they were not originally part of a non-indicative voice-marking system, but reflect affixes that were either indicative voice markers or (in Ross’s interpretation) nominalizers. Comparative evidence for imperative marking in a number of other AN languages suggests that the Puyuma *-an* imperative in fact reflects **-en*, with neutralization of contrast between the two suffixes (Blust 2013: 502–503).

To maintain the claims of the NAn hypothesis, then, it appears impossible to avoid concluding that PAn **-en* was both: (a) a formative for patient nominals, and (b) a marker of imperative (but not indicative) verbs. This is a decidedly strange duality of function, and one that must be justified by some type of argument. If **-en* had only function (a) in PAn we need to explain how *any* kind of verbal function might have evolved from it. Structures like **kaen-en ni X* which under the interpretations imposed by the NAn hypothesis could mean only ‘X’s eating/X’s food’ would almost certainly evolve *first* into indicative verbs meaning ‘is eaten by X’, since they have the same structure and present an inherent ambiguity, whereas the imperative constructions with **-en* are structurally quite different. In conclusion, adopting the perspective of the NAn hypothesis for the sake of discussion, the imperative functions of **-en* would most likely develop *later* than the use of this suffix in indicative verbs (or at least not before them!). On the other hand, if PAn **-en* was both a voice marker in indicative verbs and a nominalizer, its evolution into an imperative marker would be unsurprising.

3.2 The history of Malayo-Polynesian, 2: The Ergative Austronesian hypothesis

Aldridge (2016: 31) adopts Ross’s Nuclear Austronesian hypothesis whole cloth, but rejects his view that PAn had a Philippine-type voice system that was lost in Rukai. Rather, like Starosta (1995) she maintains that Rukai is a primary branch of the AN family as against all other languages, and that PAn had an accusative alignment system similar to that of Rukai. The basis for this interpretation is that Rukai lacks the structure of a Philippine-type voice system (in Aldridge’s terms a system of ergative alignment), as well as lacking many of the morphological components

of widely-distributed Philippine-type languages. She thus posits an AN family tree that splits at the top into 1. Rukai, and 2. the rest, which Starosta (1995) called 'F1' and she calls 'Ergative Austronesian.' However, the claim that because Rukai does not have a Philippine-type voice system its ancestral forms could not have had such a system is a by now familiar 'absence of evidence = evidence of absence' argument, and we have already seen that arguments based on negative evidence can never be definitive.⁵ Aldridge's arguments for an Austronesian family tree with two primary branches – Rukai and Ergative Austronesian – are entirely theoretical, and while these may be of some interest in their own right, they *assume* rather than demonstrate the subgrouping proposed, and simply provide a road map of how the transition from an accusative to an ergative alignment system could have taken place, if in fact that transition actually happened. The problem for the historical linguist is knowing with any certainty that it did happen.

As one reason for skepticism about subgrouping arguments based on negative evidence, we can cite the fact that many descendants of PMP, which had an undisputed Philippine-type verb system, have lost nearly all traces, or in some cases, all traces of that system. In using MP as a test case for the directionality of change at a higher node in the AN family tree we can start with the accusative Chamic languages of mainland Southeast Asia. As a result of centuries of acculturative adaptations to their Mon-Khmer neighbors (and others on Hainan island), these languages preserve none of the PMP voice markers as active affixes. The once centrally important affixes *Si-, *-an and *-en have disappeared, so far as has been determined to date, without a trace. The same can be said for the high-frequency actor voice marker *<um>, which plays such a central role in many Formosan and Philippine languages, and in others outside the central block such as Toba Batak (van der Tuuk 1971), Old Javanese (Zoetmulder 1982), and Chamorro (Topping 1973; the role of *<um> in Malagasy, by contrast has been almost completely lost). However, a slight trace remains, as Thurgood (1999: 298) reconstructs Proto-Chamic *minum; *minam 'to drink' as a monomorphemic word which reflects PMP *um-inum 'AV-drink'. In addition, although Thurgood reconstructs Proto-Chamic *patah 'vomit', which cannot regularly reflect PMP *um-utaq 'to vomit', Haroi *matah* 'vomit' appears to reflect the PMP form, and similar traces of *<um> in languages where it has otherwise completely disappeared are found in reflexes

5. Reflexes of *Sa-, *<in> and *-an are in fact used in Rukai, but only to form nouns. However, the other key elements of typical Philippine-type voice systems, namely reflexes of *Si-, *<um> and *-en, are absent, as are the affixes in (6). The grounds for arguing that the first split within the AN language family separated Rukai from the rest are thus partly structural (absence of a Philippine-type voice system), and partly functional (so-called 'second generation forms' used only as nominalizers). In either case, the appeal is to negative evidence (absence of 'focus', and absence of so-called 'second-generation forms' used as verbs).

of PMP *um-utaq (AV-vomit), as monomorphemic words in various languages of eastern Indonesia, Melanesia, and Micronesia. One might argue that the Chamic languages present a special case, since they have undergone centuries of areal adaptation to their Mon-Khmer neighbors, most of which have minimal morphology. However, the particular factors that cause affixes to be lost are not restricted to contact-induced changes, as will be seen in other examples below.

In finding cases like these it is easy to nourish the hope that a language like Rukai, which is situated in the heartland of Philippine-type voice systems, will show at least a trace or two of being descended from a language that was typologically much more like most other Formosan aboriginal or Philippine languages than Rukai is today. However, there is no guarantee that this will be the case if the typological divergence of Rukai from more typical languages of its area is due to a history of extreme innovation. To show this we need only look at MP languages that retain no known traces of any kind of an earlier Philippine-type voice system.

The Polynesian languages provide a good example. These languages are among the most intensively studied members of the Oceanic subgroup, with grammars and dictionaries available for many of them. Yet there is not a trace, even in fossilized morphemes of *Si-, *<um>, *<in>, *-an or *-en in Polynesian languages. The closest thing we can find to trace elements of earlier voice markers is perhaps the very tenuous morphological relationship that Clark (1991: 78–83) proposed between the Proto-Polynesian nouns *fiŋota ‘shellfish’ and *faŋota ‘fishing?’, the first of which he traced to Proto-Oceanic *piŋota ‘shellfish’ (misprinted *piŋonta in the original), hypothesized as being from *p<in>aŋota with idiosyncratic deletion of *-na-*. However, attributing this reconstruction to Proto-Oceanic is clearly problematic, since Proto-Oceanic *<in> was preserved as an active VC morpheme, and had not evolved into the pattern of ablaut seen in these forms. The only other even remote possibility that any PMP voice marker is preserved as a fossil in some lexical item is the Proto-Polynesian word *kano ‘flesh, kernel, seed’, which may reflect PMP *kaen-en ‘be eaten by someone; cooked rice’.

Other MP languages that preserve neither active nor fossilized morphological evidence of descent from an ancestor with a Philippine-type voice system are not hard to find. A few chosen at random would include the following:

1. Enggano

This is a highly aberrant language of western Indonesia, the phylogenetic position of which is still being debated (Edwards 2015). It retains none of the voice-marking affixes of PMP, and appears to have no lexical items which contain fossilized traces of them (Kähler 1987).

2. Taba

Taba, in the South Halmahera-West New Guinea (SHWNG) group (Bowden 2001) retains none of the voice-marking affixes of PMP, and so far as we have been able to determine, has no lexical items with fossilized traces of them. The same apparently can be said for all of the 20–30 languages in this group, including Buli and other languages of southern Halmahera, and Numfor and other SHWNG languages of western New Guinea.

3. Yapese

This is an Oceanic language spoken in western Micronesia, which is both typologically aberrant and very difficult to classify (Ross 1996). No PMP voice marker survives in Yapese, even in trace form in lexical items such as ‘drink’ or ‘vomit’ (Jensen 1977a, b).

4. Mota

This Oceanic language of the Banks islands in northern Vanuatu has no known traces of the PMP voice markers, missing even a reflex of *um-utaq ‘to vomit’ (Codrington & Palmer 1896; Tryon & Hackman 1983). The same may be true for all of the more than 100 languages of Vanuatu for which we have basic descriptive data if some apparent reflexes of *um-inum ‘to drink’ (Nasawa *min*, Apma *mni*, Rano *mini*, Vao *muni*, Fali *minu*, etc.) turn out to have another explanation, given that no Oceanic language outside Vanuatu contains *m-* in this form.

Over 100 other languages in eastern Indonesia have no active reflexes of PMP voice markers, and are devoid even of fossilized traces except in reflexes of *um-utaq ‘to vomit’, which provides the sole evidence of the earlier presence of *<um>.

5. Manggarai

In still other languages the very process of affixation has been lost. Verheijen (1977) reported this for Manggarai, spoken in western Flores in the Lesser Sunda islands of Indonesia, and a similar reduction to zero-affixation apparently is found in other languages of the area, as Keo (Baird 2002). In the case of Keo, *minu* ‘to drink’ and *muta* ‘to vomit’ provide evidence that this language is descended from one in which *<um> was once an active affix. However, in at least the standard dialect of Manggarai even traces like this are lacking.

What conclusion can we draw from this brief survey of the history of Philippine-type voice affixes in MP languages? There seems to be no way to reasonably deny that a number of languages have lost not only the structural properties of Philippine-type languages, but also the morphemes themselves, and hundreds of others have lost all traces of the voice markers except where *<um> has been fossilized in a single word or two, most commonly a reflex of PMP *um-utaq ‘to vomit’, or *um-inum ‘to drink’.

Once this is recognized we must ask why Rukai, which shows a similar lack of structural similarity to Philippine-type languages, could not also have diverged as a result of a long history of innovation. It cannot be emphasized too strongly that the only thing that sets Rukai apart from syntactically innovative MP languages is the testability of the claim that it never had a focus system. For MP languages, no matter how much they differ from PMP, the matter is closed: they are descended from an ancestor with a Philippine-type voice system, whether they show any trace of it today or not. As explained in diagram (ii) in (3), on the other hand, the matter is different at the ‘top’ of a phylogenetic grouping. There is nothing about Formosan languages that would cause them to differ from MP languages in terms of their susceptibility to change, but if a Formosan language were to undergo radical syntactic change of a degree similar to that in MP languages like Enggano, Taba, Yapese, or Hawaiian the claim that its verb system is conservative rather than innovative would simply be untestable.

3.2.1 *The temporal asymmetry of gains and losses*

Before concluding this section it will be worth raising one more point of method in connection with the phylogenetic position of Rukai. In claiming that Rukai is innovative we need only assume that a complex but interconnected system of voice marking was lost either by the disappearance of morphemes that once manifested this function, or by the loss of the verbal properties of those original voice markers that were retained. That this is possible has been demonstrated empirically with several MP cases cited above. It cannot be overly stressed that the value of these cases is: (1) that the reconstruction of a Philippine-type voice system in PMP is non-controversial; and (2) that the complete loss of this system is attested in at least dozens of daughter languages, with a nearly complete loss (leaving only a trace of *<um> in the verb ‘to vomit’) in hundreds of others. By contrast, the claim that Rukai is conservative makes at least the following assumptions:

- (14) Implications from Rukai as the first offshoot of PAN
 - a. *<in> evolved from marking perfectivity in nouns to marking perfectivity in verbs and in deverbal nouns that describe the product of an activity.
 - b. *-an changed from a marker of locative nouns to a marker of locative voice or locative nouns.
 - c. The affixes *<um>, *-en, *-aw, and *-ay were innovated *ex nihilo* in what Aldridge (2014, 2016) calls ‘Ergative Austronesian’ (EA).

This last point is important enough to justify some emphasis. Surprisingly, neither Starosta (1995) nor Aldridge (2014, 2016) comments on the historical implications of their claim that Rukai is conservative. The affixes *Si-, *<um>, *-en, *-aw,

*-ay and *-anay are not found in Rukai, but must be assigned to Ross's Nuclear Austronesian, and at least some of them (*<um>, *-en, *-aw, *-ay) to Aldridge's Ergative Austronesian, since reflexes are found in two or more primary branches of her EA group (*<um> in Tsou, Puyuma and NAn, and *-en, *-aw and *-ay in Puyuma and NAn (see § 4.1.1 below).

As already noted, Rukai reflects none of these forms in a verbal function, and this has been used as 'evidence' for a primary division of AN into *Rukai vs. the rest*. However, unless it was lost the missing voice morphology of Rukai (*Si-, *<um>, *-en) also shows that this language provides incomplete evidence of the raw material from which the Philippine-type voice system could have been built. Needless to say, languages do not innovate affixes spontaneously. Rather, affixation is the long-term result of free morphemes gradually losing their independent status and meaning and becoming part of the same phonological word as the free morphemes to which they are attached. This process first requires the raw material out of which affixes might be created, and then enough time for this process to work its way through to completion. This is particularly true for infixes, which may require even more time than other affixes to acquire their status as part of the morphology of a language.

If Rukai separated from Proto-Ergative Austronesian (PEA) at least 5,000 years ago and PEA possessed the affixes *<um> (Tsou, Puyuma, NAn), *-en (Puyuma, NAn), *-aw (Puyuma, NAn), and *-ay (Puyuma, NAn) in *any* function one must ask where these key grammatical markers came from: if they were already present in PEA immediately after the split from Rukai they must have also been present in proto-Rukaic and subsequently lost. Without this assumption we are left with what is clearly an unsatisfactory theory of spontaneous generation of key grammatical elements.

Stated differently, there is an important and previously ignored temporal asymmetry between the implications of assuming that Rukai is syntactically *conservative* vs. assuming that it is syntactically *innovative*. If Rukai is innovative it has had over five millennia of independent history to evolve and change until no trace of an earlier Philippine-type voice system remains – even longer than the morphosyntactically innovative MP languages mentioned above. If it is conservative, however, one must somehow explain how the full set of what Ross (2009, 2012) has called 'second generation' voice affixes could have appeared essentially out of nothing at the time PAn split into Proto-Rukaic and PEA, or shortly thereafter. Needless to say, the latter scenario runs against virtually everything linguists have learned about grammatical change.

What this section has shown is that no morphosyntax-based argument for higher-level subgroups in AN proposed so far stands up to close scrutiny. While positive arguments showing the errors of the Nuclear Austronesian and Ergative

Austronesian hypotheses are welcome, strictly speaking they are not necessary, since both of these hypotheses are based on negative evidence, and it is known that this method of inference can never be conclusive. Moreover, the assumption that PAN could not have been a Philippine-type language because Rukai lacks this type of structure is inconsistent with the clear evidence that PMP was a Philippine-type language and that many of its descendants have not only lost the structural properties of a Philippine-type verb system, but have also lost much – or in some languages all – of the morphology commonly associated with these properties.

4. *-an before *-en? Implications of the morphosyntax-based subgroupings for Austronesian historical morphology

In what follows, we reconsider the basic assumptions of the EA and NAn hypotheses using Formosan data. Particular attention is devoted to the chronology of *-en as implied by the two hypotheses, and it is demonstrated through this case that diachronic inferences built on negative evidence are inescapably inconclusive. Before discussing *-en, we repeat the main assumptions of the two hypotheses in (15), with the evidence that motivates them summarized in (16).

(15) Main assumptions of the EA and NAn hypotheses

- a. The EA hypothesis: The Philippine-type voice system emerged after Rukai split off from PAN. The nominalizers <in>, -an, and sa- in Rukai have never been reanalyzed as voice affixes.
- b. The NAn hypothesis: Prior to N-into-V, the AN voice system was expressed by the so-called ‘first-generation affixes.’ At this stage, the ‘second-generation affixes’ *<in>, *-en, *-an, and *Si-/Sa- had exclusively nominalizing functions.

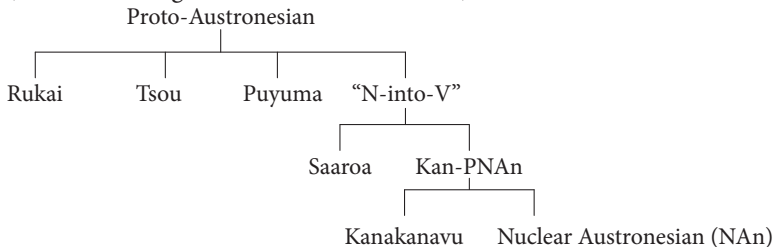
(16) Evidence that motivates the EA and NAn hypotheses

- a. The EA hypothesis: All higher-order AN languages except Rukai exhibit traits of a Philippine-type voice system.
- b. The NAn hypothesis: Higher-order AN languages except Rukai, Tsou, and Puyuma employ the second-generation affixes (reflexes of *<in>, *-en, *-an, and *Sa-/Si-) both nominally and verbally, while members of this set of affixes that occur in these three exceptional languages are exclusively nominalizers.

As already noted, based on the observations in (16b) Ross (2009, 2012) and Aldridge (2016) argue for the separation of Rukai, Tsou, and Puyuma from the rest of the AN languages. This proposal implies a family tree with three single-member primary branches plus an innovative branch that contains all other languages, which we will call the “N-into-V” subgroup in the following discussion.

More recent research has shown the functional variations of the ‘primary affixes’ in Formosan languages to be more complex than previously thought. Zeitoun & Teng (2016) report that Saaroa and Kankananavu exhibit partial noun/verb homophony in their voice paradigms, Saaroa reflecting only *<in>, and Kankananavu reflecting only *<in> and *-en with verbal functions. To accommodate these facts they propose that N-into-V developed in the order *<in> > *-en > *-an, *Si-/Sa-, with these two languages splitting off from Proto-“N-into-V” before this transformation was complete. In this view, Saaroa is assigned to a first node under the “N-into-V” subgroup, defined by the putative reanalysis of *<in> as verbal, and Kankananavu at the next level, defined by the putative reanalysis of *-en as verbal. The NAn subgroup is placed at the bottom, defined by the putative reanalysis of the nominalizers *-an and *Si-/Sa- into voice affixes (Zeitoun & Teng 2016). This revised proposal is illustrated in (17).

- (17) Austronesian highest-order subgrouping under the N-into-V hypothesis (Zeitoun & Teng 2016, based on Ross 2009)



N-into-V as a subgrouping criterion has profound implications for early AN morphology, in particular, the chronology of the second-generation affix *-en. Given its presence in most primary branches of the family as defined by exclusively shared phonological innovations, it was commonly assumed that *-en was reconstructable to PAN (Wolff 1973; Starosta et al. 1982; Blust 2013). However, despite the claim that noun/verb homophony in NAn languages derives from the reanalysis of the full set of second-generation affixes, i.e. *-en, *-an, *Sa-/Si-, and *<in>, into voice affixes, a reflex of *-en is not attested in Rukai, Tsou, or the most well-studied Puyuma dialect, Nanwang. A reflex of *-en is also unattested in Saaroa, the first offshoot of the “N-into-V” subgroup (Zeitoun & Teng 2016). This leaves us with little positive evidence to reconstruct PAN *-en under the morphosyntax-based subgrouping.

The absence of *-en* in non-NAn languages can be interpreted in either of two ways, as shown in (18):

- (18) a. [*Hypothesis A*]: *STEM-en was a post-PAn innovation.⁶
 b. [*Hypothesis B*]: *STEM-en existed in PAn, but was independently lost in Rukai, Tsou, and Puyuma, as well as in Saaroa.

Both Ross (2012) and Aldridge (2016) adopt Hypothesis A, assuming that PAn *-an was a general undergoer nominalizer used in both patient and locative nominalizations prior to the emergence of a specific patient nominalizer *-en, as illustrated in (19).

- (19) PAn morphology under Ross (2012: 1264, 1268–1269) and Aldridge (2016: 36–37)

	AV	PV	LV	CV
Neutral (N)	-	*-an		*Sa-/Si-
Perfective (N)	-	*<in>..-an		*Sa-/Si-

Under this reconstruction, PAn and all daughter languages that split off before *-en emerged employ only four primary affixes: actor voice (AV) *<um>, general undergoer nominalizer *-an, instrumental nominalizer *Sa-/Si-, and the perfective marker *<in>.

The goal of the following subsections is to examine the validity of Hypothesis A by reconsidering evidence for *-en as a secondary innovation, drawing on comparative data from both non-NAn and NAn languages.

4.1 Against *-en as a post-PAn innovation

As is well known, in nearly all Philippine-type languages the PV affix *-en* has a zero allomorph in perfective forms, and as a result in these constructions <in> functions as a portmanteau affix with both aspectual and voice-marking functions, as stated in (20).

- (20) Distributional restriction of *-en*
-en is obligatorily realized as a zero allomorph when combining with the perfective marker <in>.

This special constraint is exemplified in (21) and (22) with data from Seediq and Tagalog. It can be seen that the patient nominalizer *-en* in irrealis patient nominalizations is overt, and combines with other morphology (i.e. Ca-reduplication), but in perfective patient nominalizations it is obligatorily null.

6. Given the findings of Zeitoun & Teng (2016), Hypothesis A can be restated as: *STEM-en was an innovation after Saaroa split off from the “N-into-V” subgroup.

(21) Seediq

Verbal	Irrealis patient nominalization		Perfective patient nominalization	
√-un (<*-en)	C.RED-√-un		<in>√-Ø	
sebuc-un 'to beat'	s-sebuc-un	'thing to be beaten'	n-sebuc-Ø	'thing beaten'
huqil-un 'to kill'	h-huqil-un	'one to be killed'	n-huqil-Ø	'one killed'
bari-un 'to buy'	b-bariq-un	'thing to be bought'	b<n>ari-Ø	'thing bought'

(22) Tagalog (English-Tagalog Online Translator)

Verbal	Irrealis patient nominalization		Perfective patient nominalization	
√-in (<*-en)	CV-√-in		<in>√-Ø	
bilh-in 'to buy'	bi-bilh-in	'thing to be bought'	b<in>ili-Ø	'thing bought'
linis-in 'to clean'	li-linis-in	'thing to be cleaned'	ni-linis-Ø	'thing cleaned'
lutu?-in 'to cook'	lu-lutu?-in	'thing to be cooked'	l<in>utó?-Ø	'thing cooked'

Given its appearance in different primary branches of AN, it is reasonable to assume that this morphological restriction was present from the time that *-en first appeared. With this special constraint in mind, particular attention will be paid to patient nominalization morphology in Puyuma, Rukai, and Saaroa in the following subsections, followed by a discussion of their implications for the N-into-V hypothesis.

4.1.1 Patient nominalization in Puyuma: Tamalakaw and Katripul vs. Nanwang

Nanwang Puyuma shows no evidence of *-en. However, some instances of -en are found in the Tamalakaw and Katripul dialects, based on limited data in Tsuchida (1980), Teng (2009), and Ross (2009), as seen in (23):

(23) Patient nominalization in Tamalakaw (TK) and Katripul (KP)*

Root	Perfective patient nominalization		Irrealis patient nominalization	
√	<in>√		Ca-√-en	
TK kerutr	k<in>erutr	'thing dug'	ka-kerutr-en	'thing to be dug'
ekan	in-kan	'thing eaten'	a-kan-en	'thing to be eaten'
KP kerutr 'dig'	k<in>erutr	'thing dug'	^ka-kerutr-en	'thing to be dug'
kezeng 'pull'	^k<in>ezeng	'thing pulled'	ka-kezeng-en	'thing to be pulled away'

* Due to gaps in attestation, words marked with ^ are manufactured forms based on the descriptions in Ross (2009: 308–309), and Teng (2009: 825), where Ca-√-en is the general pattern for the Katripul non-perfective patient nominalization. The forms ka-kerutr-en and a-kan-en are from Tsuchida (1980), k<in>erutr is from Teng (2009: 825), and ka-kezeng-en is from Ross (2009: 308).

As shown above, in these dialects the patient nominalizer *-en* appears only in irrealis forms, while perfective nominalizations show a zero allomorph, as in *k<in>erutr* ‘thing dug.’ This *-en*: zero contrast between irrealis and perfective nominalizations is strikingly parallel to the similar relationship in Seediq and Tagalog (21)–(22), and strongly suggests the presence of **-en* in pre-Puyuma.

Ross (2012) discusses the presence of *-en* in the Katripul *Ca-√-en* pattern, and notes that it is unique outside NAn languages. To account for this observation, he proposes (1) that productive use of the realis patient voice affix **-en* in PNA derives from pre-PNA **Ca-√-en*, which had a low functional load, and (2) that PNA **√-en* was initially a finite verb formed by analogy from a pre-PNA irrealis patient nominalization **Ca-√-en*, which was reanalyzed again into a realis patient nominalizer via (independent) analogical back-formations in languages like Paiwan and Pazeh (Ross 2012: 1268–1269).

However, the presence of *-en* in the Katripul Puyuma *Ca-√-en* pattern and its common use in NAn languages does not require such a complex historical scenario, with its unlikely sequences of nominalizer to verb to nominalizer, or its speculative parallel analogies in distinct languages. Instead, it follows straightforwardly from Hypothesis B that PAn **-en* was retained in pre-Puyuma, and then lost in Nanwang. With this analysis, the distributional asymmetry between the irrealis and perfective morphology of Tamalakaw and Katripul *-en* follows directly from the morphological constraint in (20), just as in NAn languages.

The advantage of the present proposal lies not only in its simplicity and consistency with other higher-level languages, but also in its ability to explain why *<in>√* as a structure that lacks an overt patient nominalizer is employed for perfective patient nominalizations in Tamalakaw and Katripul Puyuma.

If this proposal is not adopted the *<in>√* pattern in Tamalakaw and Katripul would have to be analyzed with no zero allomorph of **-en*, but merely a combination of the perfective marker *<in>* with the root. However, this analysis is difficult to maintain, as it implies that non-NAn languages employed a zero morpheme for patient nominalization prior to the emergence of **-en*, which is highly problematic, since: (i) it suggests a morphological ambiguity in a bare stem, representing both an Actor voice imperative, and a patient nominalizer; (ii) it implies that all types of nominalizations and voice morphology are derived by adding an additional affix to patient nominalizations; and (iii) it forces the analysis that *-en* in Katripul and Tamalakaw is an allomorph of the zero-formed patient nominalizer, which only appears in irrealis nominalizations.

In sum, treating the contrast between perfective and irrealis morphology in Tamalakaw and Katripul as evidence for pre-Puyuma **-en* provides a simpler and more coherent account for the observed phenomena. In what follows, we turn to

patient nominalization morphology in Nanwang Puyuma, which lends additional support to this analysis.

4.1.2 *Tamalakaw and Katripul -en vs. Nanwang -an*

Synchronically, Nanwang Puyuma has no reflex of *-en. Given the presence of -en in Tamalakaw and Katripul, this gap in Nanwang is naturally explained as a product of loss. That the absence of -en in Nanwang is not a retention but rather an innovation is supported by its productive use of an <in>√ pattern in perfective patient nominalizations, as reported in a series of works, including Ogawa & Asai (1935), Huang (2000), Teng (2007, 2009), and Chen (n.d.), as exemplified in (24).⁷

(24) Two interchangeable structures of patient nominalization in Nanwang Puyuma

Root		Perfective nominalization		Irrealis nominalization	
√		<in>√		Ca-√-an	
deru	'cook'	d<in>eru	'thing cooked'	da-deru-an	'thing to be cooked'
trima	'buy'	tr<in>ima	'thing bought'	tra-trima-an	'thing to be bought'
salem	'grow'	s<in>alem	'thing grown'	sa-salem-an	'thing to be grown'
base	'wash'	b<in>ase	'thing washed'	ba-base-an	'thing to be washed'

Given the assumption that the <in>√ pattern involves a zero-marked patient nominalizer as discussed above, the <in>√ structure in Nanwang is best accounted for as *fossilized morphology* that derives from the constraint in (20), which gave rise to an <in>√-∅ pattern before the loss of *-en. This analysis is further supported by the presence of -en in Tamalakaw and Katripul and their use of the same <in>√-∅ pattern in perfective patient nominalizations, as previously shown in (23).

A careful comparison of patient nominalization morphology in Tamalakaw and Katripul with that in Nanwang suggests that the loss of -en triggered the locative nominalizer -an to fill the resulting gap: Nanwang exhibits a special morphological alternation in perfective patient nominalizations, namely that the <in>√ pattern can freely alternate with an <in>√-an pattern without semantic consequences (Teng 2007: 199–200; Chen n.d.), as shown in (25)⁸:

7. Ogawa & Asai (1935: 301) report five examples of perfective patient nominalization in Nanwang, with the structure <in>√: *in-i-kan* 'food eaten', *d<in>awai* 'thing made', *b<in>rai* 'thing given', *b<in>usu?* 'clothes washed', and *p<in>ana?* 'thing shot'.

8. Nevertheless, it is noteworthy that two Nanwang consultants felt that <in>√-an (as opposed to <in>√) refers to an object produced by an event that was earlier in time. For example, *d<in>eru-an* 'what had finished cooking awhile ago' (e.g. the object had been cooked and put

(25) Two interchangeable patterns in patient nominalization in Nanwang

Root		Perfective nominalization		Irrealis nominalization	
na'u	'see'	n<in>a'u(-an)	'thing seen'	da-deru-an	'thing to be cooked'
trakaw	'steal'	tr<in>akaw(-an)	'thing stolen'	tra-trima-an	'thing to be bought'
kerang	'roast'	k<in>erang(-an)	'thing roasted'	ka-kerang-an	'thing to be roasted'
karatr	'bite'	k<in>aratr(-an)	'one bitten'	ka-karatr-an	'one to be bitten'

In perfective nominalizations, then, Nanwang allows both <in> $\sqrt{}$ and <in> $\sqrt{-an}$ patterns, while in the irrealis equivalents, it exhibits a *Ca-* $\sqrt{-an}$ pattern, which corresponds to the *Ca-* $\sqrt{-en}$ pattern in Tamalakaw and Katripul. Such a consistent replacement of *-en* with *-an* in Nanwang patient nominalization patterns is illustrated below by comparisons with Tamalakaw (26) and Katripul (27).

(26) Patient nominalization morphology in Tamalakaw (TK) and Nanwang (NW)⁹

	Root		Perfective nominalization		Irrealis nominalization	
TK	ekan	'eat'	in-kan	'thing eaten'	a-kan-en	'thing to be eaten'
	kerutr	'dig'	k<in>erutr	'thing dug'	ka-kerutr-en	'thing to be dug'
NW	ekan	'eat'	in-kan(-an)	'thing eaten'	a-kan-an	'thing to be eaten'
	kerutr	'dig'	k<in>erutr(-an)	'thing dug'	ka-kerutr-an	'thing to be dug'

(27) Patient nominalization morphology in Katripul (KP) and Nanwang (NW)

	Root		Perfective nominalization		Irrealis nominalization	
KP	kerutr	'dig'	k<in>erutr	'thing dug'	^ka-kerutr-en	'thing to be dug'
	kezeng	'pull'	^k<in>ezeng	'thing pulled'	ka-kezeng-en	'thing to be pulled'
NW	kerutr	'dig'	k<in>erutr(-an)	'thing dug'	ka-kerutr-an	'thing to be dug'
	kedreng	'pull'	k<in>edreng(-an)	'thing pulled'	ka-kedreng-an	'thing to be pulled'

This observation implies the presence of the patient nominalizer **-en* in pre-Puyuma, and therefore indicates that a distinction between patient nominalization and locative nominalization was present in pre-Puyuma. This hypothesis is supported by evidence from Tamalakaw, which shows a clear morphological distinction

on the table) : *d<in>eru* 'what had just finished cooking' (e.g. the object was still in the pot). Intriguingly, a Seediq informant offered a similar judgement for the two forms.

9. For *in-kan* and *a-kan-en* under (26) cf. Tsuchida (1980: 203, 229). Under (27) cf. Teng (2009: 825) for *k<in>erutr*, and Ross (2009: 308) for *ka-kezeng-en*. The manufactured forms *^ka-kerutr-en* and *^k<in>ezeng* are extrapolated from the data and statements in these two sources. All Nanwang data is from Chen (n.d.).

between patient and locative nominalizations. As illustrated in (28), all uses of *-an* in Tamalakaw Puyuma as reported in Tsuchida (1980) denote locative nominalizations, while patient nominalizations are encoded distinctively with an *<in>√-Ø* pattern.

(28) Patient nominalization and locative nominalization in Tamalakaw Puyuma¹⁰

Root		Perfective nominalization		Locative nominalization	
kerutr	'dig'	k<in>erutr	'thing dug'	k<in>erutr-an	'place dug'
ekan	'eat'	in-kan	'thing eaten'	kan-an	'utensils for eating'
riputr	'wrap'	ni-riputr	'boxed lunch' (lit. 'thing wrapped')	–	–
pu-alak	'impregnate'	k<in>i-pu-alak	'bastard'	–	–
nimun	'bathe'	–	–	na-ninum-an	'bathing place'
tra'i	'feces'	–	–	u-a-tra'i-an	'lavatory'
rumay	'rice plant'	–	–	pu-a-rumay-an	'rice granary'
pazek	'carry'	–	–	pazk-an	'a carrying basket'
sirap	'sweep (dust)'	–	–	sirap-an	'dustpan'

To summarize, the morphological variations in these three Puyuma dialects point to the generalizations in (29):

- (29) Implications from patient nominalization in Tamalakaw, Katripul, and Nanwang
- Pre-Puyuma had **-en*, which functioned at least as a patient nominalizer, and was subject to the morphological constraint in (20).¹¹
 - "The second-generation affix" **-en* can be traced back to PAN, given its presence in both Puyuma and NAn languages.¹²
 - When **-en* was lost, *-an* assumed the function of *-en* as a patient nominalizer in Nanwang.

10. For the Tamalakaw forms listed under (28) cf. Tsuchida (1980: 202, 203, 207, 229, 234, 235, 240, 256).

11. As it is difficult to determine whether **-en* in pre-Puyuma was restricted to nominal use, we can only assume that it functioned as least as a nominalizer.

12. See Sagart (2010, 2013) for a similar claim based on (i) fossilized reflexes of **-en* in Nanwang numerals, and (ii) tentative verbal use of *<in>* in perfective nominalizations. Teng & Ross (2010) reject both claims, arguing (i) that the fossilized suffix in Nanwang numeral terms reflects PAN **-N* rather than **-en*, and (ii) that the asserted verbal uses of *<in>* are exclusively nominalizations. We maintain that the presence of **-en* in PAN is supported by the morphological contrast between perfective and irrealis patient nominalizations in a number of high-order AN subgroups.

In what follows, we turn to Rukai and Saaroa, each presenting a striking parallel to what has just been observed in the three Puyuma dialects.

4.2 Rukai as a case of extensive loss: Budai Rukai <in>√-Ø

No available description of Rukai has a reflex of *-en. According to Ogawa & Asai (1935) and Li (1973), patient nominalizations in Tanan Rukai are formed with -an, as shown in (30):

- (30) Perfective patient nominalization in Tanan Rukai
- | | |
|--|--------------------------|
| <i>k<in>ani-an-li</i> ‘my eating’ (‘what I ate’) | (Li 1973: 105) |
| <i>ni-buLu-buLu-an</i> ‘person-instructed’ | (Li 1973: 208) |
| <i>b<in>aay-an ~ ni-baay-an</i> ‘what was given’ | (Li 1973: 197, 206) |
| <i>b<in>aað-a</i> (< *<in>...-an) ‘give’ | (Ogawa & Asai 1935: 137) |
| <i>a-kani-a’</i> (< a-√-an) ‘food’ | (Ogawa & Asai 1935: 337) |
| <i>a-kani-an</i> ‘food to be eaten’ | (Li 1973: 156, 208) |

As seen in (30), Tanan employs <in>√-an for perfective patient nominalizations, and a corresponding a-√-an structure for irrealis patient nominalizations. This pattern resembles that shown for Nanwang Puyuma in (25), suggesting that the <in>√-an pattern in Tanan may well be a product of the loss of *-en.¹³

Further support for this hypothesis is found in Budai Rukai, which represents a clear case of the <in>√ pattern in perfective patient nominalizations, as seen in (31):

- (31) Fossilized <in>√ morphology in Budai Rukai (Online Dictionary of Formosan languages; hereafter cited as ODFL)

Root	Perfective patient nominalization
√	<in>√/<in>√-an
cabu ‘wrap’	c<in>abu (<in>√) ‘food made with wrapped taro and meat’
	c<in>abu-ane (<in>√-an) ‘food made with wrapped taro and meat’
	c<in>abu-cabu (<in>√-RED) ‘food made with wrapped sticky rice and millet’

In this example Budai employs both an <in>√ pattern (c<in>abu) and an <in>√-an pattern (c<in>abu-ane). Another word derived from *cabu* ‘wrap’ which also employs an <in>√ pattern in perfective patient nominalizations is c<in>abu-cabu ‘food made with wrapped sticky rice and millet’ (lit. ‘thing wrapped’). Other words

13. One might argue that the <in>√ pattern in Tanan Rukai is a product of Puyuma contact. However, this is unlikely, since Puyuma verbs are formed by the first-generation affixes: PV -aw, LV -ay, and CV -anay, none of which co-occur with the perfective marker <in>.

in Budai that reflect the same fossilized morphology include *p<in>asu payvaval-riane* (made.by.NMZ white.gauze) ‘thing made with white gauze’, *s<in>aylri* ‘rice soup made with fried taro’, and *t<in>alriki* ‘decayed tooth’. Given the pattern of *-en/an replacement documented above, these examples strongly suggest the existence of *-en in Proto-Rukai, and imply that Rukai has undergone an innovation like that in Nanwang Puyuma, which lost the reflex of *-en, with only fossilized evidence preserved. The observation that Budai Rukai employs <in>√-an morphology as the general pattern for perfective patient nominalization (Zeitoun 2016; ODFL) further suggests that -en was replaced by -an, just as in Nanwang Puyuma.

That Rukai is morphosyntactically innovative rather than retentive is additionally suggested by its limited use of *<um> and *<in>. Synchronically, Rukai shows little evidence of the Actor voice *<um>. However, at least one Proto-Rukai form appears to contain a reflex of *<um>, suggesting its presence in Proto-Rukai. This is illustrated in (32), taken from Blust & Trussel (ongoing):

- (32) Proto-Rukai **m-aLa* as a reflex of PAN *um-ala ‘fetch, get, take’
- | | | | | |
|-----------|-------------|-----------------------|---------------------------------|----------------|
| Formosan: | Pazeh | <i>m-ara</i> | ‘to take, get, obtain; marry’ | |
| | Tsou | <i>m-aro</i> | ‘to take’ | |
| | Kanakanavu | <i>um-á-ala</i> | ‘to take’ | |
| | Saaroa | <i>um-a-ala</i> | ‘to take’ | |
| | Proto-Rukai | <i>*m-aLa</i> | ‘to take’ | (Li 2004: 616) |
| WMP: | Ilokano | <i>um-ála</i> | ‘resemble in looks, take after’ | |
| | Bontok | <i>ɾ<um>ála</i> | ‘to get’ | |
| | Ifugaw | <i>um-ála</i> | ‘to get, take’ | |
| | Wawonii | <i>um-ala-o</i> | ‘to take’ | |
| | Buginese | <i>m-ala</i> | ‘to get, fetch, obtain’ | |

The fossilized *<um>/*-en morphology in modern Rukai on the one hand, and the absence of active reflexes of these affixes on the other, strongly suggests that Rukai is evolving in the direction of *extensive loss*. The distribution of the perfective marker *<in> in Rukai varieties supports the same inference. Synchronically, only the southeastern Rukai dialects, Tanan and Budai, preserve limited reflexes of *<in> (Li 1973; Zeitoun 2000, 2016). In the northwestern dialects Mantauran, Maga, and Tona, <in> has disappeared without a trace (Zeitoun 2007, 2016), which shows a pattern of loss similar to that of *-en.

The strong possibility that *-en existed in Proto-Rukai reveals the weakness in using N-into-V as a subgrouping criterion – since the absence of *-en in modern Rukai does not necessarily imply its absence in Proto-Rukai, or in PAN. Given this analysis, the possibility that Rukai was once a language with a Philippine-type voice

system and noun/verb homophony is, in fact, impossible to rule out.¹⁴ A similar argument holds for the presence or absence of *-en in Saaroa, as discussed below.

4.3 Saaroa patient nominalization and its implications for the NAn hypothesis

Modern Saaroa has no attested reflex of *-en (Zeitoun & Teng 2016). However, its patient nominalization morphology is strikingly parallel to that of Nanwang Puyuma and Tanan Rukai, in that both a *lhi*-√ (< * <in>√) and *lhi*-√-*an* pattern are used in perfective patient nominalizations (Pan 2012; Teng & Zeitoun 2016).

As with the similar arguments for Puyuma and Rukai, the use of *lhi*-√ implies the former presence of *-en in Saaroa, which gave rise to an <in>√-∅ pattern before replacement of the suffix by -an. The data under (33) summarizes the patient nominalization patterns in five Formosan languages, from which it can be seen that non-NAn languages with no reflex of *-en share the same morphological constraint in perfective patient nominalizations as languages that have *-en reflexes.¹⁵

(33) Patient nominalization in Seediq, Saisiyat, two dialects of Puyuma, and Saaroa

Language	Perfective patient nmz (* <in>√-∅)	Irrealis patient nmz (*Ca-√-en)
Seediq (NAn)	k<n>ari 'thing dug'	k-kari-un 'thing to be dug'
Saisiyat (NAn)	k<in>oih 'thing dug'	ka-koih-in* 'thing to be dug'
Tamalakaw Puy. (non-NAn)	k<in>eRutr 'thing dug'	ka-keRutr-en 'thing to be dug'
Nanwang Puy. (non-NAn)	k<in>erutr 'thing dug'	ka-ketutr-an 'thing to be dug'
Saaroa (non-NAn)	lhi-kali** 'thing dug'	a-kali-a 'thing to be dug'

* Note that the ka- prefix in the Saisiyat example *ka-koih-in* 'thing to be dug' (ODFL) is not an instance of Ca-reduplication, but an irrealis marker (Yeh 2000: 59; Zeitoun et al. 2015: 482, 498) combined with the patient nominalizer -in, an allomorph of -en.

** According to Teng & Zeitoun (2016: 149, Table 8), Saaroa *lhi-kali* may be used in verbal environments as well. If such a bifunctional use of <in>√ reflects the general pattern in Saaroa, it implies that pre-Saaroa *-en functioned as both a patient voice affix and a patient nominalizer, thus posing a serious challenge to the N-into-V hypothesis in general, as well as to the criterion defining the putative internal layers of the "N-into-V" subgroup (17).

14. Ross (2014) raises another important piece of evidence for Rukai being morphosyntactically innovative rather than retentive based on fossilized evidence of a genitive/ergative category reflected in the Proto-Rukai oblique pronominal forms: PR **naku-an* (1SG.OBL) and **nami-an* (1PL.EXCL.OBL), which suggest a direct derivational relation with PAN **naku* (1SG.GEN) and **nami* (1PL.EXCL.GEN), respectively.

15. Three examples of the <in>√ pattern in perfective patient nominalizations appear in Pan (2012): *lhi-tapai* 'thing drawn' (318), *lhi-timalha* 'thing heard' (68), *lhi-kita-kita* 'what was seen' (201). These examples are either found in the presupposed clause of pseudo-clefts or normal argument positions, and hence are uncontroversial cases of patient nominalization.

Given the difficulty of knowing whether this suffix was monofunctional or bifunctional, the likelihood that pre-Saaroa had *-en poses a serious challenge to adopting N-into-V as a subgrouping criterion for this language.

4.4 An East Formosan parallel: Amis *-en* vs. Kavalan *-an*

We have seen that some languages have <in> but not *-en*, a situation that on widely-shared structural grounds implies the earlier presence of the suffix. We have also identified a recurrent tendency for *-en* to be replaced by *-an* in different languages. In what follows we discuss a similar drift within the East Formosan subgroup, which presents a striking parallel to what has been observed in Puyuma, Rukai, and Saaroa.

Modern Kavalan has no reflex of *-en, and exhibits only a three-way contrast (AV/LV/CV) in its voice system (Li & Tsuchida 2006). It is classified as an NAn language under the East Formosan subgroup in Ross (2009, 2012), primarily based on exclusively shared phonological innovations with Amis, Siraya, Basay, and Trobiawan (Blust 1999). As a reflex of *-en is found in both Amis and Siraya (Wu 2013; Adelaar 2011), the absence of *-en* in Kavalan is uncontroversially a case of independent loss.

Crucially, although it lacks a reflex of *-en, modern Kavalan employs an <in>√ pattern in both perfective undergoer verb forms and perfective patient nominalizations, showing indirect evidence for *-en, similar to that observed in Nanwang Puyuma, Budai Rukai, and Saaroa.

(34) Kavalan (Li & Tsuchida 2006; ODFL; Lin 2011)

Root	Perfective verb/patient nmz	Irrealis patient nominalization
√	<i>ni-√-∅</i>	<i>sa-√-an</i>
spaw ‘put down’	s<in>paw ‘have put down’	^sa-spaw-an ‘thing to be put down’
qRas ‘call’	q<in>Ras ‘have called’	sa-qRas-an ‘thing to be called’
tnun ‘weave’	t<in>nun ‘have woven’	sa-tnun-an ‘thing to be woven’
sangi ‘make’	s<in>angi ‘things done’	^sa-sangi-an ‘thing to be made’

The loss of *-en* in Kavalan triggered a compensatory use of the locative nominalizer *-an* in irrealis patient nominalizations, showing yet another independent replacement of *-en* by *-an*. That the *sa-√-an* pattern in Kavalan is innovative is clear from the Amis examples in (35), which employ a canonical *Ca-√-en* pattern in irrealis patient nominalizations parallel to that in Sediq, Tagalog, and other NAn languages, as was shown in (26) and (27).

(35) Amis (Wu 2013: 120–121)

Root		Irrealis patient nominalization	
√		Ca-√-en	
'adup	'hunt'	a-'adup-en	'thing to be hunted'
efeng	'abandon'	a-efeng-en	'thing to be abandoned'
licay	'ask'	la-licay-en	'thing to be asked'
ca'it	'hang'	ca-ca'it-en	'thing to be hung'

4.5 Summary

We have shown that three non-NAN Formosan languages which lack a reflex of *-en preserve indirect evidence for this suffix, implying that *-en is reconstructable to PAN. We have further shown that higher-level AN languages that lack a reflex of *-en have adopted a parallel strategy to fill the morphological gap created by loss of *-en. Having now discussed the evidence for the “loss” hypothesis, it will be worthwhile to briefly revisit the competing hypotheses in (18) and argue for Hypothesis B from a different angle.

It is non-controversial that PV and LV can be reconstructed to PAN as two distinct verbal categories, given that this distinction is attested in Tsou, Puyuma, and the majority of NAN languages. Proposing that patient nominalization and locative nominalization were not distinct in PAN, as Ross (2012) and Aldridge (2016) do, thus creates an apparent asymmetry between PAN verbal and nominal morphology, as shown in (36a). By contrast, the “loss” hypothesis maintains the symmetry between PAN nominal and verbal categories, and offers a simple account for the fossilized <in>√ pattern of (36b), further strengthening the view that this analysis is preferable.

(36) a. PAN (pre-NAN) morphology under Hypothesis A

	AV	PV	LV	CV
Neutral (N)	*ta-	*-an		*Si-/Sa-
Perfective (N)	-	*<in>.-an		-
Irrealis (N)	-	*Ca-.-an		*Ca-
Opt./Hor. (V)	*M-	*-aw	*-ay	*-anay
Imperative (V)	∅	*-u/-i	*-u/-i	*-ani

b. PAn morphology under Hypothesis B

	AV	PV	LV	CV
Neutral (N/V)	*ta-	*-en	*-an	*Si-/Sa-
Perfective (N/V)	–	*<in>..-Ø	*<in>..-an	–
Irrealis (N/V)	–	*Ca-..-en	*Ca-..-an	*Ca-
Opt./Hor. (V)	*M-	*-aw	*-ay	*-anay
Imperative (V)	Ø	*-u/-i	*-u/-i	*-ani

5. Conclusion and implications

Recent approaches to AN higher-order subgrouping identify Rukai, Tsou, and Puyuma as historically distinct from all other AN languages based on two pieces of negative evidence: (1) that Rukai lacks a Philippine-type voice system; and (2) that Rukai, Tsou, and Puyuma lack noun/verb homophony. This paper questions the methodological soundness of these hypotheses, first by showing that inferences in any branch of science that are based on negative evidence are doomed to inconclusiveness, and second by showing that extensive loss in morphosyntactic change clearly happened in Malayo-Polynesian languages, and therefore cannot be ruled out with Formosan languages that may form primary branches of the AN family. In addition, it is shown that some traces of an earlier Philippine-type voice system can still be found in both Rukai and Puyuma, and that these cannot easily be reconciled with the view that PAn was an accusative language, or that the initial function of the so-called ‘second-generation’ affixes was solely to form nouns. More generally, this paper is a reminder that the Comparative Method of linguistics remains the most reliable tool for both reconstruction and subgrouping, and that novel methods of inference based on syntactic data remain both speculative and controversial.

Acknowledgements

Sections 1–3 were written by Blust, § 4 by Chen, and § 5 jointly. Data from Tgdaya Seediq and Nanwang Puyuma in this paper are from Chen’s fieldwork during 2015–2016. We are indebted to Ilka Wild for helpful information and discussion of Kananavu, and to Elizabeth Zeitoun, Malcolm Ross, and two anonymous reviewers for their overall assessment of the paper. Any remaining errors of fact or interpretation remain our sole responsibility.

References

- Adelaar, K. Alexander. 2011. *Siraya: Retrieving the Phonology, Grammar and Lexicon of a Dormant Formosan Language*. Berlin & Boston: De Gruyter Mouton.
doi: 10.1515/9783110252965
- Aldridge, Edith. 2014. Ergativity from subjunctive in Austronesian languages. *Online Proceedings of The 14th International Symposium on Chinese Languages and Linguistics (ISCLL-14)*, 289–318. Available at <http://iscll-14.ling.sinica.edu.tw/files-pdf/Papers/Session6/Aldridge.pdf> (accessed 15 September 2016).
- Aldridge, Edith. 2016. Ergativity from subjunctive in Austronesian languages. *Language and Linguistics* 17.1: 27–62. doi: 10.1177/1606822X15613499
- Anceaux, J. C. 1952. *The Wolio Language: Outline of Grammatical Description and Texts*. (VKI 11.) The Hague: Martinus Nijhoff.
- Anceaux, J. C. 1987. *Wolio Dictionary (Wolio-English-Indonesian), Kamus Bahasa Wolio (Wolio-Inggeris-Indonesia)*. (KITLV XVIII.) Dordrecht & Providence: Foris Publications.
- Baird, Louise. 2002. *A Grammar of Kéo: an Austronesian Language of East Nusantara*. Canberra: The Australian National University dissertation.
- Bellwood, Peter. 2011. The checkered prehistory of rice movement southwards as a domesticated cereal – from the Yangzi to the Equator. *Rice* 4.3: 93–103.
doi: 10.1007/s12284-011-9068-9
- Blust, Robert. 1999. Subgrouping, circularity and extinction: some issues in Austronesian comparative linguistics. *Selected Papers from the Eighth International Conference on Austronesian Linguistics*, ed. by Elizabeth Zeitoun & Paul Jen-kuei Li, 31–94. (Symposium Series of the Institute of Linguistics (Preparatory Office), Academia Sinica, No. 1.) Taipei: Academia Sinica.
- Blust, Robert. 2003. *Thao Dictionary*. (Language and Linguistics Monograph Series A5.) Taipei: Institute of Linguistics (Preparatory Office), Academia Sinica.
- Blust, Robert. 2013. *The Austronesian Languages*. 2nd revised edition. (Asia-Pacific Open Access Monographs.) Canberra: College of Asia and the Pacific, The Australian National University.
- Blust, Robert, & Stephen Trussel. Ongoing. *The Austronesian Comparative Dictionary*. Web edition. Available at www.trussel2.com/ACD (accessed 15 September 2016).
- Bowden, John. 2001. *Taba: Description of a South Halmahera Austronesian Language*. (PL 521.) Canberra: Pacific Linguistics.
- Bulbeck, David. 2008. An integrated perspective on the Austronesian diaspora: the switch from cereal agriculture to maritime foraging in the colonisation of island Southeast Asia. *Australian Archaeology* 67: 31–51. doi: 10.1080/03122417.2008.11681877
- Cauquelin, Josiane. 2015. *Nanwang Puyuma-English Dictionary*. (Language and Linguistics Monograph Series 56.) Taipei: Institute of Linguistics, Academia Sinica.
- Chang, Henry Y. 2006. Rethinking the Tsouic subgroup hypothesis: a morphosyntactic perspective. *Streams Converging into an Ocean: Festschrift in Honor of Professor Paul Jen-kuei Li on His 70th Birthday*. (Language and Linguistics Monograph Series W-5), ed. by Henry Y. Chang, Lillian M. Huang & Dah-an Ho, 565–584. Taipei: Institute of Linguistics, Academia Sinica.
- Chen, Victoria. n.d. Fieldnotes on Nanwang Puyuma and Tgdaya Seediq.

- Clark, Ross. 1991. Fingota/fangota: shellfish and fishing in Polynesia. *Man and a Half: Essays in Pacific Anthropology and Ethnobiology in Honour of Ralph Bulmer*, ed. by Andrew Pawley, 78–83. Auckland: The Polynesian Society.
- Codrington, R. H., & J. Palmer. 1896. *A Dictionary of the Language of Mota, Sugarloaf Island, Banks' Islands*. London: Society for Promoting Christian Knowledge.
- Davis, Karen. 2003. *A Grammar of the Hoava Language, Western Solomons*. (Pacific Linguistics 535.) Canberra: Pacific Linguistics.
- Dyen, Isidore. 1990. Homomeric lexical classification. *Linguistic Change and Reconstruction Methodology*. (Trends in Linguistics Studies and Monographs 45), ed. by Philip Baldi, 211–230. Berlin & New York: Mouton de Gruyter. doi: 10.1515/9783110886092.211
- Edwards, Owen. 2015. The position of Enggano within Austronesian. *Oceanic Linguistics* 54.1: 54–109. doi: 10.1353/ol.2015.0001
- English, Leo James. 1986. *Tagalog-English Dictionary*. Quezon City: Congregation of the Most Holy Redeemer.
- English-Tagalog Online Translator. Available at <http://www.tagalogtranslate.com/> (accessed 15 September 2016).
- Ferrell, Raleigh. 1969. *Taiwan Aboriginal Groups: Problems in Cultural and Linguistic Classification*. (Institute of Ethnology, Academia Sinica, Monograph No. 17.) Taipei: Academia Sinica.
- Greenberg, Joseph H. 1957. *Essays in Linguistics*. Chicago: University of Chicago Press.
- Huang, Lillian M. 2000. *Beinanyu Cankao Yufa [A Reference Grammar of Puyuma]*. Taipei: Yuan-Liou Publishing Co., Ltd.
- Jensen, John Thayer. 1977a. *Yapese Reference Grammar*. (PALI Language Texts: Micronesia.) Honolulu: University Press of Hawai'i.
- Jensen, John Thayer. 1977b. *Yapese-English Dictionary*. (PALI Language Texts: Micronesia.) Honolulu: University Press of Hawai'i.
- Kähler, Hans. 1987. *Enggano-Deutsches Wörterbuch [Enggano-German Dictionary]*. Berlin: Dietrich Reimer.
- Lewis, M. Paul, Gary F. Simons, & Charles D. Fennig (eds.) 2016. *Ethnologue: Languages of the World*. 19th edition. Dallas: SIL International. Available at <http://www.ethnologue.com> (accessed 15 September 2016).
- Li, Paul Jen-kuei. 1973. *Rukai Structure*. (Special Publications No. 64.) Taipei: Institute of History and Philology, Academia Sinica.
- Li, Paul Jen-kuei. 2004. The internal relationships of Rukai. *Selected Papers on Formosan Languages*, Vol. 1, ed. by Paul Jen-kuei Li, 559–624. (Language and Linguistics Monograph Series C-3.) Taipei: Institute of Linguistics, Academia Sinica.
- Li, Paul Jen-kuei, & Shigeru Tsuchida. 2006. *Kavalan Dictionary*. (Language and Linguistics Monograph Series A-19.) Taipei: Institute of Linguistics, Academia Sinica.
- Lin, Dong-yi. 2011. A syntactic account of interrogative verbs in Kavalan. Paper Presented at the 47th Annual Meeting of the Chicago Linguistic Society, April 7–9, 2011. Illinois: University of Chicago.
- Ogawa, Naoyoshi, & Erin Asai. 1935. *Gengo ni Yoru Taiwan Takasago Zoku Densetsu Shū [The Myths and Traditions of the Formosan Native Tribes (Texts and Notes)]*. Taipei: Taihoku Imperial University Institute of Linguistics.
- Online Dictionary of Formosan languages (ODFL). Available at <http://e-dictionary.apc.gov.tw> (accessed 15 September 2016).

- Pan, Chia-jung. 2012. *A Grammar of Lha'alua, an Austronesian Language of Taiwan*. Cairns: James Cook University dissertation.
- Pawley, Andrew, & Lawrence A. Reid. 1979. The evolution of transitive constructions in Austronesian. *Austronesian Studies: Papers from the Second Eastern Conference on Austronesian Languages*, ed. by Paz Buenaventura Naylor, 103–130. Ann Arbor: Center for South and Southeast Asian Studies, University of Michigan.
- Ross, Malcolm. 1996. Is Yapese Oceanic? *Reconstruction, Classification, Description: Festschrift in Honor of Isidore Dyen*. (Abera Network Asia-Pacific 3), ed. by Bernd Nothofer, 121–166. Hamburg: Abera Verlag Meyer & Co.
- Ross, Malcolm. 2002. The history and transitivity of Western Austronesian voice and voice-marking. *The History and Typology of Western Austronesian Voice Systems*, ed. by Fay Wouk & Malcolm Ross, 17–62. (Pacific Linguistics 518.) Canberra: Pacific Linguistics.
- Ross, Malcolm. 2006. Reconstructing the case-marking and personal pronoun systems of Proto Austronesian. *Streams Converging into an Ocean: Festschrift in Honor of Professor Paul Jen-kuei Li on His 70th Birthday*, ed. by Henry Y. Chang, Lillian M. Huang & Dah-an Ho, 521–564. (Language and Linguistics Monograph Series W-5.) Taipei: Institute of Linguistics, Academia Sinica.
- Ross, Malcolm. 2009. Proto Austronesian verbal morphology: a reappraisal. *Austronesian Historical Linguistics and Culture History: A Festschrift for Robert Blust*, ed. by Alexander Adelaar & Andrew Pawley, 295–326. (Pacific Linguistics 601.) Canberra: Pacific Linguistics.
- Ross, Malcolm. 2012. In defense of Nuclear Austronesian (and against Tsouic). *Language and Linguistics* 13.6: 1253–1330.
- Ross, Malcolm. 2014. Major morphosyntactic changes in early Austronesian. Lecture at the Advanced Linguistics Workshop on Language Change, July 14–24. Taipei: Institute of Linguistics, Academia Sinica.
- Sagart, Laurent. 2010. Is Puyuma a primary branch of Austronesian? *Oceanic Linguistics* 49.1: 194–204. doi: 10.1353/ol.o.0070
- Sagart, Laurent. 2013. Is Puyuma a primary branch of Austronesian?: a rejoinder. *Oceanic Linguistics* 52.2: 481–492. doi: 10.1353/ol.2013.0027
- Sagart, Laurent. 2014. In defense of the numeral-based model of Austronesian phylogeny, and of Tsouic. *Language and Linguistics* 15.6: 859–882. doi: 10.1177/1606822X14544623
- Sagart, Laurent, Roger Blench, & Alicia Sanchez-Mazas (eds.) 2005. *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*. London and New York: Routledge Curzon. doi: 10.4324/9780203343685
- Sapir, Edward. 1921. *Language: an Introduction to the Study of Speech*. New York: Harcourt, Brace and Company.
- Schachter, Paul, & Fe T. Otanes. 1972. *Tagalog Reference Grammar*. Berkeley: University of California Press.
- Starosta, Stanley. 1995. A grammatical subgrouping of Formosan languages. *Austronesian Studies Relating to Taiwan*, ed. by Paul Jen-kuei Li, Cheng-hwa Tsang, Ying-kuei Huang, Dah-an Ho & Chiu-yu Tseng, 683–726. (Symposium Series of the Institute of History and Philology, Academia Sinica, No. 3.) Taipei: Academia Sinica.
- Starosta, Stanley, Andrew Pawley, & Lawrence A. Reid. 1982. The evolution of focus in Austronesian. *Papers from the Third International Conference on Austronesian Linguistics*, Vol. 2, ed. by Amran Halim, Lois Carrington & Stephen A. Wurm, 145–170. (PL C-75). Canberra: Pacific Linguistics.

- Teng, Stacy Fang-ching. 2007. *A Reference Grammar of Puyuma, an Austronesian Language of Taiwan*. Canberra: The Australian National University dissertation.
- Teng, Stacy Fang-ching. 2009. Case syncretism in Puyuma. *Language and Linguistics* 10.4: 819–844.
- Teng, Stacy Fang-ching, & Malcolm Ross. 2010. Is Puyuma a primary branch of Austronesian? A reply to Sagart. *Oceanic Linguistics* 49.2: 543–558.
- Teng, Stacy Fang-ching, & Elizabeth Zeitoun. 2016. The noun-verb distinction in Kanak
- Thurgood, Graham. 1999. *From Ancient Cham to Modern Dialects: Two Thousand Years of Language Contact and Change*. Honolulu: University of Hawai'i Press.
- Topping, Donald M. (with the assistance of Bernadita C. Dungca). 1973. *Chamorro Reference Grammar*. (PALI Language Texts: Micronesia.) Honolulu: University Press of Hawai'i.
- Tryon, D. T. & B. D. Hackman. 1983. *Solomon Islands Languages: an Internal Classification*. (Pacific Linguistics C-72.) Canberra: Pacific Linguistics.
- Tsang, Cheng-hwa. 2005. Recent discoveries at a Tapenkeng culture site in Taiwan: implications for the problem of Austronesian origins. *The Peopling of East Asia: Putting Together Archaeology, Linguistics and Genetics*, ed. by Laurent Sagart, Roger Blench & Alicia Sanchez-Mazas, 63–73. London & New York: Routledge Curzon.
- Tsuchida, Shigeru. 1976. *Reconstruction of Proto-Tsouic Phonology*. (Study of Languages and Cultures of Asia and Africa Monograph Series, 5.) Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa, Tokyo University of Foreign Studies.
- Tsuchida, Shigeru. 1980. Puyuma (Tamalakaw dialect) vocabulary, with grammatical notes and texts. *Kuroshio no Minzoku, Bunka, Gengo [Ethnology, Cultures and Languages Along the Black Current]*, ed. by Kuroshio Bunka no Kai (Black Current Cultures Committee), 183–307. Tokyo: Kadokawa Shoten. (In Japanese)
- Tuuk, H. N. van der. 1971[1864–1867]. *Tobasche Spraakunst. [A Grammar of Toba Batak]*. (KITLV Translation Series 13), translated by Miss Jeune Scott-Kemball, ed. by A. Teeuw & R. Roolvink. The Hague: Martinus Nijhoff.
- Verheijen, Jilis A. J. 1977. The lack of formative IN affixes in the Manggarai language. *NUSA* 5: 35–37.
- Wolff, John U. 1973. Verbal inflection in Proto-Austronesian. *Parangal kay Cecilio Lopez: Essays in Honor of Cecilio Lopez on His Seventy-fifth Birthday*. (Special Monograph Issue No. 4), ed. by Andrew B. Gonzalez, 71–91. Quezon City: Linguistic Society of the Philippines.
- Wu, Ming-Yi. 2013. *Ameizuyu Cidian [Amis Dictionary]*. Taipei: San Min Book Co., Ltd.
- Yeh, Marie M. 2000. *Saixiyu Cankao Yufa [A Reference Grammar of Saisiyat]*. Taipei: Yuan-Liou Publishing Co., Ltd.
- Zeitoun, Elizabeth. 2000. *Lukaiyu Cankao Yufa [A Reference Grammar of Rukai]*. Taipei: Yuan-Liou Publishing Co., Ltd.
- Zeitoun, Elizabeth. 2007. *A Grammar of Mantauran (Rukai)*. (Language and Linguistics Monograph Series A4–2). Taipei: Institute of Linguistics, Academia Sinica.
- Zeitoun, Elizabeth. 2016. A comparative study of nominalization in Rukai. Paper Read at the 11th International Symposium on Taiwanese Languages and Teaching, July 12–13, 2016. Taipei: Academia Sinica.
- Zeitoun, Elizabeth, Tai-hwa Chu, & Lalo a Tahesh Kaybaybaw. 2015. *A Study of Saisiyat Morphology*. (Oceanic Linguistics Special Publication No. 40). Honolulu: University of Hawai'i Press.

- Zeitoun, Elizabeth, & Stacy Fang-ching Teng. 2016. Reassessing the position of Kanakanavu and Saaroa among the Formosan languages. *Oceanic Linguistics* 55.1: 162–198.
doi: 10.1353/ol.2016.0001
- Zoetmulder, P. J. (with the collaboration of S. O. Robson). 1982. *Old Javanese-English Dictionary*. (2 vols). KITLV. The Hague: Martinus Nijhoff.

Authors' addresses

Robert Blust
Department of Linguistics
University of Hawai'i at Mānoa
1890 East-West Road
Moore Hall 569, Honolulu, HI 96822
USA
blust@hawaii.edu

Victoria Chen
Department of Linguistics
University of Hawai'i at Mānoa
1890 East-West Road
Moore Hall 569, Honolulu, HI 96822
USA
ychen6@hawaii.edu

Publication history

Date received: 8 May 2016

Date accepted: 21 October 2016