

## EVALUATION OF EVIDENCE AND ERRORS IN AUSTRONESIAN RECONSTRUCTION

*R. David Zorc*

In my employment as a lexicographer (Philippine languages, Somali and Armenian) the evaluation of evidence on an entry by entry basis is a daily task. It is equally imperative that we Austronesianists establish criteria for evaluating the overall (and relative) quality of our evidence since it so heavily bears on the value of a reconstruction. It might also be useful, once we have worked out the reflexes and studied the cognates, if we could develop some margin of reliability for the languages with which we deal. Certain broad outlines are suggested in my concluding remarks and can be refined over time with the cooperation of colleagues.

### *1 Kinds of evidence*

In the establishment of etyma, the kinds of linguistic evidence must be considered and weighed. In previous articles (Zorc 1982b:114 and 1984:84), I reiterated Dempwolff's two classic distinctions (test and criterion) and introduced a third (witness). Two others (secondary and false) have also been mentioned. All of these deserve more thorough presentation and study.

Firstly, the direction of postulation (be it upwards or downwards) can be encapsulated in the term.<sup>1</sup> Thus a one-to-one relationship (in either direction) provides test evidence (e.g., PAN \*S > Pai /s/). However, a two-to-one relationship upwards provides criterion evidence (e.g., Akl /ʔ/ < PAN \*q ~ \*ʔ), while downward the evidence is secondary (e.g., PAN \*R > Ilk /r/ ~ /g/). A many-to-one upward relationship is proposed to provide witness evidence (Tag /l/ < PAN \*-d-, \*-D-, \*l, \*-j-, \*-Z-), while downward, false evidence (e.g., PAN \*R, \*r, \*Z > Tag \*r).

Secondly, in any given cognate set, each language contributes something to the reconstruction, and each phoneme within the word may have a different status. Furthermore, we must consider the level we are reconstructing. The weakest element in each piece of evidence should serve as the common denominator for overall quality. For example, if we are reconstructing PPH, Aklanon [h] provides test evidence for PPH \*h (but at a higher level only

---

<sup>1</sup> I owe this observation to Jean-Paul Potet (letter dated 1 May 1994) commenting on Zorc (1990) and querying my use of these terms.

criterion evidence for PAN \*S or \*H). Similarly, Kalamian [k] could be construed to provide test evidence for PPH or PAN \*q (since PAN \*k usually became Kal zero). However, since Kal [k] < \*k is found in reduplicated monosyllables, such instances would reduce it to criterion evidence. Additionally, Kal [k] is found in so many loanwords that some of the data from this language is best considered as witness evidence.

### 1.1 Test evidence

Test evidence is provided when one synchronic phoneme descends from one (and only one) historically posited phoneme, e.g.:

Paiwan /ts/ < PAN \*C  
Amis /s/ < PAN \*S  
Palauan /ch/ < PAN \*q  
Ilokano /e/ < PAN \*e

Dempwolff (1926) proposed that Ivatan /l/ derived exclusively from PAN \*r (since PAN \*l > Ivt /x/, and PAN \*d, \*D, \*Z, and \*j > Ivt /r/) and was therefore a test language for this phoneme. Previously he had reconstructed \*r on the basis of criterion evidence.

The determination of whether one is dealing with test evidence depends directly on one's hypothesis about the parent-language phonology. For example, if one proposes or accepts a single PAN voiced (or lax) bilabial, then Tagalog /b/ < PAN \*b is test evidence; if, however, one accepts a PAN (PMP or PHN) \*B, then the evidence will be criterion. If one accepts only PAN \*Z, then Malay /j/ provides test evidence; but with a hypothesis of a PAN \*Z and a PMP \*z, then Malay becomes a criterion language. When Wolff (1991) or Pejros (1994) reduce PAN \*C to an accent-determined allophone of \*t, the evidence of Paiwan, Puyuma, Rukai and some other Formosan languages must be reevaluated or reappraised.

Some languages are more blessed than others in providing test evidence. Paiwan may provide excellent and unique evidence for up to twenty-three phonemes of the PAN inventory (especially \*S, \*q, \*j, \*N, \*d<sub>2</sub>, \*e), being mute on \*R, \*H, and non-committal on \*b vs. \*w and \*d vs. \*Z. While Palauan has only ten consonant phonemes, some of which reflect multiple mergers, it supplies excellent test evidence for PAN \*q (and criterion evidence for \*R and \*j). Hundreds of Austronesian languages supply test evidence for \*m, \*k, \*a, \*i, \*u, but since these are not problematic phonemes, their value in comparative linguistics is minimal.

### 1.2 Criterion evidence

Criterion evidence is provided when a phoneme descends from and therefore relates to two or three proto-phonemes, i.e., it is the result of a merger at some stage in the development of the daughter language. However, with the evidence of additional languages, the comparativist can 'triangulate' on the most probable correspondence set, e.g.:

Akl /ʔ/ < PHN \*q ~ \*ʔ and Iban /-ʔ/ < PHN \*-ʔ ~ \*-h, therefore,  
 Akl *bukáʔ* and Iban *bukaʔ* 'open' < PHN \**bukáʔ*  
 WBM /z/ < PPH \*-d- ~ \*-j- and Knk /g/ < PPH \*j ~ \*g, therefore,  
 WBM *ngazan* and Knk *ngájan* 'name' < PPH \**ngájan*.

While obviously not the most valuable, this is the most frequent kind of evidence we comparativists bring to bear in our trade. Reconstructions so derived are generally quite solid.

### 1.3 Witness evidence

Witness evidence is provided when a synchronic phoneme descends from and relates to four or more proto-phonemes. I have treated this in earlier papers (Zorc 1982b:114 and fn. 5, 1984-1985:84, 1987, 1990:176), wherein I propose that this may be useful in determining the antiquity of an etymon — but not its phonemic shape.

Thus, Aklanon /-E-/ (an unrounded semivowel that sounds somewhat like a loosely articulated velar fricative [ɣ]) descends from no less than four and up to six PAN phonemes in intervocalic position (depending on one's acceptance of \*z or \*D):<sup>2</sup>

< PAN *-D-	<i>uEáng</i> < PAN * <i>quDáng</i> 'crustacean'
< PAN *-d-	<i>huEám</i> < PAN * <i>Sedám</i> 'borrow'
< PAN *-j-	<i>páEay</i> < PHF * <i>pájay</i> 'rice plant'
< PAN *-l-	<i>úEuh</i> < PAN * <i>qúluH</i> 'head'
< PAN *-Z-	<i>uEán</i> < PAN * <i>quZán</i> 'rain'
< PAN *-z-	<i>taEúm</i> < PAN * <i>Cazém</i> 'sharp'

<sup>2</sup> In most cases, I cite forms reconstructed by Dempwolff (using Dyen's conventions) or Blust with only minor modifications so that colleagues will recognize them. This is not the place for my own unique interpretation of the PAN inventory.

Tagalog /h/ descends from four PAN phonemes:

< PAN *S	<i>káhoj</i> < PAN * <i>káSiw</i> 'tree'
< PAN *H	<i>alupíhan</i> < PAN * <i>qaluHípan</i> 'centipede'
< PAN *-l-	<i>báhay</i> 'house' < PMP * <i>baláy</i> 'building'
< PAN *-Ø	<i>asawáh-in</i> 'be married' < PAN * <i>gasáwa</i>

Iban /-ʔ/ (glottal stop) also descends from four:

< PAN *-ʔ	<i>pukíʔ</i> < PMP * <i>pukíʔ</i> 'vulva'
< PAN *-H	<i>baraʔ</i> < PAN * <i>báRaH</i> 'live coals'
< PAN *-S	<i>kayuʔ</i> < PAN * <i>káSiw</i> , PHN * <i>káyuh</i> 'tree'
< PAN *-R	<i>ikuʔ</i> < PAN * <i>ikuR</i> 'tail'

Amongst the languages of the North Cordilleran subgroup (Ibanag, Atta, Agta, Yogad, etc.), only Malaweg differentiates \*s from \*t. It is a test language for both phonemes; whereas Ibanag, which merges the two, as well as all final voiceless stops into glottal stop, is a witness language.

With its great number of borrowings from numerous language sources (see Blust 1992a), Tiruray data on the whole might best be considered witness evidence. This is not to imply that it cannot be cited in cognate sets, but Tir data should be presented with extreme caution unless the reflexes conform in every way with those expected for this language.

Given the loss of final consonants (including their reanalysis with certain suffix types) and numerous mergers in the Oceanic languages, a word might be descended from any of several etyma within a range of formal and semantic similarities. Hence, for reconstructions with a phonotactic shape of \*CV(C)CVC, most eastern Austronesian languages are witnesses.

I expressed my concern (in Zorc 1987) about the reliability of Philippine languages in the reconstruction of PAN apicals (\*d, \*D, \*z, \*Z), especially Tagalog which has been used so extensively in the literature. Additionally, if we bear in mind Dahl's analysis of Javanese and Madurese (1976:66ff.) supporting a hypothesis of phoneme split under Indic influence, much of the evidence for PAN distinctions from the Western Austronesian branch melts away. Unfortunately, Dahl did not challenge the Tagalog data and in fact went to great pains to explain it on occasion (1976:65, 80). In the evaluation of critical evidence, Dempwolff took Tagalog to be critical, whereas Dyen (1947) did so for Javanese. There are just too many instances where the Indonesian and the Formosan evidence are in conflict for a reliable conclusion of \*d vs. \*D (or \*d<sub>2</sub> or \*d<sub>3</sub>), e.g., PAN \*[d]amaR 'torch, light', PAN \*[d]áRaq 'blood', PAN \*[d]elés 'bowstring', PAN \*[d]úRiH 'thorn', PAN \*laHú[d] 'sea(wards)'. In my opinion no solution is possible on the basis of witness evidence from Tagalog or — for that matter — from Bisayan (see Zorc 1977:211-216) or any other Philippine language (see Zorc 1987:752, 758).

#### 1.4 Secondary evidence

Secondary evidence is provided when phonemes do not manifest their regular reflexes, yet careful analysis reveals that the evidence is legitimate. Usually, some irregular changes or dialect developments occurred within the history of the language itself. For example, Pangasinan has some penult reflexes of PAN \*e with /a/ besides the regular /e/, such as, Png *andí* 'none' < PHN \**hendí?* 'not', Png *taló* 'three' < PAN \**telú*, Png *apát* 'four' < PAN \*[S]*jepát*, Png *aném* 'six' < PAN \**eném*, Png *panó* 'full' < PAN \**penúq*.<sup>3</sup> That these are internal secondary developments and therefore not loanwords is evidenced by the basic nature of the vocabulary items just presented, as well as innovations limited to the South Cordilleran subgroup to which Pangasinan belongs, e.g., PSC \**betik* 'run' > Png *batik*, PSC \**ekmun* 'swallow' > Png *akmón*, PSC \**selí* 'foot' > Png *salí*.

There is an apparent splitting of PAN \*R into Ilokano /g/ and /r/, even with some doublets such as Ilk *bágo*, *barfi* 'new' < PAN \**baqRuH*, Ilk *bibíg*, *bibir* 'lip' < PAN \**bibir*.<sup>4</sup> While the issue is far from settled, I take both to be valid, with /r/ as primary (criterion) evidence and /g/ as secondary, see also Conant (1911), Reid (1973), and Tharp (1974).

For the most part, Blust (1980, 1983-1984, 1986, 1989 and in progress) puts secondary evidence or discussions thereof in his notes following the etymologies.

#### 1.5 False evidence

False evidence is sociolinguistic in nature, but nevertheless critically important in Austronesian studies, because it involves the identification of loanwords which may demonstrate regular reflexes. Tag *tangháli?* 'noon' represents \**tengáq* + \**qaRi*, but is clearly a loan from Malay *tengah hari*. Historical records indicate that a Brunei Malay community had been established at Tondo-Manila when the Spanish arrived, and the Tagalogs of that time surely adopted many words either to fill in gaps or as a sign of erudition (as they

<sup>3</sup> A similar situation where Tagalog has an /a/ for PAN, PMP \*e is not, to my view or that of Wolff (1976), legitimate and therefore represents false evidence (see section 1.5), such as Tag *tápal* 'poultice' < Mal *tampal* or Jav *tapel* < PAN \**Ca(m)pel* (PAA#92, AE3#69); while the PAN reconstruction is legitimate, the inclusion of Tag is not. There are dozens of similar cases where loans from Malay or Kapampangan are involved.

<sup>4</sup> Tharp (1974) discusses his review of 91 occurrences of Ilokano reflexes of \*R (unfortunately the data are not provided). Of these there were 8 doublets, 56 occurrences of /g/ and 27 of /r/, i.e., /g/ is the most common altogether. However, if position is considered, then there were more initial occurrences with /r-/ (9) than with /g-/ (4), with one doublet.

have since done with Spanish and English loans or as the English did with Norman French).

Dempwolff (1938) included forty-one known loans<sup>5</sup> that exemplified regular reflexes among the languages he discussed. Although he marked them with a raised <sup>x</sup>, he rarely gave any indication of the source language (sni = source not indicated); such lacunae can be filled in thanks to the research of Gonda (1973), Jones (1978), and Wolff (1976).

- sni <sup>x</sup>*aca[r]* 'pickle' [VL3:14] (Indic or Arb)
- sni <sup>x</sup>*arak* 'liquor' [VL3:15] (Arb < Persian)
- sni <sup>x</sup>*badan* 'body' [VL3:18; Jones] (Arb)
- sni <sup>x</sup>*bazu* 'shoulder garment' [VL3:18; MBT:363] (Arb < Persian)
- sni <sup>x</sup>*baca* 'read' [VL3:21] (Skt)
- Skt <sup>x</sup>*balañza* 'sell' [VL3:21] (actually Pali or Singhalese)
- sni <sup>x</sup>*bedil* 'gun' [VL3:25] (no known source)
- sni <sup>x</sup>*Dacin* 'balance, scales' [VL3:42] (Ch)
- sni <sup>x</sup>*zaga* 'hold watch' [VL3:45] (Indic)
- sni <sup>x</sup>*kuTa* 'fortress' [VL3:85] (Skt)
- sni <sup>x</sup>*kuwat* 'be strong' [VL3:85] (Arb)
- sni <sup>x</sup>*cuka* 'vinegar' [VL3:88] (Indic, Prakrit)
- sni <sup>x</sup>*manik* 'bead(s)' [VL3:106] (Indic)
- sni <sup>x</sup>*mula* 'beginning, origin' [VL3:107] (Skt)
- Indic <sup>x</sup>*panday* 'craftsman' [VL3:110]
- Skt <sup>x</sup>*palangka* 'sedan chair' [VL3:112] (actually Indic or Portuguese)
- sni <sup>x</sup>*peTi* 'box, trunk' [VL3:118] (Tamil)
- Indic <sup>x</sup>*tambul* 'side-dish with drinks' [VL3: 125] (no known source)
- sni <sup>x</sup>*tarazu* 'balancing scale' [VL3:129; MBT:362] (Persian)
- sni <sup>x</sup>*sembeliq* 'slaughter according to Islamic rite' [VL3:150]  
(no known source)

Blust also did so in his earlier reconstructions (1970, 1973), being ever careful to indicate the donor or source language.

- Arb <sup>x</sup>*akal* 'intelligence' [PAA#02], Arb <sup>x</sup>*asal* 'source, origin' [PAA#11]
- Arb <sup>x</sup>*dakwa* 'accusation, blame' [PAA#144], Arb <sup>x</sup>*hikmat* 'caution' [PAA#427]
- Ch <sup>x</sup>*bakia?* 'wooden clogs' [PAA#13], Ch <sup>x</sup>*bakmi* 'noodles' [PAA#19]
- Ch <sup>x</sup>*lumpia* 'Chinese egg roll' [PAA#258]
- Persian > Arb <sup>x</sup>*saluar* 'trousers' [PAA#379]
- Port <sup>x</sup>*beniaga* 'trade, commerce' [PAA#28; not directly Skt]

---

<sup>5</sup> The full list was presented in my original paper and is available to interested readers on request; it is abbreviated here due to space limitations.

- Skt *\*asah* 'hope' [PAA#10], Skt *\*barita* 'news; event' [PAA#31]  
Skt *\*campaga* 'jasmine flower' [PAA#78], Skt *\*guna* 'usefulness'  
[PAA#180]  
Tamil *\*bilanggu* 'prisoner' [PAA#52], Tamil *\*kalakati* 'areca-nut cutter'  
[PA3#143]  
Tamil *\*kawal* 'guard, watchman' [PAA#199]

Unfortunately, thereafter Blust (1980, 1983-1984, 1986, 1989, in progress) ceased to do so, although he admitted that 'the identification of intimate borrowings — particularly of Malay loanwords in various languages of western Indonesia and the Philippines — remains one of the most vexing problems in comparative AN linguistics' (1980:29ff.).

I say 'unfortunately' because, besides the valuable historical implications than can be derived from an analysis of such loans (e.g., Wolff 1976), a published list would help students avoid the mistakes of reconstructing them on their own. In the early 1970's, I reconstructed PPH *\*mani?* 'peanut' until I learned it was from Mexican Spanish. In reading dissertation drafts of colleagues, I helped cull out spurious reconstructions such as *\*kudál* 'fence' (Sp *corral*) and *\*gadapún* 'water-jar' (Sp *garrafin*). In fact, the unwitting (or unmarked) presentation of loanwords is not necessarily limited to neophyte Austronesianists:

- Sp *\*adán* 'Adam' [male personal name] [PA3#251, not directly Arabic]<sup>6</sup>  
Skt *\*banga?* 'break, collide; compete' [PAA#13; MBT:364]  
Skt *\*bisa* 'poison, venom' [Blust 1992a:39; Gonda 1973:154, MBT:359]  
Indic *\*kapas* 'cotton' [VL3:75; Gonda 1973:323]  
Ch *\*kawaq* 'cauldron' [VL3:76; MBT:347]  
Tamil/Dravidian *\*kuda* 'horse' [Blust PAA#223; Gonda 1973:490]  
Skt *\*lasuna* 'garlic' [VL3:93; Gonda 1973:649]  
Skt *\*ñata* 'clear' [VL3:108; Gonda 1973:139]  
Hindi or Arabic *\*pasar* [VL3:115; Gonda 1973:80, Jones 1978:69]  
Tamil or Arabic *\*pinggan* 'plate' [VL3:118; MBT:364; Jones 1978:70]  
Tamil *\*puTu* 'rice cake' [VL3:123; MBT:363]  
Ch *\*sampan* 'boat' [VL3:149]  
Tamil *\*tunay* 'hard cash' [PAA#141; MBT:362]

If Wolff (1976) is correct, then the following are Malay loans in the Philippines, some of which may be valid reconstructions of a proto-language more immediate to Malay, e.g., PMJ, PIN, or I-WI:

---

<sup>6</sup> Jean-Paul Potet kindly called this to my attention in a letter dated 12 October 1994.

Mal *\*baladaw* 'curved dagger' [PAA#21] > Tag *balaráw*, Ceb *baláraw*, Mar *baladao*.

Mal *\*dungu* 'stupid' [PAA#157] > Tag *dungú?* 'dull, stupid; timid'.

PIN *\*keben* 'trunk' [VL3:76] > Mal *keban* '4-cornered matwork bag' > Tag, Akl, Bik, Ceb, Ilk, Png *kabán*, WBM *kavan*, Mar *kaban* 'crate; unit of dry measure' [PFL#1679; MBT:362].

Mal *\*mahál* 'expensive, dear' [PAA#269, PFL#2407] > Tag, Bik, Ceb *mahál*, Akl *maháE*.

PMJ *\*pasu* 'earthenware vessel' [VL3:115] > Tag *pasfí?*, Kpm *pasú?*, Png *pasó*.

PMJ *\*taker* 'unit of measure' [PAA#090; MBT:362] > Tag *tákal*.

Mahdi (1994) is devoted to 'maverick protoforms'. His article is a tour de force on the intra-western-Austronesian borrowing process affecting the reconstruction of *\*keRbau* 'water buffalo', *\*kurung* 'cage', *\*bari[]* or *\*besi* 'iron', *\*pirak* 'silver', *\*baLituk* 'gold', etc.

There is also the problem of reconciling posited borrowings with actual source forms, such as Dempwolff's *\*sembeliq* or *\*tambul* in the list above. Some of the following may be actual Austronesian words despite their author's claims to the contrary:

PHN *\*DaRa* 'maiden' > Tag *dalága*, Mal *dara* [VL3:42; Gonda (1973:487) disassociates from Skt, and the g::r correspondence is valid]

PWI *\*kala* 'scorpion' [VL3:73, Gonda (1973:478) disassociates from Skt]

PHN *\*luhuR* 'prayer, supplication' > Tag *lúhog* 'intensive supplication', Mal *lohor*, Jav *luhur* 'midday prayer' [Blust (PAA#260), Jones (1978:50) and Wilkinson (1959:700) cite Arabic *Zuhr* 'midday', which may account for the Mal and Jav forms, but comparison with Tag is nonetheless attractive, since the r::g correspondence is certainly without precedent for a loanword].

## 2 Kinds of reconstructions

Etymologists should be aware of five areas of potential overlap: doublets, disjuncts, monosyllabic roots, synonyms, and morphologically complex reconstructions.

Blust acknowledged that the term *doublet* is used to describe several quite distinct phenomena (1980:27): phonologically similar reconstructions, e.g., *\*bingaq* and *\*bingaR* 'volute shell', *\*baNaw* and *\*baNaR* '*Smilax*' (= true doublets) and etyma containing a monosyllabic root, e.g., *\*+pak<sub>1</sub>* 'slap, clap', *\*+pak<sub>2</sub>* 'break, crack, split', *\*+pik* 'pat, light slap', *\*+puk<sub>1</sub>* 'throb, thud,



clap, break', \*+puk<sub>2</sub> 'dust', \*+puk<sub>3</sub> 'gather, flock together' (see Blust 1988:80).

I proposed (Zorc 1990) that the terms and abbreviations used can be adapted to include: doublets (Dbl), disjuncts (Dsj), monosyllabic roots (Mon), and synonyms (Syn). Where there is still some potential ambiguity as to the mixture of types, a convention can be adopted to mark suspect root boundaries, e.g., \*ti+ku?, \*te+ku? 'bend, curve' (Dbl+Mon). At least some of the residual difficulties have to do with the quality of specific language evidence rather than problems of labeling reconstructions.

## 2.1 Doublets

Doublets, called *Nebenformen* by Dempwolff, are reconstructions that are phonologically (or formally) and semantically similar. Dyen (1951:534) defined doublets as 'synonymous and nearly homonymous meaning-forms'. They are surprisingly (and to some scholars, disturbingly) common, a necessary result of our method and the Austronesian predilection for word play. Of Dempwolff's 2,215 entries,<sup>7</sup> a full 21% (466) are doublets,<sup>8</sup> or on average one in five. The works of Blust (1970, 1980, 1983-1984, etc.) are also replete with them.

In terms of method, Dyen (1951) unified several of Dempwolff's doublets by more carefully determining the reconstruction of \*Z, and later (Dyen 1953a) of \*q and \*h:

- PHN \*a(n)daw & PMP \*ha(ñ)jaw 'day, sun' (unified as \*qa(n)jaw)  
 PHN \*aRus & PMP \*haRus 'current' (unified as \*qaRus)  
 PMP \*añud & PMP \*hañud 'float' (unified as \*qañud)  
 PMP \*angin & PMP \*hangin 'wind' (unified as \*hángin)  
 PMP \*atay & PMP \*hatay 'liver' (unified as \*qatáy)  
 PMP \*atep & PMP \*hatep 'roof' (unified as \*qatép)  
 PMP \*baRu, \*b <aq> aRu & POC \*beRu, \*b <aq> eRu 'new' (unified \*baqeRu)  
 PWI \*da'eh & PMP \*zauh 'far' (unified as \*Za(h0)uq, or possibly \*diauq)  
 PHN \*dalan & PMP \*zalan 'path, road' (unified as \*Zálan)

<sup>7</sup> The figure is from Blust (1980:3); I have never successfully counted the entries for myself. Every time I have tried, I came up with different figures because I have so personalized the data that I either deduct those rejected or add when I have split his data into two or more reconstructions.

<sup>8</sup> There may be an additional 26 which I feel Dempwolff missed, e.g., PMP \*buRuk 'spoiled' — PMP \*busuk 'rotten', PMP \*darizi — PHN \*zari 'finger', PMP \*zayu 'foreigner' — PMP \*zauq 'far', etc. These were presented in table 1 of my original paper; the full list is available upon request.

- PMP \**duRuh* & PMP \**zuRuh* 'liquid' (unified as \**ZuRuq*)  
PMP \**hu(N)Dang* & PMP \**u(N)Dang* 'crustacean' (unified as \**quDang*)  
PMP \**huzan* & PMP \**udan* 'rain' (unified as \**quZaN*)  
PMP \**hulu* & PHN \**ulu* 'head' (unified as \**qulu*)  
PIN \**pe(N)Dem* & PMP \**pezam* 'shut the eyes' (unified as \**peZem*)  
PMP \**tuha* & PMP \**tuwa* 'old' (unified as \**tuqaS*)

Recently Wolff (1993a) has done the same by his unification of \*N and \*ñ. Despite his zeal, some of the unions may be unfounded, such as \**qaLimeCáq* 'paddy leech' and PHF \**qaLimátek* 'jungle leech', which can be justified on phonological grounds by consistent accent differences in the Philippines and the \*q vs. \*k reflexes, as well as on semantic grounds in that two distinct species are involved (dry vs. wet), maintained synchronically in Tag *limátik*, Akl *alimátuk* vs. Tag *lintá?*, Akl *línta?*. Nevertheless, many of his proposals are well-argued and well-founded, such as:

- PAN \**languy* & PAN \**nanguy* 'swim' (unified as \**ñangúy*)  
PHF \**Látad* & PHF \**Nátad* 'yard' (unified as \**ñatad*)  
PMP \**ñamúk*, PHN \**namúk* & PHN \**lamúk* 'mosquito' (unified as \**ñamúk*)

In terms of the Austronesian predilection for word play, the doublets can be highly instructive. I have organized lists of them around individual features, such as initial, medial and final consonants, vowel alternations, accent pairs, morphological pairs, metathesis and 'shimmer' (Zorc 1977:58 and fn. 24). I believe this would make an excellent topic for a dissertation, and would be more than happy to send my doublet lists to and assist a degree candidate with data such as:

- initial Ø:/b/ PPH \**abáh* — PMP \**babáh* 'carry on back'  
initial Ø:/l/ PMP \**ípen* — PAN \**lipen* 'tooth'  
initial /d/:/p/ PMP \**dia* — PMP \*(*ma*)*pia* 'good'  
initial /q/:/s/ PMP \**qa+liR* — PMP \**sa+liR* 'flow'  
medial /j/:/s/ PAN \**qujing* — PPH \**musing* 'charcoal'  
medial /k/:/q/ PHF \**tákí?* — PAN \**Caqi* 'excrement'  
final Ø:/n/ PHF \**beRngi* — PHF \**beRngin* 'night'  
final /D/:/R/ PMP \**akaD* — PMP \**ukaR* 'root'  
accent pair PHN \**ha(m)bél* 'weave' — PHN \**há(m)bel* 'blanket'  
accent pair PMP \**sákay* 'mount, climb' — PMP \**sakáy* 'ride, get on'  
metathesis PHF \**dalúkap* — PNP \**dakúlap* 'palm (of hand)'  
metathesis PHF \**púnuq* — PMP \**púqun* 'tree trunk'  
morphological PHN \**babaw* — PAN \**b <al> abaw* 'rat, mouse'  
morphological PAN \**ari* — PAN \**m-ari* 'come on; let's go!'

compound PAN \*Cáu[h] — PMP \*tau-mataq 'person, human being'  
series PAN \*a — \*ka — \*ma — \*Na 'and, with'  
series PAN \*báHi — PAN \*ba-báHi — PMP \*ba-b <in> ahi 'woman, female'  
series PAN \*asa — \*isa — \*esa — \*tasa 'one'  
nasal PHN \*qepá — PHN \*qempá 'bran, chaff'  
nasal PMP \*tangíRi — PHN \*tanggiRi 'Spanish mackerel'  
shimmer PAN \*biRaS — PHN \*piRah 'roe, fish eggs'  
shimmer PHN \*ubak — PHN \*ú(m)pak 'bark (of tree); husk'  
vowel PAN \*SebaN — PHF \*SabaN 'carry on the back'  
vowel PMP \*tiDuR — PAN \*túDuR 'sleep'  
syllabic PHN \*baquR — PHN \*baweR 'spring or trigger trap'  
syllabic PAN \*Siáq — PHF \*Seyáq 'shame, shyness'

There is also the danger of overreliance on a formulaic approach to unify or reconcile conflicting data sets. For example, Paz (1981) reconstructs PPH \*supsép 'suck' or \*dagú? 'blood' to reconcile in a single formula data which is more appropriately assigned to two (\*sepsep vs. \*supsup, \*dáRaq vs. \*duRúq, Zorc 1981).

How chary must we be of doublets? In my opinion, not very. Once we have accounted for legitimate phonological discrepancies (and thus the forms are or are not unifiable), they must be a diachronic fact of life. Sometimes they may represent a 'cute' way of reconciling the unreconcilable, such as my PHN \*kamding for uniting \*kambing and \*kanding 'goat', all of which may point to a maverick anyway. A certain amount of solace may be gained from the realization that there are many instances synchronically available in the study of virtually any well-described Austronesian language. The *Tagalog Slang Dictionary* (Zorc 1991) is replete with examples of how words are shaped and reshaped. As it is now, so it well may have been then.

## 2.2 Disjuncts

Disjuncts are reconstructions that critically depend on an overlap of cognate sets. Blust (1970:112ff.) introduced this very valuable term:

In at least one case Dempwolff assigned a form in a daughter language to each of two variant shapes of a reconstruction (as with the Fijian cognate in Tag, NgD *gumi*, Fj *kumi* < \**gumi* 'beard'; TB *gumis*, Jav, Mal, NgD *kumis*, Fj *kumi* < \**kumis* 'beard'). While both of these variants were reconstructible on Dempwolff's subgrouping assumptions without the Fijian cognate, it is possible that the reconstruction of either variant might have depended crucially on this item. Imagine, for example, that the only evidence for \**gumi* were Tag *gumi*, Fj *kumi*. As Fj *kumi* could reflect either \**gumi* or \**kumis*, there would be no

way to know whether Tag *gumi* represented an innovation, or together with Fj *kumi* constituted evidence for a variant of \**kumis*. In such a case, the association of Fj *kumi* with Tag *gumi* would be open to serious objections. Let us refer to this type of situation as the overlap of cognate sets. Cognate sets are said to overlap if the reconstruction of the prototype of one or more of them depends crucially on a member that can be assigned to more than one set.

In order to avoid arbitrariness in this and similar cases, the following principle has been adopted: where the reconstruction of variants is possible only by allowing the overlap of cognate sets, 'disjunctive alternatives' rather than variants are proposed.

He has maintained this in his subsequent publications (see Blust 1980:25ff., 1983-1984, etc.). The following are some examples of disjuncts; the critical evidence on which this kind of doubling is based has been indicated:

PMP \**babaw* ~ \**wawaw* > Mar *oaoao* 'to weed' [PAA:440]  
PHN \**latak* ~ \**latek* 'sediment, dregs' > Jav *latak* ~ *latek* [PAA:230]  
PHN \**lungzag* ~ \**rungzag* 'jump' > Mal *lonjak* ~ *runjak* [PAA:262]  
PHN \**reñay* or \**rinay* 'aftermath of a storm' > Tag *lináy* 'cessation of wind after a storm' [PAA:357]

The next two are better classified as instances of witness evidence:

PMP \**sampen* ~ \**sampet* > POC \**sampo* 'take hold of, catch' [PAA:384]  
PMP \**sulut* ~ \**sulung* 'to dress' > POC \**sulu* 'sarong, put on sarong' [PAA:87]

Because we are characterizing the extension of language evidence to two or more etyma, a kind of 'double-dipping' (if you will), I would prefer strictly to characterize the reconstruction as a disjunct (as does Blust), but more loosely to call the language evidence disjunctive. I accept that a disjunct is reconstructed when the evidence is crucial and could point to either (i.e., going upwards). On the other hand, when any language form is proposed to derive from more than one etymology, then it is handy to have a label such as 'disjunctive' to characterize the competing etymologies (i.e., going downwards).

Tag *gatá?* 'coconut milk' < PMP \**Rataq* [AE1#382], PHN \**gátéq* [AE4#201] or PHN \**getáq* [VL3:55]  
Tag *ibay* 'giddiness' < PHN \**ibay* 'nausea, giddiness' [VL3:66] or via Kpm < PMP \**ibeR* 'desire, crave' [CDF 4:173]

The following exemplify my characterizations of this phenomenon (not necessarily those of Blust); supporting data can be found in the references cited:

- PHN \*+bir — PHN \*+biR 'edge, rim' [see AE1#065note, Bey]  
 PMP \*puláh — PMP \*pulaq 'red' [see AE2#296]  
 PAN \*qê+Ret — PAN \*Se+ReC 'tight' [see Tsuchida 1976; ACD#h71; Zorc 1982b:133]  
 PHN \*Ribu — PHN? \*ribu 'thousand' [see VL3]  
 PHF \*Sêmay — PHF \*Hêmay 'rice' [see ACD, #S25]  
 PAN \*Siup — PSP \*heyúp 'blow (on)' [see ACD, #S33]

### 2.3 Synonyms

Synonyms are reconstructions that occupy the same 'semantic space'. It is critical, however, that they be assigned to the same time depth or have persisted to another reconstructional level, otherwise we are dealing with innovations, i.e., one form that replaces another in the development of a subgroup. The following exemplify such replacements (and cannot appropriately be called synonyms):

fetch water	PMP *a(ng)+suR	→ PPH *sakedu
gills	PMP *há(ng)sang	→ PNP *hadang
year	PMP *taqún	→ PPH *dagʔun

Scholars have tended to shy away from using this term to characterize their reconstructions. This is probably because a particular semantic feature could have separated forms, especially those for flora and fauna, where different species could be represented. Thus the following are not good candidates for synonymy:

bone	PAN *CuqelaN (condylar), PAN *dúRi (fishbone)
disaster	PMP, PHN *tiwas, PMP, PHN *walat, PHN *sahul [see VL3]
kill	PAN *paCéy > PMP *patéy, PMP *bunúq (with weapon)
pound (fibers)	PAN, PMP *báyuH (rice), PAN, PMP *bak+bak PMP *Tuk+Tuk (prepared food)
water	PAN *Danum (potable), PMP *wahiR (fresh)
snake	PAN *SúlaR (overlap with 'worm'), PHF *bulai, PMP *anipa (large), PMP *nipay, PMP *sawáh (python)

worm PAN \*qúlej (overlap with 'snake'), PMP \*bulati (intestinal), PMP \*kalati (earth)

Given the above reservations, the following are presented as possible synonyms:

above, on top	PAN *bábaw, PHF *i-Caʔas
bite, gnaw	PAN *kaRáC, PAN *kat+kat
bite, nibble	PMP *ket+ket, PMP *kit+kit
buy	PAN *belí, PHF *sáliw
count	PHF *Hiáp, PHF *bilang; PMP *qitung, PMP *wilis
father	PAN, PMP, PHN *ama, PHN? *bapa?
finger nail	PAN *kuSkus > PMP, PHN *kukúh, PHN *sulu
fireplace, hearth	PMP, PHN, PPH *DapúR, PPH *hapuy-an
forest	PMP *qútan, PMP *halás
forest	PPH *ka-kayúh-an, PPH *kâlásan, PPH *Rubat
hack, hew	PAN *tek+tek, PAN *taRáq
handspan	PMP, PHN *dangan, PHN *zangkal
iron	PHN? *besi, *bari
night	PAN *RabiʔiH, PAN *beRngi
now, already	PHF *da, PHF *Na, PHF *la
penis	PAN *qútiN > PMP, PHN *qútin, PHN *bútuq
root	PMP *Ramút, PMP *wákaR, PMP *waRet; PMP *dalij (butress)
rotten	PAN, PMP, PHN *buRuk, PMP, PHN *busuk, PHN *baRiw
salt	PAN *qasiRa, PHF *timus; PAN *qasiN (flavor)
smoke	PHF *qébél, PMP *anús, PMP *qasúh
tooth	PAN *ngípen, PAN *ngísi[h]
white	PAN, PMP *putiq, PMP *burak, PMP *bud[ea]q

## 2.4 Monosyllabic roots

Monosyllabic roots are equivalent to *phonesthemes* (e.g., \*+buk 'pound', \*+suk 'enter'), not functors (e.g., \*si 'name marker', \*ni 'agentive, possessive', \*na 'ligature, apposition', etc.). Dyen (pc) suggested that these be called radicals since the term 'root' is so well established with the normal results of the application of the comparative method, and might lead to confusion. However, if used consistently in a phrase, such as 'monosyllabic root', 'root candidate' (etc.), the meaning should be clear and ambiguity avoided. What is important is that the isolation procedures of root candidates

must be stated explicitly and followed rigorously (and scientifically), otherwise 'methodological chaos' is probable.

There are clearly many etyma that contain a monosyllabic root (see Blust 1988, Zorc 1990), such as PAN \*+keC, PMP (etc.) \*+ket 'sticky, adhesive' in:

- PAN \*de+keC 'adhere, stick to'
- PMP \*ca(ng)+ket 'sticky'
- PMP \*da(ng)+ket 'stick, adhere'
- PMP \*di+ket 'adhesive'
- PMP \*li+ket 'adhesive'
- PMP \*pe+ket 'viscous'
- PMP \*si(ng)+ket 'stick, adhere to'
- PHN \*la+ket 'viscous'
- PHN \*le+ket 'stick, adhere'
- PHN \*pi+ket 'bird lime' [all examples from Blust 1988:106ff.]

or the root series \*+ngaC, \*+ngeC, \*+ngiC, \*+nguC 'gnash the teeth (as in anger or pain)':

- PAN \*Re+ngeC 'angry, annoyed' [AE1#371; Tsuchida 1976:179]
- PAN \*re+ngiC 'grimace' [AE2#313]
- PMP \*ha+ngit 'anger' [AE1#405]
- PMP \*ngat+ngat 'gnaw' [VL3:109]
- PHN \*bi+nget 'moody, irritable' [PAA#57]
- PHN \*bu+nget 'anger, angry' [AE1#81]
- PHN \*nget+nget 'gnash the teeth' [VL3:109]
- PPH \*ngut+ngut 'biting pain' [Zorc 1971]

These have been treated in great detail by Blust (1988) and other scholars have since reacted, e.g., Nothofer (1990, 1991, who suggests a broader range of radicals than just CVC sequences) and Dyen (who has reevaluated Dempwolff's evidence).<sup>9</sup> I proposed (Zorc 1990) that some symbol be used to mark off such roots, such as a plus sign, as I have also done herein. I also suggested that some reconstructions might contain roots not limited to the terminal -CVC, e.g., PMP \*tang+an 'hand' (\*tang 'grasp' + -an). There may also be instances of compounds, e.g., PMP \*ka?+ngah 'fissured, slightly cracked' (AE1#118) or PMP \*la?+bak 'wide open' (AE1#247), each with my revisions (Zorc 1984-1985:88). These are certainly areas worthy of disciplined research.

<sup>9</sup> Personal communications (1974, 1988) and in a paper presented at the ICAL 7 conference.

### 2.5 Morphologically complex reconstructions

Morphologically complex reconstructions involve attributing a morpheme break to the etymology. Blust (pc) speaks of a 'benign slash', an ad hoc tool to get synchronic language data to conform with a given reconstruction. This was often apparent in Dempwolff, such as his PMP \**uRita* [VL3:160] for what should have been doublets PMP \**kuRita*, an irregular PIN \**gurita*, and a GCP \**puRita* 'octopus, polyp', without any suggestion of a prefix \*\**k-* or \*\**p-*.

In diachronic studies, when we are dealing with established affixes, such as \**pa-* [causative], \**ma-* [adjective], \**-en* [direct passive], \**-an* [locative], \**<in>* [perfective], \**<um>* [active], morpheme divisions should be straightforward, and there can be little question that a data set belongs to the etymon under consideration. However, in other instances, word division may be arbitrary or arguable, e.g., PHN \**qali-peqip* 'scapula' (Blust AE1#10) may well have been \**q<al>ip+qip* or PHF \**qati-mela* 'flea' (Blust AE1#23) could have been \**qa-timel-a* (see Zorc 1984-1985:88), while \**qalimaH* 'hand' is probably \**qa-limáH* 'the five', although \**q<al>imaH* is morphologically justifiable.

Blust (in progress) reconstructs inflected forms, e.g., PAN \**Sapúy* 'fire', PHN \**ma-hapúy* 'fiery, on fire', PHN \**mang-hapúy* 'set on fire', PIN \**mang-hapuy-i* 'apply fire to', PHN \**maR-hapúy* 'build a fire', PPH \**pa-hapuy-an* 'firewood, what is used to make a fire', PHN \**paR-hapuy-an* 'fireplace', PHN \**pang-hapuy* 'what is used to make a fire', PHN \**paR-hapuy-an* 'fireplace', PHN \**h<in>apúy* 'fired, exposed to fire', PHF \**S<um>apúy* 'start a fire', PPH \**hapuy-an* 'fireplace, hearth', PHN \**hapuy-hapuy* 'firefly', PHN \**hapuy-hapuy* 'tree sp.'. I have expressed reservations<sup>10</sup> when homonymous suffixes are involved, e.g., PAN \**-an* [locative focus verb] > Tnt *api-an* 'bring or apply fire to', PAN \**-an* [locative noun] > Tag *ápuy-an* 'stove, fireplace', Muk *puy-an* 'hearth', PHF \**-án* [instrumental imperative or aorist] > BtdIfg *apuy-an* 'that burned (as wood); that revived (i.e., coals)', Tag *ápuy-an* 'lighter'. Blust replied by citing the Starosta, Pawley & Reid hypothesis and concluding that 'cross-linguistically [...] a very good case can be made for the nominal and verbal uses of \**-an* reflexes as having a common source.'<sup>11</sup>

In the conclusion of another paper (Zorc 1982b:133), I pointed out that grammatical derivations may ultimately affect a reconstruction, e.g., Iban *dua* 'two', *be/dua?* 'divide', *se/duay* 'you-two'. How, if at all, does this affect the reconstruction of PAN \**DuSa* 'two'? Since the reconstruction of inflections and derivations is only now being actively pursued, we have yet to work out consistent criteria and ineluctable conclusions. What is important for future

<sup>10</sup> In a letter to Blust dated 6 July 1993.

<sup>11</sup> In a letter from Blust dated 20 August 1993.



work is the establishment of PAN, PMP, or lower order affixes, such as Blust (in progress) is doing so admirably. In order to do this well, we need to come to grips not only with the details of the synchronic morphology of the languages we use (beyond the phonology), but also the diachronic morphological developments in lower order subgroups.

### 3 Conclusions

I have reviewed the kind and reliability of evidence we use in our work. I believe we should state the degree to which we can rely on specific language evidence. For example, Malay is an excellent witness for PAN \*q in all positions, but a poor one for PAN \*S — and this only in initial position (see Zorc 1982b:124ff., footnotes 25-27). In working out the reflexes of PPH \*h for Hanunoo, I computed that Han retained initial \*h- on 57.6% and lost it on 42.4% of the data, and is therefore only a fair criterion language. Tharp (1974) worked out indices for the split of PAN \*R into Ilokano /g/ and /r/. No one has had the time to compute such figures for each An language, but surely a minimal statement along a scale of 'excellent — good — fair — poor' is in order. A list of test and criterion languages for various PAN phonemes (as provided by Blust 1980:16-18) is useful, but does not indicate the degree of reliability of some languages, e.g., Samal and Gorontalo are at best poor witnesses for initial \*S. Paiwan may turn out to be a far more critical test language than many of the others listed in evidence for \*d, \*D, \*j distinctions, yet it was omitted (or was it rejected?).

We need to view some evidence with caution if a language is either a provable donor or a major receiver of loanwords. For all of its importance in Austronesian reconstruction, Malay has been a major player in western Austronesian affairs since 200 BC (see Mahdi 1994). It has had a strong influence on Tagalog (Wolff 1976) and either through Tagalog or directly on other Philippine languages. Hence, reconstructions based upon Malay-Philippine correspondences that involve conservative phonemes or any irregularities should probably be dismissed or withheld, e.g., PHN? \*bintul 'k.o. fish or crab trap' (AE1#70), Mal *endal* > Tag *andál* (AE1#120; see Zorc 1984-1985:86 for additional citations). While these do have value, they put the burden upon researchers to decide if the \* should be changed to an <sup>h</sup>.

In several instances the question must be raised: Are we labeling the data or the reconstruction? One answer lies in what direction are we going: up or down? If in an initial comparison two reconstructions are possible due to ambiguities in the evidence, then we are reconstructing disjuncts. We could, of course, opt to cite a clear-cut lower level etymology (PHN \*gumi, PIN \*kumis 'beard') and limit the ambiguity to a note (e.g., possibly PMP if Fj *kumi* cognate). Once we have our etymologies, if we discover forms that compete for more than one, then we have disjunctive evidence. The optimum

situation is one wherein we discover disambiguating evidence, thereby resolving either a doublet or a disjunct.

One final issue that should be raised is the quantity of evidence required. Apart from startling cases of agreement with every possibility of borrowing excluded, I hold up as an ideal a 'rule of five', i.e., five non-contiguous languages supporting a reconstruction. There are several reconstructions, some going back as far as Dempwolff (1938, e.g., PHN *\*sauq* 'anchor', PHN *\*daqen* 'leaf') or Blust (1970, e.g., PHN *\*balung* 'pond, spring', PMP *\*utan* 'cargo') which have not acquired additional cognates. I am concerned that further supporting data have not come to light, especially with the immense quantity of research done recently on languages with excellent lexicographic studies. Do these attest to selective retentions (see Zorc 1986), fragile innovations, limited loans, or inadequate research methods?

Any kind of evidence can be valuable, once we understand its implications. BotSbl *gh̄ma?* 'sheath, scabbard' provides false evidence for PSP *\*Rúmaq* (a semantic innovation of PAN *\*Rumaq* 'house', i.e., for knife, *bolo*, etc.). If inherited, it should be Sbl *\*\*yúma?*; so it is a loan. But from which language? Tagalog has *kalúban*, which is unrelated to this etymon and probably a loan from a NPh language (Cas, Kpm, Png, Ilk, etc.). The Sambal form may, however, provide evidence that the early Tagalogs said *\*\*gúma?* upon their arrival in southern Luzon, but later gave that form up in favor of a then current local term.

As research continues and our knowledge grows, the status of evidence must get reevaluated. Thus, accent phenomena have led to reinterpretations of the PAN system (Dahl 1981:108-117, Zorc 1983, Wolff 1991, 1993b). Wrong turns and errors may still have important implications for other branches of our science. For example, if Wolff (1974) is correct in rejecting PAN *\*r*, Dempwolff's treatment of Ivatan as a test language for that phoneme (1926) may be reinterpreted as a treatise on the sociolinguistic phenomenon of intimate borrowing.

#### Abbreviations

ACD	Blust (in progress)	CDF	Zorc (1979-1985)
AE1	Blust (1980)	Ceb	Cebuano
AE2	Blust (1983-1984)	Ch	Mathew Charles (ms)
AE3	Blust (1986)	Ch	Chinese loanword
AE4	Blust (1989)	Dbl	doublet
Akl	Aklanon	DS	Zorc data system (ms)
An	Austronesian	Dsj	disjunct
Arb	Arabic loanword	Fj	Fijian
Bey	Blust (1988)	GCP	Greater-Central-Philippines (Blust 1991)
Bik	Bikol		
BotSbl	Botolan Sambal	HLC	Dyen (1990)
Cas	Casiguran Dumagat	Hov	Malagasy (data from VL3)

*Evaluation of evidence and errors in Austronesian reconstruction*

Iban	Iban (Sea Dayak)	PMP	Proto-Malayo-Polynesian
Ilk	Ilokano	Png	Pangasinan
Ivt	Ivatan	PNP	Proto-Northern-Philippine
Jav	Javanese	POC	Proto-Oceanic
Kal	Kalamian	Port	Portuguese loanword
Knk	Kankany	PPH	Proto-Philippine
Kpm	Kapampangan	PSC	Proto-South-Cordilleran
Lar	Dyen (1953a)	PSP	Proto-Southern-Philippine
Mal	Malay	PWI	Proto-West-Indonesian
Mar	Maranao	Rd	Reid (1971)
MBT	Wolff (1976)	Sbl	Sambal(ic)
McF	McFarland (1977)	Skt	Sanskrit loanword
Mlg	Malagasy (data from Dahl)	sni	source not indicated
Mon	monosyllabic root	Sp	Spanish loanword
NACD	note in Blust (in progress)	SPh	southern Philippine
NgD	Ngaju Dayak	Syn	synonym
NPh	northern Philippine	Tag	Tagalog
OJav	Old Javanese	TAG	Ferrell (1969)
PAA	Blust (1970)	Tamil	Tamil loanword
PA1	Blust (1972a)	TB	Toba Batak
PA2	Blust (1972b)	Tir	Tiruray (Blust 1992a)
PA3	Blust (1973)	Tum	Blust (1992b)
Pai	Paiwan	VL3	Dempwolff (1938)
PAN	Proto-Austronesian	VRR	Blust (1981)
PANN	Wolff (1993a)	WBM	Western Bukidnon Manobo
pc	personal communication(s)	Yap	Yap (1977)
PFL	Zorc (1971)	x	a loan or maverick reconstruction
PHF	Proto-Hesperonesian-Formosan	*	a reconstruction
PHN	Proto-Hesperonesian (West Austronesian)	**	form not known to occur
PIN	Proto-Indonesian	<X>	an infix
PMJ	Proto-Malayo-Javanic	-	morpheme break
		/	suspect morpheme break
		+	monosyllabic root

REFERENCES

- Blust, Robert A.  
 1970 'Proto-Austronesian addenda', *Oceanic Linguistics* 9:104-162.  
 1972a 'Proto-Oceanic addenda with cognates in non-Oceanic Austronesian languages', *Working Papers in Linguistics* 4.1:1-43.  
 1972b 'Additions to "Proto-Austronesian addenda" and "Proto-Oceanic addenda with cognates in non-Oceanic Austronesian languages"', *Working Papers in Linguistics* 4.8:1-17.  
 1973 'Additions to "Proto-Austronesian addenda" and "Proto-Oceanic addenda with cognates in non-Oceanic Austronesian languages" II', *Working Papers in Linguistics* 5.3:33-61.  
 1980 'Austronesian Etymologies I', *Oceanic Linguistics* 19.1:1-181.  
 1983-1984 'Austronesian Etymologies II', *Oceanic Linguistics* 22-23:29-149.

- 1986 'Austronesian Etymologies III', *Oceanic Linguistics* 25:1-123.
- 1988 *Austronesian Root Theory: An Essay on the Limits of Morphology*. Amsterdam: John Benjamins.
- 1989 'Austronesian Etymologies IV', *Oceanic Linguistics* 28:111-180.
- 1991 'The Greater Central Philippines hypothesis', *Oceanic Linguistics* 30.2:73-129.
- 1992a 'On speech strata in Tiruray', in: M.D. Ross (ed.) *Papers in Austronesian Linguistics* 2. Pacific Linguistics A 82:1-52.
- 1992b 'Tumbaga in Southeast Asia and South America', *Anthropos* 87:443-457.
- in progress *Austronesian Comparative Dictionary*.
- Conant, Carlos E.
- 1911 'The RGH law in Philippine languages', *JAOS* 31:181-196.
- Dahl, Otto C.
- 1976<sup>2</sup> *Proto-Austronesian*. Scandinavian Institute of Asian Studies Monograph Series 15. London: Curzon Press.
- 1981 *Early Phonetic and Phonemic Changes in Austronesian*. The Institute for Comparative Research in Human Culture, Series B: LXIII. Oslo: Universitetsforlaget.
- Dempwolff, Otto
- 1926 'Ivatan als "Test-Sprache" für uraustronesisches \*I', *Zeitschrift für Eingeborenen-Sprachen* 16:298-302.
- 1934-1938 'Vergleichende Lautlehre des austronesischen Wortschatzes', *Zeitschrift für Eingeborenen-Sprachen*, Beihefte 15, 17, 19. Berlin: Dietrich Reimer.
- Dyen, Isidore
- 1947 'The Tagalog reflexes of Malayo-Polynesian \*D', *Language* 23:227-238.
- 1951 'Proto-Malayo-Polynesian \*Z', *Language* 27:534-540.
- 1953a *The Proto-Malayo-Polynesian Laryngeals*. William Dwight Whitney Linguistic Series 9. Baltimore: Linguistic Society of America.
- 1953b 'Dempwolff's \*R', *Language* 29:359-366.
- 1990 'Homomeric lexical classification', in: Philip Baldi (ed.) *Linguistic Change and Reconstruction Methodology*. Trends in Linguistics: Studies and Monographs, 45. Berlin: Mouton De Gruyter.
- Ferrell, Raleigh
- 1969 *Taiwan Aboriginal Groups: Problems in Cultural and Linguistic Classification*. Institute of Ethnology, Academia Sinica, Monograph no. 17. Taipei.
- Gonda, J.
- 1973 *Sanskrit in Indonesia*. New Delhi: International Academy of Indian Culture.
- Jones, Russell
- 1978 *Arabic Loan-Words in Indonesian*. London: School of Oriental & African Studies.
- Mahdi, Waruno
- 1994 'Some Austronesian maverick protoforms with culture-historical implications I', *Oceanic Linguistics* 33.1:167-229.

- McFarland, Curtis D.  
1977 *Northern Philippine Linguistic Geography*. Study of Languages and Cultures of Asia and Africa Monograph Series 9. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.
- Nothofer, Bernd  
1990 'Review Article' of Blust 1988. *Oceanic Linguistics* 29.2:132-152.  
1991 'More on Austronesian Radicals (or Roots)', *Oceanic Linguistics* 30.2:223-258.
- Paz, Consuelo J.  
1981 *A Reconstruction of Proto-Philippine Phonemes and Morphemes*. Publication 3 of the Cecilio Lopez Archives of Philippine Languages and the Philippine Linguistics Circle. Diliman: University of the Philippines.
- Pejros, I.  
1994 'Some problems of Austronesian accent and \*t ~ \*C (Notes of an outsider)', *Oceanic Linguistics* 33.1:105-127.
- Reid, Lawrence A.  
1973 'Kankany and the problem \*R and \*l reflexes', in: A. Gonzalez (ed.) *Parangal Kay Cecilio Lopez*. Quezon City: Linguistic Society of the Philippines, 51-63.
- Reid, Lawrence A. (ed.)  
1971 *Philippine Minor Languages: Word Lists and Phonologies*. Oceanic Linguistics Special Publication 8.
- Starosta, Stanley, Andrew K. Pawley & Lawrence A. Reid  
1982 *The evolution of focus in Austronesian*. Pacific Linguistics C 75:145-171.
- Tharp, James A.  
1974 'Notes on the Ilokano reflexes of Proto-Austronesian \*R', *Working Papers in Linguistics* (University of Hawaii) 6.6:47-51.
- Tsuchida, Shigeru  
1976 *The reconstruction of Proto-Tsouic phonology*. Study of Languages and Cultures of Asia and Africa, Monograph Series, No. 5. Tokyo: Gaikokugo Daigaku.
- Wilkinson, R.J.  
1959 *A Malay-English Dictionary* (romanized). London: Macmillan & Co.
- Wolff, John U.  
1974 'Proto-Austronesian \*r and \*d', *Proceedings of the First International Conference on Comparative Austronesian Linguistics*, *Oceanic Linguistics* 13:77-121.  
1976 'Malay borrowings in Tagalog', in: O.D. Cowan & O.W. Wolters (eds.) *Southeast Asian History & Historiography Essays Presented to D.G.E. Hall*. Ithaca: Cornell University Press, 345-367.  
1991 'The Proto-Austronesian phoneme \*t and the grouping of the Austronesian languages', in: Robert Blust (ed.), *Currents in Pacific Linguistics: Papers on Austronesian languages and ethnolinguistics in honour of George W. Grace*, Pacific Linguistics C 117:535-549.  
1993a 'The PAN Phonemes \*ñ and \*N', *Oceanic Linguistics* 32.1:45-61.  
1993b 'Proto-Austronesian Stress', in: Jerold A. Edmondson & Kenneth J. Gregerson (eds.) *Tonality in Austronesian Linguistics*, Oceanic Linguistics, Special Publications 24:1-15.

Zorc, R. David

- 1971 'Proto-Philippine Finder List'. Cornell University: typescript.
- 1977 *The Bisayan Dialects of the Philippines: Subgrouping and Reconstruction*. Pacific Linguistics C 44.
- 1979-1985 *Core Etymological Dictionary of Filipino*. Fascicles 1-4 (A-H). Manila: Linguistic Society of the Philippines.
- 1981 Review of Consuelo J. Paz, *A Reconstruction of Proto-Philippine Phonemes and Morphemes* (1981). *Philippine Journal of Linguistics* 12.2:44-50.
- 1982a 'Micro- and Macro-subgrouping: Criteria, problems, and procedures', in: Rainer Carle et al. (eds.) *GAVA': Studies in Austronesian languages and cultures dedicated to Hans Kähler*, Band 17. Berlin: Dietrich Reimer Verlag, 305-320.
- 1982b 'Where oh where have the laryngeals gone? Austronesian laryngeals re-examined', in: Amran Halim, Lois Carrington & S.A. Wurm (eds.) *Papers from the Third International Conference on Austronesian Linguistics 2: Tracking the travellers*. Pacific Linguistics C 75:111-144.
- 1983 'Proto Austronesian accent revisited', *Philippine Journal of Linguistics* 14.1:1-24.
- 1984-1985 'A Guide to the Filipino Linguist in Reading and Appreciating Robert A. Blust's "Austronesian Etymologies"', Review Article in *Philippine Journal of Linguistics* 15.2-16.1:81-93.
- 1986 'The genetic relationships of Philippine languages', in: Paul Geraghty, Lois Carrington & S.A. Wurm (eds.) *FOCAL II: Papers from the Fourth International Conference on Austronesian Linguistics*, Pacific Linguistics C 94:147-173.
- 1987 'Austronesian apicals (\*dDzZ) and the Philippine non-evidence', in: Donald Laycock & Werner Winter (eds.) *A world of language: papers presented to Prof. S.A. Wurm on his 65th birthday*. Pacific Linguistics C 100:751-761.
- 1990 'The Austronesian monosyllabic root, radical or phonestheme', in: Philip Baldi (ed.) *Linguistic Change and Reconstruction Methodology*. Berlin: Walter de Gruyter, 175-194.
- 1991 *Tagalog Slang Dictionary*. Kensington, MD: Dunwoody Press.