

THE GENETIC RELATIONSHIPS OF PHILIPPINE LANGUAGES

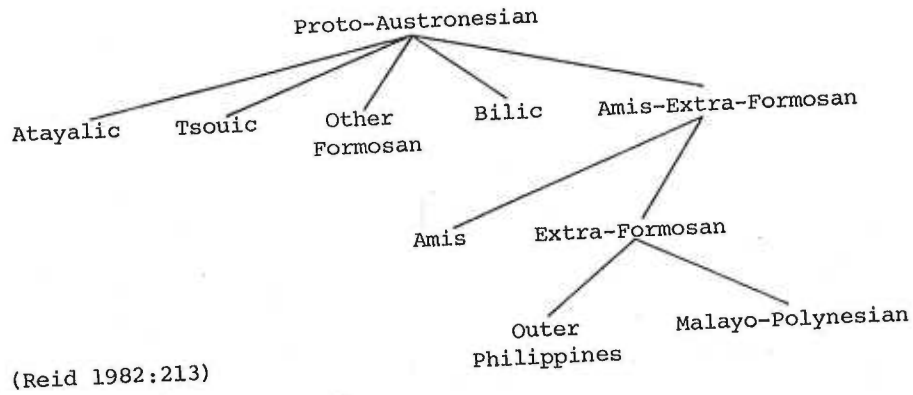
R. David Zorc

0. BACKGROUND

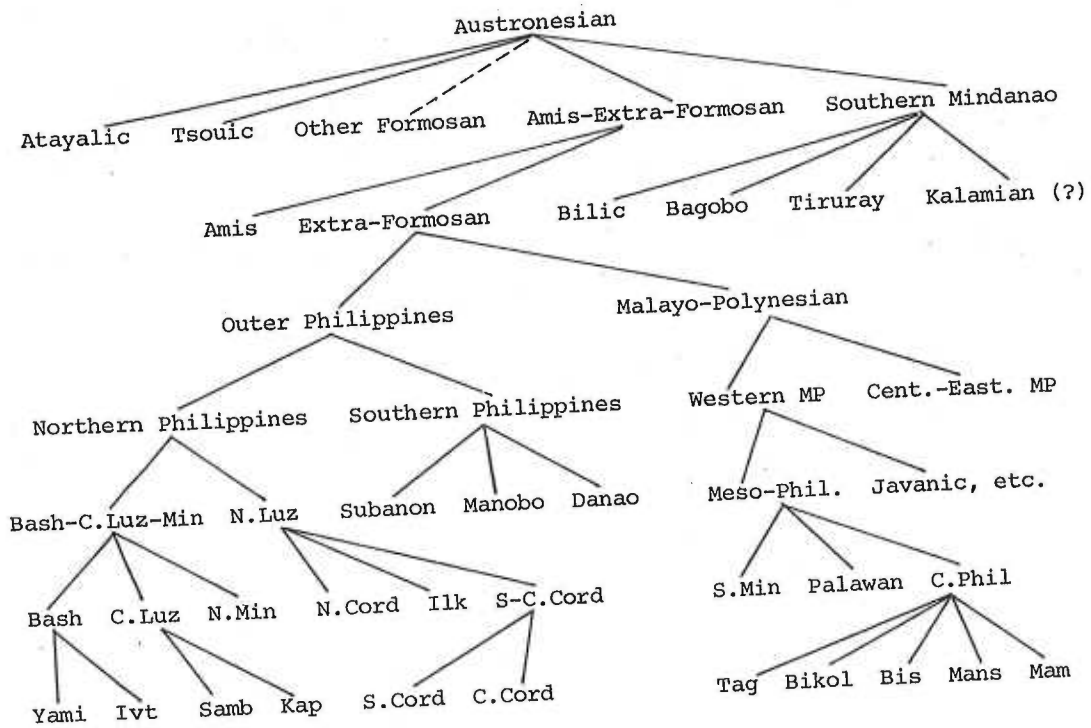
At TICAL (the Third International Conference on Austronesian Linguistics, Bali, 1982) Reid challenged the assumption that most, if not all, of the languages spoken in the Philippine archipelago descend from an immediate proto-language (PPH). The evidence he presented suggested that those languages which share in nasal infixation into root words (*CV_nCV[C]) form a subgroup of AN - the northern Philippine and Bilić languages were shown not to have *nC cognates of PMP etyma and were thus excluded from the subgroup (Reid 1982:204,211ff). Reid therefore assigns to nasal infixation the status of a highly qualitative phonological innovation, which subgroups Central and some other Philippine languages with Malay/Indonesian and Oceanic languages. He tentatively drew Tree 1 (1982: 213) and Tree 2, which delineates the various Philippine subgroups (p.c.).¹ An analysis of the latter tree reveals a genetic chasm between Tiruray and Manobo, Kalamian and Palawan, or Bisayan and Ilokano. Tagalog is portrayed as closer to Malay than to Bontok, Cebuano to Fijian than to Kapampangan, and Sambal to Amis than to Mamanwa. A corollary to this hypothesis is that all of the exclusively shared agreements amongst Ph languages are "the effects of thousands of years of language convergence" (p.c., Reid to Ruhlen, 27.8.1982) or, those that are genuine must be relics or retentions attributable to PAN or pre-PMP. These genetic implications must be tested.

It is my purpose here to show that there are a compelling number of lexical and other innovations that substantiate a Western Austronesian node more traditionally thought of as "Proto-Philippine". Because this subgroup has a high order of diversity, and due to the propensity and natural probability for lexical replacement, not all groups continue to share all innovations. Indeed, as my lexical study has continued over the past 15 years, I have been impressed by innovations that skip over micro- or lower-level subgroup boundaries and yet delineate the same macro-subgroup established by widespread innovations. These *selective innovations* do not fit a convergence hypothesis, and hence form an integral part of my paper.

Paul Geraghty, Lois Carrington and S.A. Wurm, eds *FOCAL II: papers from the Fourth International Conference on Austronesian Linguistics*, 147-173. *Pacific Linguistics*, C-94, 1986.
© R. David Zorc



Tree 1



Tree 2

Austronesian high-order subgrouping (Reid, 21.11.1981)

1. PROCEDURES FOLLOWED

(a) I began this study by organising a looseleaf edition of McFarland 1977 on the basis of the number of etyma and uniques, e.g. M5 *brain* shows only cognates of PPH *hútək *brain*, M19 only PPH *dílaq *tongue* - these were assigned a code 1-Ø (one etymon, no uniques). M36 has cognates of PPH *kukúh *finger nail*, but KnKS has a unique, kiwí - this was assigned a code 1-1 (one etymon, one unique). This was continued until my last entry was M282 *lie (falsehood)* - Ø-22 (no [widespread] etymon, 22 uniques [see 3.7 below]).

(b) I then compared these with data in Reid 1971 and Yap 1977, making a card for each reconstruction that could be assigned to "PPH" or any higher level up to PAN. Examples of the resultant cards are:

<i>eye</i>	PAN *maCá	>	PPH *matá	Cha oho(s)
M- 07	[1-1-Ø]			Sina bul'ga?
Y- 253	[1-1-4]	CCrN	*?atá	?Bot mulátot
R- 96		Mn	*bərən	Tas ?igh?a
B- 45	[12345]			
D- 03	[77.2%]			
F- 206	[1+3?]			

Tree: PAN > PPH > PNP/PSP
 Note: continuance into *all* major subgroups.

<i>fire</i>	PAN *Sapúy	>	PPH *hapúy	Cha fuwego
M- 159	[1-Ø-Ø]	NMg	*báRah	Mam gabok
Y- 20	[1-5-2]	Bs+Bk	*kaláyu	Tag ^x ?apóy<SLZ
B- 143	[12345]	Mk	*?atulun	
D- 46	[15.05%]	IB1	*lipo[]	
F- 88	[1+4?]	Sn	*putu	
R- 112	[1+1+1]	Gor+Moŋ	*súluq	

Tree: PAN > PPH > PNP/PSP
 Note: SPh intrusions with multiple replacements (except Pl, Kl, SMg, Sb, Dn, Mb, OBl, Mn)

<i>left</i>	PAN *ka+wiRí	>	PPH *ka+wiRí	Cha iskierda
		Iv	*g(a)-uri	Gad dawí
M- 32	[1-5-4]	NCr	*dimiRí	Btd wodhoŋ
Y- 171	[1-5+-5]	Ilk+CCr	*kǎ(N)+tigíd	Bot ?úki
R- 161		CCr	*?igíd	Sbl ^x wiri
B- 02	[1235]	SPh	*()-iban	Sina háyin
D- 91	[8.17%]	CPh/SPh	*waláh	Kpm ^x kayli?
F- 383	[1+4+]	Gor+Moŋ	*kələ+wiri	Ttb ^x ka?mbiri
				Tsw kolo?di

Tree: PAN > PPH > PNP/PSP
 Note: multiple replacements in several major subgroups, such that there are selective retentions in Cas, Mlw, Png(NPh) and Sn, Mn(SPh).

tongue	PAN *Səma	
"lick"	PHF *dɫlaq	> PPN *dɫlaq tongue
M- 19	[1-ø-ø]	PPH *dɫlaq
Y- 97	[1-ø-2]	
R-323		Cha lenwa
B- 32	[13]	Ilt limut
D- 31	[25.8%]	
F-239	[2+5?]	Fm *li[d]am

Tree: PAN *Səma > PPH ø (loss); PHF *dɫlaq
 lick > PHN tongue (sem.shift) > PPH > PNP/PSP
 Note: continuance into all major subgroups after
 semantic shift, with apparent loss of PAN etymon.

By this method, the earliest known etymologies were established, and subgroups that differed from these in form or meaning were set up. The code was expanded to show: PPH/PAN etyma - subgroup etyma - uniques and loans.

(c) I went through all the etymologies reconstructed for discrete Ph subgroups to verify and/or expand the list of proposed innovations. There is a slight imbalance in these results. Studies by Sneddon of Minahasan (1978) and Sangiric (1984) or by Reid of Central Cordilleran (1974) employ quite a large data base, while other works [e.g. by Allison of Danaw (1979), by Gallman of Mansakan (1979), or by Thiessen of Palawanic (1981)] are limited to the SIL 372-meaning list. Hence, number of innovations should not be taken as indicative (at this stage) of historical conjectures (e.g. time-depth of separation, status as an "innovating" language, etc.) beyond the subgrouping hypotheses presented.

(d) I checked all posited innovations against all available AN reconstructions, withdrawing those that can currently be proven to be relics (selective retentions). Further research will undoubtedly reduce the various lists. However, my own research is still far from complete and more candidates will be forthcoming. Some innovations may well be shifted upwards (e.g. PPH < PHN or PMP), but could still be of value unless they can be established at the PAN level.

Scholars' continuing additions to AN etymologies (and Blust is to be commended in particular for his untiring efforts) increase our corpus of reconstructions to beyond double that of Dempwolff's time. Although few have been given in support of specific subgroups, the time is nigh when many of them could be. Of Blust's 443 "Austronesian etymologies II" (n.d.), only 50 were cause for withdrawal from or upwards revision of my entire list (but rarely more than three for any proposed subgroup within the entire Philippine family); a few are now put forward as PHN innovations (Table 3), along the criteria outlined below (in section 4).

2. METHODOLOGICAL PRELIMINARIES

Although there are problems involved in the isolation of lexical innovations, there are means of dealing with them. I outlined and applied certain precautionary measures (Zorc 1977:234f) and refined them (Zorc 1982:313f) as follows.

2.1 Limit forms to basic vocabulary and avoid items of trade or culture that could freely pass from one language to another

I have almost exclusively confined this study to the basic meanings found in McFarland 1977, Reid 1971, Yap 1977, Ferrell 1969 and Ray 1911, and checked these against my own data files and the data and reconstructions published by Dempwolff, Dyen, Blust, Tsuchida, Nothofer, Sneddon, Mills, et alii. [Forms considered non-basic or cultural are marked with a code "6" in the tables; see also section 4.]

2.2 Dismiss forms with phonological irregularities

That is, not in conformity with the standard reflexes worked out for a given language, e.g. h in a language that loses *H or *S, r in a language where *R > g, *R > y, *d/*D > d, etc. Note that *morphophonemic changes* (e.g. metathesis, syncope, assimilation, etc.) are not taken as phonological irregularities, and may be treated as innovations (e.g. Mn+Sn *leRe? *neck* < PHF *liqəR) provided that any such systematic changes cannot be counted more than once. That is, one lexical item may be included as representative of the phenomenon establishing a phonological (rather than lexical or semantic) innovation, e.g. Bs *C| < PHN *|C; Sn *s...t < PMP *t...s, Buh -wa- < PAN *-a(qSH?ø)u-, etc.

2.3 Reconstruct, wherever possible, an etymon for a given meaning at the earliest possible stage

For example: *blood* was PAN *Da:Raŋ, so PSP *duRúŋ or SMB *dipanug *blood* are well-established innovations (but see 2.5). The 49 reconstructions in Blust 1981a (coded "12345") that have cognates in all major AN subgroups or the 51 meanings in Dyen, James and Cole 1967 that have a retention rate above 14% were considered highest in quality, and were coded "1" in the tables (viz: highly qualitative innovations). Where other etyma can be established with confidence (in both form and meaning) at the PMP or PAN level, and appear to have been replaced, they were coded "5" in the tables (see 4).

2.4 Consider the character and quality of each proposed innovation, including its geographical and linguistic distribution, potential spread, etc.

While it is difficult to distinguish a common from a spread innovation, and, in the case of conservative phonemes to isolate a borrowing, linguistic geography (such as McFarland 1977) greatly assists in showing how forms may stay within or creep over proposed subgroup boundaries.

In determining the quality of an innovation, several criteria will be discussed in section 4. Two, however, are noteworthy. Dyen, James and Cole (1967: 168) suggest that meanings of lower productivity (or low retention rates) contribute more information than those with high retention rates. The common retention of PAN *limá *five* (D1 - first on Dyen's list) or PAN *maCá (D3) indicates little more than that the languages compared are Austronesian. But, the sharing of Iv *<um>tək by Itv and Itb, or of Mk *balləŋ by Mansakan dialects in the meaning *play* (D196) or of IMb *hagsil or Iv *rukmeŋ in the meaning *cold* (D183) should be highly informative and therefore indicative of subgrouping. However, I have

noted that since these meanings are highly subject to replacement, they are equally highly susceptible to borrowing. In evaluating such innovations, critical judgments become necessary.

More important are those meanings that have a high probability of retention but are nonetheless replaced. These appear to yield highly qualitative innovations, such as Mn *bərən *eye* (D3), Bk *gədaʔán or Iv *líman *die* (D7), or Bs+Bk *kaláyu, Mk *atulun, Ibl *lipo[], Sn *putuŋ *fire* (D46).

2.5 Determine if the innovation is a formal or semantic one, and if the change could happen independently

Often an old form can be established as having changed meaning (PMP *bǎRéq *to swell (as abscess)* > SBS *bagáʔ *thick*, PHF *báRaH *embers* > Iv *red*, NMg *fire*, PHN *sídəŋ *blinded by glaring light* > SMg *day*). Note that each of these semantic shifts involves the replacement of well-established etyma: PMP *(ma)kapál *thick*, PMP *ma-íRaŋ *red*, PAN *Sapúy *fire*, PAN *qal(ə)jaw *day* respectively. New forms have also been coined from previously unknown material (e.g. NCr *busáli *abscess*, PHN *lú(n)tuŋ *cook*), while others may be the result of reshaping (e.g. PPH *bulbul *feather* < ?PMP *bul(u)-bul(u), PHN *laqlu *pestle* > PHF *qaSəlu). When the same innovation has happened independently [Ilk nala-bága *red*, Bon ballaan *red yam* :: Iv *red* (above), or Tsw baha *fire* :: NMg *fire* (above)] further evaluation, including a re-look at the semantic assignment of the etymon, is necessary. This process could involve either rejection of the candidate(s) or a devaluation of the quality assigned.

Even after applying these measures, any proposed innovation may be a relic lost everywhere else, or as yet undiscovered in another language. However, as the number of such candidates increases, there is a strong probability that the majority will survive even protracted research for outside cognates. I take heart in the fact that of the 85 innovations posited by Reid (1974) to establish the CCr subgroup (or groups within CCr), only 18 need qualification or revision, e.g. ten may have been borrowed by neighbouring Kly, Kyp, Ibl, or Gad groups. Meanwhile, my own research has uncovered additional forms that bring the CCr total to 98; these will be published in subgroup-specific studies continuing the present paper.

3. TERMINOLOGY

3.1 Highly retentive cognates

Highly retentive cognates are found in most (if not all) major subgroups under discussion. Examples amongst Ph and AN languages include: PAN *Zálan *road, path*, PAN *qaCáŷ *liver*, PAN *súsu *breast*, PAN *matákut *to fear*, PAN *Cáŋis *to cry*, PAN *káʔən *eat*, PAN *ŋájan *name*, PAN *búlan *moon*, PAN *quZÁN *rain*, etc.

3.2 Selective retention cognates

Selective retention cognates are limited to a single subgroup (that can be established along comparative arguments) but are then found in distantly related languages. For example, WBS *dahíʔ *forehead* < PAN *daqíS is limited to all

members of the proposed WBs subgroup, yet no cognate has been found in any other Ph language researched to date. Other interesting examples include:

- Dn+Mb *?atiŋ *perspiration* < PMP *atiŋ [Blust 1980:#24]
 Dn *nipay *snake, serpent* < PMP *nipay [Blust 1981a:#106]
 Iv *sayap *to fly* < PHN *sayap *wing* [Blust n.d.:#355]
 Bl *?ikoŋ *tail* < PMP *i(ŋ)kuŋ [Blust 1980:#167]
 PSP *qəbəl *smoke* (Mk,Sb,Dn,Mb,Bl,Pon,Tsw) < PHF *qəbəl [Tsuchida 1976]
 PNP *dāŋan *span [8 inches]* < PMP *dāŋan [Blust 1980:#442]

I have over a hundred additional examples in my files. Such etyma would not be allowed in the strictest applications of the comparative method because they are clearly retentions (not innovations). Nevertheless, (a) they serve as a synchronic isogloss around the proposed group in many observed instances, and (b) their retention amidst heavy pressure for innovation (as evidenced by replacement in the other groups) cannot be adequately explained but surely reflects some historically relevant phenomenon. I have not included these intentionally here, but numerous discoveries of similar forms will probably be forthcoming from the candidates in the various tables. However, each list in toto could still be used as a subgroup identifier for newly discovered speech varieties with some assurance of success. [See, for example, that in Zorc 1972:125-128 for WBs, or Zorc 1977:269-276 for Warayan and CBS.] This mixture of synchronic and diachronic material for subgrouping purposes is not without precedent: lexicostatistics counts the sum of retentions and shared innovations without distinguishing between them, and yet is of some value in formulating a subgrouping argument - especially if it coincides with the results of other methods.

3.3 Widespread innovation (w)

Cognates are limited to numerous language groups representing the most diverse nodes of a proposed tree, e.g. Table 1.

3.4 Selective innovation (s)

Cognates are found in only a few language groups that are geographically and genetically diverse, representing distinct nodes of a proposed tree, e.g. Table 2. Because of the distance (temporal and spatial) between the language groups involved, I can not see how a hypothesis of borrowing or convergence can be put forward apart from the proto-language immediately shared by those groups. The only other alternative is that the etyma in question are selective retentions, which may be proven by the discovery of outside cognates. I feel reasonably confident that a sufficient quantity will stand the test of time and continued research.

3.5 Contact innovation (c)

Cognates are found in two (or more) genetically distinct languages and are the result of common innovation after contact between the groups. Tag páwɨs and Kpm páwas clearly reflect a SLz innovation *páwəs *sweat*, but do not serve to subgroup these languages together. Such developments help establish the degree of convergence between languages and must make the researcher chary of

positing selective innovations when the distribution of forms will not otherwise warrant such reconstruction, viz: a genuine cognate between Kpm and Tag would normally yield a PPH etymon. See Pallesen 1977 for examples of Sama-Tsg convergence.

3.6 Borrowed innovation (b)

Cognates are found in genetically distinct languages, but irregularity of form or distribution suggests in a straightforward manner that one speech variety has borrowed from another. Ntg, Agy *tambək fat* (expected ***tambə?*) must be loans from Bs **tambək* because of the irregular reflex for *k (> Kl Ø) and the additional evidence of Kal *linuk fat* (which is probably itself a Kl innovation replacing PPH **tabá?*). Since virtually all Mb languages retain reflexes of PAN **Sapúy fire*, Dbw *káyu* and Ags *kaədu?* must be loans from a SBs dialect of Bk+Bs **kaláyu fire*. Such judgments must be made explicit, since they are open to criticism. Pallesen (1978:92f) was quite correct in his comments on my treatment of several Bs innovations:

When, however, a large number of putative exclusively shared innovations are demonstrated to have cognates outside the subgroup, then the boundaries of the subgroup are very much in question It is undoubtedly valid to identify similar forms in language B as borrowings from language A which is known to be influential, but not if the forms themselves are the main evidence of the influence of language B.

Exclusion of all of the forms Pallesen cites would have resulted in a more "air-tight" Bs group, but would have been tantamount to "sweeping the problems under the rug". As it is, it is probably best to list all data, facts, and hypotheses so that scholars may productively engage in debate.

3.7 Uniques (u)

Forms are limited to a single speech variety (or dialect group). Such forms do not enter into this survey, but it is noteworthy that genuine uniques are not nearly as numerous as they at first appear. Many are selective retentions. Thus, Ilt *pəsit blind* is cognate with Bin *pəsit* < PHN **pəsit* [Blust n.d.: #278], Ilt *pandək star* with Miri *fatak star* (and possibly Moŋ *pandək spot, speak* < PHN **pandək star; speak (of light)* [Blust n.d.:#260]). Ilt *tambian five* is a semantic innovation from PHN **sa+n- one + *bɬaŋ count* - nor is Ilt unique in having a quinary numbering system [Dahl 1981a:fn.5, via Blust]. Sneddon (p.c.) has encouraged me to treat San and Snl as a single witness since they form a dialect chain; r or h reflexes < *R are evidenced in dialects of each.

4. CRITERIA: TYPE, QUALITY, AND NUMBER OF INNOVATIONS

It may seem a commonplace, but it is an often overlooked fact that every innovation means something. Interpreting each innovation requires isolating the types outlined above: widespread, selective, contact, or borrowed. Furthermore, the quality of an innovation must be assessed. I suggest the following measures as a rule of thumb to weigh the innovations proposed in the various tables. The

ordering should be considered as *relative* (not absolute), so that the higher the rating, the higher the *overall* quality, e.g. 1 is better than 5, but it does not follow that 3 is *necessarily* better than 4, etc.

- 1 - Replaces a well established PAN or PMP form in a highly-retentive meaning. [See 2.4 and 3.1.]
- 2 - Formal innovation (e.g. a change in or addition of morphological material) not attested outside the group.
- 3 - Semantic innovation (i.e. the *form* may be quite old, but a definite shift in meaning has occurred replacing the etymon most closely established as having that meaning). [See 2.5.]
- 4 - A phonological or morphophonemic innovation. [See 2.2.]
- 5 - Replaces an earlier (PAN, PMP, PHN) form, but in a meaning of low retention rate or with a high probability of replacement. [See 2.4.]
- 6 - Although an item of trade or culture, distribution suggests a special kind of innovation is involved (e.g. semantic shift, pre-historical contact, etc.).
- 7 - Currently known distribution suggests innovational status, but continued research is required to establish this, viz: "none of the above".

Reactions from several colleagues have made it clear that these seven categories are perhaps better characterised as *kinds of innovations* to which further judgments concerning *quality* must be added, e.g. H (high) ... L (low). I agree that a complex morphophonemic innovation such as cluster metathesis (4H) is of greater significance than a mildly deviant semantic shift (3L) or the addition of common affixes such as *si or *i [name markers] to pronominal stems (2L). Even if this requires "going back to the drawing board", scholars should attempt to rate the quality of proposed innovations and make their own criteria explicit.

Quantity, while relative to the state of current research, must support any subgrouping hypothesis. It would be remarkable indeed if a genuine subgroup left its evidence in one linguistic area (e.g. phonology) but not in any other (e.g. lexicon or grammar). The subgrouping proposed by Reid (1982) based on nasal infixation (or the lack thereof) suffers by its singularity and the lack of additional supporting evidence. The appearance of diverse cognates of *ŋC forms could be a selective retention, a contact innovation (Ml/In influence in the central and southern Ph has been strong - see Wolff 1976 for numerous examples), or a borrowed innovation. Conversely, the non-appearance of cognate *ŋC forms could be the result of independent loss, or complex parallel developments (in Formosa and in the northern Ph or Bilic). In any event, failure to share in an innovation is not of itself proof of exclusion from a subgroup.

The origin and status of *ŋC forms needs further study and evaluation. Of the 22 forms cited from Bontok which do not show a medial nasal cluster (Reid 1982:205f), only four have CPh and SPh cognates which unequivocally reflect a nasal (*ampil *favour one person over another*, *dampilas *cliff*, *kempit *press, clamp*, *kindat *wink, open up eyes*) whereas eight have not been observed in these latter groups with any nasal (*apu *grandparent/child*, *ma-hátaq *unripe*, *hútek *brain*, *lútuq *cook*, *tábun *cover up*, *tahép *winnow*, *t'ipun *assemble*, *túbuq *grow*). Thus, CPh and SPh languages are intermediate on a cline between heavy nasal infixation (Oc/Ml) and little to nil nasal infixation (NPh/Fm). Note that

NPh languages do have nasal clusters in etyma that are not likely to be loans: PNP *andu *long* (by syncope < *anaduq), PNP *hiŋpis *thin*, Iv+Ncr *tuŋduq *point*, ICS *suŋbat *answer*, NCr *siŋpət *good, kind*, Ilt pandək *star* (see 3.7). Nasal infixation (or its loss) is far from being established as a highly significant qualitative innovation.

5. THE EVIDENCE FOR A PHILIPPINE SUBGROUP

The number of exclusively shared lexical innovations that I have gathered thus far suggests that the languages of the Philippine archipelago (exclusive of the Sama-Bajaw group) form a single AN subgroup. This "Philippine" (or a less geo-politically prejudicial label of *Eastern Hesperonesian*) Group includes Yami (of Botel Tobago Island, within the Bashiic/Ivatanic subgroup of NPh) and the languages of northern Celebes (including Minahasan, Sangiric, Mongondow, and Gorontaloic within SPh).

Widespread innovations in support of this group are presented in Table 1, and selective innovations in Table 2. Constraints of both time and space have not permitted the inclusion of the data (which can be found in the sources cited), but languages or subgroups that have cognates of the etyma under consideration are listed. The format adopted gives the following information:

- etymon number
- type of innovation (widespread, selective, etc.)
- quality [kind] of innovation, using a numerical code (section 4)
- level of reconstruction (e.g. PPH, PHN)
- reconstructed shape
- semantic assignment
- data sources (McFarland, Reid, Yap, Ferrell, Blust, etc.)
- subgroups or languages that have cognates
- (irregularities of any kind) [e.g. (+Isq) = form probably borrowed by Isneg.]
- languages reflecting semantic shifts are put after a semicolon, along with the meaning [e.g. ;Akl *stop* = the cognate means *stop* in Akl]
- [any additional information]

I am reluctant to draw a tree at this stage and feel that Ph developments were more like amoebic colonisations than absolute splits. In general, I subscribe to the tree drawn by McFarland (1980:62) for the *upper nodes* (viz: PHN > PPH > PNP/PSP) and to that drawn by Reid (included herein as Tree 2) for the *lower nodes* (e.g. NPh includes [Iv+SLz+NMg] + [Ncr+ICS]; SPh would include Bl+[Mb+Dn+Sb][CPh+SMg+Pl+Kl], Moŋ, Gor, Sn+Mn).

Because Reid's hypothesis puts Bilic and the NPh languages closer to Formosan than to Malayo-Polynesian languages, I include Table 3 to show the affinity of *all* Ph languages to Western Austronesian (PHN), i.e. the subgroup of next highest order.

6. FUTURE DIRECTIONS

This paper represents the first edition of a study intended to bring together evidence (published and unpublished) for each Philippine micro-subgroup. The macro-subgroup (PPH or Proto-Eastern Hesperonesian) is dealt with here.