Extra-Austronesian Evidence for Formosan Etyma[1]

Paul K. Benedict

The Formosan languages have long been known to show unusual phonological features, marking them off from the Malayo-Polynesian, and to present many distinctive roots. The present paper presents evidence that both the features and many of the roots have connections with those of other Austro-Tai languages, Japanese as well as Kadai and Miao-Yao. Additionally, many of the root connections extend also to the Philippine languages, pointing to the ancient migration routes taken by the ancestral Austronesians, coming from the mainland to Taiwan and (probably in part via Taiwan) to the Philippines, then on down to Borneo, Java and eventually out into the Pacific.

The extra-AN evidence supplied by Japanese as well as the mainland Austro-Tai families offers much that ties in with the Formosan languages. The evidence relating to the special Formosan phonological features has

---

1 The evidence presented in this paper, from Japanes as well as the mainland (Kadai, Miao-Yao), attempts to bring up to date the earlier findings set forth in *ATLC* and *JAT*. For both latter families, in fact, there has been an unprecedented flow of linguistic material, the bulk of it in Chinese, at times on newly uncovered languages; on the comparative side, monumental studies have appeared on Miao (Wang Fushi, in Chinese) and Tai (Li Fangkuei) while two of the main groupings within Kadai: Kam-Sui (South-Central China) and Hlai (formerly Li, on Hainan) have received some much-needed attention, hence we are now in a far better position to work out reasonable reconstructions at both the P-KD and P-MY levels.

The Appendix lists roots with special extra-AN/Philippine as well as extra-AN/ Formosan connections, the former on the basis of citations/discussions by Blust, Dyen, Li and Wolff in the Symposium papers.

already been presented in considerable detail, with new data continuing to come in from the recently recorded (mostly in Chinese) languages of the Kadai group; see the (recently updated) Table I of all the KD reflexes for the AN/KD *p + r/l/ḷ/ clusters. The parallel clusters with *m present greater problems but here also new findings have occurred, notably for SIX, with Jp. and KD (Laha/Gelao) both supporting PAT-level *$^?$umləm, with the *l now specified by Formosan evidence: Makatau (one dial.) ulum, Sir. (one dial.) tuluman < *t-ulum-an; Bunun: Ishbukun $^?$abnum < *$^?$əmləm (with reg. *l > /n/), rightly called 'a problem' by Paul Li. This evidence for setting up *ml at an early (Formosan) level in fact strongly indicates that *pl (in DIE), perhaps also *pr (in EYE), may well have existed along with *pḷ, which yielded Paiwanic *C_1 along with Atayalic *l (see App.:ENCLOSURE; also BEAR for velar cluster). There is also strong extra-AN support for the recognition of widespread Formosan nasal increment (NI) shifts of Kadai type: *mp > /b/, *mb > /m/, etc.; see especially ANT-HILL/, with Pai. b < *mp ~ p < P-Pai. *(m)p; also 'chest/heart' (below); also App.: BEAR; BOTTOM/FOOT; MOUNTAIN: NOTE; VISCERA; also HOLE (Pai. b < *mb for reg. v < *b); also BANANA (in ATLC): Formosan: all of *bulibuli > *bilibili type[2] apart from Fav. bilpil, pointing to an earlier *(m)puli(m)puli, nicely confirmed by PMP *punti < *pulpli (reg. *pl > /t/ as well as *l > /n/) < *pulipuli as well as by P-Tai *(m)pliᴬ 'banana flower/bud'; note also within AN Pai. pi-naqup ~ mi-naqup 'wash face'; N. Phil. *mi-(n)da$^?$up 'id.'. Even Japanese, which has reduced the PAT consonant inventory by over half, serves as the key witness in two etyma containing the rare *c (> Jp. /t/):

---

2   BANANA shows widespread DS and A changes but note Kan tabunəbunə, Kvl bə ni:nã$^?$ (both /n/ < *l); also Sir bulbil agong with the cited Fav bilpil; *u confirmed by MP, *i by both MP and Tai.

SEA: P-Pai. (Sai./Pai./pazeh/Fav.) *(qa-)wacal; P-PN *wasa; Jp. wata; also RECITE: WMP *ʔucap 'speak' ~ (Ml.) 'recite' ~ (Jav.) 'enumerate [=recite] good deeds'; Jp. uta '[recital:] song, poem'; uta-i, Old Jp. utaF-i 'sing, chant, recite'.[3]

Phonological evidence also on occasion involves only given etyma, e.g. the final *-i of BIRD (see App.) finds Jp. support. Sir. ma-kuliaŋ provides the key clue to reconstruction in the widespread root for YELLOW: *kuli(ñ)jaŋ (contra ATLC) while Thao pitaw has a similar role in another root with the same wide distribution: DOOR: *pi(n)təw, with the *-əw (rather than *-u) rime confirmed within AN by Moklen (writer's recent fieldwork)as well as outside AN by both Jp. to and P-KD (Tai/KS/Hlai/Laha) *təwᴬ. In the case of the root for TWO, the subject of a vast Dyen/Blust et al. literature, the extra-AN (KD, MY, perhaps also Jp.) evidence requires a trisyllabic *drawasa (with SYL-I and SYL-III reflected in KD, SYL-II in MY), suggesting PAN *ḍəwasa > *ḍəwsa (Dyen) and readily explaining Tag. dalawa < *ḍa-ḍawa [sa] as well as Sed. daha < *dwatsa (/h/ < *ts) < *d[w]asa (A ts/s).

The mainland languages and, to a lesser degree, Japanese all exhibit extensive syllabic reduction, with frequent loss or disguise of affixed elements, hence they provide little help with morphology; note, however, *qa- in BEE/SWEET and *-l- in FRUIT/SEED; also *ʔ (-um-) ari 'come' (Blust), nicely supported by the KD evidence (Benedict 1991a). The

---

3   P-Pai *c shows merging with *C in Sai /s/: wasal 'sea' and and Pai /ts/: vatsal 'deep pool' ~ (W.) 'lake' as well as Paz /s/: ʔawas < *[q]a-was[al] 'sea'; cognate lacking in Ami, which has dis. /ts/ for *c vs. /t/ for *C; in BUY/SELL, Blust reconstructs *saliw (= *tsaliw) but for F cites Thao θa:riw 'buy, sell' and Ata saliu 'trade' (both < *C-) along with Ami caliw (= *tsaliw) 'borrow, lend', pointing to PAN *caliw (see App.: Add-3).

mainland evidence indicates that in another root an infixed *-l- has remained disguised in AN; cf. *tsapaw 'field hut' (Pai. dial. 'dwelling house') ~ *ləpaw 'hut' (Kvl./Samihim 'house'), both cited by Blust in AN Etym.-II, the latter from *tsa-l-apaw, with two mainland cognates exhibiting different infixation patterns: P-Tai *law^B; 'animal/grain enclosure', from *[tsap-] l-aw; P-MY *praw^B 'house', from *[tsa] p-r-aw (note the three different infixes!). In one interesting 'culture' root set up in ATLC: *qulaw 'liquor' the underlying analysis long remained unsuspected until help came from AN: *q-ʔulaw 'sthg. intoxicating'; cf. N. Phil. *ʔulaw 'intoxicated' and the PAN root (incl. Pai.) *ʔulaw 'dizzy, giddy, dazed' (Blust in AN Etym.-II, which omits Ceb. ʔu:law 'ashamed'); for the nominalizing role of *q(a)- here, cf. (App.) under COVER/ *ʔumuk 'cover', *[q]a-ʔumuk 'hat'[4]; note Pai qa- 'nominalizing prefix, no longer productive' (Ferell 1982), established by LIQUOR root at P-Austro-Kadai level. Finally, evidence for one key kin term prefix is provided by Jp., with possible help from KD; cf. Ata. k- 'prefix for deceased persons', maintained in the following root: PAT *-amu: P-Pai. (Puy./Rukai) *t-amu(-an) 'grandparent/ancestor'~ 'grand-child' P-Tai (SW) *hmu-a(n)^A < *ʔmu-a(n), from *[k-] amu-a(n) 'grandparent (mat.)' ~ 'ancestor of 4th gen.'; Jp. kami ~ (comp.) kamu- '[the ancestors:] the gods'; Old Jp. kami < *k-amu-i, with the *ʔ-i suffix found especially with older-than-Ego terms along with the *k- prefix for deceased persons. (Jp. scholars like Yanagita Kunio have long derived kami-worship [Shinto] from the veneration of ancestors). It is unclear whether the kin term k- prefix occasionally found elsewhere with older-than-Ego items, e.g. Fijian: Nadrau k-amu 'father', an app. cognate of the above root: Pai. k-ama 'father', k-ina 'mother', is of the

---

4   Cf. also PMP *Diɣi 'stand' ~ (NPH) *qa-Diɣi 'pillar'.

same origin as Ata. k- but in any event this Jp. witness convincingly establishes this feature at an early Austro-Jp. level.

As for semantics, the mainland and/or Jp. evidence is occasionally of value in establishing the more likely 'original' meaning of a given root with conflicting glosses in AN and Formosan; of. the following:

(1) PMP *$^?$anay 'termite'; Pazeh $^?$alái 'ant'; Jp. ari 'id.', confirming the medial *-l- and indicating the rime as *-əy > Jp. -i rather than *-ay > Jp. -e, Old Jp. -ë.

(2) WMP *tipits 'thin' (~ PMP *nipits); Sed. tipix 'small' (reg. -x < *-ts); Jp. tiisa = chiisa, Old Jp. tiFisa < *tiFis-a 'id.'; this suggests that the proto-gloss should be set up as 'small' rather than 'thin' yet the WMP meaning is nearer the *pis = *pits 'thin, tenuous, fine' set up as a 'root' by Blust, who failed to include /tipis/ in his listing for *pis.

(3) PAN *balu- < *mpalu-: WMP (N. Phil.) *balu-kuŋ 'chest'; Ami falo-haŋ 'id.', falo-cok 'heart'; P-MY *plou^B < *plu^B (reg. shift) 'heart'; this points to PAT *(m)palu 'heart' rather than ~ chest', with the latter to be analyzed as 'heart-place' or the like.[5]

On the lexical side, both the mainland languages as well as Japanese/ Ryukyuan at times reveal links with Formosan-only forms/roots, providing a basis for reconstructions at the earlier PAT, PAK or PAJ levels; see Appendix. Significantly enough, there are additional linkages with Formosan/ Philippine and Philippine-only roots as well; see Appendix. There is a contrasting dearth of linkages with 'other MP', indicating that there was

---

5   KS has three possible cognates here, all for 'heart': Laha lul[6] (h.t.), perhaps from *[q-]lu-l[u] < *[pa]lu-l[u]; P-Hlai *hlaaw^B, a possible VT derivative: *[p]alu > *law^B, subsequent *qa- prefixation: *qa-law^B > *hlaaw^B (thru VT); Gl forms of /hlu/ ~ /hlau/type, app. cognates of the Hlai series; there are parallels for the indicated /l/ for *l after *q(a)-.

indeed an attenuation of lexical material as the early AN movements continued south and east of the Philippines. This furnishes further, crucial support for the thesis of an AN homeland, if not on the mainland, at least at the northern, Formosan/Philippine pole of the widespread AT distribution.

A final point concerns culture. All the evidence to date indicates that the ancestral AT peoples created the 'high culture' of the Far East, with the cultivation of rice and millet a key component. One of two RICE roots - one can debate their original referents - apparently has AN representation only in Formosan and, similarly, the roots for WEED and PLOW, strongly suggesting that the main early cultural push off the mainland was to Taiwan, with later movements to the Philippines and on southward. the latter perhaps bypassing Taiwan, at least in part.

APPENDIX

ANT-NEST/TERMITE *ta(m)pulak Pai ꞇabulək ~ (Tju) ꞇapulək 'large arboreal ant-nest' from *ta(m)pulak (DS ə/a in SYL-III).

P-KD (Tai/Be/KS/Hlai) *pluak 'termite' (thru VT).

BAMBOO[1] *batakan P-F/PH id. (Dyen: B-39).

Jp take, Old Jp takë < *takai < *takan (reg. shifts; also reg. CRL in trisyllabic root).

BAMBOO[2] *kalabu P-Pai *kalabu-labu: Kan kanabunabu 'large bamboo'; P-Ruk *balo-balo < *balu-balu (M b-l/l-b); P-Li (1977) reconstructs *baləbalə, citing Man valəvalə (also in Tsu. n.d.) but the earlier O/A (1935) cites Man

valovalo, pointing to P-Ruk *balo-balo.

P-MY *hlau[B] < *[k-]la[b]u. *[k-]la[b]u; for *-b- > [zero] here, cf. NOTE no MOUNTAIN.

BARK, n.   *pulak P-Tsouic (Saa) *puɬakə.

P-Tai *pɬiak (thur VT after DS ə/u) ~ *plaak (A a/u, then VT).

BEAR, n.   *kḷu(m)bay Sed kŭmay ~ sŭmay (Ata lacks cognate); P-Tsouic *Cumayi; P-Pai *Cumay, all with *mb > /m/.

Jp kuma < *k[r]uma-i (typical CRR, with reg. *mb > /m/).

P-KD (Tai/Be/KS/Lq) *k[r]umay[A], with var. VT forms, with or without DS; Lq maintains *k- < *k[r]- before /m/; reg. *mb > /m/.

P-MY *kḷop < *kḷub[ay] (with reg. CRR; /o/ < *u).

NOTE Tsu. cites only the Tsouic and Paiwanic forms, without so much as a fn., by this omission treating the Sed forms as unrelated. Even more remarkably, Dyen (A-120) cites only one of the two dialectical forms: Hogo sŭmay, shunning the other: Iboho kŭmay! This linguistic excommunication yielded a PAN-level *Cumay rather than the indicated *kḷumay or the like, as firmly supported by the comparative evidence.

BEE/SWEET *(qa-)walu P-Ata/Taouic *walu 'honey-bee'; P-Pai *(qa-)walu 'honey-bee' ~ 'honey' ~ 'sugar' ~ 'sugarcane': Pai: Tachaban (O/A) alu 'sugar' ~ qalu 'sugar-came'.

P-KD *qwaal[A] < *qa-wal[u] (thru VT) 'bee' (Hlai) ~ 'sweet' (Tai/KS).

NOTE P-Hlai *kua[l][A] 'bee': dis. WS Li reflex lacking; final -uai form for anticipated *-uaɬ (Hlai lacks this final); possible early loans in this /kuai/ form to Viet-Muong: *k'way 'bee' and/or to Tibeto-Burman: P-TB *kwa:y

'id.' (*STC*).

BELLY *ba[r/R] aŋ P-Ruk *baraŋə.

Jp hara, with reg. *b > /h/, *-ŋ > ø.

NOTE *baraŋ is indicated here in view of *baRaŋ 'rib', q. v.

BIRD *tari P-F *tari-, as compounded in *tari-kuk 'fowl', retained in earlier recordings of Kan: tariku:ka (O/A - 1935) and tarikūka (M. Yen - 1962-63, cited by F.) but later exhibiting the anticipated A u/i: tarukūuka (Tsu. - 1968-69); both Tsou and Saa show this feature (Tsu. cites P-Tsouic as *taru-) while Paiwanic forms show either A u/i or A a/i (note Sai ta-taḷa), the latter also with DS ə/a (Li cites as *təra-).

Jp tori, Old Jp töri, from *təri (parallel DS ə/a).
NOTE For the *-kuk, cf. WMP *kukuk 'cackle'; P-Tai *kuk 'id.'. The Tsouic forms, reflecting a suffixed *-a, have a parallel in RABBIT, q. v. and an additional parallel in Tro manukka 'bird' (Bas manuk), indicating that these Ketagalan forms (Li cit.) do not represent borrowing, as generally supposed (see the discussion in Li), but rather are to be viewed as cognates of PMP *manuk, from a PAN-level *manuk or perhaps *mamruk or the like (cf. the discussion of 'six' in text), the cluster analysis indicated by P-KD *(ʔ)mrok (Lq mlok) and P-MY *(ʔ)m[rl]ɔʔ as well as by Jp -me < *-mai (see JAT).

BIRD (OF PREY) *taka P-Pai *taka-, as represented by Pai takaŋa 'eagle'; Sai takako 'falcon'.

Jp taka 'hawk, falcon'.

NOTE Both Paiwanic forms remain unanalyzed; the alternative

reconstructions: *takaŋa/takaku are less likely in view of the fact that Jp handles trisyllabics of this kind thru CRL, so that roots of this shape would have yielded Jp /kana/ or /kaku/.

BOARD/BEAM *balihiR P-F/WMP id. 'board' (Tsu. *baNiR;-Dyen: B-19 *baH₁NiR/paH₁NiR) (Tag pa:nig 'panel, as side of a board').

Jp hari 'beam, girder' (*b/p- > /h/; *h > ø).

NOTE This root presents severe recon. problems in AN, quite apart from the apparent Tag doublet: Tsu. analyzes P-Ruk *ba²ali as from *b-aɤ-aliɤ; also, an -h- is present in Ami vahliḷ, Bun banhil and (not cited) Sai balihḷah, the last with an infixed *-ɤa- (/l/ < *ɤ), with final -h app. for *-R (see NOTE on COVER/); finally, both Bun and Sai (vs. Ami) support *balihiR.

BOAT *baŋkaq PAN *(qa-)baŋkaq: PMP *baŋka²; Kvl, Bas baŋka, Tro vaŋka 'canoe'; also Blust (fn. 5) P-F-/PH/Moken *qabaŋ (typical CRR after prefixation); also Fav abak < *[q]a-baŋ[kaq] (A k/ŋ)

P-KD (Tai/Laha) **²baaŋ^A < *qa-baŋ (thru VT; /²/ < *q-) 'clf. for boat, canoe, raft', et al.

NOTE Blust (ibid.) comments on the recent addition of Mok kabaŋ to the AN set (he reconstructs only *qabaŋ): ... "thus extending the evidence for a continuous cultural transmission of at least one type of boat in the Austronesian world (including Taiwan) for a period of several millenia"; the Fav cognate requires an early dating also for the *qa-baŋkaq form (to explain the A) while the KD evidence pushes the dating even further back (hardly suprising when dealing with the world's greatest navigators).

BOTTOM/FOOT *(n)til(til) P-Pai id.: Pai dil < *ntil 'base, foundation; [bottom:] buttocks' (F.; Ho glosses 'anus'); Sir tiltil 'foot'.

P-KD (Tai/KS/Laha) *tiin^A 'foot', (Shan also 'bottom, foundation'; Laha in comp. 'mts. for grazing' = 'foothills'), from *tiltil (thru VT; *-n < *-1).

BREAST *tśi(tśi) Bun tsitsi (Davidson 1903, cited by F.).

Jp titi [tśitśi].

P-KD (KS/Hlai/Gl) *tśi(tśi).

NOTE For the initial *tś- see the analysis in JAT: 88-92.

Davidson also recorded the doublet: Bun tsutsu 'breast', representing PAN *tśutśu; both *tśitśi and tśutśu appear to have been derived from a proto-level *tśitśu, as represented by Kvl sisu, Ketagalan: Bas/Tro tsjitsu.

BUBBLE/FOAM *(m)puq(m)puq Bun puqpuq 'foam, bubble'.

P-Tai (N-Tai/KS) *(k-)buk 'foam/froth', from *[bu]k-buk < *[mpu]q-mpuuq (reg. /b/ < *mp, -k < *-q).

BUBBLE/FOAM/FLOAT *(m)bu(m)buŋ Pai id.: bubuŋ < *mbumbuŋ 'air bubble', qa-vuvuŋ < *-bubuŋ 'viscera' = 'sthg. bubbly, esp. lungs, intestines'.

P-KD *^?(m)bu^A 'foam/froth (Laha); float (KS/Lq)> P-Tai *^?vu^A: Saek vu^1 < *^?vu^A (see Benedict 1991b), from *q(a-)bu; P-KS also *(^?)(m)buŋ^A 'float'.

BUY/SELL *caḷiw P-F/PH *caliw (IV-542 *saliw; Wolff *ʃaliw); for the initial *c- here, see Note 3.

P-Hlai *ḷiu^C 'sell'.

NOTE Cf. P-Hlai *ḷoi < *ḷəḷi under CICADA; also *ḷwa^A < *[mb]uḷa

under PLANT; also *lou < *lu 'eight' < *walu.

CARRY ON BACK/SHOULDER *tsabik P-F id.: Sai h-in-abik 'carry athwart the back'.

P-KD (Tai/Hlai/Laha) *(s-)biak < *(ts-)abik (thru VT; /s/ < *ts): P-Hlai *ʔbiak 'carry on shoulder'; Laha bik (h.t.) 'id.'; P-Tai *ʔbeek 'id.' (typical leveling): Nung beek (h.t.) 'carry on shoulder sthg. on pole', also peek (h.t.) 'carry on back' the latter form reflecting A p/b after *s- rather than the regular /ʔ/ < *s-, also maintaining the apparently earlier meaning: 'carry on back'.

CHEEK *(m)pi(N)Gi PAN id. Kvl piŋi < *piNGi; Sai piʔiʔ < *piGi-G[i]; (part. redup.); Bun: Ish piʔiŋ < *piGi-NG[i]; WMP *pipi < *pi[G]i-pi[G]i (redup., with *G > ø).

Jp hige, Old Jp Figë 'cheek (Fi-) hair (-gë)' (see under HAIR), from *pi[G]i (*G > ø).

P-KD *ʔ(m)piŋᶜ 'cheek (Laha), buttock (Hlai)', from *q(a-)(m)piNg[i] (cf. Kvl).

NOTE Dyen: 2-17 *piŋi, citing only Kvl piŋi, SubS si-piŋi, rejected by Blust on basis that latter appears to be a cognate of Tag pisŋi et al., reflecting pis[ə]ŋi [app. from a metathesized *sə-]; other PH forms show a prefixed *qa-, as in KD: Ceb, Pan ʔapiŋ, while Mnb piŋpiŋ is the NI parallel of WMP *pipi.

The reconstruction of PAN (< PAT) *G is provisional (cf. the discussion in JAT: 53-54), based here on the /ʔ/ ~ /ŋ/ variation; note, moreover, that Bun:Ish /ʔ/ reflects P-Bun *h (F. Li 1988), the indicated *G > /h/ shift paralleling Ish *q > /X/.

CHILD *(ʔu-)(N)ʔalak PAN id.: P-F/WMP *(ʔu-)alak: Tsou óko < *ohoko (/h/ < *1; D k/ʔ after *h); Sai ʔalʔalak 'young person': also P-F *wawa (Dyen: A-129 *w₃aw₃a[) 'child (Ami), baby (Thao), young of animals (Sed)', from *ʔu-a-ʔu-a[lak] (part. redup., with CRR); also PPH (Zorc) *ʔanak < *ʔalak 'child (offspring)' (Isn ʔanʔanaʔ), *ʔanak ~ *ʔuŋa 'child (young)' (Zorc *ʔŭŋaʔ), from *ʔuNʔa[lak with CRR after prefixation.

Jp wara- < *ʔu-ala[k].

P-KD *(ʔu-)alak (see Benedict 1988 for details).

P-MY *h(w)aᴮ/ᶜ < *(u)ŋa (var. *ʔu- prefixation → VT), the NI form of the root (cf. PPH) with reg. CRR.

NOTE For the *ʔu- marker with kin terms, see also GRANDCHILD. Alternatively, a distinct PH/MY root can be set up; there is some evidence, however, from the Miao data already on hand, that the MY forms reflect *ŋ(w)a[la]k.

CICADA *laḷi PAN *lali (IV-414 *Nali); P-F *la-lali (Li:37 *NaNaLi).

P-KD *laiᴬ (Be), thru simple VT; also *ḷoiᴬ (Hlai), thru VT following DS ə/a (/o/ < *ə).

CLOSE EYES/SLEEP *qiləp PAN id. 'sleep' (Li 1983: 11); also Sai ʔiləb < *ʔiləpʔiləp (A b/p) 'close'.

P-KD (Tai/Be/KS/Hlai) *qi(-m-)l[oa]p: P-Tai *hlap 'close the eyes; (+ SLEEP) sleep'; Be lop ~ lap (h.t.) '(usu. + SLEEP) sleep'; PKS (except Kam) *qhlap 'close the eyes', Kam nap < *mlap < *[qi-]m-lap 'id.'; Hlai: SD ñïap 'id.', from *mlïap < *[q]i-m-lap (thru VT).

NOTE For the vocalism of KD, see NOTE on EXTINGUISH.

Sai appears to have maintained, thru part. redup., an earlier, generalized

meaning of this root: 'close'; if this be the case, however, one might wonder about the striking parallelism in development shown by AN and KD.

Unlike COME/GO, showing complete incorporation of the *-m- actor focus-marker, the incorporation in this root is only partial, represented both in KS (Kam) and Hlai (SD).

CLUSTER/SWARM *qupuŋ P-F/WMP id. (I-438; Wulff) 'bunch, cluster' (Pai 'swarm of bees').

P-Tai *fuuŋ^A 'swarm, herd, flock, crowd', from *qupuŋ (reg. /f-/ < *q-p-, with VT).

COME/GO *?(-um-) ari PAN ?ari (IV-17 ari; Dyen: B-13 *]ari[).

PKD (Tai/Be/KS/Lq/Hlai/Laha/Pub/Cl) *(?)ma[rl]^A 'come' ~ 'return (come back ~ go back)' ~ (Laha) 'go' (app.'go back'), from *(?)-m-a[rl] (for details, see Benedict 1991a).

NOTE Blust, who gives the basic gloss as 'come; let's go', notes that PMP-level *maɣi (Blust describes some *-r- ~ *-ɣ- variation) 'towards the speaker' (Ml 'come here'!) includes an incorporated *m- (cf. NOTE on WEED); note that the (non-phonemic) initial *?- is variably reflected in KD.

Note the complete incorporation of *-m- ~ *m- in this root as contrasted with the partial incorporation in CLOSE EYES/.

COVER/CLOUD *ɣəhəm P-Pai *ɣəhəm(ɣəhəm) 'cloud': Bun luhum; Sai: Taai ləmləm; P-Ruk *əmə:mə.

P-Tai *hom^B 'cover (esp. with cloth or other covering)'; Saek hom (A: l.t., poss. reflecting voiced *-h- as medial) '[to cloud up:] steam, smoke'; Laha hom^B 'cover with blanket'.

P-MY *^{2}ɔm 'cloud', with reg. /ɔ/ < *ə before labials and /^{2}/ < medial *-h-, leading to CRL rather than CRR (see NOTE on HAIR).

COVER/(HEAD) COVERING *Rumək P-F id. 'to cover': Sed *g-um-əmuk (M ə-u/u-ə); Ata *^{2}-um-umuk (A u/ə), a men's lang. form with typical ^{2}-for g- (Li 1983: 12): Squ umuk (cover (as a lid), cover the head, wear on the head,; Sai hömək 'to cover'.

P-KD *hmu[a/ə]k, thru VT, with *R/ɣ > h-: P-Tai *hmuak 'hat, cap' (Ahom also '[to cover the body:] attire, clothe'); Laha muək (h.t.) < *hmuək 'hat'.

NOTE This root presents evidence for PAN *R vs. *ɣ, in Ata merging with /g/ as reflex but yielding Sai /h/ rather than the anticipated /l/ for *ɣ; cf. also Sai -h for final *-R in BOARD/, with Ruk ø rather than /r/ or /^{2}/ (possibly thru D ø/^{2}).

DEER[1] *(qa-)luwaŋ P-F/PH id.: P-F id. (Tsu. *Nuwaŋ; Li *(qa-)Nuwaŋ) 'female deer (Kan); female muntjac (Saa); deer, carabao (Bun/Thao); carabao (Pai/Ruk/Paz); elk, ox, horse (Sir)'; P-Ph (Zorc) *(^{2}ə)nuaŋ 'carabao', from the DS *qə- < *qa- prefixed form: *qəluaŋ, with reg. -n- < *-l-.

P-KD *ka-lu(w)aŋ (ka- is reg. DS form of *qa-; cf. P-PH *^{2}ə-): Tai/Laha *kwaaŋ^A via (*klw- lacking) *klwaaŋ^A (with VT); also (Nung) *klook < *kluak (typical leveling) < *kluaŋ (A k/ŋ) < *k-luaŋ, a proto-level doublet without *-w-.

NOTE PAN *luwaŋ rather than *luaŋ finds some comparative support in this root, with a P-KD-level (variable) *-w- having shaped the development to *kwaaŋ^A via *klwaaŋ.

DEER[2] *banan P-Pai (Pai/Fav) *bənan (DS ə/a).

P-KD)(qa-)banan^A: P-Tai *naan^B 'deer (chamois)' (thru VT); P-KS id. 'meat/flesh': Mulao '(comp.) deer'; Be nan$^{B/C}$ < *naan$^{B/C}$ 'meat/flesh'; also Hlai: WS va^B 'deer', from *q(a)-ba(nan), parallelling va^C 'shoulder', P-Tai *$^?$ba^C, from PAT *q(a)-baɤa.

DIE *dradraw P-F *[ḍ]a[ḍ]aw: Bas/Tro (ma)lalau.

P-Hlai *[tr]aaw^C < *[tr]a[tr]aw (thru VT).

NOTE For the F initial, cf. Bas lusa 'two'; Hlai reflexes also as in 'two': *traw^B < *draw[asa] (other KD *sa) but dis. (dial.) reflexes lacking here, as in F.

ENCLOSURE/VILLAGE *qaplaŋ P-Pai *qaC₁aŋ: Pai qatcaŋ 'pigpen'; Bun $^?$atsaŋ ~ $^?$asaŋ ~ $^?$ašaŋ, Sai $^?$äsaŋ 'village'; P-Tsouic *$^?$acäŋə: Kan $^?$acaŋə 'stone walls', Saa $^?$acaŋə 'walls of pigpen'; P-Ata *qalaŋ 'village'; also (Sed)*q-n-alaŋ 'fence'.

P-Hlai *fraan^A = *praan^A 'village' (thru VT; D n/ŋ after *pl- > *pr-).

P-MY *ra(a)ŋ^B 'village' (var. VT).

NOTE Li: 117 *qaCaŋ, omitting the Bun form with /ts/ ~ /s/ vs. /t/< *C while attributing the Ata /l/ to an unmotivated A l/c.

The Sed infixed derivative for 'fence' has a possible parallel in MP: NPH: Gad/Yog $^?$ala:saŋ 'fence' < *q-al-asaŋ, with /s/ for the anticipated /t/ after the infix (*q-l-apl-); cf. also Cham saŋ '[enclosure:] house'.

EXTINGUISH *$^?$ədzəp P-F/PH id. (Dyen: B-108 *]eD₂₃ep).

P-KD (Tai/Be/Hlai) *$^?$dz[oa]p: P-Tai *$^?$dap; Be zɔp (h.t.) < *$^?$zɔp < *$^?$dzɔp; P-Hlai *$^?$dzop.

NOTE P-KD *$^?$dz- dis. by Hlai reflexes: ts- ~ s- ~ r- (h.t.), yielding P-Tai *$^?$d-, Be z- (h.t.). P-KD /o/ < *ə, with some evidence now on hand from the svailable data for the reconstruction of *ə in some circumstances; the phonology remains to be worked out but before final labials there is some merging with /a/, as in this root and in CLOSE EYES/, q. v.

FAR/LONG *$^?$a(n)dzawil P-F *$^?$adzawil 'far' (Li: 47 *dawiN; Dyen: B-97 *Zaw$_{12}$il) but note Puy adawil, Ruk:Bd adáili.

P-KD *($^?$)a(n)zaw[i]l: Laha yaaw^B < *azaw[il] thru VT; /z/ < *dz) ~ yaal^B < *azal[il] (A l/w) 'long'; P-Tai *yaaw^A ~ *ñaaw^A ~ *naaw^A 'id.' (Ahom also 'far'), from *a(n)zaw[il] (var. VT); P-Hlai *$^?$naaw 'id.', from *$^?$anzaw[il] (thru VT).

NOTE Laha y- for z- in TU and BB dial's; cf. PAN *qudzal 'rain'; Laha yal^C but BB zal, TU kzen (see ATLC: 187-8).

FAT/GREASE/OIL *simaɤ P-F/PH id. (I-409 *SimaR; Wolff *simag).

P-KD (Tai/KS/Laha) *ma[rl]^A < *[si-l-]ma[rl] ~ (Buy/Laha/Pub/Lq/Gl) *mla[rl]^A < *[si]m-l-a[rl]; Laha has both mal^A 'fat/grease/ oil' and mnal^A (D n/l) 'fat/stout'.

P-MY *hmei^A, from *hmi (reg. -ei < *-i) < *s-mi < *s[i]mi[i] (A i/a; reg. -i < *-ɤ).

NOTE PAT *sim- > P-MY *hm-, as anticipated; P-KD *m- rather than *hm-, pointing to the alternative infixation: *s-l-m-.

FEAR/FEARFUL *talaw P-F/WMP (I-413 id.; Wolff id.) 'timid, fearful, cowardly'.

P-KD (Tai/KS) *$^?$laaw$^{A/B}$ 'fear', from *[t]alaw (thru VT; reg. /$^?$/ for

*[t-]).

NOTE Contrast the development with NI shown in STAR.

FISH *(m)bulaw P-Pai *bulaw: Puy:Chihpen vulau 'fish'; Pai vulau 'loach'; P-Ruk *bolaw 'type of fish'; Ami pudaw 'small river fish, minnow' (p- for *b- unexplained).

P-MY *mbraw[B] < *mblaw[B].

FLEA/LOUSE *timula P-F *qa-timula 'flea' (Tsu. *qatimula; Li *qatimuLa; I-23 *qati-mela (DS ə/u); Wolff *qatimela); also the CRR form: *timel (IV-665) 'id.' (Kel 'clothes louse'); add Ifg ti:mol di manuk 'chicken (manuk) flea'.

P-KD (Tai/Be/KS/Lk *([?])mla[rl][A] 'body louse' (Tai also 'flea/louse of fowl'), from *[ti]m[u]la-l[a] (part. redup.; dis. Hlai reflexes lacking; /[?]/ for *t-).

P-MY *[?]mula 'flea', (/[?]/ for *t-; reg. CRL in trisyllabics).

FLESH/MEAT *(N)Gayam(Gayam) PAN *qayam(qayam): P-F *qayam 'bird' (generally), also 'any omen bird' (Pai-F.); 'fowl' (Ami: Tap/Tau); 'animal' ~ 'meat/pork' (Sai); 'flesh/meat' (Kan); add Ata *si > yam 'pork'; also *qaya(m)qayam 'bird' (Pai/Sir/Ruk:Tan/Bud); 'animal' (Puy: Rik); WMP *[?]ayam 'fowl (Ml/Jav); animal (NPH: Ibg/Isg/Mlw/Ita/Gad/Yog); dog (Tag/Bik); fish on the line (Tbt)'.

P-KD (Tai/LK/Hlai/Lq/Lt/Gl) *(N)Ga(y)am(am)[B] (part. redup.) 'flesh, meat' (Tai also 'prey' ~ 'food' ~ 'bait' ~ 'pulp').

P-MY *(N)Ga(y)[A/B] 'flesh, meat' (Yao also 'deer' ~ 'pulp') (reg. CRR, with var. *-y-: *Gay-am vs. *Ga-yam).

NOTE NPH:Gad-A ʔəggaːyam 'animal' app. represents a part. redup. root, with -gg- for *-G-, a distinctive reflex maintained in this form; PAN also has *q- for G- in Pat *(N)Giluɣ 'flow': PMP *ʔiluɣ < *[q]iluɣ; P-MY (Miao) *NGl[ui]ᴮ (-i < *-ɣ), with NGl- for[lacking] *NGl̠-; prob. also N-Tai: Wum ɣiu² < *[l̠]iuᴬ (thru VT); P-Yao *l̠youᶜ < *l̠yu, perhaps a [back] loan from Ch (流 liu).

Hlai: SD/HT *ma(a)mᶜ < *[Gaya]m-am (var. VT, with tone shift); Lk momᶜ < *məm (as in Hlai, with DS ə/a); Matisoff (1988: 305) has suggested a connection with P-Tai *maamᴮ 'spleen'.

The polysemy of this etymon is remarkable, with parallel developments from a core 'flesh/meat' via 'game' to fauna of various kinds as well as to 'food' and 'bait' as well as 'pulp'; the fauna group misled Dempwolff into setting up the basic WMP gloss as 'tame/domesticated' (Zahmsein); it is probably significant that the core sense of 'flesh/meat' had been entirely lost at the PMP stage, having barely survived even at the PAN stage, this sharply marking off F from PH; as a cap on the polysemy, note the curious Pai p(aə)- denominals: caqi 'feces', pəcaqi 'defecate'; isiq 'urine', puisiq 'urinate' (incorporating the *ʔu-marker!); alak 'child', pualak 'give birth' (see CHILD for the *ʔu-); also qayam ~ qayaqayam 'bird', pəqayam 'taste', hardly what a speaker can can only guess at: 'ejecting the bird' but historically rather more understandable as 'after half-swallowing a piece of meat, bringing it back up in mouth to taste it'; these are Tju/Tja forms; infixed *-əm- appears in Pai: But/Sti, the latter in the meaning 'look at' (here that piece of meat has been removed from the mouth for a look).

FLOAT *qaludz PAN id. (Tsu. *qañu[zZ]): P-Ata *qaliuc: gen. *qaluic (A u /i) but Ataː/Squ lui? (active) ~ qliu?-an (passive), the latter preserving the

/-lɨ-/ < *ĺ.

   P-KD (Tai/Laha) *(k-)looyᴬ < *(q-)luay (typical leveling), from *(q-)aluy (thru VT) < *(q-)alus (reg. -ɨ < *-s, via *-z < *-dz), with /l/ for (lacking) *ĺ.

FLOWER *baŋal Pai vaŋal < *[b]aŋal 'fruit' ~ 'flower' (F.; six dialects in O/A).

   Jp hana (/h/ < *b, /n/ < *ŋ, ø < *-l).

   P-KD *ʔbaŋal < *q(a-)baŋal: Laha baalᴬ (h.t.) < *ʔba[ŋ]al; Hlai:JM ŋɨaᴬ, P-Gl *-ŋauᴮ from *[ba]ŋa[l].

   P-MY *biaŋᴬ, (thru reg. CRR), from *b <i> aŋ; see NOTE on RICE[1], citing parallel men's lang. <i> in RIB[1].

FLY/WING *pikpik Ami pikpik 'fly', sa-pikpik 'wing'.

   P-KD (Tai/Be/Hlai) *piik 'wing, fin', from *pikpik (thru VT).

   NOTE CF. Sika kəpik 'wing, fin', under *pik 'pat, light slap' in Blust 1988; also Ami pihpih < *piqpiq 'to fan', sa-pihpih 'fan'; fikfik < *bikbik 'shake off'; P-Ruk *sa-bikibiki 'fan'.

   Blust 1988 cites a contrasting *pak 'slap, clap', incl. *kapak 'beat the wings'; *papak 'id.' (NPH 'wing'); add Ata papak '[head-wing:] ear'; the core sense of 'beat' is maintained, however, in NgD papak 'drive in a nail' as well as on the mainland: P-KD (Tai/Be/Hlai) *pa(a)k < *pa(k)pak (var. VT) 'drive in/into, stick in/into, plant, prick (Tai: gen.); drive in with hammer blows (Tho); pound, slap (Be); stick on, paste (Hlai)'; Yao (ch. lang.) baʔ (h.t.) < *mpaʔ < *mpak 'beat'.

FROG *ʔop(ʔop) P-F *ʔupʔup (Li:48) (Puy 'bull-frog' in O/A).

P-KD (Tai/Be/KS) *$^?$op 'frog (Kam); small frog (Be); bull-frog (Shan)'; also (Tai) *$^?$oop < *$^?$op$^?$op (thru VT) 'croaking of frogs'.

FRUIT/SEED *(m)bu(-ḷ-)ay Ata:Squ buai ~ boai 'fruit'; Pai (Southern) buabuay < *mbua-mbuay 'flower'; Sai:Taai boḷay < *bu-ḷ-ay 'fruit'; Kvl mu:lay < *mbu-ḷ-ay 'id.'.

Jp mi, Old Jp mī 'fruit, nut, berry, seed', from *moi < *muai < *mbuay.

P-KD (Tai/KS) *$^?$m[rl]uay^C 'seed; clf. for seed, fruit and other spherical things', from *q(a)-mbu[rl]ay (thru VT, with reg. /m/ < *mb).

NOTE The alternative recon.: P-F *(m)buɣay (Dyen: A-14 *bur₅ay, with Kvl m/b 'by analogy'), works very well with the KD correspondence and is not vitiated by the Jp (*-ɣ- > ø) but Ata reg. *ɣ > /g/ (rarely > ø). In addition, there is solid evidence for PAT -level *-ḷ- infixation in other 'flower' ~ 'fruit' etyma, along with evidence for *-i suffixation. Tsuchida has suggested that *(m)buay is related to PAN *(m)buaq 'fruit' despite the (Tsu.) 'inexplicable' loss of *-q (but if suffix properly written *-$^?$i, then *-q + *-$^?$ > ø, with parallels in ST; see *STC*: 123), exhibiting infixation in Kadai: P-KD *$^?$b[rl]ook < *$^?$b[rl]uak (typical leveling), from *q(a)-bu-ḷ-aq; note also the Sai infixation in another root in this group: PAN *buŋah 'flower', Sai:Taai poŋlah (A p/b).

GO *$^?$usa P-F id. (Li: 19 *kuS₁a; Dyen: A-129 *uS₁a]).

P-KD (Buy/Laha/Pub/Pup/Lq/Gl) *s(w)a^B < *$^?$usa (var. VT).

GOURD *luRi Pai lui < *lu[ɣR]i.

Jp uri 'melon' (*l- > ø; *-R- > -r-).

NOTE PAT *ɣ and *R largely merged in AN but note evidence for *R in COVER/; medial *-R- > Jp -r- vs. *-ɣ- > Jp -y-.

GOURD/DIPPER *(m)buas Pai vuas 'gourd (while still on vine)'.

P-Tai *ʔ(m)buayᴬ 'dipper, ladle', from *q(a)-(m)buas (-y = -i < *-s).

GRANDCHILD *lah(lah) P-F *ʔu-läh: Sai oläh; P-Ata *ula < qi 'child', with incorporated *ʔu- marker.

P-KD (Tai/Be/KS/Lq/Hlai/Pub/Gl) *qlaalᴬ < *qa-lal[a] (thru VT, with part. redup. (details in forthcom. work on Saek/P-KD *-l).

NOTEFor the *ʔu- marker with kin terms, see also CHILD; this marker is also reflected in a third younger-generation term: Jp uma 'grandchild', from a basic PAT-level *-ma 'father/child' (signature AT self-reciprocity), which yielded PAN *ʔama 'father' (see *JAT*: 145 for the *ʔa-) and referential *t-ama 'id.' (Jp tama '[deceased father:] ghost/spirit') but in PN the general term for 'child': Maori tama 'child/term of address to man'.

HAIR *buhəkas PAN *buhək[aə]s (var. A ə/a) 'hair (head)' (Isi also 'pubic hair' and 'feather'; Paz also 'body hair' and 'feather/ down').

Jp ke, Old Jp kë 'body hair, feather' (reg. -ë < *-ai < *-as); also -gë (A g/k) in hige (see CHEEK); also -ga < *-ga-i (typical CRR) in siraga 'white/ gray (sira-) hair (-ga)'.

PAN shows the anticipated /ə/ in SYL-III, here with the help of A ə/a, but /a/ app. retained in at least some forms of Kvl, with Moriguchi recording bu:qas 'hair' (-əs recorded by F., also by Tsu: p. c., 1/93).

Jp kami 'head hair' perhaps includes a similarly reduced ka- but the analysis of this form remains undetermined.

Blust (Table 2) P-F/PMP *bukəS/buSək, with M k/S, as favored also by other ANists (cf. WINNOW); the present analysis explains both the Jp cognate (< *-kas, not < *-kəs) and the early recorded (F-K Li) form for Thao: hūkiš, later and presently: fukiš; it seems preferable to attribute this variation to the known dialectical divisions in Thao (Paul Li: 1983a) than to a 'misrecording' (by F-K Li!), with fukiš < *bukəs < *bu[h]ukəs (A u/ə), as in F generally, vs. hukiš < *[bu]hukəs (A u/ə); contrast PMP < *buhək[əs].

Thao also has maintained PAT/PAN *-h- in *buhat 'work/cultivate field': PMP id. 'produce, perform' (Ml also 'cultivate'; Kad 'till the soil': NPH: Sam '[dry] field'; PO *puat-a 'harvest'); Thao bū:hat 'field (wet, rice)', mu-bū:hat 'work' (cf. Sir mu-uma 'work' < *qumah 'field'); Pai vavua ~ vavuavuan ~ ka-vavuan ~ ka-vuavuan < *(ka-)buabuan 'field (dry)' (A ø ~ -n/-t) (cf. Sai ʔömʔömäh, Ruk:Bd omãoma 'field (dry)'); P-Tai *het 'do, make, work', *het naᴬ 'cultivate ricefields (*naᴬ)? (P-KD *naᴬ 'ricefield') (CRL, with reg. -et < *-at); P-MY *ʔaiᶜ 'do, work, cultivate (ricefields)', with reg. /ʔ/ < *-h-, with CRL (cf. COVER/CLOUD) and reg. -ai < *-at; note that MP, unlike F, has retained the basic 'work/cultivate fields' sense of this root, with 'field' a secondary development in each, the source of the Thao denominal: mu-bū:hat; DAC app. retained the 'cultivate' sense, along with a specialized deverbal; cf. ArCh 拔 b'wât (< *bu[h]at: /h/ > ø 'to plow, furrow' (Guoyu); 伐 b'ɪwăt (reflects typical ArCh palatalization) 'earth thrown up by a plowed furrow' (Zhouli).

HAIR(BODY) *gumul P-Pai id.: Puy humul 'body hair, feather/ down'; Sai kumul 'pubic hair'.

P-KD *hmu[u]n (var. tone; not dis. for vowel length): Be munᴮ/ᶜ (h.t.) 'pubic hair'; P-Hlai *hŋunᴬ (hŋ- < [lacking] *hm-) 'body hair, fur', from

*[g]umu[l] (-n < *-l).

HAIR(BODY)/FEATHER *kupuḷ P-Pai id.: Thao kupur 'body hair, feather/ down'.

P-KD (Tai/KS/Be) (k-)pul[A] 'body hair, feather', with *(k-) indicated by Be vun[A]: *k-p- > v- is reg. shift in KD; cf. P-Tai *fu- < *qup- in CLUSTER/.

HAIR(PUBIC) *qu(m)bis P-F id. (Tsu *qubis$_{16}$; Li: 56 *qubis); also (not cited) the NI forms (*mb > /m/): Thao qu:mis, Kvl ʔu:mis.

P-KD *(ʔ)umi[A] > *mui[A] (thru VT) 'body hair (Buy); [eye]brow (Be)' ~ *ʔmoi[A] (DS ə/u, then reg. *ə > /o/, followed by VT) 'pubic/auxillary hair (Tai: Shan also 'beard'), [eye] brow (KS: Maonan', from *(q)umbi[i] (/m/ < *mb; -i < *-s)

NOTE Jp kami 'hair (head)' perhaps includes a reduced -mi < *-mbi[s] (reg. shifts) but the analysis of this form remains uncertain; for ka-, see NOTE on HAIR.

Dyen: 1-22 *qubiS$_{16}$/qibuS$_{16}$, the latter on basis of Saa (MP) ihu 'hair, feather' ('and those associated with it'); this can be viewed as further support, along with the KD cognates, for an early (PAK/PAN) status for this root.

Cf. also F and KD forms for 'beard' (note this gloss in Shan); note also Ruk:Man ubusi 'pubic hair' (O/A) (A u/i): P-F *ʔ[]mumus < *[qu]mumus 'beard': Kvl mu:mus, Tsou mʔũmʔu (M m-ʔ/ʔ-m); P-KD (Tai/Buy/KS/Laha/ Pub/Pup/Lq) *(ʔ)mum[B/C] 'beard' (C-Tai also 'body hair') from *(q-)mum [u], with reg. CRL, paralleling Tsou, a part. redup. form; these forms appear to have been derived from the above root, after A u/i (as in Ruk:Man):

< *q[u]mbumbus but the specification: 'beard' requires the setting up of a distinct root at the PAK level.

HAIR/FEATHER/WING *palíd PAN *palí[dj] 'wing, feather' (Blust 1.1 *paNid/paNij) 'wing'): Ata, Sed, Bun 'wing' ~ 'feather'; Ruk 'wing', specified 'with feathers'.

Jp ha, Old Jp Fa 'feather; (comp.) wing', from *pa[l'id] (thru reg. CRR),

P-KD (N-Tai/KS/LK/Laha) *k-pa[y]it 'wing', from *q[a]-pa[lí]it (-t < [lacking] *-d).

P-MY *(qa-)pled 'hair, fur, feather', from *p[a]líd (reg. /e/ < *i), with *l for [lacking] *í.

NOTE It has recently become clear that voiced stops must be reconstructed for P-MY, here *-d yielding Miao *-n, Yao *-i.

In addition to a 'front (dento-alveolar) *l (= *t, *L, *N) and a 'back (velar) *l̠ (= *l), a palatal *í must also be reconstructed for PAN where P-F *l corresponds to PMP *ñ rather than to *n, as in this root; before *i, Ruk has a dis. reflex: -r- (P-Ruk *paridi 'wing') while before *u Ata has -li- (see FLOAT), hence *í can be reconstructed at the P-F level in both roots.

This root nicely illustrates the at times almost antipodal separateness of MY from the rest of AT; in MY it covers 'hair' of all sorts, both animal and human, as well as 'fur' and 'feather' but not 'wing'; in Jp and AN the range has been greatly diminished to 'bird hair' = 'feather' ~ 'wing' while in KD it has been further restricted to simply 'wing'; note how Jp is grouped here with AN vs. KD.

HIT (WITH STICK) *pa(N)Guḷ Pai p-ən-aŋuḷ, from *paNGuḷ.

P-Tai *ɤool[C] 'strike (esp. with stick for beating/hammering)' ~ *ɤool[B] 'stick (for beating), club, hammer' (C-Tai only < tone *B and as verb), from *ɤual (typical leveling) < *[p]aGuḷ (thru VT).

NOTE CF. also PAN *pu(ŋ)ku[ḷ/l] (app. A u/a): WMP *pu(ŋ)kul 'throw, beat' (Ml 'strike, hit, beat, knock'); P-F *puku[l]: Kvl puqun 'hit (with fist)'; Bas pukun 'hit; hamer'; app. represented also in DAC; cf. Ch 棍 kuən (Mand gūn) 'stick, rod' (not attested early).

HOLE/CAVE *(m)boloŋ P-Pai *(m)buḷuŋ: Pai (buḷu)buḷuŋ ~ bəluŋ ~ bəruŋ 'hole ~ (Mkz) bəloŋ 'cave'; Tao bulun 'hole'; Sir varyng 'den'.

P-Tai *brooŋ[C] 'empty, hollow; hole, cave', from *boḷoŋ (thru VT; /r/ < *ḷ).

HORN *waqa P-Pai *waqa 'horn (Ami/Bun/Thao), deer (Sai)'; P-Ata *waqa < nux 'deer'.

P-KD *waqaw[A] < *waqa-w[aqa]: gen. (Tai/Be/LK/Hlai/Laha/Pub/Lati/ Gl) *qaw[A] (reg. CRL) but P-KS *qwaaw[A] (M q-w/w/q, with VT).

HOUSE *[dẑd]aɤan P-Ruk *da²an < PAN-level *[dẑd]aɤan.

Jp. ya, from *ya-i < *yan (/y/ < *ɤ; typical CRR after -i < *-n).

P-KD *-aran[A] > *raan[A] (N-Tai/KS: simple VT) ~ *rɨan[A] (SW/C-Tai) (DS ə/a, than VT ɨa) ~ *rɨɨn[A] (Hlai:HT) (typical leveling) ~ lon[A] (Laha) < *ron[A] < *rəən[A] (leveling, with reg. /o/ < *ə) ~ lan[A] < *ran[A]/zɨan[A] < *rɨan[A] (Be: dial. var.).

HUNDRED *ɤi(m)bəw WMP *ɤibu '1,000'; P-F *ɤatsibu '100': Kvl ɤasibu, Ket latsibu, Bas latsebo (DS e/i), from 'one (tsa-) hundred (-ɤibu)', with complex M ts/ɤ and a/i.

Old Jp -bo (< *-bəw) in '500' and '800'; Jp momo < *mbəwmbəw.

NOTE The F analysis follows that of Dahl (1976: 132), who accepts Wolff's *ɤ- (over *r-) for WMP *ɤibu. Dahl, however, under the (common) misapprehension that he was dealing with "true" primitives, added the following comment:

It is quite possible that a primitive society of Proto-Austronesians had no need for so high a numeral as "thousand", and that the original meaning of the work was 'an extraordinarily high indefinite number'. From this vague meaning it has developed into "hundred" in Kv. and in the higher cultures of Jv. and Ml. into "thousand".

It now seems, rather, that these early "primitives", having created rice cultivation and the like, knew how to count, after all, at least up to 100, and that it was the ancestral (and entrepreneurial!) WMP's who hiked the amount to 1,000.

HUNGRY *ʔu[r/l]ay P-Ata *mu-ʔuray PAN-level *-ʔu[r]ay.

Jp ue 'hunger', from *u[l]ai (*-l̩- > ø after *u; -e < *-ai).

A possibly related Paiwanic form supports *-l̩- in this root; cf. Pai qaul̩ay 'dried up (fruit, veg's)': 'stomach dried up' = 'empty stomach' = 'hungry'.

HUSK *qə(m)pa PAN *[q]ə(m)pa: WMP *ʔə(m)pa 'husk of grain, chaff'; Ata pa < qiʔ 'rice husk' (Dyen: 3-5 cit.).

P-KD (Buy/KS/Lk/Laha/Gl) *k-baᶜ < *q-mpa 'husks, bran'.

INFANT *hubuq Bun hubuq 'baby less than one year old' (Jeng).

Jp ubu '[infanthood:] simplicity, naiveté; (comp.) infant'.

IN-LAW *(n)tuɣaŋ PAN *[Ct]uɣaŋ: P-Pai id.: Bas torang 'man with illicit [in-law tabooed] sexual relations' (P.Li 1993) ~ 'wanton (woman, man)' (Asai); also Bas setolang, Tro matolang 'adultery'; WMP *turaŋ 'in-laws (Blust).

P-KD *(n)trooŋ[A] 'in-laws (esp. thru marriage of children)', from *(n)truaŋ (typical leveling) < *(n)turaŋ (thru VT).

JAW *qabay Ata qabay.

P-Tai (Lao/Nung) *(ʔ)waay[A], from *(q-)abay (reg. intervocalic *-b- > -w-, with VT); cf. SPEAK/CALL.

JUICE/WATER *(m)bidzuq P-Ata *biyuq 'juice; fluid (sap of trees)', etc.

Jp mizu, Old Jp midu 'water, juice' (/m/ < *mb; /d/ < *dz; ø < *-q all reg.shifts).

NOTE PAN medial *-dz- > P-Ata ø (with -y- as *i-u glide); cf. P-Ata qual < ax 'rain' < PAN *qudzal (see NOTE on FAR/).

LOUSE[1] ba(m)bulay Paz babulay 'head louse; (comp.) body louse'.

P-KD (Tai/KS/Lk) *(ʔ)(m)bulay[A/B] 'louse (esp. bird/fowl)', with VT yielding /rw-/ (< *lu-) forms in Tai but with reg. /l/ reflex maintained in Lk blei[B] (-ei < *-ay); *(ʔ) for *(b-).

LOUSE[2] *jaɣatru P-F *jaraCu (Li: 73 *DaRaCu) 'body louse'.

P-KD (Tai/KS/Hlai/Laha/Gl) *-atru > *tru[A] ~ (thru VT) *traw[A] 'head louse'.

P-MY (Yao) *[tr]ou^C (dis. Miao reflexes lacking) 'body louse' (reg. -ou < *-u).

NOTE For the initial *j-, cf. the analysis in Tsu.: 158.

Both KD and MY give evidence of 'cluster attraction' in this root, with retention of SYL-III.

This cognate set indicates that PAT *tr marged with labial clusters (Table 1), yielding P-F *C (cf. NOTE on WINNOW), whereas PAT *dr > PAN *ḍ = Ḍ, as in *drawasa 'two' > PAN *ḍawasa > *ḍ[w]atsa (A ts/s) > Sed daha$^?$ (/h/ < *ts) ~ (part. redup.) *ḍa-ḍawa[sa] > Tag dalawa ~ (DS ə/a, with CRM) > *ḍəwsa (Dyen).

MANY *[l/ḷ]iaw Pai ḷiaw.

Be liaw^{BC}.

MARROW/BRAIN *luqəs P-F id. (Dyen: A-55 *lu[q$_{14}$Q$_2$]eS$_{13}$).

P-MY hlui$^{A/C}$, from *slui < *[lu]s-lus (/u/ < *u-u; -i < *-s) *lu[qə]s-lu[qə]s, thru CRM (*q > ø), with redup.

MORNING *sanu P-F *sa-sanu (Dyen: a-97 *SeSa[/SaSa[): Ata sasa-n 'morning'; P-Ruk *səsanə (DS ə/a ~ u) 'today'; add Thao (F.) ša:ša:nu$^?$ 'morning'; Paz (Tsu.) sā:sāun-an (M u/n) 'id.'.

Jp asa, from *[q]a-sa (typical CRL after prefixation).

P-Tai *hnaɨ^A < *s[a]na-ya (A a/u; reg. /ɨ/ < *ya).

NOTE For the Tai suffixed *-ya, cf. Ruk:Tan suffixed -a in 'noon', 'today', 'tomorrow' and 'yesterday'.

Jp asu 'morrow, tomorrow' perhaps also belongs in this cognate set,

from an assimilated *[q]a-su[nu].

MORTAR *lutsuŋ PAN id.

Jp usu (reg. *l- > ø, *ts > /s/, *-ŋ > ø).

NOTE Dyen: 3-4 *D₅esuŋ but Sed duhuŋ, with unexplained d- for *l-, is a likely men's lang. form (see NOTE on RICE[1]).

F reg. reflects *lutsuŋ whereas WMP reflects *lətsuŋ, the Jp cognate indicating DS ə/u in WMP rather than A u/ə in F.

WMP also (in PH) reflects prefixed *qa- (e. g. Itn/Ilk ʔalsuŋ), yielding thru DS forms reflecting *ʔətsuŋ (Blust: f. *esuŋ), the loss of *l after *qa- paralleled in PAN *qa-liCu 'evil spirit' (Tsu.: 166); add Ata:Mas ʔali > utux 'ghost'; WMP *ʔanitu 'departed soul'; P-PN *aitu 'ghosts, spirits', all with incorporated *qa-, as shown by the cognate Old Jp itu 'divine power', Jp itu-k-i 'deify', et al. (see *JAT*); cf. also the loss of *l in PAN-*lima 'hand/arm' (Blust: f.), app. after a *qa-l- > *qa-l- shift; cf. Ton/Uve nima < *[qa]lima 'id.'.

MOUNTAIN *bu(bu) Ata: Squ buʔ ~ bəʔbuʔ 'peak'.

Jp fumoto, Old Jp Fumötö 'foot (-mötö) of a mountain (Fu-)'.

P-KD (Tai/Lq) *buᴬ.

P-MY (Miao) *b[ou]ᴮ (reg. -ou < *-u).

NOTE Cf. also Squ qbubuʔ '[peaked] hat'; Egerod compares with bubuʔ 'breast', a cognate of Kan mumu 'id.', from PAT-level *(m)bu(m)bu; cf. Jp. mune 'chest, breast', from *mbu-ne (-ne for '[sternum] ridge' (cf. yane 'roof' < ya 'house' + -ne 'ridge[-pole]'; cf. also KD: Laha mum < *mbum(bu) 'breast'.

Curiously, the Sed homophonic bubuʔ 'mother' represents the similar

PAT-level root: *(m)bəw((m)bəw) 'female (human), woman, mother'; cf. WMP *ʔib[u/əw] 'mother'. with prefixed *ʔi- for 'female (human/animal)'; Jp. imo 'younger sister (male sp.)', from *imbəw (reg. shifts) < *ʔi-mbəw; Eastern Miao *bu <P-MY-level *bou 'woman', from PAT-level *b[uəw]. The Jp. -o points to PAT-level *-əw rather than *-u and this finds support in the early loan into ST, from DAC or a related donor source, reflected in ArCh/MCh 婦 b'i̯wəg/ b'iu: (< *bəwᴮ) 'woman, wife', which in turn yielded an early loan to MY, viz. P-Yao *bwaŋᴮ 'daughter-in-law', with reg. -aŋ for -əg (Downer in *Asia Major* 18 [1973]:21]) along with the anticipated tonal agreement. MY also reflects a PAT-level *ʔabəw, with the basic *ʔa- kin term prefix (*JAT*: 145): P-Yao *au < *a[b]əw 'married woman, wife'; for *-b- > [zero] here, cf. BAMBOO[1]. The linking of 'mother' with 'younger sister' in (basically 'female') kin etyma of this genre is matched in two PST roots: *ma, usually 'mother' ~ 'female', the loan source of Yao mua/muo/mu forms for 'younger sister'; *mow, which gave rise to ArCh/MCh 母 məg/məw: 'mother' (loan to Yao muaŋᴮ 'id.' - Downer cit.) and also, thru the widespread ST kin term *-i suffixation, paralleling that found in AN (see Benedict 1990: fn.5), to ArCh/MCh 妹 mwəd/muâi- 'younger sister', from *mow-i.

The *i- 'female' marker appears also in PAN *ʔina 'mother' ~ 'female'; cf. also Kan ʔi-nuaŋ 'female deer', ta-ʔi-nuaŋ, Saa ta-i-ɬuaŋ 'female pygmy deer', contrasting with the unmarked-for-gender Ruk: Tan ʔunuaŋ 'deer' (see DEER[1] for this root). This suggests a proto (pre-AT) *ʔi- 'small/female' vs. *ʔu- 'large/male' contrast, with the latter having lost its size/gender specificity at some pre-PAT level. This contrast can be set up at the P-Austro-Tai level; cf. P-Tai *ʔiᴮ 'small', *ʔiᶜ 'female', the latter represented in WT by iᶜ 'mother', also used for older sister, cousin, followed by name of person, e.g.

i^C tu^C 'sister Tu, cousin Tu'; P-Tai (Shan) *[?]u^B 'father'.

MOUTH *ŋudźuy P-Ruk *ŋodoy < PAN-level *ŋu[dźd]uy.

    P-MY *ñjui^A, from *[ŋ]udźuy (thru VT: *u-u > /u/, with A ñ/ŋ).

NET *[?]aray P-F id. (Li: 83)

    P-KD (SW-Tai/ Be) *[?]raay (var. tone) < *[?]aray (thru VT).

OTTER *sanaq P-F id. (Dyen: A-91 *S₆anaq).

    P-KD *([?])anak < *[s-]ana[q] (-k for [lacking] *-q) *naak (Tai/Be)
(thru VT) ~ *[?]niak (Hlai) (DS ə/a, then VT -ɬa-).

PLANT, v. *(m)buḷa P-F/PH *muḷa (Dyen: 310 *muLa[?]), from *mbuḷa.

    P-KD *mbuḷa^A: P-KS *mbya^A < *mbla^A (thru CRM); P-Hlai *ḷwa^A <
*[mb]uḷa (thru VT, with CRL).

PLOW, v. *tiḷay Sai pa-tiḷay 'to plow', p-in-atiḷay 'wet field'.

    P-KD (Tai/Be/KS/Laha/Lq) *thlay^A (P-KD lacks *thḷ-).

    NOTE The app. early loan to Ch: 犁 liər ~ lịər < *lə(:)y^A, the (ultimate)
source of many KD and MY forms, points to a PAN-level *-əy rather than
*-ay rime in this root.

PULL OFF *suɤut P-F/PH id. (IV-606 *SuRut; Wolff *sū́ɤut) 'pull/ draw';
Bontok 'remove rice from panicles by pulling'.

    P-Tai *ruut < *-urut (thru VT) 'pull, detach or strip off grains from
the stalk, etc.' (F K Li gloss in *HCT*).
P-MY (Yao) *hrut 'pull, defoliate', from *srut < *suɤut (*-u-u > reg.

/u/). (app, Tailoan).

RABBIT *li(n)tok/lo(n)tok P-F *lituk/lutuk (Dyen A-45 *lituk/ *lutuk; Li: *Lituk and (A u/i) *Lutuk).

P-KD (Be) *$^?$[l]on$^{B/C}$: Be lən$^{3/5}$ ~ zɨən$^{3/5}$ (h.t.) < *$^?$rən, from *$^?$[l]on (dis. Hlai reflexes lacking) < *q(a)-lon[tok].

P-MY (Miao) *$^?$l[ɔ]^B < *$^?$lo (reg. shift < *q(a)-lo[tok].

NOTE Haudricourt (1985) reconstructed the Be /r/ here and in RICE[1], rightly calling both forms 'precious'.

Both KD and MY reflect prefixed *qa- as well as the doublet form of the root, with A o/i, to be assigned to the PAT level.

The suffixed *-a sometimes found with faunal terms (see NOTE on BIRD) occurs here in P-Tsouic *lituk-a; DAC app. also exhibited this suffix here; cf. ArCh 兔 t'o- 'rabbit', the (ultimate) source of many forms in KD and MY, incl. P-Tai *tho($^?$)^B, the /$^?$/ here and the sandhi tone (-) of ArCh pointing to a SYL-III *-a after /k/; the 'rabbit' of the Ch animal cycle is of the same origin: 卯 mlôg: < *m-lo^B (with ST *m-'animal prefix').

REED *qaɤəlu P-F. id. (Dyen: A-74 *qaRelu$^?$).

P-Tai *law^A = *lau^A (reg. *-ɤ- > ø, then VT, with CRL).

REST/RECLINE *saŋay P-F id. (Dyen: A-98 *[S$_{16}$XX$_2$]aŋay) 'to rest'.

P-Tai *hŋaay < *saŋay (thru VT) 'lie on back, recline'.

RIB[1] *tagəɤaŋ P-F/PH id. (Tsu.; Wolff id. 'chest cavity'): Sed, Paz, Sir (regang, thru M r-g/g-r) 'rib'; NPH: Gad 'chest' but others all 'rib': Isi taggəŋ < *tag[ə]ɤəŋ (A ə/a).

P-Tai (Si.) *grooŋ^A 'rib', from *[ta]gəʁəŋ (A ə/a; *ə > /o/; VT).

P-MY (Miao) *ta[g]^B (dis. Yao reflexes lacking) 'rib', with CRR in trisyllabic root because of the *-ə-: *tagəʁəŋ > *tagʁəŋ.

RIB² *baRaŋ P-Pai *baʁaŋ: Bun balaŋ, Thao faːɬan, Kvl baʁaː<i> ŋ (see NOTE on RICE1); P-Tsouic *várəŋə.

Jp abara < *[q]a-bara.

NOTE Medial *-R- indicated by Jp -r- (vs. -y- < *-ʁ-; see HOUSE).

Two other possible AN correspondences have been uncovered, both represented by isolated forms for 'rib': Puy balaba (via *bala-ba[la])'; NPH *barak: Its barak, Ibg bara?.

RICE¹ *(m)bəʁaw P-F *bəʁaw 'hulled rice': Ruk:Man və?ao; Kan u-burau (A u/ə to the *?u- marker) 'hulled rice'; also Ruk:Tan boáo, Maga bro: 'cooked rice'; also P-F/MP (gen.) *bəʁa > ts 'hulled rice'.

Jp mo-, Old Jp id., combining form (typical for Jp; see *JAT*) for 'rice', from *mbo-, with m- *mb- by far the best represented of all the NI reflexes in Jp, in no fewer than 15 roots (*JAT*:80-81), incl. BEAR, FRUIT/, HUNDRED, JUICE, (see also NOTE on MOUNTAIN), from *mboyo (reg. /y/ < *ʁ, -o < *-aw), with A o/ö (reg. /ö/ < *ə), in moti = mochi 'rice cake', Old Jp motii < motiFi, the latter for 'small [cake]', from *tipits, with reg. loss of final *-ts; cf. (above: introductory remarks) Jp tiisa, Old Jp tiFisa 'small' (/s/ < *ts before suffixed -a); also in momi 'unhulled rice', Old Jp id., with -mi of obscure origin.

P-KD *(m)b(-)raw$^{A/B}$ 'rice' 'food/meal' (cf. Ch 飯 fàn 'rice'→'meal')':Si kăbraw^B 'unhulled rice'; P-Tai *braw^A 'evening meal'; P-KS *mbraaw^A 'id.', from the part. redup. *[mbr]ambraw (thru VT); P-Hlai *vrap = *brap

'hulled rice', also from a part. redup. prototype but without NI and with different syllabic division: *brab[raw] (P-Hlai lacks *-b); Be lau⁴ < *[r]aw$^{B/C}$ '(comp.) rice': vo⁴ lau⁴ 'rice hull (vo⁴)'; also phia⁴ (i.t.) < *bra$^{B/C}$ (reg. shift) 'cooked rice/food', from *bra[braw] (init. cluster CRR); also lop/zop < *r[a]p (see NOTE) 'hulled rice', from *[b-]rap (cf. Hlai); this unusual triplet is the product of (1) plain vs. redup. and (2) cluster (< *br-) vs. dyad (< *b-r-) lines of development (Benedict 1989).

P-MY **mblaw^B 'rice plant/unhulled rice'.

NOTE Haudricourt (1985) reconstructed Be /r/ (for lop/zop) here and in RABBIT, rightly calling both forms 'precious'.

DAC perhaps also had this root: cf. ArCh 稻 d'ôg: < *daw^B 'rice plant/ paddy', from *braw^B; a shift of this kind is commonplace in KD and is not without parallel in DAC.

The PAT/PAN *ʔu- marker is represented both in *bəɣaw: Kan u-burau and in *bəɣa > ts: Saa ə-vəra θ (A ə/u); Tsouic has *ʔu- also in Saa ʔu:ŋu, Kan ʔuʔuŋu 'horn'; Kan kunutsu, Saa ʔukuɬutsu 'body/animal hair'; Kan ta-kûisi, Saa ʔukui 'goat'; Saa ʔususu 'breast' while Ruk:Man has ʔulusu 'brain' (see MARROW/BRAIN) and Ruk:Tan has ʔunuŋ (F.)/onoaŋ (O/A) 'deer' (see NOTE on MOUNTAIN); cf. also Kvl u- prefixed to numerals in counting objects.

The general PAN *bəɣa > ts, which has prevailed everywhere except for the tiny Kan/Ruk pocket in Taiwan (perhaps influenced by its almost symbolic cultural value), shows that something along the lines of the Atayalic men's language is far earlier than the 1,600 years estimated for it in Atayalic (Paul Li 1983b: 16). An earlier dating is also suggested by the widespread 'reformations' described by Wolff (ibid.: p.c. in fn. 18):

These deliberate reformations of the vocabulary are found all over the

Austronesian area for various purposes... One thing worth investigating would be secret languages - they are so common in so many aeras of Indonesia and the Philippines that I wouldn't be surprised to find that they also exist in Formosa (or at least at one time existed).

As shown by Li, these alterations in form take many shapes, including semantic distinctions, e. g. Sed:Ina bluku$^?$ (women's lang.) 'small winnowing basket' ~ blu < hiŋ (men's lang.) 'large w. b.'. As regards the present root, P-Ata *-x (< *-ts) is on record as a substitute phoneme (ibid.: 13) while final *-w is frequently subject to substitution (ibid.: 3), hence *bəɣa < ts makes very good sense men's lang.-wise. In other roots, under RIB^2 Kvl baɣa:<i>ŋ precisely matches the Ata infixed -i-, as in May luhu <i> ŋ 'mortar'; in this root (see MORTAR), Sed has d < uhuŋ; cf. WMP *laŋuy 'swim' ~ *daŋuy = *d < aŋuy 'id.' (IV-140); /$^?$/ in Ata can replace a medial consonant (ibid.: 12); in the following root it appears to have replaced an infix: PAN *tamtam 'smack the lips, taste' (II-420); add Ifg ma-tamtam 'tasty'; also P-F/PH t-al-am 'taste' (Tsu. *[tT]aLam); Blust makes note of his earlier recon. with medial *-$^?$-, based on Mar taqam, WBM ta$^?$am, app. the replacement of an infix (Blust wisely eschews explanation).

This sort of thing threatens to drive comparative linguists mad and, even sadder, to inspire a good deal of linguistic abuse; on the positive side, it might discourage recon's such as *bəɣa[w/ts]. At still earlier levels, even the ATist has to feel concerned in as much as one key root provides evidence that MY also can be involved - and one can't get any earlier than that in AT; cf. PAN *(m)paqa 'thigh', also *pa > qi, represented by WMP *pa$^?$i (cited by Demp. as db. of *pa$^?$a) as well as Thao pa:qi$^?$ 'buttocks' (note the semantic shift); here the model has been the replacement of the final syllable by -qi, paralleling Ata -qi$^?$; cf. Ata luqus/lu > qi$^?$ 'marrow' (see MARROW/);

the NI form: *mpapa yielded P-KD *(b-)qa^A (> P-KS *qwa^A); P-MY (Miao) has *paay^A = *paai^A 'id.', from *pa[$^?$]a[$^?$]i ($^?$ + $^?$ > ø, as in ST; see *STC*: 123), < *pa[q]a-[q]i, with *-qi as an addition rather than replacement; note another NI form of the root: P-Ruk *vagisi 'thigh', from *bagis (A g/q) < *mpa > qis, with an -s added for good measure; Jp hagi 'shank' (another semantic shift) can be analyzed as ha > gi (h- < *p-) but the medial voicing is unexplained (as often in Jp) and the final -i cannot be dis. for *-i/-is in view of the reg. -i < *-s. Additionally, MY also parallels Formosan in providing evidence for <i>; see RIB for Kvl, FLOWER for MY.

The speakers of a proto-language had a proto-culture, we sometimes forget, and the PAT-speakers cultivated ricefields, probably married their cross-cousins (Benedict and Blust for PAN by different, complementary routes) and, it now appears, also had the zany institution of a men's language, with *-qi as one of the markers. This has survived to the present day in Atayalic, where it is gradually losing ground to the 'standard' women's speech, and perhaps to fairly recent days in Kvl (note above: app. infixed *i; baɤa: < i > ŋ 'rib'), but outside this marginal northern Taiwan area this ancestral men's language has long since lost out, leaving in its wake only marginal bits and pieces, curiosities for the inquiring linguist. It also seems likely that /i/ was a principal vowel in added segments in this men's language, as it is in the present Atayalic pattern, although this hardly meshes with the presence of an underlying *$^?$i for 'female/feminine' (see NOTE on MOUNTAIN); cf. also P-Tai *$^?$ic 'little', also used as prefix for 'little animals'; perhaps originally *$^?$i- 'small/fem.' vs. *$^?$u- 'large/masc.', the latter losing gender designation and becoming simply the *$^?$u- marker; the phonoesthetics are clear here but why the choice of /i/ as a favorite vowel in the men's language segments$^?$ The anawer here may well be tied in with

the answer to the more basic question: why the men's language?

RICE[2] *qasal P-Pai *qasal: Puy ?asal 'hulled rice'; Pai qasal/qasan 'unhulled rice'.

   P-KD (Tai/Laha) *saal[A] 'hulled/cleaned rice', from *[q]asal (thru VT).

ROAD/PATH *qəloɣ P-F *qəluɣ (Dyen: A-77 *q$_{12}$eLuR) 'road (Sed); path (Ruk); trail (Pai)'.

   P-KD (Tai/Be/Laha/Gl) *qəlo[rl][A] (dis. reflexes lacking) 'road, path, way' > P-Tai *xron[A] (*xr-for [lacking] *xl-); Gl maintains *qəl-.

   NOTE Ruk shows A ə/u and, as alternative, the root can be reconstructed as *qəluɣ, with the same A ə/u (reg. /o/ < *ə) for P-KD.

ROOT/DIVISION *ɣamuc/*ɣamic P-F *ɣami[Cs] (Li: 103 *Rami[Cs] 'root'; add P-Ata *gam > il 'root', also (Eg.) '[basic division:] social class' (Squ); also NPH: Iva yamit; Yami lamit '[penis rootlets:] pubic hair'; PAN *ɣamut: PPH (Zorc) id. 'root'; PO (Blust) *ɣamu 'id.'; note especially Yami (Asai 1936) yamut 'root', yamut ~ lamit 'pubic hair'.

   P-Tai *hmuat 'section, division', from *[r]amut (thru VT, with typical *r > /h/; dis. Hlai reflexes lacking for P-KD *-c).

   NOTE In final position, P-F *-C uniformly reflects PAN *-c, represented by PMP *-t, hence can be written simply *-c.

   This doublet reflects a basic 'large' vs. 'small' (rootlets) distinction, with the anticipated *u vs. *i vocalic contrast.

SALIVA *ŋalay P-F/PH id. (Dyen: C-246 *ŋaLay).

   P-KD (Tai/Be/KS/Lk/Hlai) *ŋ(a)lay[A] (with var. VT).

SALTY *qapə(ñ)jaŋ Pai qapədaŋ < PAN-level *qapəjaŋ.

P-KD (Tai/Be/KS/Lk) *$^?$(ñ)jaŋ^C, with /$^?$/ < *q-, thru CRM.

P-MY *$^?$ña(a)ŋ^B ($^?$ñ- < [lacking] *$^?$ñj-; /$^?$/ < *q-) 'bitter, astringent', thru CRM induced by *ə in SYL-II.

SCRAPE *kuskus P-F/PH id. (IV-313 *kuSkuS)

SCRAPE OFF *kitskits P-F/PH. (IV-297 *kiSkiS) but cf. Wolff, fn. 41:
'There are roots with the shape qisqis, qiʃqiʃ [ʃ = ts], *kiskis, kiʃkiʃ, kuskus and kuʃkuʃ, all with the meaning 'rub, scrape, shave' and the like. Clearly there is a process of sound symbolism at work here, but one or more of these roots may well have been inherited. *kuskus is actually a good candidate as an inherited form...'

P-KD (Tai/Laha) *khui^B < *kus (-i < *-s, with typical tone *B) ~ *gui^B < *kuikui (A g/k) 'scratch, dig up dirt (as hens) (Tai); scrape, tear, dig (Laha)'.

P-Tai *khit < *kits (-t < *-ts) 'rub, wipe'; also *khiit < *kitskits (thru VT) 'strike with sliding notion, scratch, mark with lines, rub against sthg.'; also (Si.) *kriit (< infixed *-r-) 'draw a line, e.g. with a knife'.

NOTE As Wolff has suggested, *huskus appears to have the best early (at least PAK) credentials of the above group of roots, significantly along with *kitskits, indicating that both *kutskuts and *kiskis are analogical formations.

The basic gloss should be extended to include 'wipe', as in P-Tai *khit; note also NPH:Kap kuskus < *kutskuts 'wipe'.

DS forms occur both in AN and KD, the AN with final *-s, the KD with

*-ts: Ami mi-kəskəs 'scratch (with fingernails)'; P-Tai *khoot < *kotkot (thru VT; /o/ < *ə) 'scrape (bone)'; also *got < kotkot (A g/k) 'shave'.

SHADOW/SPIRIT/GOD *(m)pili Saa (O/A) pili 'shadow'.

P-KD *phriᴬ (*ḷ > /r/ in cons. clusters) 'spirit, demon, god'.

Jp hi, Old Jp Fi 'sun' < 'sun-god'; Old Jp also the earlier 'spirit' and 'god' (compounded in deity names in the Kojiki), from *pi[ḷi] (with reg. CRR); also mi- < *mpi- '[godly/holy:] exalted', an honorific represented inter alia in Jp mikado '[exalted (palace-) gate:] Emperor (the Mikado)' and (derived meaning: 'sun') in minami 'South' (-nami 'waves'); also -ri, the 'split cognate', from *-ḷi (reg. /r/ < *ḷ): Inari 'rice (ina-) god (-ri) 'god of harvests'.

NOTE For Saa 'shadow', cf. Eng. *shade* 'shadow; disembodied spirit; ghost'.

Jp maintains the earlier meaning of 'god/holy' in compounds: hijiri 'god (hi-) knower (jir-i)'; hiko 'god (hi-) child (-ko) = 'prince' (= miko < *mpi-ko); hime 'god (hi-) female (-me) = 'princess'. For further details, see *JAT*.

SHELL (SEASHELL) *(-)li(m)pa Pai kalipa < *ka-lipa. 'shell (gen.)'.

P-tai (SW) *ˀbiaᴮ 'cowrie (shell)', from *ˀmpia < *q(a)-[l]impa (thru VT).

SKIN/HIDE *kaba P-F *kaba: Kan káva 'skin'; Pai kava '[hides:] clothing', sə-mu-kava 'take off clothes'; Bas kaba 'jacket, coat'.

Jp kawa, Old Jp kaFa 'skin; hide, fur'.

SNAIL *munaḷ Sai monaḷ < *munaḷ.

Jp nina (A n/m); Old Jp mina (D i/u; *-ḷ > ø)

P-KD (Tai/KS/Lk) *ʔnual 'maggot, worm' from *[m]unal (thru VT; *-1 [Saek] < *-ḷ), with prefixed *q(a)- > /ʔ/.

SPEAK/CALL *qibu Pai qivu < *qi[b]u 'speak, call'.

Jp. i-i 'speak' Old Jp iF-i, from *iFɩ < *iFu-i (reg. ɩ < /ui/).

P-Tai (SW) *kiuᴬ ~ *ʔiuᴬ 'call', from *[q]i(w)u < *[q]i[b]u (reg. medial *-b- > -w-); cf. JAW.

SPIT/SPITTLE *tuɤipəs P/F/PH id.: P-F *tiɤipəs 'spit' (Li: 109 *tiɤpəs) (A i/u): Paz tixipəs, Kvl t-m-iɤpəs, Sed t-um-əpəs (A ə/i); add Sed təpəs 'spittle' (*ɤ > ø); NPH (Ilk/Itn/Mnb/Luba/Knk/Ini/Isi/Gad) *tuɤpa[ ] 'id.', thru CRM, with reflexes for final *-as rather than *-əs (see discussion of this point in Blust: 5.1,4).

P-KD (Si/Hlai:QD) *thuiᶜ 'spit', from *tu[ɤ]i[pəs] (*-ɤ- > ø).

P-MY *thuiᶜ 'spit', with development as in KD.

NOTE This rare agreement in the KD and MY developments, here involving even the tone (*C), can be traced to the 'weak' *i in SYL-II → CRR in KD rather than the typical CRL (*ATLC*: 151-2).

CF.also Laha thowᶜ < *thuᶜ 'spit', perhaps from *tu[ɤ]u[pəs] (A u/i), contrasting with P-F (A i/u).

SPLIT (OFF) *tsi(m)pak((m)pak) WMP *si(m)pak 'split' (Blust 1970-407 *si(ŋ)pak): Iban 'chipped, a chip'; also the doublet: *tsibak < *tsimpak 'cleave' (ibid.-402 *sibak); also WMP (Demp.) *bak < *[tsi]mpak 'split off'; PMP (Demp.) *bakbak < *[tsi]mpakmpak 'debark'.

Jp hag-i 'strip off, tear off, flay, skin', from *pakpak (A g/k).

P-Tai *baak < *mpakmpak (thru VT; /b/ < *mp) 'to split into halves (Ahom); to skin (Shan)'.

P-MY *si(m)pa$^?$ < *tsi(m)pak '[sthg. split off; grain chip:] bran'.

NOTE A rare disyllabic P-MY recon. necessitated by the variable NI:Miao *sp[ia$^?$] but Yao *$^?$bia$^?$ (h.t.) < *[s]mpia$^?$; since both show VT they present strong evidence that this process (VT) was an ongoing one in MY, as in KD.

SPREAD OUT *sa(m)paɤ PAN id. (Tsu. *S$_{13}$apaR) 'spread out (esp. mats); mat'; for F/PH, Wolff cites *sapaɤ 'woven bamboo or palm leaves'; add Pai (w.) s-m-apa 'spread out (on ground); put pad of leaves between burden and one's back'; PMP *ha(m)paɤ 'spread out (as mats); mat'; PO *$^?$ampa ~ (DS ə/a) $^?$əmpa 'lie like a mat; mat'. Jp har-i 'spread, stretch'; har-a '[spread of land'] field', with Jp nominalizing *-a (ef. Texan 'spread of land').

P-KD (Tai/Be/Hlai) *phɨan$^{A/B}$ 'spread (mat, coverlet); clf. for mats, woven objects', from *[s]əpa[ɤ] (DS ə/a; cf. PO, then VT ɨa.

P-MY phaan^A 'clf. for mats, coverlets, nets, etc.$^?$ (thru VT).

STAR *(n)talaw NPH (Luba/Knk/Ini) *talaw < *[Ct]alaw.

P-KD (Tai/KS/Lq/Hlai) *$^?$draaw^A < *q(a)-ntalaw (thru VT; *ḷ > /r/ in cons. clusters).

STRETCH *biyac P-F/PH id. (Dyen: B-52 *biaC) 'draw a bow to its full extent (Puy); drawn, stretched (of bow) (Ilk)'.

P-KD (Tai/Be/Hlai) *$^?$yia[ct] [dis. Hlai:BD reflex lacking) 'to stretch (esp. by pulling), stretch out, stick out', from *[b]iya[c] (thru VT; /$^?$/ for *b-); P-Hlai also *ya^B 'stretch, stick out', perhaps from a part. redup. *[bi]ya[yac].

STRIKE REPEATEDLY (WITH SHORT, QUICK STROKES) *tsaktsak PAN id.
(IV-538 *saksak; Wolff *ʃakʃak) 'hack, chop up'; Blust notes: 'Possibly equiva-
lent to Dempfolff's *saksak [= *t́akt́ak] 'prick, pierce, stab' while Wolff
expresses similar doubts about any connection between these reconstructed
forms.

P-KD *sak (< *tsak): Tai: Shan "strike repeatedly with short quick
motions, peck (as a fowl), strike at (as a snake)'; Ahom 'thrust a pointed
instrument'; Kh 'to paw (as a pony)'; Si 'bore, pierce (as with an iron),
tattoo, dig (as in ground with sharp point), strike (with sword)'; Lao 'pierce,
print, prick, tattoo'; WT/BT/Dioi 'prick (skin), tattoo'; Hlai 'chop up'.

P-MT (Miao) *ts[a?] < *ts[ak] 'chopper for cutting firewood'.

In both AN and KD this root presents the same problem: how to recon-
cile 'pierce/prick' with 'hack/chop up'? The answer here is provided by Tai,
esp. Shan: set up as a proto-gloss: 'strike repeatedly with short, quick strokes
(whether in piercing or cutting)'.

SUCK/SIP/INHALE *(n)tsuptsup P-F/PH *tsuptsup (IV-588 *supsup; Wolff
*ʃupʃup) 'sip, suck [blood]'.

Jp su-i, Old Jp suF-i (/s/ < *ts; /F/ < *p) 'suck, sip, inhale'.

P-KD (Tai/Laha) *suup 'suck, inhale, scent', from *tsuptsup (thru VT;
/s/ < *ts).

P-MY (Miao) *nts[op] 'suck (blood)' (/o/ < *u).

The doublet: *tsəptsəp is represented both in AN and KD; cf. PAN
*tsəptsəp (Tsu. *θəpθəp) 'suck'; P-Tai (Shan) *soop < *tsəptsəp (/o/ < *ə)
(thru VT) 'to scent'; prob. also P-Tai *sop (< *tsəp) '[inhaler/sucker:]
mouth'.

SWEEP\WINNOW *ta(m)pi((m)pi) PAN id.: PMP *ta(m)pi 'remove dust and chaff, winnow' (NgD tepe = tempe 'pound rice'): P-PN *tafi < *tapi 'sweep' ~ *tapi < *tampi 'wipe off'; P-Pai *tabi(bi) < *tampi(pi): Puy:Hinan tabi '[rice-pounder:] mortar' (cf. NgD); Ruk wa-bi:bi: 'wipe'.

P-KD (Tai/Be/KS/Hlai) *bi(bi)ᴬ 'winnow, fan', from *[ta]mpi(mpi); also P-Tai (Lao) *ˀbianᴬ 'winnowing basket', from *[t-]mpi-an, with P-KD nominalizing *-an, with /ˀ/ for *t-.

P-MY (Yao) *peiᶜ < *piᶜ < *[ta]pi[pi] 'skim off (rice water)'.

NOTE Cf. the parallel forms in AN with focus-marker *-an: Tag tahip 'up and down movements of grains being winnowed', tahip-án 'winnowing basket'; Sai:Taai h-öm-aləp 'winnow', kapä-häləp-an 'machine to winnow cereals'.

THORN *(n)tsuqaɤ P-F/WMP id. (I-404 *suqaR; Wolff *ʃuqagh).

P-KD (Tai/KS) *(n)suˀanᴬ (SW/C-Tai *sianᴬ, with > /i/ before ˀ).

TONGUE *səma (III-328 *Sema).

P-KD (KS/Lq/Laha: comp.) *(k-)maᴬ < *(q[a]-)ma.

TURN AROUND *[m]a-liuts(liuts) P-Pai *(ma-)liuts(liuts): Ruk. (F.) ma-liuliūs; Puy (Ting) mu-lius ~ mu-liulius 'turn (sthg.) around'; (F.) mu-liūs 'id. (person)' ~ mu-liuliūs 'id. (thing)'; (Tsu.) m-u-a-riyus (unexpl. /r/ for *l).

P-Tai *hliawᴬ 'turn around (esp. head to look)', from *[m]a-hliu (thru VT) < *-sliu < *[liu]s-liu[s] (Procr. loss of final).

NOTE This root provides rare evidence for prefixed *ma- in KD, here paralleling Ruk ma-.

TURTLE (SOFT-SHELLED) *(n)Ci(m)pa P-F/PH *qa-(n)Cipa (Tsu. id.; Wolff *qatipa); add Chamic *tə(m)pa(-pa) (DS ə/i), as reflected in Jar təpa; (loans to MK:) Suei tpaa, Chrau dapa < *tapa (A a/ə), Khmu tmpa; also Vn baba < *papa (b- < *p- is reg. shift).

P-KD *q-[t]ipa > (Tai: comp./Laha) *pa^A (reg. CRL) ~ *p(i)a (var. VT > (prefixing *qa-) > (Tai: gen.) *f(ɨ)a^A (reflecting DS ə/a, then VT ɨa, along with reg. f < *q-p-).

NOTE Tai and KS also have (poss. back) loans from Chinese: MCH 鱉 pɹät 'soft-shelled turtle', a likely loan from DAC from a part. redup. prototype (cf. Vn baba): *[t]ipa-t[ipa] > *piat (thru VT, an established DAC feature) > pɹät (typical Ch shift before *-t).

TURTLE/TORTOISE *draulu P-F *ḍaulu (Dyen: A-24 *D$_{14}$aulu[]) 'an inedible mt. turtle (Ata); tortoise (Bun)'.

P-Tai *dlaw^A (Saek) ~ *tlaw^C (gen. Tai, with typical unvoicing) ~ *thaw^C (< *thlaw^C) (Hlai, with typical sec. aspiration) 'turtle (land), tortoise', from *dlau[lu] (A l/r); init. cluster → CRR).

TUSK/BOAR *walis P-F id. (Tsu. *W$_3$aɨiS$_1$; Li: 116 *waNiS) 'tusk/tooth' (Tsouic/Ruk/Pai/Puy/Ami/Paz/Sir) ~ 'tusk/wild pig (Bun) ~ 'wild pig' (Thao); Sai wališ 'fang', wališ-an 'wild pig (male)'.

Ryukyuan: Shuri wā 'pig' (reg. CRR); Jp i, Old Jp wi 'boar', from *wi [lis] (A i/a).

TWO *putša Pai -pusa-, Bun pusa-n (W. Li 1990) 'two (in series)'; P-Tsouic *-pusa 'id.', from PAN-level *pu[stš]a.

Jp futa- 'two'.

For the recon. of *tš, see *JAT*: 88-91).

VISCERA *(m)paŋ(m)paŋ Pai vaŋvaŋ < *[b]aŋ[b]aŋ 'viscera', from *mpaŋ mpaŋ.

P-Tai *paaŋᴮ 'spleen (human)', from *paŋpaŋ (thru VT).

NOTE Cf. also WMP *limpa 'spleen': Ml limpa 'liver', limpa kĕtil 'little liver' = 'spleen', perhaps part. redup. from a root: *[ll]i(m)paŋ via *[ll]impampaŋ, the source also of this F/KD root: *(m)paŋ(m)paŋ.

VULVA *pa(n)ti P-Ruk (Tan/Bud) *pati

P-KD *-a(n)tiᴬ/ᴮ: P-Hlai (BD/WS Li) *tha(a)iᴮ (reg. sec. aspiration along with VT); Pub toiᴬ (l.t.) < *doiᴬ, from *dəiᴬ (reg. /o/ < *ə) < *-ənti (DS VT; /d/ < *nt); P-KD *ñiᴬ (Mak) ~ *ña:iᴬ (Kam; Sui has the db.), from *-andi (var. VT; sec. nasalization: /n/ < *nd, with palatilization before *-i); poss. also P-SW/C-Tai *hiᴬ < *hñiᴬ although only *h- < *hŋ- is well attested for Tai.

WASH *sawsaw PAN id. (IN-594 *SawSaw; Wolff *sawsaw) 'wash (Ami:gen. term); rinse (clothes) (Tag)'.

P-Hlai (XF/Cun) *saawᴮ 'wash (veg's) (XF); bathe (Cun)', from *sawsaw (thru VT).

NOTE Prob. from *[l[i/u] saw (saw), with part. redup. paralleling that in the doublet: *l[i/u] (n)tsaw(tsaw): Puy mə-lisaw ~ l-əm-isaw 'wash (utensils)' (/s/ < *ts); WMP *lutsaw 'rinse out, wash out/away'; P-KD *daawᴬ ~ *zaawᴬ 'wash (esp. rice, also for gold, silver); bathe', from *dzaaw < *ntsawt-saw (thru VT), with underlying sense of 'washing by movement in water' (cf. Tag l-ag-usao 'movement in water'); P-MY ntsaawᶜ 'wash; bathe', from

*ntsawtsaw (thru VT).

WEED, v. *[rɣ]ə(m)bay Puy (Ting) *mə-rəbay/r-əm-əbay

P-KD (Tai/KS) *(ʔ)(m)b[rl]aay^A (M rl/b; A a/ə VT; /ʔ/ for *rɣ).

NOTE As suggested by the Puy cognate, it is possible that the P-KD *(m) represents an incorporated *-m- infix rather than NI; see the NOTE on COME/GO.

WIND/TYPHOON *bali(bali) P-F/PH id. 'wind' (Dyen: B25); Tag bali-baːli 'strong, changeable wind' (Dyen cit.); add Kal baːli 'typhoon'.

P-KD (Buy/Be/Hlao/Laha/Gl) *w(ɨ)an^A 'wind' (Hlai 'typhoon'), from *[bal]i-wal[i] (A w/l), with DS ə/i → var. VT ɨa).

WINNOW *tahu(m)pəs PAN id.: Blust (Table 2) cites P-F *tapeS ~ PMP *tsSep = *tahəp < *tahəp[əs] (A ə/u, with CRR); Tsu. *tapəS$_{13}$ < *ta[hu]pəs; also *tapuS$_1$ < *ta[hu]pus (A u/ə), both with CRM; add P-Ata *t-um-apəs ~ *t-um-abus (< *t-um-ampus); P-F also *Capus: Puy ṭapu-i 'winnow', Sai sapöh 'sweep' (see SWEEP for semantics), from *trapus (see LOUSE for /C/ < *tr) < *t-r-apus.

Jp mi, Old Jp mï 'winnower', from m[ou]i < *mp[ou]i (-i < *-s).

NOTE Inasmuch as Old Jp /ï/ is not dis. for *oi/ui it is uncertain whether or not'it reflects A u/ə (cf. the F assim.).

For the bifurcate F/MP development here, cf. HAIR.

WORM (EARTHWORM) *qa-(n)tulal Pai:But qa-túlaĺ < *-tulal.

P-KD (Tai/KS) *(ʔ)(n)tu[rl]al^A (Saek *tlual^A, thru VT) < *(qa-) (n)tul al, with *q(a)- as elsewhere represented by /ʔ/.

APPENDIX - ADDENDUM-1

PLAIT/WEAVE/SEW *(n)(tra)trahiqis PAN *Cahiqis 'sew': Blust (Table 2) cites P-F *Caqis/PMP *taSiq, the former thru CRM from *Ca[hi]qis. the latter (= *tahi$^?$) thru CRR from *tahiq[is]; note also for F (Ruk/Paz) *CaCaqis; note also the Tsu. cit.: Tongan sia$^?$-i (= *tia$^?$-) (M ia/ai) 'weave (net or web)'.

   P-Tai *thak 'plait', from *tra[h]k (A r/h) < *tra[i]q[is] (-k < *-q-); also (C/N-Tai) *dak (/a/ < *nt) 'plait (mats); knit, braid'; Be da$^?$ (h.t.) < *taak 'sew (when making new clothes)', from *tra[h]ak (A a/i); Hlai:jM taak 'weave' ~ *ta < *t[h]ak 'plait'; P-Hlai (gen.) *thrɨak 'weave', from *tri[h]ak (M ia/ai; cf. Tongan) 'weave'.

   P-MY *ntat 'weave', from *ntrat[rahiq] (A t/tr).

   NOTE The medial *-h- in this root has led to contrasting CRM vs. CRR developments in F and MP; cf. also HAIR and WINNOW:

	PAN	P-F	Thao	PMP
HAIR	*buhək[aə]s	*buk[aə]s	fukiš	*buhək
		*buhukəs	hukiš	
WINNOW	*tahu(m)pəs	*tapəs	—	*tahəp
		*tapus		
PLAIT	*cahiqis	*Caqis	š-m-aqiš	*tahiq

APPENDIX - ADDENDUM-2

RICE[1] NOTE (cont'd): Likely scenario for development of men's language in AT: early on, a group of adolescent males began to play around with their language, esp. with the $*^{?}i$- prefix marking females/femininity. They followed a simple, contrariant rule: convert everything to the opposite: 'bad' becomes 'good' (recent U.S.A. illustration). Make /i/ a marker for males and their language rather than for females and suffix it or infix it but never, never prefix it. Later on, things got a little out of control, with other elements used, some with vowels other than /i/, but the basic rule against prefixation has held fast to the present day in Atayalic (and from what is known of Kvl).

It would appear that the AT peoples played with their speech to an unprecedented degree. The Austronesians, esp. the Malayo-Polynesians, went in for "blooting", involving variations of various kinds in the first syllables of words, leaving the second syllables as landmarks, so to speak, collected by Blust (1988) as a series of "roots", distinguished here from roots proper by the convenient portmanteau: "bloot". There is evidence for both kinds of games in MP, including survivals (text) such as the doublet for 'thigh': $*(m)pa^{?}a \sim *pa^{?}i$, but it seems clear that the "blooting" game has long since taken over, making for the extensive collections presented by Blust in great detail. Can it be that the playing of games such as this, even some of the rules, is attributable to a proto-language at times? On the evidence of AT, the answer is: yes.

APPENDIX - ADDENDUM-3

The Amis Connection (fn. 3)

The Amis reflexes, paralleling those of MP, appear to provide significant support to the Amis/Extra-Formosan grouping posited by Mark Harvey (1982). PAN *c is meagrerly represented in initial or medial position but the few cognate sets available show that Amis, in contrast to Atayal and at least three Paiwanic languages (Pai/Sai/Thao), has distinct reflexes for PAN *C and *c; cf. the following, incl. Old Jp.:

	PAN	Paiwan	Sai. Ata.	Thao	Amis	WMP	PPN	Old Jp.
die/end	*maCay	matsay	masay (Sai.)	maθay	mapatay	*matay	*mate	
	*paCay					*patay		Fate
sea	*wacal	vatsaɬ	wasal (Sai.)	-	-	-	*wasa	wata
recite	*ucap	-	-	-	-	*ucap	v/osa (Fij.)	utaF-i
united	*cahiŋ	tsaiŋ	-	-	-	*cahiŋ	-	-
rise	*caka	-	-	-	-	*caka		taka-
	*ca(ŋ)ka-i				-	*ca(ŋ)kay	hake	takë
	*ca(ŋ)ka-t				tsakat	*ca(ŋ)kat		kat-i
peak	*qapucuk	-	-	-	ʔapotsok	*ʔapucuk	-	-
cut up	*calcal	-	-	-	tsaltsal	*calcal	-	-
broken	*p[ə]caq	-	-	-	ptsaʔ	*pəcaʔ	voza (Fij.)	-
trade	*caliw	-	saliu (Ata.)	θariw	tsaliu	*[c]aliw	-	-

NOTES:

die/end PAN (gen.) *maCay 'die' ~ *paCay 'kill' but note Amis mapatay
'die'; Toba-Batak mate 'dead' ~ pate 'come to an end'; Old Jp. Fate (<
*patai) 'end', Fate-ri 'end, be finished, die'.

sea Sai wasal 'sea', wasalwasal 'lake/pond'; Pai 'deep pool; (Western) lake';
the root appears to be unrepresented in WMP.

recite 'speak, converse' (Demp.): Malay 'recite'; Jav '[recite =] enumerate
good deeds'; Old Jp. utaF-i (< *utap-) 'sing, chant, recite'.

united Pai 'united, connected'; WMP 'together (NgD); tar (Tag); syrup (Jav)'
(Dahl 1976 cit.).

rise Amis 'arise, go up'; Blust (AE-IV: 535) cites *sakat 'rise, climb up' at
the PAN level but the nominalized Malay caŋkat 'a [rise:] low hill'
disambiguates for initial *c- here as well as in the related forms (cf.
Buru saka-h 'ascend to', saka-t 'up'): *xa(ŋ)kay, for *sa(ŋ)kay 'ride on
something' (AE-II: 344), surely (contra Blust) etymologically the same
root ('mount') as Dempwolff's *takay 'ascend (besteigen)'; PPN hake
(< *ncakay) 'up, upwards' (Tuamotuan 'go onto, upon') and *caka:
Cebuano saka 'climb' (sakay 'ride a vehicle'); Wolff (Taipei symp. paper:
The Position of the Austronesian Languages of Taiwan within the
Austronesian Group: fn. 44) notes, 'There are also forms with reflexes
from *ʃaka [= *tsaka] in various [WMP] languages ...'; Kadai has both
the unaffixed root: Kam-Sui *khaᶜ 'climb' and an *-an derivative: Hlai
*khaanᶜ 'id.'; Old Jp. taka- 'high'; takë (< *takai) 'height'; also kat-i
'[rise above:] surpass, prevail', with regular canonical reduction-on-the-
left with the verbal suffix.

peak Amis 'the very top of a mountain'; WMP 'peak of a mountain' (Blust:
AE-IV: 33).

cut up Amis 'break [= cut] up with a large machete or ax'; WMP 'cut up; demolish' (Demp.).

broken Amis '[broken:] out of joint'; WMP 'in pieces (Demp.); broken (PPH - Zorc).; Fij. 'wrecked (canoe)'.

trade Ata 'trade'; Thao 'buy, sell'; Amis 'borrow, lend'; Blust (AE-IV: 542) cites *saliw 'buy, sell' at the PAN level but the WMP forms cited (glosses: 'exchange' ~ 'buy' ~ 'sell') do not disambiguate from *[c]aliw while Ata s-, Thao θ- are the reflexes for PAN *C-, in contrast to Ata h-, Thao t-, which are the reflexes for PAN *s-.

The key item here is trade, tying in Amis c- with Ata s- and Thao θ-. MP has numerous *s- and *c- doublets, hence items without this tie to established *C reflexes cannot entirely disambiguate. Amis has /ts/ also for PAN *s- and for broken may well reflect a *pəsaq doublet, as indicated by Puy mu-pisa?/mu-pəsa? 'broken'; note Puy /s/ < *s vs. /t̪/ < *C (m-in-at̪.ai 'die'). Alternatively, the Puy form can be viewed as evidence for a Puy/Amis rather than simply an Amis connection. Ross (Taipei symp. paper: Reconstructing Proto-Austronesian Verbal Morphology: 378) concludes, 'It seems likely that Proto-Malayo-Polynesian ... may subgroup with a small number of Formosan languages, probably in the south of Taiwan ...'. The phonological evidence can be seen as supporting subgroup membership surely for Amis, perhaps also for Puyuma, while denying it for Paiwan, which has /ts/ for both PAN *C and *c.

TABLE I.
AN/KD *p + r/l/ḷ Reflexes

	*(m-)pr	*pl	*(m)pḷ init.	*(m)pḷ med.
P-Paiwanic	*C	*C	*C	*C₁
Paiwan	ts	ts	ts	
Bunun	t	t	t	ts ~ s
S/C. Tai	*t	*t	*h	*th
Saek (N. Tai)	pr	pr	th	
Buyang	t	pi = py	-	
P-Kam-Sui	*mpr	*pr	-	
Kam	t	t		
Mulao(Dawu)	mɤ < *mpɤ-	pɤ		
Lakkia	pl	pl	pi = py	
Be	?d < *t	?d < *t	?d < *t	
P-Hlai	*C	-	*t(h)	
Jiamao	t		d < *t	
Cun-hua	h		-	
Other	*tsh		*th	
Laqua	t	t	-	
Laha	t	ph	p	
P-Lachi	*mp(h)y	*p(h)	*[c]	
Lati	mc	ph	c	
Lachi	-mti	p	-	
Jinchang	phĭ	ph	-	
P-Gelao	*C	*pl	*(n)t	
Gao:Wanzi	t	p	-	
S. Gelao	t	pl	t	
Pudi	-	-	nt	
Longjia	s	-	-	

NOTE Table from The wild in Kadai, *Kadai* III: 67-70, q. v. for further details; PAT *mapra 'eye', *ma-play 'die ~ *pa-play 'kill'; *plaŋits 'weep'; *qapḷaŋ 'village'; also PAK *qu(m)pḷal 'wilderness; forest'.

ABBREVIATIONS

> suffixed element (men's lang.) < > infixed element

A assimilation Ar Archaic AN Austronesian AT Austro-Tai Ata Atayal(ic) *ATLC* Austro-Thai: Language and Culture Bas Basay BB Ban Bung (Laha) BD Baoding (Hlai) Bik Bikol BT Black Tai Bud Baudi (Rukai) Bun Bunun But Butanglu (Paiwan) Buy (Buyang) C Central Ceb Cebuano Ch Chinese Chi Chihpen (Puyuma) CRL canonical reduction on the left CRM id. in the middle CRR id. on the right Cun Cunhua (Hlai) D dissimilation dis. disambiguate/disambiguating DAC D[onor] to ArCh DS distressing (other vowel to /ə/) F Formosan F. Ferrell Fav Favorlang Gad Gaddang Gl Gelao *HCT* Handbook of Comparative Tai HT Heitu (Hlai) Iba Ibanag Ifg Ifugao Ilk Ilocano Ina Inago (Sediq) Ini Inibaloi Ish Ishbukun (Bunun) Isi Isinai Isn Isneg Ita Itawit Itn Itneg Jar Jarai Jav Javanese *JAT* Japanese/Austro-Tai JM Jiamao (Hlai) Jp Japanese Kad Kadazan Kan Kanakanabu KD Kadai Kel Kelabit Ket Ketagalan Kh Khamti Knk Kankanaey KS Kam-Sui Kvl Kuvalan Lk Lakkia Lq Laqua Lt Lati M metathesis Man Mantauran (Rukai) Mand Mandarin Mar Maranao Mas Maspazi? (Atayal) May Mayrinax (Atayal) MCh Middle Chinese Mkz Makazayazaya (Paiwan) Ml Malay Mlw Malaweg Mnb Manobo MP Malayo-Polynesian MY Miao-Yao N North(ern) NgD Ngaju-Dayak O/A Ogawa & Asai P(-) Proto- Pai Paiwan(ic) Paz Pazeh PH Philippine PN Polynesian PO Proto-Oceanic Pub Pubao Pup Pupeo Puy Puyuma Ruk Rukai Saa Saaroa Sai Saisiyat Sam Sambal SD Southern Dai (Hlai) SEA Southeast Asia(n) Si Saimese (=Thai) Sir Siraya Squ Squliq (Atayal) ST Sino-Tibetan *STC* Sino-Tibetan: A Conspectus SW Southwest Tag Tagalog Tan Tanan (Rukai) Tao Taokas Tap Taparon (Ami) Tau Tauran (Ami) Tja Tjavuali

(Paiwan) Tju Tjuabar (Paiwan) Ton Tongan Tro Trobiawan Tsu. Tsuchida TU Than-Uyên (Laha) Uve Uvean Vn Vietnamese VT vocalic transfer (from preceding syllable) W West(ern) WBM Western Bukidnon Manobo WS White Sand Li (Hlai) WT White Tai Wum Wuming (N-Tai) XF Xifang (Hlai) Yog Yogad.

BIBLIOGRAPHY

Benedict, Paul K.

    1988   Kadai Linguistics: The Rules of Engagement. In Edmondson, J. A. and D. S. Solnit (eds.). Comparative Kadai: Linguistic Studies beyond Tai. The University of Texas at Arlington.

    1989   KD Clusters/dyads: PT *pl/*p-l/*phl. Kadai I: 19-14.

    1990   The gods of Sino-Tibetan. Acta Orientalis 51: 161-72.

    1991a   Kadai Incorporated *-um- Infix. Kadai III: 71-72.

    1991b   Proto-Tai preglottalization. Ibid.: 65-66.

Blust, Robert A.

    1970   Proto-Austronesian Addenda. Oceanic Linguistics 9:104-62.

    1988   Austronesian Root Theory. Armsterdam: John Benjamins.

    Forthcoming *Austronesian Comparative Dictionary.*

Dahl, Otto C.

    1976   *Proto-Austronesian.* Second revised edition. London: Curzon Press.

Ferrell, R.

    1982   Paiwan Dictionary. Canberra: The Australian National University.

Harvey, Mark

    1982   Subgroups in Austronesian, in Halim, Amrab, Lois Codrington and S. A. Wurms, eds., *Papers from the Third International Confer- ences on Austronesian Linguistics,* Vol. 2: *Tracking the Travellers,* Pacific Linguistics C75, Dept. of Linguistics, Research School of Pacific Studies, The Australian National University, Canberra.

Haudricourt, André-Georges

    1985   Du nouveau sur le Bê. In Thurgood, G. et al. (eds.). Linguistics

of the Sino-Tibetan Area: The State of the Art. Canberra: The Australian National University.

Li, Paul Jen-kuei

1983a   Notes on Thao Dialects. BDAA 43: 48-50.

1983b   Types of Lexical Derivation of Men's Speech in Mayrinax. BIHP 14.3: 1-18.

1988   A Comparative Study of Bunun Dialects. BIHP 19.2: 479-508.

1993   New Data on Three Extinct Formosan Languages. BIHP 63.2:301-323.

Li, Wensu

1990   The Word-formatives Function of the Morpheme -an in Bunun. MZYW 1990, 4: 25-28 (in Chinese).

Matisoff, James A.

1988   Proto-Hlai Initials and Tones. In Edmondson, J. A. and D. S. Solnit (eds.). Comparative Kadai: Linguistic Studis beyond Tai. The University of Texas at Arlington.

Ting, Pang-hsin

1978   Reconstruction of Proto-Puyuma Phonology. BIHP49.3:321-392.

References to roots cited in Symposium papers are given under the names of the scholars: Blust, Dyen, Li, Wolff, with numbers as they appear in the papers.

I-, II-, III-, IV- plus numbers are references to roots cited by R. Blust in his *Austronesian Etymologies*.

Borrowing and Inheritance in Austronesianistics

Isidore Dyen

Yale University (Emeritus)

This paper is a reply to Blust's ISASRT paper that argues for the Formosan hypothesis: that Taiwan was the Austronesian homeland. The essential counter-argument, supported by appended cognate set lists, is based on the homomeric method and concludes (1) that the numerous Formosan-only cognate sets point to a single Formosan subgroup and (2) that the numerous Philippine-Formosan-only cognate sets select the Philippine languages as the Formosan next-of-kin. Blust's hypothesis that all the proposed Formosan-only cognate sets can be explained as borrowings is baseless. Furthermore the closer relationship of Philippine (than, say, that of Polynesian) subverts the hypothesis that the non-Formosan languages, as Malayopolynesian, formed a unity against the Formosan languages, however subclassified.

Currently borrowing hypotheses have been offered to explain data that contraindicate promulgated theses. One thesis is that the Formosan languages (i.e. the Austronesian languages of Taiwan) do not constitute a single subgroup. The other thesis is that the closest relative of the Formosan languages is not the Philippine languages. Blust 1992 was presented at the ISASRT to support both theses and this paper is offered in reply. Li 1992 also supports the first thesis. Since conference papers can be revised before publication, references made to the original papers may appear to be unclear and perhaps even inaccurate, but that can not be helped. This paper too is a revision of one circulated at the conference.

Blust (1992.26 ff.) supports the first thesis by criticising in detail a list

of 37 Formosan cognate sets in Dyen 1967 with both an Atayalic and a non-Atayalic member indicating the unity of the Formosan languages. He accepts six (1.10, 1.13, 1.14, 1.19, 1.25, 1.26) marked by 'YES.' He objects to 31, marked by 'NO,' on various grounds. Nine of the latter (1.1, 1.7, 1.12, 1.16, 1.17, 1.22, 1.24, 1.30, 1.36) are justifiably excluded because extra-Formosan cognates exist. Three more (1.8, 1.15, 1.18) can be excluded on the same basis.

However one set (1.21) is valid after the removal of words showing objectionable matchings. Another (1.32) is valid after the correction of an incorrectly recorded form. One more (1.27) is valid; it was improperly excluded because borrowing could not be ruled out. Seven (1.2, 1.4, 1.5, 1.6, 1.23, 1.29, 1.33) are valid because their exclusion was due to Blust's unfamiliarity with the correspondences of the Formosan languages.

One more (1.3) is valid despite the failure of Atayal ramu<? (ramu in the text) 'blood' fails to exhibit the expected correspondence (q or h) to a Paiwan -q (in jamuq 'blood'); the failure is regarded as a distortion due to 'men's speech.' The profound importance of Li 1982 is difficult to over-estimate. It deals with the eccentricities of Atayalic lexemes and marks a turning-point in their treatment. It proposes that deliberate tampering with the shapes of words led to distortions attributed to 'men's speech' and thus offers a simple explanation of queer deviations in Atayalic words otherwise suggesting likely cognates. These distortions, where detected, take the form of 'affixes;' they are marked here by arrows (< >) pointing toward the inherited portion of word: e.g. (as numbered in Appendix A) Aty snaq<uy 'otter' (A78), Aty[SE] ll<i>uŋ 'river' (A56). Sometimes an inherited portion is elided, as in SedTd rkl<ic 'leopard' (A47) from *Lukelaw. These distortions bear on some of Blust's exclusions in the 1967 list and in the

Formosan-Philippine list: e.g. he calls the segmentation of Sed bql-it (1992.33, here bql<it, Appendix B 69) 'leg' arbitrary. Li's discovery may have even more far-reaching importance; the tampering may have extended to word replacement, reducing the lexicostatistical score.

Since 6 were unchallenged and 11 mistakenly invalidated, 17 of the original 37 are justifiable in terms of the original purpose (and a total of 27 of the sets remain cognate sets if their distribution is disregarded). Notwithstanding two others (1.29, 1.33) are excluded as concerning only the relationship of Atayalic with Pazeh. Li (1985.259 f.) proposed a North Formosan, here called Pazic. Pazic is one of two branches of Formosan; it includes Atayalic (Atayal and Seediq), Saisiyat, Pazeh, and some defunct languages. The other branch is Paiwanic; it includes Paiwan, Puyuma, Amis, Bunun, Tsouic (Saaroa, Kanakanabu, Tsou), Rukaic (Tanana, Budai, Maga, Tona, Mantauran), Kvalan, Thao, and other defunct languages. The argument for this latter grouping, at least of the named languages, will appear in a work by Tsuchida and myself as part of a treatment of the Formosan subgroup and its position in the Austronesian family.

Only sets including a Pazic and a Paiwanic member (and no members from a non-Formosan language) bear on a Formosan unity. Li (1992) appears to recognize additional Formosan branches, but the evidence is not (or perhaps not yet) conclusive; in any case it is not taken into account here. The two sets above thus bear on the proposed Pazic unity, not on a Formosan unity. Since Atayalic is the Pazic member required in the list, only cognate sets containing an Atayalic member along with a Paiwanic member (and no extra-Formosan member) are now valid for the intended purpose; they number 15.

Appendix A is the collection of cognate sets assembled by Tsuchida and

myself that can now be regarded as the present state of that original list. The 15 that have survived appear there, each marked by the number of their position in the original list. An index to these follows with A preceding the number in the appendix: 1.2 - A18; 1.3 - A20; 1.4 - A48; 1.5 - A9; 1.6 - A86; 1.10 - A69; 1.13 - A131; 1.14 - A46; 1.19 - A133; 1.21 - A103; 1.23 - A80; 1.25 - A128; 1.26 - A100; 1.27 - A10; 1.32 - A101.

The original list was drawn up to show that Atayalic belongs with the other Formosan languages despite the low scores of Atayalic members in a lexicostatistical comparison with the other Formosan languages and in fact with all other Austronesian languages. The sources were the lists given in Ogawa and Asai (1935, in an appendix pp. 2-35), and Swadesh lists collected from various contributors. Despite misgivings about the accuracy of the data and my interpretation, it seemed particularly important at that time to publish this evidence of Atayalic's close relation with the other Formosan languages precisely because others might be misled into believing that the time of divergence of Atayalic was so remote that it should be regarded as a prime branch of Austronesian and that Taiwan should be regarded as the likely Austronesian homeland. This consequence was one of the hypotheses-- given third in rank--that I had presented in 'A Lexicostatistical Classification of the Austronesian Languages' (1965.57).

As the author of the lexicostatistical classification and of a proposed method of inferring a homeland from the distribution of daughter languages (1956), I felt it necessary to present the evidence that contraindicated these inferences. The original list treated as 'cognate sets' such sets of similar words as could be judged to be reasonably likely to turn out later to satisfy the requirements of systematic correspondences. Of the 37, only eight (1.9, 1.11, 1.20, 1.28, 1.31, 1.34, 1.35, 1.37) seem now to be unlikely to turn out

to correspond systematically. The warning involved a risk worth taking even if it did go largely unheeded. In any case it is surprising to have the list criticised for its lack of rigor by one who readily admits sporadic change in the assembly of cognate sets, and gives weight for the same purpose not only to non-dialectal 'variants,' but also to connections with and similarity to non-cognates (cf. Blust 1992.36).

If the 15 are added to the 14 that appear in Tsuchida 1976 (pp. 313-320, cf. Blust 1992.30), the total reaches the more respectable 29. Those in Tsuchida's publication are marked with (T) in Appendix A (100, 131 also occur in the original list): 32, 57, 63, 64, 71, 73, 76, 88, 91, 96, 100, 105, 119, 129, 130, 131.

Inadvertently, Blust added one more by his correction of one, number 9 (now A62) among those sets offered as evidence of the Formosan-Philippine relationship discussed below, raising the final number to 30. [Similarly, but in the opposite direction, I excluded two in the Formosan-only list above (1.8, 1.30) because Philippine cognates were found; the sets consequently belong rather to the Formosan-Philippine collection. They are now Appendix C 150 and 442 respectively. It is of some interest that Blust excluded 1.8 because it might be due to borrowing. He likewise excluded 1.15 because the Tsou word did not systematically correspond, ignoring the fact that the remaining words constituted a satisfactory cognate set; however a Philippine cognate was found and the set is now 39 in Appendix B. Still another (1.28, cf. Blust 1992.29) is Appendix C 353.]

Blust (1992.30) disregards the fact that 'Dyen (1991.92) claims that Formosan is "supported by over 97 sets" on the grounds that 'no more than 20 (6 plus 14) convincing cognate sets have been proposed in print.' The impression is given that he expected the justification of that claim to be of

lower quality than those in my first attempt 25 years ago. He was aware (1992.25) that that work was based on very meager sources; yet the small number presented and the even smaller number of survivals could reasonably be taken to imply a much larger collection would be uncovered. He further notes (ibid.) that these materials have 'been superseded by a wealth of new studies over the past two decades...' and proceeds to list over 20 works dealing with the Formosan languages, nearly all, if not all, of which, among others as well as the results of Tsuchida's fieldwork, Tsuchida and I have been using in our continuing investigation of the Formosan languages. Apparently he has not absorbed all that is in his bibliography. His surprising errors in some of his exclusions among the 1967 list could have been avoided or minimized if he had observed carefully the systematic correspondences in Tsuchida 1976, to which he refers in his discussion, and those in Li 1985 (pp. 261-266), also in his bibliography.

Blust emphasizes the point that in criticising the evidence for a Formosan unity he is dealing only with published material, presumably to show that it is insufficient for its purpose and thus unreliable as an indicator. One can only speculate as to why he did not take the scientific path of testing the hypothesis himself directly and objectively by seeing whether the cognate sets presented in the original list were the only ones available in the publications that he considered (1992.25) 'For our purposes the most important...,' but chose--wisely, as it turns out in view of his difficulties with the Formosan material--rather to play the 'objective' critic unconvinced by what we had published. Perhaps unintentionally, but inescapably, he leaves the false impression that he found that these other important publications contained no information contraindicating his thesis or supporting ours.

This impression is strengthened by his announcement of his planned

dictionary of Austronesian cognates (1992.36); he does not indicate that his materials include Formosan-only sets other than those in my list. In view of his intended publication it is hard to see how he could have missed the many cognate sets limited to Formosan languages in the published materials he refers to. Presumably all Formosan-only cognate sets are regarded due to borrowing--not merely as possibly so (cf. 1992.22)--and therefore ignored. One conclusion seems unavoidable: he felt his Primary Branch thesis was most secure if he restricted the opposing evidence to what was published by Tsuchida and myself and thus evaded considering the available contraindicative material in the mass of publications in his list.

The original list was drawn up before I met Tsuchida, who was already deep into a rigorous study of the Formosan languages. But that is not the only reason that the original list contains errors. It will be some time before all errors are removed in the construction of cognate sets, if it is ever accomplished, since much Austronesian material remains to be collected and investigated. On the other hand that is not a reason to hesitate to indicate trends that appear as these sets accumulate. The trends become clearer as new languages are brought into comparison, old cognate sets are expanded or corrected, and new cognate sets are formed.

That original list is outdated and so to some extent is its discussion. Though its warning was ignored, its conclusions have received much additional support. Much more and better data is now available, due in no small measure to the efforts of Tsuchida and Li and their students as well as to the study of the other publications referred to above. Li (1992.269-276) has been able to draw up a list of well over a hundred 'Formosan-only' cognate sets. It contains many valuable additions, though some of the sets lack an Atayalic member and thus do not meet the requirement of the original list.

It is unfortunate that, despite the size of his collection, Li (1992.276) seems to regard his entire list as likely to be due to intra-Formosan borrowing, even though he presents them as without exception exhibiting only systematic correspondences. Presumably he views the hypothesis of borrowing as necessary to his migrational hypotheses, which bring various languages into Taiwan in succession. These hypotheses are weakened by the absence of a careful association of each of the migratory groups with a clearly inferable point of origin.

The 135 cognate sets in Appendix A that Atayalic now shares with Paiwanic are restricted in the following way. One Atayalic and one Paiwanic member are required. A Saisiyat and/or Pazeh member are/is also admissible on the basis of Li's proposed Northern Formosan. The presented sets are those shared by Paiwanic with Pazic, if the latter is represented by an Atayalic member. This restriction must be kept in mind, for if one wishes to calculate the number of Pazic-Paiwanic sets, those sets must be added that Saisiyat alone shares with Paiwanic (now 45), that Pazeh alone shares with Paiwanic (now 43), and that just Saisiyat and Pazeh share with Paiwanic (now 8). The total is then an additional 96 that increases the number of cognate sets that Pazic shares exclusively with Paiwanic to 231. These numbers are of course only current and provisional and, in this sense at least, only approximate. New cognate sets can be expected to be found that increase the size of collections of the same distribution and new cognates can be expected to be found that change the membership of cognate sets in collections.

All of the sets are believed to satisfy the systematic correspondences of Austronesian cognates thus far constructed with certain exceptions. These exceptions are reasonable and based on experience in applying the compara-

tive method. Not all scholars will necessarily agree in admitting such excep-
tions, just as they do not necessarily agree on the systematic correspondences
themselves. Where it makes sense we admit supplementary hypotheses of
assimilation, dissimilation, metathesis, and analogical change to explain the
fact that one or another phoneme is not anticipated according to the corre-
spondences appropriate to a given word. Such changes are known to occur
with some frequency in linguistic change. No attempt has been made to
discover whether any of the first three types of changes are systematic,
though such systematic changes are known to occur. The assumption is made
that if they are not systematic, the words involved indicate borrowings from
dialects in which they were systematic; dialect borrowings are not invalidated
for subgrouping purposes, since they occur within a language. These excep-
tions concern the method of inference, not infringements on the law of the
regular change of phonemes. All of the exceptional instances are marked as
to type and the phoneme involved. No appeal to sporadic change is made or
admissible. A phoneme otherwise unaccounted for in a comparison for what-
ever reason is marked as inexplicable; the expectation is that it will become
explicable.

Each cognate set has a distribution that is exhaustive as far as is known.
Given the present state of our investigation of the Austronesian languages
and the number of those that remain to be investigated, a cognate set is actu-
ally open-ended.

A collection of cognate sets with the same distribution over languages is
a homomerous collection, or a homomery. Homomerous collections vary in
size. Experience has shown that large collections tend not to dissipate when
they lose sets, partly because of their size, but also because changes of
membership elsewhere bring additions. After some large proportion of the

available cognate sets in a family has been formed, the large homomerous collections are expected to remain large despite any changes in membership, as they tend to do even now.

Although the reconstructons in the appendices are constructed carefully, the emphasis is on the cognate set rather than the reconstruction. The sets provide the evidence for subgrouping. Errors in reconstruction do not invalidate the implication of the set. Similarly the invalidation of a member or the discovery of a new member does not necessarily invalidate the entire set for its collection unless the invalidation or addition changes the set's distribution so that it must be removed from its collection and assigned elsewhere.

A separate cognate set is occasionally constructed when some members exclusively share a meaning even though they are all also included in a larger set with other cognates that have a similar, but distinguishable, meaning. Thus the words for 'head' form a Formosan-only set (cf. Blust 1992.33), but at the same time are cognate with Tagalog and Bikol words for 'skull.' Such included sets are justified on the grounds that the exclusive semantic agreement can validly be regarded as having been limited to a subgroup proto-language, whereas the original meaning-- 'head' or 'skull,' or, less likely, some third meaning--is to be located in a still earlier proto-language. As with other cognate sets no attempt is made to distinguish directly between inheritances and innovations.

The terms 'innovation' and 'borrowing' are commonly used in subgrouping, as I have, as though they were mutually exclusive despite the fact that a borrowing is one type of innovation. When used in subgrouping the term 'innovation' is intended to refer to an internal rather than an intrusive change. The term 'mutation' is used in the following for an 'internal change' mutually exclusive with a 'borrowing.'

What are called cognate sets here are sets that have systematically corresponding phonemes in the sense indicated above. There is no way of knowing, in the absence of records, whether in every instance these sets contain continuations from Proto-Austronesian and are thus prime cognates or are shared mutations or, for that matter, undetectable borrowings or owe their correspondences to convergence. Nevertheless the term 'cognate set' is unambiguous for our purposes in the sense of 'likely cognate set' because, as we see it, these sets, taken as a group, are more likely to contain cognates at some level than to be due to some other factor such as borrowing or convergence.

Like nearly everybody else we consider mutations to be the basis of the inference of a subgroup. However comparatists widely believe that, aside from extremely special cases, the only directly inferable mutations are phonemic mergers and splits. Even these are often not helpful, because many are of the sort that occur independently. Furthermore a merger or split conceivably occurred in the proto-language as part of its dialectalization. In rare cases such changes are homomerous and numerous, and then are determining, as they are in subgrouping the Germanic languages among the Indoeuropean languages. Much more can be said on this subject, but this brief statement will have to do. It is consequently not surprising that Blust's claims to discover mutations, but offers scrappy bits that have not been freed of reasonable competitive or even superior interpretations that make them useless for his aim.

Solid direct evidence of mutations with a higher likelihood of postdating the end of PAN is for all practical purposes non-existent. Vocabulary can yield evidence of the presence of lexical innovations, but this evidence must be indirect. Shared inherited mutations are inferred to be the factor that

makes for large homomerous collections. Given the observation that languages are constantly changing, particularly in lexicon, and that the lexicon of a language is always very large, the exclusive sharing of a large excess of likely cognate sets by a subset of a language family favors the hypothesis that that excess is due to, and thus is evidence of, many shared lexical mutations. It is those included mutations that favor the chances that the sharing languages might also share a more recent common proto-language than any earlier one shared with other members of the family. Our inferences of a subgroup are therefore made when the collections of cognate sets with exactly the same exhaustive distribution--i.e., homomerous cognate sets--are significantly larger than any competing collection involving one of the members of the putative subgroup than can be reasonably explained in any way other than by attri-buting them to a privately shared proto-language. The expectation is that the resultant hypotheses will introduce satisfactory order in the hypotheses regarding the history of the systematic correspondences, the morphology, and the syntax of the Austronesian languages.

The basic assumption is that immediate daughters of the same most recent proto-language will share a significantly larger number of cognate sets with each other than with other related languages. This expectation follows from the premise that they would have been most recently the same language; dialects of the same language share overwhelming amounts of words, obviously when they are mutually intelligible, but even when they are not. The shared cognate sets in the sister languages will include both lexical inheritances from any proto-language prior to their most recent proto-language and any lexical mutations peculiar to their immediate proto-language. The number of inheritances from ancestral proto-languages (i.e. those prior to the last one) tends to decrease with time in collections of exclusively shared

cognate sets and thus the percentage of mutations inherited in common from the last proto-language, though subject to the same attrition, tends to be prominent in such collections. The value of the collections for subgrouping arises from the likelihood that they contain mutations inherited in common. The likelihood that lexical mutations inherited in common are present within a homomerous collection is regarded as roughly proportional to the number of the sets in the collection. The number of such mutations pointing to a subgroup is taken to be roughly proportional to the excess in a homomerous collection.

In the homomeric method a cognate set in a homomery is <u>not</u> regarded as a mutation, though it may be one. Blust's repeated references to our cognate sets as innovations (1992.25 and passim) is an uncomprehending misrepresentation of our intent; the sets consist only of likely cognates. Further research will increase, decrease, or leave standing the likelihood that the sets are truly cognate sets, that their distribution is proper, and that, if their collections are excessively numerous, it is likely that those collections contain mutations inherited in common numerous enough to warrant inferring a subgroup. The claim that mutations are involved in a collection used to argue for a subgrouping is based on its excessive size, not on the claim that each set marks a mutation.

Blust is a proponent of Taiwan as the Austronesian homeland. The basis of this inference is that there are three primary branches of Austronesian on Taiwan: Atayalic, Tsouic, and Paiwanic (1992.22). If this classification and geographical distribution are taken as a basis, any other homeland hypothesis is more complex. Normally any agreement of any pair of these three--exclusive or otherwise--should then be expected to be attributed to Proto-Austronesian, particularly if they have been diverging for many millennia.

Instead he follows a consistent policy of subjecting Formosan-only lexical agreements to borrowing as an alternative hypothesis to taking them as indicating inheritances. The Formosan languages are viewed as close enough geographically to borrow from each other, without the event of borrowing being detectable, but too distant from each other genetically to form a subgroup (cf. Blust 1992.27). Such borrowing hypotheses are thus essential to the general hypothesis that the interrelationships of the Formosan subgroups are of the highest order in the Austronesian family. As Blust puts it, '...no lexical reconstruction could safely be assigned to Proto-Austronesian if its known distribution is confined to the Formosan languages, since the latter have been in close geographical proximity and hence in a potential borrowing relationship for perhaps six millennia.' What is left unclear is how they diverged in the first place if their geographical proximity was an insurance of the likelihood of borrowing.

It seems that Blust inadvertently involved himself in a quandary. He says, 'In later publications (as Blust 1982, and 1983/4a) the Formosan languages were treated for purposes of lexical reconstruction as a single primary branch. This difference of treatment did not reflect a commitment to the hypothesis that the Formosan languages constitute a subgroup.' The statement also does not reflect a commitment to the subgrouping of the Formosan languages into three primary branches. It does suggest an investigator either in search of a freedom to interpret unhindered by the implications of the data or uncertain that there are three primary branches on Formosa.

The borrowing hypotheses he offers as alternative hypotheses concern sets of words whose interrelation is formally indistinguishable from that of inherited words and thus, if borrowed are undetectably so. In the compara-

tive method a lexical comparison is judged to contain inherited words if the phonemes of the words exhibit systematic correspondences with each other and the words are unlikely to result from borrowing. Therefore if only systematic correspondences are required for a 'cognate set,' some sets may have resulted from borrowing; such unidentified borrowings are undetectable. Detectable borrowings are marked by deviations from systematic correspondences, which upon discovery lead to easy identification as borrowings.

The time interval indicated above by Blust, whose basis is not presented, is exaggerated in relation to the interval suggested by the Swadesh list percentages. The latter fit roughly with a period nearer 4 millennia ago for the time when the Formosan languages began to diverge. In any case there is a patent confusion between a potential borrowing relationship and an actual event of borrowing. It is simply not true that if people can borrow words, they do so in large amounts, just because they are on the same island. Taiwan has 13,885 square miles and a fair number of natural obstacles in the form of mountains and rivers. If the Formosan languages differentiated from a single language on Formosa, the different languages must have originated in groups that had become completely separated, so that the island must be viewed as large enough for this to occur. Completely bounded languages--i.e. completely separate languages without mutually intelligible dialects--do not arise from the same original language unless the chain of mutually intelligible speakers is broken. Since a chain of mutual intelligibility is not dependent on geographical contiguity, it can hardly be expected to be broken for some centuries after the groups have come to be separated. It follows that the kind of re-contact necessary for interlinguistic borrowing must have post-dated the separation by an interval long enough for phonological changes among the new distinct languages to make most borrowings, even though

they are from related languages, stand out from among inherited words by their discrepant reflexes and (as a consequence) correspondences.

Although normally borrowings are inferred from discrepancies in the correspondences, it is conceivable that there will be a few instances that escape notice, some because the systematic correspondences themselves present problems, and others because the changes--or the lack of changes--in an inherited word would yield the same shape as a borrowing. However, there is no reason beforehand to believe that the number will be large. An individual set might be due to borrowing or convergence, though many sets will not really lend themselves to such a claim. The requirement that the phonemes must systematically correspond keeps instances of undetected borrowings and convergence, if any, to an inconsequential number. The likelihood is low that undetectable borrowings or convergence, taken together, are a significant factor in any large homomerous collection.

There are even indications that Blust may be preparing to apply such (undetectable) borrowing hypotheses to the Formosan-Philippine-only cognate sets. If done on a large scale, such an extension tends to make the selection between cognates and borrowings the arbitrary decision of the investigator. There is no limit to the range of undetectable borrowing hypotheses, particularly if one disregards the impediments presented by mutual unintelligibility and geography. The only restraint is the adjudged unlikelihood of such hypotheses if a large number is required to explain compared lexemes with systematically corresponding phonemes otherwise associated with inheritance. Borrowing hypotheses where lexemes correspond systematically are best reserved to explain that only one reflex is be found in an inherited form where two or more candidates appear to reflect the same proto-phoneme in the same phonemic environment.

An interesting consequence of Blust's blanket exclusion of Formosan-only cognate sets is that it applies to such sets as show the same correspondences as those shown by Formosan words in sets with a non-Formosan member. Furthermore it is not uncommon for a Formosan-only set to acquire a non-Formosan member and thus change its distribution, as Blust himself shows. It would appear furthermore to follow that a borrowing hypothesis should apply whenever more than one Formosan member is involved in any set. If this application is disallowed, then as the number of Formosan-only sets increases, it becomes proportionally difficult to see how the Formosans could have kept their borrowings restricted to the Formosan-only sets and thus separate from the inheritances that are the evidence for the phonological distinctions found only among the Formosan languages, perhaps the strongest support for the Formosan homeland hypothesis. Finally, it seems hardly reasonable to have a set converted from one invalid for closer relationship into one valid for distant relationship upon the discovery of a new member when the likelihood of being shared inheritances is central to both treatments.

Blust's attempt at a blanket invalidation of Formosan-only sets for inferring PAN lexicon covers a wider area than perhaps he anticipated. If such sets are invalid for inferring PAN lexicon because they might be due to borrowing, they are likewise invalid on the same grounds for inferring a Proto-Formosan unity. It is difficult to see how the sets could be borrowings for Proto-Austronesian without being at the same time being borrowings for the purpose of subgrouping the Formosan languages. Despite this consequence he felt compelled to deal in detail with the evidence of the original list. In this sense a Blust 'NO' is intended to invalidate what had already been invalidated. Perhaps he was aware that such a wholesale exclusion with-

out direct evidence supporting a borrowing hypothesis in each case is not compatible with the comparative method as that method has been formulated and practised for more than a hundred years.

Unlike Li, Blust is generally careful to claim only that (undetectable) borrowing is an alternative hypothesis for the Formosan-only sets and thus to invalidate them (for a second time) as evidence for a closer relationship. The effect is however the same as Li's more forthright view; the evidence of Formosan-only cognate sets is to be regarded as invalid for closer relation-ships. The consequence is that in both cases there is an implicit demand at every turn that proof be offered that a given set is not due to borrowing, a demand that can not be met directly except by showing that it is absurd and thus has no scientific standing. Its implied introduction nevertheless does involve the 'negativism and special considerations' that Blust disclaims (1992.36), but apparently allow him to maintain his view of the Austronesian subgrouping that unfortunately has gathered more adherents than evidence.

The requirement can not be that first the sets supporting a subgrouping must be proved not to be due to borrowing. Rather the justification of the inference that borrowing is not involved lies in the systematic correspon-dences of the sets and the unlikelihood of borrowing as an explanation. This somewhat subtle difference appears in the elementary textbooks on the comparative method (e.g. Meillet 1954.2 ff.) and is precisely the reasoning that supports the use of large homomerous collections in subgrouping. Accepting the demand for proof that systematically corresponding words are not borrowed before using them in a genetic argument simply means giving up the comparative method.

Since there has been so much recourse to hypotheses of borrowing, some remarks on the subject of borrowings are in order. The textbooks that deal

with historical linguistics like Bloomfield's (1933.425-475) and Hockett's (1958.402-407) classify borrowings chiefly into two types. Bloomfield uses the terms 'cultural' and 'intimate' for these types, whereas Hockett prefers the respective purposeful terms 'need-filling' and 'prestige-seeking.' Cultural borrowings are words for novelties that speakers of one language have adopted from another language with whose speakers they are in contact. Intimate borrowings on the other hand are not associated with novelties. They displace words of the borrowing language that originally indicated that particular meaning. English borrowed the words 'mountain,' 'river,' and 'flower' among others from French, not because they signified novelties, but because, to follow Hockett's indications, they tended to give a better impression of the speaker to his interlocutor.

Evidence of intimate borrowing is thus to be sought in the ordinary vocabulary. But intimate borrowing is characteristically associated with a language whose speakers are dominated, either politically or socially, by those of the lending language. The dominance relation is taken to explain not only some Norman French loanwords in English, but also some Germanic loan-words in French and Italian. The assertion of a dominance relation in connec-tion with the alleged Formosan-Formosan loanwords has not been made, with perhaps the exception of Amis loanwords in Kvalan suggested by Tsuchida in our investigation.

Since the Formosan languages are interrelated and like languages, the need to borrow from each other hardly exists, excepting words for novelties. For this reason semantically similar words in Formosan languages that satisfy the requirement of systematic correspondence have an overwhelminghly greater likelihood of being cognate than of being due to borrowing. Recourse to positing intimate borrowings seems unjustified.

There are further considerations. In the absence of direct observation the assertion that a lexical agreement results from a borrowing is a hypothesis of the same type as a hypothesis of cognation. As such it must explain something; that is, it must be justified by evidence. A borrowing hypothesis is not a 'wild card' to be played at will. Bloomfield says (1933.328), 'Every word has its own history.' Other things being equal, a borrowing hypothesis is more complex than one of inheritance precisely because the former requires the act of borrowing over and above the changes to which words of any source are subject. An inherited word is already in the language continuity and does not require a separate hypothesis to explain its presence.

The view that borrowing hypotheses can explain the systematically corresponding words of Formosan languages becomes increasingly less tenable as the accumulation of such sets grows because the increasing size militates against the patchwork required if the implications of the lexical agreements for a genetic interrelation among the Formosan languages are ignored.

The question for the investigator is whether a borrowing, if it occurred, can be determined; the answer requires careful, persistent study, principally because the Formosan systematic correspondences themselves present many difficulties. It is suspicious if one deserts the duty to distinguish between sets likely to be inherited and those likely to be due to borrowing as evidence mounts contraindicating his thesis.

It is not yet clear that the phonological arguments by themselves are strong enough to provide a basis for the hypothesis that Formosan is a primary branch of Austronesian. The possibility exists that the contrasts *t: *C, *s:θ were only dialectally present in the last stage of PAN, even though originally it is likely that these contrasts were universal over PAN territory. Likewise the distribution Proto-Formosan S corresponding to 'MP' h, Ø might

have been dialectal in PAN, though the correspondence reflected the same phoneme at an earlier stage, the division between the different dialects being analogous to the centum-satem division in Proto-Indoeuropean. It would be a little surprising that the Formosan languages were the only retainers of the sibilant articulation. Nevertheless it can be conjectured that there might have been two (or three) steps: 1. *S > Ø in the PAN dialect antecedent to Oceanic at the same time as (or prior to) the same event in the PAN dialect antecedent to East Indonesian and was retained in the dialect antecedent to the western tier of languages which I call Indo-Formosan; 2. *S > *h in the Proto-Indo-Formosan dialect antecedent to Proto-Philippine and Proto-West-Indonesian, but was retained in the Indo-Formosan dialect antecedent to Proto-Formosan. In any case the argument for subgrouping based on *S (> h) > Ø is weak because of the frequency with which this change occurs, as in the history of Iranian, Greek, Puluwat, among others, including Puyuma.

Shared inherited mutations are not sought directly in the homomeric method. Identifying such mutations directly has been successful only in rare cases except mergers and splits. It is not that it is absolutely impossible. The difficulty lies in the fact that elsewhere all alternative competitive hypotheses must be evaluated and eliminated.

A case in point is Blust's hypothesis that *mu[second singular enclitic genitive' is a 'Malayopolynesian' (i.e. extra-Formosan) mutation on the grounds that it is not reflected by Formosan languages. The non-appearance of a cognate in a related language is usually due to its replacement by an analogical change, a mutation. He posits (1977.8) that the Malayopolynesian *mu[is the same as PAN *mu['second person plural enclitic genitive' which has been substituted for the PAN 'second person singular genitive,' presumably *Xu[, by an analogical change. The history he suggests is that the substi-

tution occurred in a 'politeness' replacement of the singular form by a plural form in the ancestor of the 'Malayopolynesian' languages, as in the replacement of 'thou' by 'you' in English. At issue is whether a simpler hypothesis is available to explain the phenomena, for if there is, his selection of the particular alternative mutation that favors his thesis becomes arbitrary and contraindicated since it is the more complex of competitive explanations.

Perhaps the chief difficulty with his hypothesis is the need to explain the 'politeness shift' of *mu[from singular to plural in such a way that plausibly permits PAN *kaXu['second singular nominative' to remain in place, for the latter appears not only in Formosan languages, but also in the non-Formosan languages that have *mu[as the 'singular genitive.' As the first step, Blust constructs the hypothesis that the *ka- of *kamu['second plural nominative' is the same as the *ka- of *kaXu[. Then he identifies *ka- as a plural marker (1977.8) that was extended to the singular *Xu[, 'probably in Pre-Proto-Austronesian' (ibid.), in a 'politeness shift,' but requiring that it, in the form *i-kaXu[(Kan ii-kasu, Tag qi-kaw 'thou [nominative],') be a PAN 'polite' form beside *iXu['id. familiar'. Blust says (ibid.), 'Since the nominative and genitive forms of the 2nd pl. (*kamu : mu) differed just in the phoneme sequence *ka..., it is not at all unlikely that this portion (i.e. *mu - I.D.) was transfered to the singular as a means of distinguishing familiar (*i-Su) and polite (*i-kaSu) uses.'

Needless to say, since *kaXu[must be attributed to PAN, there is no obligation to attribute a 'politeness' element to it there, since no direct evidence for that meaning exists. PAN ka- can at least as well be regarded as a prefix attached to first and second person pronouns, alternating with the *a- attached to the first singular *]a-ku[and the *ki- of *ki-ta[(cf. Ami ki-mi 'first plural exclusive nominative.')

The status of *i-Xu[as a PAN nominative is suspect because it could have been formed secondarily by adding the common pronominal prefix *i- to *Xu[. Precisely because it is a Formosan-only cognate set it is clearly ineligible under his blanket rule for reconstruction as a nominative in PAN. It corresponds formally to Tag qiyo 'second singular oblique' and thus is likely to have a different history from the one needed by his hypothesis; it is more likely to have replaced a form of *ka-Xu[by analogy in a Proto-Formosan. Even if PAN status were granted to *iXu[, his hypothesis would be topsy-turvy because, as Blust recognizes, evidence points to *kaXu[as the older form; in a 'politeness shift' it is the older form that has the 'familiar' meaning and loses its 'polite' aspect to a younger one, not the reverse.

There is however no Austronesian evidence of a meaning difference between *kaXu[and *]iXu[indicated. His hypothesis that there was is thus ad hoc. It is not required by the evidence; it is necessary only to his further conclusions.

He concludes (1977.9), 'There is thus no reason to attribute *mu[in this function (i.e. 'second singular genitive'--I.D.) to Proto-Austronesian...Since *mu[is unambiguously reconstructible as a PAN plural pronoun (matching *kamu[), its use as a singular pronoun must have been an innovation.' But he has failed to show that it is reasonable to suppose that *mu[would have acquired this function without being part of a 'politeness shift' that included *kamu[. The 'politeness' quality attributed ad hoc to *kaXu[was necessary to keep it in place against any encroachment by *kamu[. His hypothesis also requires that *]i-Xu[be PAN, whereas as a 'nominative' it is more easily interpreted as a dialectal replacement of *ka-Xu[, probably in Proto-Formosan.

A simpler scenario can be achieved if the non-Formosan *mu['singular'

is associated with such a form as Kan -musu 'second singular genitive' under a PAN muXu[, regardless of the fact that the latter is not unlikely to be a dialectal form reached by an assimilation (u/i) from a *miXu[(cf. Ami misu 'second singular genitive.') On the other hand the Formosan and extra-Formosan forms leading to a PAN mu['second plural genitive' can be associated as rapid speech forms with those which variously suggest a PAN meyu (with assimilation Sai, Sambal muyu), miu (A: i/e and then with metathesis: muy), miw (Dyen 1974.24 f.). This hypothesis would explain at the same time the interspersed distribution of *kamiw (and associated forms) and *kamu[and the fact that some Oceanic languages retain *mu['plural.' Under this hypothesis the relatively complicated 'politeness shifts' become unnecessary, since the two *mu[forms now have come to be alike by convergence.

A different, even simpler hypothesis was available from the beginning, namely that *mu['singular' was PAN and was replaced in Proto-Formosan by *Xu[, originally associated with oblique functions.. This hypothesis alone is preferable to the involved hypothetical changes based on ad hoc meanings needed to bolster the Primary Branch hypothesis. It follows too that, if his hypothetical mutation is necessary to the Primary Branch hypothesis (1992.38) and it is a mutation only if the Primary Branch hypothesis is posited, a vicious circularity has been achieved.

The distribution of *mu['singular' does not coincide with that of the phonological mergers (i.e., *t : *C among others), which do not coincide with each other, and thus robs these arguments of reciprocal support. It does not even coincide with the distribution *S > h > Ø since *S > Ø appears in Puyuma, thus required to be a change independent of the one attributed to 'Malayopolynesian.'

In this connection Blust's claim that I misrepresented his classification (1992.38) is carelessly stated. I simply presented my evaluation of the basis of his classification and I have seen no reason to change it.

In connection with the list of Formosan-Philippine cognate sets involving words with meanings associated with body parts Blust suggests additions to 6 cognate sets (6, 7, 20, 21, 25, 26) from outside languages that remove the sets from their assigned collection; these are welcome positive contributions. To these can now be added two more (4, 8). It should be added that only one (6) has been found thus far to have an Oceanic member. One set (9) was corrected as a result of Blust's comment (see above) and moved to Appendix A. One (11) has had to be set aside as due to an error. The remaining 16 sets of the original 26 remain satisfactory for the purpose, though, to be sure, some require unessential modification.

The homomeric Formosan-Philippine sets reconstructed with an initial before k are listed in Appendix B. Those remaining from the 'body part' list are marked with their number in my original list. Those remaining that have an initial k or thereafter are listed in Appendix C after their number in our current collection and are also marked with their number in the original article. The following is an index to the sets that have persisted and are list-ed in those two appendices: 1 - B68; 2 - B69; 3 - B104; 5 - C143; 10 - C246; 12 - C252; 13 - C264; 14 - C265; 15 - C266; 16 -C267; 17 - C271; 18 - C273; 19 - C281; 22 - C331; 23 - C353; 24 -C451.

Here and there Blust is led into ill-considered rejections because of an inability or unwillingness to deal with the different transcription that I follow even though those differences have been published. In addition his treatment of some of the Formosan-Philippine cognate sets (1992.33 ff.) is conspicuously wide of the mark. He goes too far in suggesting on one hand

(1) that it is not legitimate to associate a word meaning 'head' with a word meaning 'skull' because the meanings are disparate (the additional meaning 'skeleton' of the Bikol cognate is irrelevant as is his complaint about the correspondence), and on the other hand (2) that the association of a set of words with the same base meaning 'wing' is tainted in the presence of a formally identical base with meanings like 'sound of patting or tapping'. In the first instance the requirement is a high degree of semantic similarity, whereas in the second, in an obvious contradiction, the association of sets with remotely similar meanings, if any, is required. Furthermore he claims (3) a set must be regarded as 'qualified,' whatever that means beyond its nega- tive aura, because 'similar (but non-cognate forms) occur outside the proposed subgroup'; it is difficult to associate such a statement with a prac- tising comparatist, but it is also hard to distinguish it from the hypotheses of sporadic change which appear passim in his publications, though not, to his credit, explicitly in the one presented here. He regards as similarly 'quali- fied' (4) comparisons whose Philippine member is in a North Philippine language, presumably because borrowing might be involved. That possibility can not be ruled out, but the necessary contact for borrowing can not be regarded as easily and frequently accomplished over the body of water that separates Taiwan from the Philippines before contact with the west and only slightly eased thereafter, and so must regarded as in principle unlikely.

My list was specified to be limited in scope, but dealt with body parts, recognized by many, if not most, comparatists as indicators of relationship since they are not likely to have been borrowed. Blust furthermore treats the list as though it were the sum and substance of the Formosan-Philippine sets available even though he admits he has found 9 with either i or w as their initial (1992.45); not all of them could be identical with those in my list.

Furthermore those 9 taken at face value should have been seen to imply a total many times larger for his entire collection. This could hardly be regarded as 'the groundwork for a more objective (emphasis mine - I.D.) indication of the relative sizes of homomeric corpora linking Philippine languages with Formosan languages.' (ibid.).

The Formosan-Philippine collection, which I have often heretofore referred to as having 400 entries, now has 475, though that number should be regarded as subject to variation. As an arbitrary sample of these cognate sets I have listed the first hundred or so (ending before initial k) in Appendix B in the alphabetical order that we have adopted, trusting that it will be realized that the total is about four times as large. Although reasonable efforts have been made to take outside cognates into account, no guarantee can be offered that such will not be found, nor, on the other hand, that additions to the collection will not be discovered as our work progresses.

Let us however be clear. Our view is that the Formosan languages are an Austronesian subgroup whose immediate relationship is with the western languages as a member of the Hesperonesian subgroup. The evidence at hand points to the Philippine subgroup, whatever its membership turns out to be in detail, as its closest relative. It is not unlikely that the Philippine subgroup on the other hand is just as closely related, as far as we can tell, to the West Indonesian languages as it is to the Formosan languages and that the three subgroups constitute a chain. A careful calculation of the number of cognate sets restricted to the Philippine and West Indonesian languages has not been made; they appear to outnumber by far the Formosan-Philippine sets, but the application of the cohesion index (described elsewhere) and consideration of the different magnitudes of the sources available for the various languages can be expected to revise their value severely downward.

The Formosan languages now share 145 cognate sets with, and only with, both the Philippine and West Indonesian languages. The Formosan languages now share 81 cognate sets with only West Indonesian languages. The latter number fits with the suggestion that a chain is involved. For if there were a subgroup with just the Philippine and the West Indonesian languages as members, the just Formosan-Philippine cognate set collection and the just Formosan-West Indonesian cognate set collection would be expected to be approximately the same.

The Formosan languages now share 167 cognate sets with East Indonesian languages and additional membership from a Philippine and/or a West Indonesian language along with a few (now 11) that Formosan languages share with just East Indonesian languages. These sets lack an Oceanic member, as far as we know, as do the other sets referred to above, so that, including the sets listed above, there are now 879 cognate sets all told pointing to the membership of the Formosan languages in Hesperonesian, the western subgroup of Austronesian. The number of these sets militates against the subgrouping that aligns the extra-Formosan languages as one subgroup of Austronesian against the Formosan languages, however the latter are subgrouped.

Blust contrasts his procedure (1992.47) as 'the comparative method of linguistics' as against the homomeric method, which is thereby implied not to be. This merely calls attention to his failure to comprehend the nature of the homomeric method and the linguistic comparative method. The comparative method in linguistics is the procedure of drawing inferences from the systematic correspondences of languages that concern their relationship, their subrelationship, and their past history, including that of borrowings where they become involved. The homomeric method is a scientific application of

the comparative method as it has been applied for over a century, combined with mathematical, thus far mainly statistical and probabilistic, considerations.

Blust used what is essentially the homomeric method in constructing exclusively shared cognate sets that grouped a certain few East Indonesian languages with the Oceanic languages rather than with other East Indonesian languages and thus with Hesperonesian. He carried out the first step of collecting homomerous cognate sets rather commendably. Unfortunately he failed to take the next important and critical step of comparing his collection with the markedly larger collection that associates those same East Indonesian languages with Hesperonesian.

The fault here, as in the case of the PAN 'second person genitives,' was failing to consider and deal with competitive hypotheses which might explain the same set of observations more simply or at least as simply. Unless competitive hypotheses are excluded, a hypothesis remains at best a mere speculation and thus a treacherous basis for further reasoning. One of the important tasks of science is to evaluate competitive speculations to achieve an inference for the past and the best prediction for the future.

Appendices

In the following lists certain conventions are followed. If a phoneme is not implied by the reconstruction, the type of change believed to explain its appearance is indicated in the following ways: A = assimilation, D =dissimilation, M = metathesis; analogical change is written out. The unanticipated phoneme is first indicated followed by a slash and then either the anticipated phoneme itself or the reconstructed phoneme, marked by an asterisk, which implies a different outcome.

When different reconstructions are equally satisfactory, the alternatives are given separated by a single slash; for the sake of clarity a double slash is used if one of the alternatives contains an infix, which is marked by two single slashes. If the choice between two or more phonemes can not be made, both or all are given in sequence between brackets. If Ø is an alternative reconstruction to a phoneme, usually to *?, then brackets are used:] in initial position, [in final position, and [] in medial position.

A cognate in parentheses is indicated to be written non-phonemically. The other citations of cognates should be regarded as our best approximation to the proper sequence of phonemes; the degree of accuracy in the approximation is affected by the way it is recorded in the source.

In an Atayalic entry with a distortion attributable to "men's speech," arrowheads are used to segregate the added element. The arrowhead chosen points toward the cognate element. A suffixed element is preceded by < and an infixed element is surrounded by < and >. A prefixed element has not yet been identified.

When appropriate, an abbreviation in brackets is appended to the abbreviation for a language name to indicate the source of the form. The following are the abbreviations so used with their meaning:

[Br] - as reported in W. H. Brown 1951-6.

[Dl] - non-standard dialect.

[Fer] - as reported in Ferrell 1969.

[JCK] - as reported in J. C-H. Kuo 1986.

[JHH] - as reported in H-H. Jeng 1971.

[N] - as reported in Y. Nihira 1983.

[O] - as reported by N. Ogawa.

[OA] - as reported in N.Ogawa and E. Asai 1935.

[PLi] - as reported by P. J-K. Li.

[SE] - as reported by S. Egerod.

[ST] - as reported by S. Tsuchida.

[TTh] - as reported in T'-H. Tung 1964.

Appendix A

1. *]apa[, SedTk ?apa? 'load, to carry', Sed m-apa?, Tso ma-afo, Tha m-apa? 'to carry on one's back', Paz m-apa?, AmiSk mu-apa? 'to carry a child on one's back'.

2. *]a[S_{136}X]?i[, AtySq s-?asiy, Aty[ST] t-?asiy, BunS m-as?i? 'to cough'.

3. *]atuk, Aty -atuk, Sar m-aaki-atuku? 'to peck', Aty ?a-?atuk 'beak'.

4. *bajiw, Aty bai?, Pai vadiw, Puy padiw (p/b by analogy), Ami vaniw, Kvl baniw 'a type of white mushroom'.

5. *b/aR/ija[, AtyMx bagisa?, SedTn bgiya?, Kvl brina 'reed of the loom', Pai vayda 'blade of the loom', Hlg bali:la 'reed of the loom', Mal bəlira [D: l/r] 'weaver's sword'.

6. *ba[S_{136}X]ay, AtyMay basay 'a vine used for binding', RukMg bsée 'a type of vine', Tso fsoi 'a type of vine with edible roots', BunN, BunC, BunS basað 'a type of vine, Puereria hirsuta Matsum.'.

7. *baw$_2$a[, SedTk bawa?, AmF Li-fawa?, Tha lum-fa:wa? 'to swell', RukMn u-vava 'to swell (as dough)', SaiTa maL-bawa? 'a swelling'.

8. *baw$_3$a[, SedTk bawa? 'ferment, to froth', Pai vawa 'millet beer, alcoholic beverage', RukTa, RukTo bavá, RukBd báva, RukMg bvãa, RukMn va-vaa 'wine, alcohol'.

9. (1.5) *bilebil, SedTd, SedTrk -blbil, SedTkd -blebil, RukTa bilibíli, RukBd, RukTo bilibili, RukMg blibli, RukMn -vilivili 'to pull', Pai -vilvil 'to

pull away from (as one's hand)'.

10. (1.27) *bi[rR]ewa[, Sed bruwa? (A: u/*e), SaiTa biLwa?, BunN, BunC bilva, BunS bilva?, Bun[JHH] baliva (A: *a/e and then M: a-i/i-a) 'thunder'.

11. *bukaw, Aty[Dl] bukaw, AmiN, AmiT, AmiV vukaw 'tree fern, Cyathea spp.', Pai mukav (analogical m/b) 'a type of fern, bracken'.

12. *buŋan//*b/aL/uŋan, Aty sin-buaŋan, Sed sin-braŋan (both Atc have M: u-a/a-u), BunN, BunC, BunS buŋan, b/a/uŋan, Kvl snu-buŋan 'spear'.

13. *buqu[/*puqu[, Aty bq<ni?, AtySq q<ni? (both Aty have an analogical loss of u), Tha pu:qu? 'bone'.

14. *buR₅ay, Aty buay, SaiTa buLay 'fruit', PaiQ bua-buay, Kvl muLay (m/b by analogy) 'flower'. Aty's loss of *R is one of a small number of such instances.

15. *buRia[, Sed bgia? 'a type of wasp', SaiTa buLia? 'hornet, wasp', Kan vuria?, Sar vuria? 'ground-nesting wasp', Tha fuliya? 'a type of wasp'.

16. *butul, AtyMx butul, Ami[Dl] vutul 'masked palm civet, Paguma larvata', Sai butul 'big fox', Pai vučul 'flesh, lean meat', Puy vutul 'wildcat', RukTa, RukTo butūlu, RukBd būtulu, RukMg btūlu 'pork', Kan vutūnu?, Sar vutulu? 'deer'.

17. *buyuQ₂, Sed b-buyu? 'forest, woods', AmiSk buyuq 'mountain'.

18. (1.2) *D₁₂₄aH₁u[, AtyTk[OA] rahu?, Sai[OA] ra-raho 'big', Paz dahu? 'many', RukTa ma-Dau, RukBd ma-Dau, RukMg ma-Doo, RukTo ma-Dau, RukMn ma-dau 'big, many'.

19. *D₄akeS₁₆, AtyMx rakus, Aty rk<n>us, Aty[SE] k<n>us, Sai rakeš, Paz dakes, Pai Dakes, RukTa, RukBd, RukTo akése, RukMg Dkə́sə, RukMn dake?e, Kan cakése?, Tso c?osə, Ami rakes, BunN, BunC, BunS dakus, Tha ša:

kiš (A: š/s), Kvl zaqes 'camphor laurel, Cinnamomum camphora'.

20. (1.3) *[dZ]amu[qQ], Aty[SE] ramu<?, Sai ramu?, Paz damu?, Pai jamuq 'blood', Kvl m-zamu 'having menses'. The final consonant of Puy damuk 'blood' can be explained as having originated in a borrowing from Paiwan by a dialect in which *q > Puy H had already occurred. Cf. PLi 1992.269, where Aty ramuux (analyzed as ramu<ux) 'id.' is also cited,

21. *daŋaR, AtySq raŋa<y, AtyMx raŋa 'trap', Sed daŋar, BunC, BunS da-ŋal 'dead-fall trap'.

22. *D₁apa[, Aty m-rapa? 'shallow (of a plate)', Tso po-rapo 'shallow'.

23. *D₂apa[, AtyC ra-rapa? 'water buffalo', Sed dapa?, AmiN ra-rapa? 'cow'.

24. *[D₄Z]aRa[, AtyMxw raga?, Sed dara?, SaiTa raLa?, Paz daxa?, BunN, BunC, BunS dala? 'maple tree, Liquidambar formosana Hance'.

25. *D₂aRiŋ, Sed -s-dariŋ, RukTa, RukBd daaŋe, RukMn -da?aŋe (all Ruk have A: a/i), Sar tara-a-sariŋi?, Tso troe-creŋi, Ami mi-LaLiŋ (A: L/r), BunS da-daliŋ 'to moan'.

26. *D₁₄aulu[, Aty rawlu? 'an inedible mountain turtle', Bun daulu? 'tortoise'.

27. *D₁₄emaŋ/*D₁₄ameŋ, Sed ramaŋ (A: a/e), Pai Da-Demeŋ (A: e/a), Puy za-zemeŋ (A: e/a) 'quail'.

28. *gijiR₁₂, Aty kiiy, Pai gidi, Kvl kiniL 'side, flank'.

29. *H₁awH₁aw, Aty h/m/awhaw, Ami mi-hawhaw 'to whisper'.

30. *]ini[, Aty, Sed ?ini?, Paz ini? 'not', Sai ka-yni? 'don't', Pai ini 'not, no', Puy ini-yan 'not exist', RukTa, RukMg ini, RukBd ini, RukTo ?a-ini 'no'.

31. *]ima[, Aty, Sed ?ima?, Paz ima?, Pai ima 'who'.

32. (T) *-imaH₁, Sed m-imah, Kan mi-ima?, Sar m-i-ama? (< *m-ima? by

analogical wrong analysis, cf. Tsuchida 1976.89), Tso m-imo 'to drink'.

33. *ka[, Sed ka 'general nominative particle', Sai ka 'nominative and oblique particle', RukTa ka ~ a 'common nominative particle'.

34. *kaD₄uS₂, Aty k/m/aruh, Ami mi-karus 'to scratch'.

35. *kaRuS₃, Sed k/m/agux, Bun ma-kalus 'to scratch'.

36. *kaCu[/*ka-kaCu[, AtySk kkasu?, Paz ka-kasu?, Ami, Bun kakatu?, BunN, BunC katu-katu, BunS katu-katu? 'spider'. PLi (1992.275) also cites Tso kakatu, Tha kakatu? 'id.', both of which are borrowings, perhaps from Bunun.

37. *keRiw, Aty[SE] kgiy, SaiTa ka-kLiw, Paz kixiw (A: i/e), Puy keriw, Tha kliw, Kvl qriw 'hemp, ramie', Ami keLiw 'string, thread, ramie'.

38. *ki[, Aty ki?, Sai, Bun[Dl] ki 'and'. Cf. PLi 1992.271, who adds Paz ki 'because' to the set.

39. *kiLet/*kiteL, Aty k/m/iyuc (kyut-an), Ami mi-kiteL 'to pick (fruit)'.

40. *kiS₁₃₆ekiS₁₃₆, AtySq k-kskis 'a file', Ami mi-kiskis 'to *crape off dirt with a knife', AmC kiskis 'to scrape off the surface, strip', Tha kiškiš 'to shave, scrape'.

41. *kuluban/*kulaban, Aty kluban 'brass', Sai kuluban, Kan kunavane? 'copper'.

42. *kuCun, SedTn kusun, BunN[N], BunC[N], BunS[N] kutun 'tomorrow'.

43. *LapiC, AtySq yapit, Sed rapic, SaiTa Lapis 'flying squirrel', Pai[Dl] La-Lacip (M: c-p/p-c), Puy LapiT, La-LapiT, Sar ?a-Lapic-a? 'bat'.

44. *Lilep, Aty zizup, Aty[SE] ziup (both Aty have A: *L/l), Ami Lilep 'hawk'.

45. *L₂iluk, Aty[Dl] ziluk, Kvl rinuk 'berry', Aty iluk, Sai, RukTa Liluk,

Kvl rinuk 'plant sp., Rubus parvifolius Linn., Rubus taiwanianus Matsun'.

46. (1.14) *Lituk/*Lutuk, Aty[Dl] rutuk, SaiTa Lutuk, Pai Lučuk, RukBd Lutuku, RukTo utūku, Kan Litūk-aʔ, Sar Lituk-aʔ, Tso[PLi] rtuk-a, BunS ʔa-utuk, Sir (routock) 'hare, rabbit'.

47. *Lukelaw, Aty[SE] kl<iʔ, SedTd rkl<ic, SaiTa Luklaw, SaiTu eklaw (A: e/u), Kan ukunauʔ, Sar Lukuluʔ, Tso rʔuho (all Tsc have A: u/e), Ami Luklaw, BunN, BunC huknav, BunS ʔuknav, Tha rukðaw, Kvl Luqnaw 'leopard'.

48. (1.4) *Luluŋ, AtySq yuluŋ, SedTd, Paz[ST] ruluŋ, RukMg[OA] (k-roroŋo), Sar luuluŋ-a (A: l/L) 'cloud'. PLi (1992.270) records Paz ruruŋ 'id.', which suggests an assimilation (r/l).

49. *LuneR₂, SedTr runug, Ami Lunen (A: n/L), BunN, BunC hunul, BunS ʔunul 'earthquake'.

50. *lalai[, AtyMx lalaiʔ 'a type of cicada', AtySq lleʔ Paz lalaiʔ, Pai[Dl] lalai 'cicada', Sai lalaiʔ 'a type of cicada'. A Bun lalai 'a type of cicada' (PLi p.c.) is evidently a borrowing.

51. *lawi[, Aty lawiʔ, Pai lavi 'carp'.

52. *layaD₂, Aty layaʔ, Sed layac, Sai layar, Pai, Puy layaz, RukMg lrāDɨ, RukTo laāDɨ, RukMn laLade, Kan naLāceʔ, Tso hzocə, BunN, BunC, BunS naðaʔ, Kvl layas 'a shrub, Sambucus formosana, Ebulus formosana'.

53. *lebeʔ, Sed l-lebuʔ, RukTa ma-lébe, RukBd aa-lébe, RukMg ma-lɨbɨ, RukTo maʔa-lébe, RukMn mu-leve, Pap (malbu) 'low'.

54. *leH₁ib, AtyMx luxi<ux (A: x/h, A:u/e), Sed lh<ŋaw (A: e/i with subsequent loss), Sai ka-lhib 'a cave under a rock', Pai liv (A: i/e with contraction), AmC lihif (A: i/e), Kvl lihib (A: i/e) 'cave', PaiW liv (A: i/e with contraction) 'shelter under an overhanging rock', RukBd libi (A: i/e with contraction) 'to place under a rock for shelter, space betwen sofa and

the floor'.

55. *luber, AtyMx lubug, SedTr lubug, Pai la-luver-an, Puy luver, RukTa lebére (A: e/u), RukBd lebere (A: e/u), RukMg lbɨrɨ, RukMn luvere 'jew's harp'.

56. *luŋuluŋ/lu(u)luŋ, Aty[SE] ll<i>uŋ ('river', Sai luluŋ-an, Kan nuŋúnuŋu?, Sar luuluŋu? 'creek'.

57. (T) *lu[q₁₄Q₂]eS₁₃, AtyMxw luqus, Sed luq<i?, Sai lo?eš, Pai luqes, RukMg ulūsu (M: *q-l/l-q with loss of *q; A: u/e), Kan ne?ese? (A: e/u), Sar lee?e? (A: e/u; M: *e?/?e), Tso həsə, Ami luqes, BunN, BunC nuqus, BunS nuhus 'bone marrow', RukTa, RukBd, RukTo lūusu (A: u/e), RukMg ulūsu, RukMn lu?u 'brain'.

58. *najip, Aty p-naip, Pai ma-nadip, RukMg -ngépe 'to angle', Pai pa-nadip 'fish-hook', Bun ka-naðip (< -*nayip < *-naip) 'fish-hook'.

59. *maDa[S₁₆X], Aty maras, Sed, Bun madas, Sai maras, Kvl mazaš 'to bring'. Cf. PLi 1992.271 who suggests the reconstruction *ma[dD]aS.

60. *maqaw/*paqaw, Aty maqaw, Sai ma?aw, Bun maqav, Tha maqaw 'a tree with small, edible, spicy fruit, Litsea cubeba', Pai maqaw 'an alder, Alnus japonica', Puy maHaw, RukMn amau (M: *q-m/m-q), Ami p/in/aqaw 'black alder, Alnus formosana'. Cf. also PLi 1992.274.

61. *mas?al, Sed maxal 'ten', Kan máane?, Sar maale?, Tso maskə, BunN mac?an, BunC, BunS mas?an 'ten (non-human)'.

62. *mujiŋ Sed muhiŋ, Paz muziŋ, Tha mu:ðin 'nose', Pai, Puy mudi?an 'face', BunS miuŋ (M: iu/ui) 'eyebrow'. Cf. Blust 1992.34.

63. (T) *paQ₂eju[, Sed paxu<ŋ, Sai pä?zu?, RukBd pagu, RukMg págo, RukTo pagu, RukMn pahu, Tso pau, BunN, BunC paqau (A: a/*e) 'bile, gall bladder'. This set is regarded as associated with PAN Q₂apeju[, Pai qapdu 'bile, gall bladder' under a hypothesis of a metathesis (*p-Q₂/Q₂-p) that

arose in a Proto-Formosan dialect.

64. (T) *pat$_1$aS$_1$, Aty pataš, Sai patas 'tattoo', Sed patas 'letter', RukTa ?acase, RukBd pacáse 'to write, draw', RukMg ptásɨ, RukMn -paca?e 'to embroider', Kan tapáse?, Sar taa-tapa-a?, Tso ta-tpos-a (all Tsc have M: t-p/ p-t) 'pattern, design', Tso t/m/opsə 'to write', tposə 'book', BunN, BunC, BunS patas-an 'letter', ma-patas 'to write', Tha pata:š-an 'book', Fav (atas) 'letter', (pattas) 'to write'.

65. *pesek/*puResek, Aty, Sed puhuk (A: Q/e), Pai petek, PaiTb putek (< *puRsek), Puy persek (A: e/u), RukMg lŋu-psɨkɨ, Tso prəs?ə, BunN pulcuk, BunS pusluk-an (M: sl/ls), AmiSk, AmiN purcek, Kvl puseq (s/*Rs) 'an herb sp., Oxalis corniculata'.

66. *pika[, Aty ma-pika?, Sed m-pika?, Tso pi?o 'lame'. Cf. PLi 1992.273.

67. *puD$_4$ul, Aty p/m/urur (A: r/l), Ami mi-purul 'to wind (thread) on a reel', Pai puzul 'cocoon, spool, stick on which cord is wound'.

68. *pula[S$_{16}$Xx$_2$], Aty, Sed, Kvl pulaS 'dandruff', Sai pulaš 'dandruff, inflamed and itchy body', Ami pulas 'dandruff, scabies'. Cf. PLi 1992.272.

69. (1.10) *puS$_6$al, Aty[SE] pusal, m-pusal 'twice, twenty', Sed m-pusal, pusal, RukTa ma-?usále, RukBd, RukTo ma-púsale, RukMg ma-púslə, RukMn ma-ma-pu?ale, Kan ma-pusane?, Sar ma-puale?, Tso ma-pusku (A: u/ə), BunN, BunC, BunS ma-pusan, Tha ma-pusað 'twenty'.

70. *putuŋ, Aty, Sed, BunN putuŋ 'tinder, Gnaphalium multiceps (an herb with wooly hairs used as tinder)'.

71. (T) *q$_2$ajiS$_6$, AtyMxm qais, Sai ?äzis (partial A: s/š), RukTa, RukBd s-agi-agisi, RukMg agisnə, RukTo agisi, RukMn adi?-ae, Kan ?ais-ane?, Tso es-a, BunN, BunC qais, BunS hais 'boundary', Sar ais-a? 'between, middle'.

72. *[q$_{1234}$Q$_2$]a[lN]ian, Aty[SE] qlian, BunN[N], BunC[N] qanian 'noon,

daytime', BunS hanian 'noon, day', Kan ca-aʔaniaʔ 'daytime'. Cf. PLi 1992.272.

73. (T) (a) *[qQ$_2$]ariD$_{14}$aŋ or (b) *[qQ$_2$]aRiD$_{14}$aŋ: Paz xaidaŋ (M: *[rR]-[qQ$_2$]/*[qQ$_2$]-[rR]) 'legume' Kan ʔaricaŋeʔ, Sar arisaŋe, Tso recŋi, BunS halidaŋ 'pigeon pea, Cajanus cajan', Bun[JHH] qalidaŋ 'bean' are ambiguous in being associable under (a) with Aty[Dl] qagiraŋ 'beans, peas', Aty qayraŋ 'legume', Sir (agisaŋ) 'broad beans' and under (b) with Pai qarizaŋ 'string beans'. Cf. also PLi 1992.271.

74. *qaReluʔ[, AtyMx qagluʔ, AtySq qoluʔ, Sed gluʔ, AmiN qarduʔ 'a reed, Phragmites communis'.

75. *Q$_2$aCu[, Aty[SE] qasuʔ, Sed, Paz ʔasuʔ, BunN, BunC qatuʔ, BunS hatuʔ 'boat'.

76. (T) *Q$_2$eLuD$_2$, SedTd ʔəruc, Sed ʔruc, SaiTa kä-ʔLur, Paz ʔurut (A: u/e), RukMn ʔuLudu, Kan ʔuũcuʔ, Sar ʔuLusuʔ (all Ruk and Tsc have A: u/e), Tha qru:s 'pillar'. Cf. also PLi 1992.274.

77. *Q$_{12}$eLuR, SedSka, SedRa ʔlug 'road', RukTa ka-ele-a (A: e/u) 'path', Pai qelu 'animal trail'.

78. *Q$_2$emu[, Aty qumu<liʔ, qumuʔ (A: u/e), SedTd ʔemuʔ 'rice-cake', AtySq qmuʔ 'pieces of rice or millet broken in pounding', Pai -qemu 'to pound into flour', qemu, Puy Hemuʔ, Kan ʔũumuʔ 'flour', Kan ʔũumuʔ 'powder', tará-ʔumuʔ, Tso t-umu (all Tsc have A: u/e) 'to pulverize', Ami qemuʔ 'cake of rice or millet'. Cf. PLi 1992.271.

79. *Q$_2$epu[, SedTk p-sopuʔ (< *-sa-upuʔ), Pai -qepu, Ami ma-sa-qupuʔ (A: u/e) 'to gather, collect'.

80. (1.23) *Q$_2$iD$_5$a[S$_{16}$Xx$_2$], Sed ʔidas, Sai ʔilaš, Paz ʔilas, Pai qilas, Fav (idass) 'moon'.

81. *[q$_{123}$Q$_2$]iluŋ, AtySq t-qiluŋ, Pai ra-quluŋ, Puy Huluŋ (both Puyumic

have A: u/i) 'an aromatic shrub, Artemisia capillaris'.

82. *$q_1uD_1aS_1$, AtyMx quras, SedTd qudas, Pai qudas, RukTa, RukBd, RukTo uDásk, RukMg uDásɨ, RukMn uda?e, Kan ?usáse? (A: s/c < *D_1), Sar ?usae?, BunN, BunC qudas, BunS hudas, Fav (oeraas, oras) 'gray hair'. PLi (1977.54) recorded RukTo ?uDase, RukMn ?uda?e 'id.' whose initial ? is difficult to explain.

83. *qumeRi[, AtySq qgmi? (M: g-m/m-g), Sed qmegi?, Puy Humri? 'a creeper whose tubers yield a soap substitute', Kvl qa-mri (< *ka-qumeRi[with A: a/u and contraction a/aa) 'soapberry, Sapindus mukurossi'.

84. *Q_2umuR, SedTk (umul) 'to suck up and keep in the mouth', Pai qumu 'water held in the mouth', -qumu 'to hold water in the closed mouth', Puy -Humur 'to cram the mouth with food or water, to suck on in the mouth'.

85. *$quCux_1$, Aty -qusu? 'trunk, stem', PaiQ qucus 'stalk, stem'.

86. (1.6) *$Q_2uya[S_{16}Xx_2]$, AtySq m-quas, Sed ?uyas, BunS ka-huðas 'to sing', Tha qu:iš (A: i/a) 'song'.

87. *$RabaR_5$, AtyMs ?abag-an, AtySq bag-an, AtyMx gabag-an, SaiTa LabaL-an, BunN, BunC, BunS ta-labal 'summer', Paz rabax-an 'south'.

88. (T) *ReH_1ap, AtySq ghap, AtyMx ga-ghap, SedTd rəhak 'seed for sowing', AtyMx g/um/hap, Kan r/um/a-ráape?, Sar taru-a-rape?, Tso mi-rapo (rap-i) (all preceding Tsc have A: a/e with contraction in some cases) 'to sow', Sar erape? (M: *H_1-*R_6/*R_6-*H_1), RukTa eá?e, RukBd eápe, RukMn ?epe 'grain, seed', Sir (gaap) 'seed'.

89. *$Rila[S_{136}X]$, AtyMx gila<quŋ, Sed gla<quŋ, BunN, BunC, BunS linas, Tha liðaš-al 'pheasant', SaiTa Lilaš-an, Paz xilas-an 'male pheasant'. Cf. PLi 1992.274.

90. *Ri?ri?, Aty g-giri?, Kan ta-riri?, Sar ta-riri?, Tso riri, BunN, BunC,

BunS li?li? 'bracken, an edible fern sp.'.

91. (T) *Ritu[, Aty gitu?, SaiTa Litu?, Kan riitu?, Sar ritu?, Tso rtuu, BunN, BunC litu, BunS litu? 'loquat', Pai iču 'fruit of the kaLujil tree, Eriobotrya deflexa'.

92. *Rubu[(1), AtySq ?ubu? 'nest', PaiW uv-uvu(< *uvu-uvu), Paz xubu? 'nest'.

93. *Rubu[(2), Aty[Dl] ma-?ubu?, Paz, Pai ka-uvu 'entangled', Puy ma-ruvu-ruvu 'matted (of hair)'.

94. *[sθ]aŋel/[sθ]aŋal, Aty m-haŋal, Sai haŋal, Puy -saŋal, RukMn -aŋele 'to carry on one's shoulder'.

95. *θebu[, Aty hbu?, Sed m-hbu<raw, AmiSk, AmiN cebu?, AmiT, AmiV cevu?, Kvl sebu 'lower abdomen', Kan sũubu? (A: u/e) 'the part below the navel (of boars)', AmC cefu? 'bladder'.

96. (T) *S₆anaq, Aty snaq-uy, Pai sanaq, RukBd sána, RukMg snãa, RukTo sanã, RukMn ?ana, Kan sanã?e?, Sar sana?e?, Tso snoo, Ami sanaq, Kvl sani, Fav (channa) 'otter, Lutra lutra', Sir (hanna) 'fox'.

97. *S₆amaq₁, Sed sama?, Paz sama-binayu? (binayu? 'mountain'), PaiQ samaq, Puy amaH, RukTa, RukTo samã, RukBd sáma, RukMg smãa, RukMn ?ama, Kan samã?e?, Sar sama?e?, BunN, BunC samaq, BunS samah, Tha ša: maq, Kvl sami 'an edible herb, Lactuca indica', AmC samaq 'a tender edible grass resembling the dandelion'.

98. *[S₁₆Xx₂]aŋay, Sed m-saŋay, Kan taru-saŋãi? 'to rest'.

99. *[S₁₃₆X]aRuyu?, AtySq sguyu?, BunN, BunC, BunS saluðu? 'basket fish-trap'.

100. (1.26, T) *S₂eD₅a?//S₂/uR/eD₅a?, Aty hla<qiy, Sed huda? (A: u/e with contraction), Sai hä-holä? (l/*Ll; A: ä/a), Paz ha-hela?, Pai sula (l/*Ll), Puy urla?, RukMn u?ula (A: u/e), ?ula (M: ?u/u? and contraction), Kan eré

na? (A: e/u), Sar urula (A: u/e), Tso ruhọ (A: u/e), AmiSk, AmiN suLda?, AmiT, AmiV suLla?, Tha ?ulða?, Kvl surna?, Sir (ougla) 'snow, ice', Fav (oela) 'dew, hoarfrost'.

101. (1.32) *S_{136}eLik/*LeS$_{136}$ik, AtySq szik, Tha rišik 'liver'.

102. *$[S_{136}X]e[S_{136}X]em$/*$[S_{136}X]eme[S_{136}X]em$, Aty ssm-an, Tha ma-humhum 'evening', SaiTa ?ya-š/aL/emšem, RukTa, RukBd ma-s/aL/emeséme, RukMg ma-s/rə/msɨmɨ, RukTo ma-s/a/eméséme 'dusk'.

103. (1.21) *$[S_{16}Xx_2]$epuR, Sed S/m/puw, SedTk spg-an (A: e/u), SaiTa š/um/epeL (A: e/u), Pai -supu (A: u/e), Kan -sepére? (A: e/u), Tso s/m/uprə (spər-a) (M: u-ə/ə-u) 'to count', Kvl supr-an (M: *u-e/e-u) 'to know'.

104. *SeSa[/SaSa[, Aty sasa-n 'morning', RukMg[OA] sasa, ssa, RukMg sɨsáa, RukTo sesá-ne 'today'.

105. (T) *S_{136}i[, Aty s/m/i?, Tso mo-si 'to put'.

106. *$[S_{16}Xx_2]$ibu$[S_{16}Xx_2]$, SedTk, Bun[JHH] sibus 'sugarcane'.

107. *$[S_{136}]$ina[, Aty sina?, Sai šina?, RukBd lasia-sina 'plant sp., Erechtites S.P.'. Cf. PLi 1992.274.

108. *S_1iS_1il, AtyMx sisil-iq, Aty[SE] sil-iq, ssil-iq, Sed sisil, Sai šišil, Paz, Pai, AmiT, AmiV sisil, Kan sisiini?, Sar iini?, Kvl sisin 'omen bird, Alcippe brunnea brunnea Gould, A. nipalensis morisonia Swinhoe'. Cf. PLi 1992.271.

109. *$[S_{16}Xx_2]$uni[, SedTk suni? 'just before, a little while ago', AmC ?i-suni? 'a little bit ago'.

110. *[SXx]uruR, Aty[SE] suruw 'back, behind', Kan surúru? 'spinal column'.

111. *tana?, Aty, Paz, Puy, BunN, BunC, BunS tana? 'a prickly shrub, Aralia decaiseneana Hance'.

112. *ta[sθ]uk, Aty t/m/ahuk, Pai pi-natuk, Bun mi-nasụk 'to boil'.

113. *tiD₁iH₂ul/tiD₁uH₂ul, Aty triu<ŋ, Sed tdiu<ŋ 'wasp, hornet', Paz tiduhul 'ground-nesting wasp', Pai čiDiul 'large hornet', Puy tizul 'bee', RukMg tdíulu (partial A: d/D), RukTo tidũlu (partial A: d/D), RukMn ka-tidulu 'wasp, hornet'.

114. *tikuRa[sθ], AtyMx tiku<xiʔ, Sed tku<rih, Puy tikuras, Ami tikuLac, BunN tikulac, BunC, BunS tikulas 'partridge', RukMg tkuə́sɨ, RukTo tikuʔasə 'pheasant'.

115. *tiRebeS₁₃/*teRebeS₁₃, Aty tgb<il, Pai čeves, RukBd teébese, RukMg tibsɨ, RukTo tiʔébese, RukMn teeveʔe, Kan teréveseʔ, Sar tereveeʔ, Tso trəfsə, BunN, BunC, BunS tulbus, AmiC, AmiS tiLeves (middle *e inexplicably retained) 'a tree sp., Zelkova formosana', AmC tiLfes 'Chinese cork oak'.

116. *tiRebu[/*tiRubu[, AtyMx tiʔubuʔ, Sed tgubuʔ, Kan turuvuʔ, Sar turuvuʔ (both preceding Tsouic have A: u/i), Ami tiLevuʔ, BunN, BunC, BunS tilbuʔ 'a type of reed', Tso p-tivru (M: vr/rv) 'a reed used in making arrows'.

117. *tunan, Aty ms-tunan, BunN bis-tunan, BunS bus-tunan 'bracelet, armlet'.

118. *turu[dD₃Zz], AtySq turuʔ, Kvl turuz 'back'.

119. (T) *Cabu[, AtyMx -cabuʔ, Aty, Sai -sabuʔ, Pai -cavu, Puy T/em/-avuʔ, RukTa, RukBd, RukTo cabu, RukMg cboo, Sar -cavuʔ, Tso -cofu, Ami -tavuʔ, Kvl t/m/abu 'to wrap', Kan cavu-cavu-a 'millet or rice cake wrapped in ginger leaves'. Cf. PLi 1992.271. Now invalidated by the finding of Saa ähu 'to wrap'.

120. *CaLi[, AtyMx caiʔ, SedTr sariʔ, RukTo tai, Tso u-cri, Ami taLiʔ, BunN, BunC, BunS taiʔ, Tha θariʔ 'taro'. Cf. PLi 1992.275.

121. *CapaH₁, Sed sapah 'house', RukMn capa, Bun tapaʔ, Tha θa:paʔ 'barn, granary'. Cf. PLi 1992.271.

122. *Cekaj, AtyMx ma-cka-cka?, AtySq s-ska?, Sed ska?, PaiQ cekad-an 'middle'.

123. *CekeS₁₆, Sed sku<y 'bamboo sp.', Pai cekes 'a thin bamboo sp. (used for pipe-stems, arrow shafts)', RukMn ceke?e 'a thin bamboo used for arrows', Ami tekes 'a medium-sized bamboo sp.'.

124. (a) *Cep?aŋ/Cap?aŋ or (b) *tap?aŋ: Bun ma-tap?aŋ 'to patch' is ambiguous in being associable under (a) with AtyMx c/um/paŋ, SedTn c/m/apaŋ, SedTk s/m/apaŋ, Puy -Tapaŋ 'to patch' and under (b) with AtySq t/m/apaŋ 'id.'.

125. *Culuŋ, Aty p-suluŋ, Ami mi-tuluŋ 'to imitate'.

126. *Cumay, Sed, Sai sumay, Pai cumay, Puy Tumay, RukTa, RukTo cumái, RukBd cūmai, RukMg cmée, RukMn cumai, Kan cumái?, Sar cumi?i?, Tso cmoi, Ami, Kvl tumay, BunN, BunC, BunS tumað, Tha θu:may, Fav (choma), Sir (toumey) 'bear, Selenarctos thibetanus, var. formosanus'.

127. *-uH₁a[, Aty[SE] m-uah, uah, wah-an (M: ah/ha) 'to come', SedTk m-xa? (e/u by analogy), RukTa, RukBd, RukTo m-wa, RukMg, 1RukMn m-ua, Tao, Hoa (muha) 'to go'.

128. (1.25) *]uka[, Sed, SaiTa, BunN, BunC, BunS, Tha ?uka? 'not exist', Kan ma-ara-ūka?, Sar ma-ar-uka? 'to lose'.

129. (T) *-uS₁a[, Aty, Sed m-usa?, AtyMx ma-usa?, Paz mu-k-usa?, Puy mu-k-uwa?, Sar m-alu-kua?, Tha m-u:ša?, Tao (mausa), Bab (musa), Hoa (usa) 'to go', AmC usa 'approach', Kan m-u-ạ-k-usa?, Tso uso 'to go toward', Puy uwa? 'Go!', Fav (m-ossa) 'to go away', (pa-usa) 'to give permission to go'. Cf. PLi 1992.27O.

130. (T) *w₃aL₂ay, Aty[SE] wayay, Sed waray, Pai aLay, PuyPi waLay, RukTa, RukBd vaLái, RukMg vrée, RukTo vaái, RukMn aLai, Kan aLai, Sar aLaLi (< *w₃/aL/aL₂ay), Kvl waray 'thread', SaiTa waLay 'thread for weav-

ing', Paz waray 'hemp yarn for weaving', AmC waLay 'string'. By metathesis this set belongs with Jav, Mgr lawe 'thread'.

131. (1.13, T) *w₃aluH₂, Aty[SE], SedTn, Paz walu?, Kan aănu?, Sar alu?u?, AmiSk wadu?, BunN va-vanu, BunC vanu, BunS vanu?, Kvl wanu? 'honeybee', RukTa, RukTo valū, RukBd vălu, RukMg vlóo, RukMn alu 'bee, honey', Pai alu 'honey, sugar', Sai, Puy walu? 'honeybee, sugar', Sir (houalou) 'bee'.

132. *wa[qQ₂]iC, Aty, Sed waqic 'tooth, fang', Puy waHiT 'barb of spear or arrow'.

133. (1.19) *w₃a[S₁₆Xx₂]aw, Sed wasaw, Pai asaw, RukTr[OA] vasau, RukBd vasau 'leaf'. Cf. PLi 1992.270.

134. *w₃aw₃a[, SedTk wawa? 'young (of animals)', Ami wawa? 'child', BunC[N], BunS[N] (vava), Tha ?ă?a 'baby'.

135. *w₁ili[, Sed wili?, Pai vili, RukTa, RukBd, RukTo vili, RukMg vlii, RukMn ili, Ami wili? 'water leech', Sai, Puy wili? 'mountain leech', Paz wili?, BunN, BunC vini, BunS vini? 'leech'.

Appendix B

The list of cognate sets limited to Formosan and Philippine members and having at least one of each follows:

1. *]abigi[, Bab[O] abigi, Hoa[O] ăbigi 'areca nut', Bik[Br] qabi:gi 'areca palm, Arenga tremula'.

2. *]abiki[, Kan ?aviki?, Tso fi?i, Hoa[O] abiki, Sir (aviki) 'areca nut', Tag[Br] qabiki 'betel palm, Pinanga spp.', Bik[Br] qabi:ki 'a type of betel palm, Arenga tremula'.

3. *]ajem, Sai ?azem 'heart, mind', maL-?az-?azem 'to think', Kvl ?anem

'heart', Tag qa:lam, Bik, SL qa:ram 'knowledge, wisdom, sense', SL qa:ram (all Philippine have A: a/*e) 'intellect, intelligence'.

4. *]aken, Pai aken, Itb y-aken, MbI si-qaken, MbKC qakən 'I (predicative)'.

5. *]aLabeʔab, AmC ʔaLafʔaf 'lukewarm', Hlg qalabqab 'lukewarmness'.

6. *]aLun, RukMg anūru (M: n-r/r-n), MbS qəlun-an, Bkd qalu:n-aq, SubSc g-ulan-an (M: u-a/a-u) 'pillow', RukMn ta-aLun-ae 'head of the bed', Sar pa-anuL-aʔ (M: n-L/L-n) 'pillow, head of the bed'.

7. *]ali[, RukBd ali 'sister', Grt y-ali, Tag qa:li 'aunt'.

8. *]anu[, Puy anuʔ, Ami ʔanuʔ, Yam qanu 'if', Ivt qanuq 'when, if'.

9 *]amen, Pai ti-amen, Itb y-amen 'we (exclusive)', Tag qa:min 'our, us'.

10. *]amin, Pai amin, Bun ʔamin 'all, no more, finished', Ilk qa:min 'all'.

11. *]aŋsuH$_1$, AmC ʔancuh (partial A:n/ŋ), DgtC qaŋso 'smell of urine', Knk men-qaŋsu 'to stink (e.g. cooked camote leaves)'.

12. *]apu[, Paz ʔapuʔ 'grandmother', ʔapu-apuʔ 'great grandmother', ʔapu-apu-an 'ancestors', Tha ʔã:puʔ 'grandparent', Tag qa:poq 'grandparent, patriarch', Bik qa:pu 'grandmother', Knk qapu 'grandparent, grandchild', Itb qapu-qapu 'ancestors, grandparents', Mar qapu-qapuq-an 'patriarch, matriarch'.

13. *]ari[, Pai ari 'Let's go!', m-ari 'Come along with me!', Akl m-ali 'Come on! Let's go!', Bik m-a:ri na 'Come here!'. Cf. Blust 1980.41.

14. *]a[tC], Bun at, Tag qat 'and'.

15. *]awaʔay, Ami ʔawaʔay 'not exist, not have', IltK qaway 'there is no'.

16. *]awi[tC], AmF ʔawit 'to take, hold in the hand', Ilk qawit 'to carry, bear', Png qawit 'to carry, take'.

17. *babaH₁i[, RukTa a-babāi, RukBd[ST] a-babāi, RukBd[PLi] a-bābai 'woman', AmC fafahi 'wife', Tag baba:qi, baba:yi, BM bobai 'woman, female'.

18. *baju[H₁x₁], Ami vanuh, AmC fanuh, Png bagu 'body hair, down'.

19. *baH₁NiR/*paH₁NiR, Pai vali, Puy vanin (n-n/l-R), Sar valiri? 'board', RukTa ba?ali, RukBd bāali, RukMg bāli, RukTo ba?āli, RukMn va?ali (all Ruk are from *b/aR/aNiR) 'cypress, board', Tso fahri, AmiN, AmiT, AmiV vahliL, BunN, BunC banhil (M: nh/hn), BunS banil 'cypress (best for making boards)', Tag pa:nig 'panel (as the side of a board)'.

20. *bai[, SedTk pai (analogical p/b), BunC bai?, Kvl bai, Tso[TTh] bai, Tbl be 'grandmother', Tag ba:yiq 'mother, grandmother'.

21. *bakal, Pai vakal 'dagger', Puy vakal 'single-edged knife', RukTo bakāle, RukBd bākale, RukMg bkāi꜄, RukMn vakale 'small knife', Tag ba:kal 'iron', bakal 'iron-tipped stick used in cultivation'.

22. *bakaŋ, Pai v/n/akaŋ 'to walk bowlegged (intentionally)', Seb bakaŋ 'bowlegged', Ilk ba:kaŋ 'bandy-legged'. Cf. also Blust 1972.7.

23. *bakbak, Sai bakbak 'to strike with a piece of bamboo', Png bakbak 'to whip, beat'.

24. *bake[sθ], AmC fakec 'to put a belt on', sa-fakec 'a belt', Seb bakus, WBM bakes 'belt'.

25. *baLi[, Pai vaLi, Puy vaLi?, Ami vaLi?, Tha fa:ri?, Kvl vaLi?, SaiTa baLi?, Paz bari?, Sir (vare), Fav (barri), Ibl ba:li 'wind', Tag bali-ba:li 'a strong, changeable wind'.

26. *[bp]aLiw/[bp]alyu, AmC su-paLiw, tu-maLiw 'to change places, to be available as a substitute', Tag balyo 'to exchange seats', SL balyu 'to change, exchange'.

27. *balluku[, AtyMxw baluku?, RukTa baLakū, RukBd baLāku, RukMg brāku, RukTo baāku (all Ruk have A: a/u), Kvl bnuqu 'winnowing basket',

BonG balluku 'a type of small head-basket with a lid'.

28. *ba[nñ][nñ]aw, Ami vanaw 'lake, pond', IfgBy banno 'lake', Seb ba: naw 'pool'.

29. *baŋaw, AmiC, AmiS faŋaw 'bedbug', AmC baŋaw 'bedbug, flying insect', Tag ba:ŋaw, Kpm ba:ŋo 'botfly', Knk ba:ŋaw 'weevil'.

30. *baŋaw//*b/er/aŋaw//*b/a[Ll]/aŋaw, Pai quli-vaŋeraw (M: ŋ-r/r-ŋ, M: a-e/e-a), MbAD, MbI ba:ŋew, Tag, SL, Hnn balaŋaw 'rainbow', Hnn ba:ŋ aw 'stripe, striped or light-colored (as the hair of certain pigs)'.

31. *baŋ[ae]θeL/*baŋ[ae]Leθ, RukBd maŋaθeLe, RukMg maŋaθrɨ, RukTo maŋāθee (all Ruk have m/b by analogy) 'to smell bad', Ilk na-baŋles 'sour, acid, musty (due to fermentation)'.

32. *baŋibaŋ, Ami vaŋivaŋ, Itb vañivaŋ 'chisel'.

33. *baŋ[sθ]i[tC], Bun[N] baŋsit 'bad smell', Ilk baŋsit 'stench of excrement, of stale cooked rice'.

34. *baqbaq, BunS bahbah, Tag ba:baq, Png ba:ba 'chin'.

35. *ba[q₃Q₁]i[, AmC faqi 'aunt', Tso baʔi, Hlg, Png ba:qi, Ala ak-baqi 'grandmother'.

36. (a) *baretinuk or (b) *bareCinuk: Itb varatinuk 'bitter orange', Ivt valatinuk 'a kind of lemon with a bitter rind' (both preceding have A: a/e) are associable under (a) with PuyKt vartenuk (A: e/i) 'orange' or under (b) with PuyKs verTenuk (AA: i/e, a/e) 'orange'; the Puy words are incompatible with each other.

37. *ba[rR]iw, BunN[N], BunC[N], BunS[N] baliv 'to exchange', Mar bariw-an 'replacement, substitute, alternate'.

38. *baSaw, AmC fasaw 'cooled off (rice, anger, feelings)', Tag ba:haw, Png ba:qaw 'cold victuals'.

39. (1.15) *ba[tT]akan, Aty btakan, AtySq ptakan (by analogy or partial

A: p/b), BunN, BunC, BunS batakan 'a type of large bamboo', Paz patakan (p/b by analogy) 'bamboo', Tir b/el/otokan 'a spiny bamboo'.

40. *ba[tTC]u[(1), BunN, BunC ma-batu, BunS la-batu? 'to throw', Tag bato, Ilk batu 'to throw, cast'.

41. *ba[tTC]u[(2), Sir (vatto), Tag bato 'kidney', Sng bə?-batu 'kidney, testicle'.

42. *ba[tT]u[(3), AtyMx, AtySk, Paz batu?, KlnKl batuq 'egg'.

43. *bau[, AmC fau 'living thing, insect, worm', Ilk bau 'rat'.

44. *bekbek, BunS[N] ma-bukbuk 'to cut (meat, etc.)', BonG bekbek 'to quarter, cut up a butchered animal'.

45. *bekuC, Pai ma-vekuc 'to be bent down (as from a heavy burden or pressure)', AmC fekut 'to bend one's body, to double over', Png bukut (A: u/e) 'bent over, hunched, stooped (of elderly), hunchback', Tbl bekut 'a bend, fold', Tir bekut , MTb bəkkut, Tag bukot (A: u/*e) 'hunchback, hunchbacked'.

46. *beLuk/buLuk, Paz ma-buruk 'rotten (as a log)', Tag bulok 'ill-smelling from rotting', Ibg bulluq 'rottenness', BonG buluk, na-bluk 'rotten (e.g. meat)'.

47. (a) *beNuC or (b) *buqnu[tTC]: Tag bu:not 'to pull out (grass, plants)', Seb bu:nut 'to pull out from between', Ilk bu:nut 'to pull out by the end or root' are ambiguously associable under (a) with Pai v/n/eluc 'to pull out (grass, plants)' under a hypothesis that the Philippine words have undergone an assimilation (u/*e) and under (b) with AmC buqnut 'to uproot, remove from the ground'.

48. *beŋebeŋ, PaiQ veŋveŋ 'wind-screen, shelter', RukMg si-bɨŋbɨŋɨ, RukMg[PLi] si-bəŋbə́ŋə, RukMn ?a-veŋeveŋe, Ilk beŋbeŋ 'curtain'.

49. *be[tTC]ak, Ami fetak 'cracked, dry (as soil)', Bik batak 'cracked

(wood, soil)', Kpm qa-btak 'splitting, opening (of earth, reed, leaf)', BonG betak 'be cracked (dry ground, clay jar)'.

50. *be[tTC]uy, AmC qa-ftuy, ha-ftuy, Ilk butuy (A: u/e), Ibg vuti (A: u/e) 'calf (of leg)'.

51. *beCux₁, Aty bsuh, Pai vecu, Puy vuTu? (A: u/e), Ami vetu?, Mar betu, Ilk buttu-an (A: u/e) 'callus'.

52. *biaC, Puy -viyaT 'to draw a bow to its full extent', Ilk si-bi-biat 'drawn, stretched (of bows)'.

53. *biD₁₄u[, Aty biru? 'to draw, paint, write', Knk bi:du 'tattooing compound'.

54. *biLi[, Aty m-bzi<nah 'to come home', Sed m-bri<nah 'to return before reaching one's destination', RukTa -biLi, RukBd biLi, RukMg -brii, RukTo -bii, RukMn -viLi 'to return, retort', Itb vili, 'to return, retore', Ivt̩ vidi, Sng bili 'to return'.

55. *bilaur, Puy, Ami vilaur, Ami vilauL (partial A: L/r), Kvl bnaur, Yam viRauR (A: R/l) 'cucumber'. Cf. PLi 1994. 262.

56. *biNebel, Sed blbul, Pai, Puy velvel (A: e/i), RukTa belebele, RukBd, RukTo belébele, RukMg blɨblɨ, RukMn velevele, Kan ta-venevene?, Sar ta-velevéle? (all Ruk and Tsc have A: e or ɨ/i), BunN, BunCS bunbun (u from A: *e/i), Tha fiðfið, Paz belebel (A: e/i), Sir (bulbil) (? M: u-i/i-u), Itb vinivex (A: i/e) 'banana', Tso fəəfəhə (? əə/əh after *e/i) 'an inedible wild banana', Fav (bilpil) (A: *i/e) 'a plaintain tree'.

57. *b₁iq₁₃eL, Aty[Dl] biqiy (A: i/*e), SedTn biqir (A: i/*e) Pai[Dl] biqeL, Kan vi?ɨLi?, Sar vi?ɨLi?, Tso f?iri (all Tsc have A: i/e), BunS bihi?, BunN, BunC biqi, Ilk bi?el, IfgBt bi:?ol 'goiter'.

58. *biRaSu[, Puy virau? 'Miscanthus sp.', Itb viahu 'a kind of sword-grass, Miscanthus floridulus', Bik biga:hu 'Miscanthus sinensis'.

59. *bi[tT]un/*bin[tT]un, Paz bintul (D: 1/n), Paz[Fer] bintun, Ibg, Gad, Klg, SubS, SubSc bitun, DgtC biton 'star'.

60. *buar, RukMg buəri, Sar m-ii-vuare? 'to collapse, tumble down (of a pile)', Tag buwal 'to fall, fell, demolish'.

61. (a) *buka[dDZz] or (b) buka[tC] or (c) buka[Sθ]: Paz bukat 'to release a trap' is ambiguous in being associable under (a) with Tag bukad 'to open (as flowers)'; under (b) with BonG bukat 'to open (window, bottle. suitcase, parcel)'; and under (c) with Tag bukas 'open (not closed)'.

62. *bukbuk(1), Sir (bouckbouck), 'to hit, strike', Seb, Mar bukbuk 'to club, pound'.

63. *bukul(2)/*buŋkul(1), Kvl buqul (A: u/e) 'ankle', SngSi bəq-buŋkuLəq 'ankle-bone'.

64. *bulabul/*bulbul, Pai v/aL/ulavul-an 'down, small feathers', Kpm, Mar bulbul 'hair, feather'.

65. *bul[ae]y, Kan vunãi?, Sar vuli?i?, Tso fkoi, Sir (vouly, vaulei, bulai, voulei), Itb vulay, Ivt buday 'snake'.

66. *buNuq, Aty[SE] s-bul<iŋ 'to cast out', Sai ši-bulo? 'to throw away', Pai -vuluq 'to throw something at (a target)', Puy -vuluH 'to throw something small overhand', Hnn bu:nuq 'throwing, casting (lance), stabbing (with a knife)'.

67. *buŋbuŋ, AmC fuŋfuŋ 'protective cover (e.g. of burlap, straw, plastic)', Png buŋbuŋ 'a cover'.

68. *buŋuH₁, Kan na-vuŋu?, Sar vuŋu?u?, Tso fŋuu, Ami vuŋuh, BunS buŋu?, BunN, BunC buŋu, Sir (vongo) 'head', Tag buŋo? 'skull'.

69. *buqel, SedTk bql<it 'leg', WBM buqel 'knee', Tag buqol (o from A: *u/e) 'ankle', Bik, Hlg bu:qul 'heel, heel-bone'.

70. *buRbuR, AmC fuLfuL 'to break into small pieces, to crumble', Tag

bugbog 'to club, pound, repeated strong blows', Png na-bugbug 'be soft, whipped (of body), pounded (of wood)', Ilk bugbug 'clubbing'.

71. *bu[sθ]uR, PaiW vutu, IfgK bu:hul, Ilk ka-bu:sur 'enemy'.

72. *butul, SedTk butul 'acne pimples', Itb vutux 'seed (general), pimples'.

73. *[D_{123}Zz]a[, Sed da?, Klg, Ibl, Msk, TbwK da 'already'.

74. *Da?aN, AtyC ral<i? 'former', Kvl zaan, Atta da:n, Msk daqan, Bik, Ilk da:qan 'old'.

75. *[dD_{24}]ab[dD_{24}]ab, AmC rafraf 'be inflamed (e.g. with anger, against perversion)', Seb dabdab 'to set fire to, to singe off', Hlg dabdab 'to flare up (of fire)', MTb dabdab 'to set fire (with a torch)'.

76. *D_{124}aD_{124}ek/*D_{124}anD_{124}ek, Puy zazek 'body', Ban dadaq (A: a/e), Ttm raqndak (A: a/e, q inexplicable) 'chest (anat.)'.

77. *D_{124}aD_{124}em, Puy -zezem (A: e/a) 'to soak, immerse (as rice, laundry)', Ilk ra:rem 'to immerse, plunge, dip'.

78. *[dD_3Zz]ake[dD_3Zz]ak, Pai -jakajak (A: a/e) 'to kick (something)', Paz dakedak-en 'to stomp', Isg -d/ay/aqdaq 'to dance'.

79. *[dD_{234}]a[Ll]uŋ, AmC raruŋ (A: r/[Ll]) 'a hut, temporary shelter', Mar daluŋ 'to hide behind, shelter'.

80. *[dD_3Zz]aLu[sθ], Pai jaLut, Puy daLus 'slippery', Kan m-usu-calu? 'to slip, slide', Tag dulas (M: u-a/a-u) 'slipperiness', SL dalus 'to slip off one's feet gradually'.

81. (a) *D_2aLaŋeD_2aŋ/(b) *[dZz]alaŋe[dZz]aŋ: Isg dalaŋádaŋ (A: a/e) 'summer days', Ilk dalaŋdaŋ 'sun-dried large fishes (after being cut open and salted)', qi-dalaŋa:daŋ (A: a/e) 'to toast, to warm on a spit', Knk qi-dalaŋa: daŋ 'to heat, warm' (only in tales)', IfgK dalaŋdaŋ 'burnt rice sticking to the pot' are ambiguous in being associable under (a) with Pai zalaŋzaŋ

'to perspire' and under (b) with RukBd[JCK] ma-dalaŋedaŋe 'hot (e.g. weather')'.

82. *[dD₃Zz]alayap, Pai jalayap 'a citrus, Citrus depressa', Ilk dala:yap 'a citrus, Citrus lima. Lunan'.

83. *[dDZz]a[Nn]aw, Bun[N] ma-danav, Tsw dano 'to bathe'.

84. *D₁₄ameL, Pai ke-DemeL, Puy ke-zemeL, RukTa maku-DeméLe, RukBd maku-DemeLe, Tso o-cmərə 'thick (as a board)' (all Formosan show A: e/a), Hlg da:mul 'thickness', TbwK ma-damel 'thick'.

85. *jaŋaw/*ŋajaw, Pai sa-ŋadaw 'a type of light-green diamond-shaped stinking insect', Ilk da:ŋaw 'a kind of stinkbug', Gad da:ŋaw 'a kind of bug that exudes a repulsive odor'. Cf. Blust 1980.47.

86. *D₂aŋeD₂aŋ, Aty r/em/eŋeraŋ 'to boil (e.g. water)', Paz[Fer] daŋedaŋ 'to cook, boil (food)', Pai ma-zaŋzaŋ 'be made to feel hot', RukTa DaŋeDäŋe, RukBd DaŋéDaŋe, RukTo wa-DaŋeDaŋe, RukMn u-daŋedaŋe 'hot, roast', RukMg o-DŋəDŋə, Kan -caŋécaŋe?, Sar -saasaŋe?, Tso t/m/a-cŋəcŋə, Png daŋdaŋ 'to dry by the fire', AmC raŋraŋ 'to hold near the fire to dry (as hands, cloth, etc.)', Tag dandaŋ 'heated', daŋdaŋ 'heated, to dry by the fire', Bik daŋdaŋ 'to dry by the fire or hot coals'.

87. *D₂aŋeLa[/D₂eŋeLa[, Pai zeŋeLa, zeŋla 'a small hardwood tree, Vitex negundo', Tag (daŋla) 'Vitex trifolia', Ilk daŋla 'Vitex negundo', Hlg duŋu:la 'Vitex aherniana'. The Pai and Hlg retention of the *e between single medial consonants is inexplicable.

88. (a) *[dD₃Zz]ape?ir or (b) D₁₄ape?ir: IfgK dap?il 'twins' is compatible under (a) with Pai japir 'twin growth, two stems from one root, unnatural double (double banana, egg with two yolks, etc.)' and under (b) with PaiW Dapir 'twin growths (banana, taro, etc.)'.

89. *[DZ]apeR, SaiTa rapeL, Tag lapag (A: a/e; analogical l/d) 'floor'.

90. *[dD₃j]aqiR/[dD₃j]aRiq, Puy daHir, Itb rayiq 'Paederia chinensis, a vine with small fruit used to blacken teeth'.

91. *[dDZz]a[q₁₂₃₄Q₂]wi[xX], BunN[N], BunC[N] daqvis, BunS[N] dahvi, Knk qa-dawi, BonG qa-dawwi 'far'.

92. *[dD₃Z]a[rR]aŋ, PuyPu mi-daraŋ 'red', Ilk daraŋ, Png dalaŋ 'flame'.

93. *[dDZz]arapa[, BunC dalapa, BunS dalapa? 'sole (foot)', Ala darapa 'foot (including toes)'.

94. *[dDZz]aRam/[tTC]aRam, Paz daxam, Fav (ma-darram) 'accustomed', Sir (ma-dagam) 'to accustom', Mar tagam 'tame'. The differing reconstructions are connected by an analogical change. Cf. also Blust 1980.117f., 1986.33.

95. *[dDZz]aRem, Paz daxem, MbAD da:gem, Ivt dayem, rayem 'needle'.

96. *[dDZz]aw, BunN, BunC, BunS daw, Tag daw 'they say, it is said (quotative)'.

97. *Zaw₁₂il/law₁₂iZ, Puy a-dawil, RukTa daili, RukBd dáili, RukMg me-dávli, RukTo ma?a-dávili, RukMn ma?a-daili, Kan ara-daini?, Sar ma-sail-a?, Tso covhi, Btk qa-lawid, TbwA qa-laid, TbwK qa-lawid 'far'.

98. *[D₂₄Z]aya[, AmC ka-raya-n 'sky, mid-heaven', BonG, Knk da:ya 'sky, heaven'.

99. *D₂ebeD₂eb, Paz zebezep, Tag dibdib, Hnn dubdub 'breast, chest'.

100. *deLen, Sar ceLen-a, Tso crən-i 'to choke on food', Itb qa-dlen 'to have a throat obstruction', Hlg du:lun 'to choke from a lump in the throat'.

101. *[D₂₃₄Zz]eme[D₂₃₄Zz]em, Tso cməcmə (A: e/a), Itb remdem, Ivt mademdem 'raincloud'.

102. (a) *[dDZz]iLaw or (b) *[dDZz]iaw: Aty[Dl] m?-riaw, mg?-riaw, BunS ma-diav 'yellow' are associable under (a) with Tag di:law 'yellow' and under (b) with Png la-dia:w-an 'to become yellow (leaves, bamboo)'.

103. *[dD]iqal, Sai ka-ri?al 'day after tomorrow', AmC r/um/iqal 'day, daytime', Itb rial 'ray, beam (of light), sun-ray', Ivt rial 'ray of light, brilliance'.

104. (3) *[dD₃Zz]ukap//[dD₃Zz]/aL/ukap, Puy daLukap 'palm (hand), sole (foot)', Ibl šukap, šalukap 'sole (foot)', Hnn dalu:kap, Ilk daku:lap (M: l-k/k-l), SblBt dawkap 'palm (hand)'.

105. *D₁₄ukuŋ, Pai Dukuŋ 'a bend', ma-Dukuŋ 'bent, stooped over, crooked', Duku-Dukuŋ 'hunchbacked', Kvl m-quzuŋ (M: q-z/z-q) 'hunch-backed', Png dukuŋ-dukuŋ 'to walk stooped, totteringly, decrepitly, to walk very bent over with age'.

106. *[dD₃Zz]uLem, Puy -duLem 'pitch-dark', Kpm du:lum (A: u/*e) 'darkness, obscurity'.

107. *D₂uyuR, Pai zuyu 'water ladle', Ibg duyug 'something used for dipping', Agt duyug 'bailer, coconut drinking vessel'.

108. *]eD₂₃₄ep, RukBd ma/eD/eDepe, RukMg u-Dɨpɨ (ɨDp-áa), RukMn edepe, Sar -esepe?, Png qerep 'to extinguish (fire)'.

109. *gamu[sθ], PuyLik ha-hamus 'claw', Tag ga:mos 'scratch mark on the face'.

110. *garuC/*raguC, Pai garuc, Puy haruT, RukTa garúcu, RukBd gā rucu 'comb', AmC karut 'rake', Ibg pa-ra:gut-en 'to harrow'.

111. *gaCegaC, RukTa gacegace, RukBd gacégace, RukMg gcǝgcǝ 'to scratch', Tag gatgat 'groove, nick, mark of an embedding rope, etc.'.

112. *gawgaw, Pai gawgaw 'to accept, receive', pa-gawgaw 'to offer something', g/ar/awgaw 'to extend the hands as if to say "Give me!"', Bik gawgaw 'to hand, give by hand'.

113. *gemgem, Pai, Ibl gemgem 'fist', Png gemgem, Kpm gamgam 'closed fist'. Although this etymon is no doubt represented in Fi gogo (/g/

= Dempwolff's ŋg) 'to gather up in handfuls or in the arm, to clear up rubbish' and perhaps also in To kokom-i 'to press in, impress', Sam qoqom-i 'to press', the set cited has value here because of the restricted distribution of the shared meaning of its members.

114. *ge[tT]em/gemet, Puy pa-htem 'to pinch the skin till it hurts', Png ge-gemet 'hold firmly in the closed fist'.

115. *geCeL, Pai geceL 'crab's pincers', -geceL, RukTa gecéLe, RukBd -géceLe, RukMg gcɨrɨ, RukMn -heceLe 'to pinch', Png getel 'to pinch off the stem of a fruit or flower with the nails', Bik gu:tul 'to pick, cull, or cut flowers'.

116. *giLgiL, Pai -giLgiL 'to scratch or rub an animal', ki-giLgiL 'to rub against (as a boar against a tree)', MbK -gilgil 'to scratch one's back by rubbing it against a post or the ground'.

117. *guCguC, Pai -gucguc 'to scratch (itch), to weed', AmC kutkut 'to use metal to scrape', Hlg gutgut 'to grate', SL gutgut 'to cut by rubbing the cutting instrument'. The Ami word is formally ambiguous in being associable with Seb kutkut 'to scratch (as the head), Mal kukot, kokot (o/u inexplicable) 'to claw' (cf. Blust 1970.47), but is included here in this set because of its closer relationship to Paiwan.

118. *[HS$_2$x$_1$]akem/*[HS$_2$x$_1$]aŋkem, Paz mu-hakem 'to embrace', Sng maŋ-aŋkuŋ 'to embrace, care for, protect, be guardian of'.

119. *H$_1$a[sθ]uR, Ami hacuL 'milk', Itb qasuy 'soup, sap, juice, broth, milk (e.g. of coconut)'.

120. *[H$_1$x$_1$]e[sθ]ek, AmC hecek 'beam, large piece of wood, post, wooden pillar', Tbl əhək 'stick used in planting rice, dibble stick'.

Appendix C

143. (5) *kamut, Aty[SE] kam<il, k-kam<il 'fingernail', km-kam<il 'to scratch (with the nails)', Paz ka-kamut, Msk kamut, SubSc k/uy/amut 'finger', Ilk k/ar/aːmut 'to scratch'.

150. (1.8) *kawaS₁₆, AtySqSk, Sed, Paz kawas, Tha kawaš, Yam, Ivt kawa-n 'year'.

246. (10) *ŋaL[ae]y, Puy ŋaLay, RukTa ŋaLái, RukBd ŋáLai, RukMg ŋrée, RukTo hãai, RukMn ŋaLái, Kan ŋãaiʔ, Sar ŋaLiʔi, Tso ŋroi, Itb ŋaxay 'saliva', Ami ŋaLay 'to drool'.

252. (12) *ŋu[S₁₂₃₆x₁]eR, BunN, BunC ŋusul, Itb ŋuhey, ŋuhuy (A: u/e) 'nasal mucus', Ivt ŋuhey 'catarrh, cold', WBM ŋuheg 'nasal discharge'.

264. (13) *pikpik, AmC pikpik 'to fly', Ami sa-pikpik, Tau pikpik, MbAD p/ag/ikpik 'wing'.

265. (14) *piLek, RukBd sa-keLepe, RukMn keLepe 'eyebrow', RukMg s-kïrpï (all Ruk have M: k-p/p-k; A: *e/i) 'eyelid', AmF sa-peLek (A: e/i), SL piluk, WBM pilek 'eyelash'. AmC sa-peLek, sa-pelek 'eyelashes', peLek, pelek 'to go off and on, to blink' raise the possibility of a doublet with *l. It is tempting to associate Tag pilik-mata, Mgd pi-pilik 'eyelash' with this set and perhaps they do belong here, but there is another set with which they also appear to be compatible that reflects a different medial consonant: WBM ke-pirek-pirek 'face', Bik, SL piruk, Btk pirek 'eyelash'. Whether just one differing set is involved is not clear in view of TbwA pi-pirek-en, TbwK pi-pigk-en 'eyelash'.

266. (15) *pil[ae]y, Pai ma-pilay, Knk piːlay 'crippled, lame', RukTa ma-ʔilai, RukBd ma-pilai, RukMg ma-plée, RukTo ma-pilái, RukMn ma-pilai,

Itb pilay 'lame'.

267. (16) *piɲi[, Kvl piɲi, Hnn piɲiq 'cheek'.

271. (17) *-pujₗu[, Puy HaLi-pudù-an, RukBd alipugu, Bun qani-puʔ 'hair whorl', Pai quli-pa-pudu-an 'crown (head), whorl (head), cowlick', Bik qarim-pu:ru, SL qalim-pu-pu:ru 'cowlick, hair whorl'. Png qalipu:gu, qali-pugu 'to arrange the hair in a roll or bun' will also fit here. It remains to be seen whether the association of AtyMx qali-puguʔ, AtySq, AtyMs q-puguʔ, AtySk la-puguʔ 'hair whorl' with this set can be justified under a *pujₗu[. Cf. PLi 1992.273.

273. (18) *punu[q₁Q₂], Pai, Ami punuq, Puy punuH, Sar punuʔu, Tso pnuu, BunN, BunC punuq, BunS punuh 'brain', RukTa ʔunū, RukBd pūnu 'forehead', RukMg t-punū-a, RukTo punú, RukMn ta-punu-e 'forehead, brain', Tha pu:nuq, Paz punuʔ, Fav (oeno) 'head', Tag pu:noq, Hlg, SL pu: nuq, Bik pu:qun, Kpm pun (the last two with M: q-n/n-q) 'chief, head of a group', Tag pu:noq, Kpm pu:nuq, Seb punu:q-an, Mar sapi-puun (M: *q-n/ n-q) 'beginning, source, origin'. This set includes the set attributed to *punu[q₁Q₂](1), the latter being distinguished by an exclusively shared semantic agreement. In terms of their forms alone, the Philippine words could also be assigned to PAN puqu[nN] 'trunk (tree)', but their meanings fit better with those of the Formosan words. The 'source of a river' and the 'head of a group' involve metaphors commonly met with and are else-where referred to by a term for 'head' or a derivative; cf. Mal pəŋ-hulu 'chief, leader', kə-hulu 'upriver'. Cf. Blust (1992.35) who rejects this association.

281. (19) *[qQ]a[dDZz]eŋ, BunS haduŋ, Bik qa:ruŋ 'mole', Itb qadeŋ 'a large mole'.

331. (22) *rukap, Kvl ruqap 'palm, sole', Yam rukap 'palm (hand)'.

353. (1.28, 23) (a) *sebu[or (b) *q₄i-sebu[:

353a. *sebu[, Sai kä-hbu? 'urine', h/om/bu? 'to urinate', Pai tevu 'a skin bag (of animal's bladder)', RukTa, RukBd ubu, RukMg sbūu, RukTo subu, RukMn uvu (all Ruk have A: *u/e), AmC cefu 'bladder', ma-cfu 'to bed-wet'.

353b. *q₄i-sebu[, Kan iivu? (A: i/e), Sar ʔivu 'urine', Sar m-iivu (A: i/e) 'to urinate', Tso sifu, Ilk qisbu, Btk, TbwA siqbu (M: s-q/q-s) 'urine'. A double i was expected in the Saaroa word for 'urine', so that the absence of the second i may prove to be inexplicable. Curiously Ferrell (1969.244) reports an early recording of the Sar word by Tsuchida as ʔi:bo.

442. (1.30) *[tT]imus, Aty, Sed timu<?, RukTa timū, RukBd timu, RukMg tmūsu, RukTo timūsu, RukMn timu, MbI, MbKC, MbCW timus 'salt', Puy -timus 'to flavor with soy-sauce'. Sai timu? 'salt' is taken to be an Atayalic loanword. Cf. also Blust 1970.144.

451. (24) *[tTC]uLek/[tTC]eLek, AmC tuLek 'to be deaf, extract wax from the ears', Kan téeke? 'ear', SblBt teek 'ear, deaf', Png te:lek 'deaf'.

REFERENCES

L. Bloomfield.

    1933   Language. New York.

R. A. Blust.

    1970   Proto-Austronesian Addenda. Oceanic Linguistics 9:104-165.

    1972   Additions to 'Proto-Austronesian Addenda (WPL 3.1:1-107)' and 'Proto-Oceanic Addenda with Cognates in Non-Oceanic Austronesian Languages' (WPL 4.1:1-43). WPL 4.8:1-18.

    1980   Austronesian Etymologies. Oceanic Linguistics 19.1-2:1-189.

1986   Austronesian Etymologies III. Oceanic Linguistics 25:1-123.

1992   The Position of the Formosan Languages: Method and Theory in Austronesian Comparative Linguistics. ISASRT 21-55.

W. H. Brown.

1951-6 Useful Plants of the Philippines. 3 vols. Republic of the Philippines, Department of Agriculture and Natural Resources Technical Bulletin No. 10. Manila, Bureau of Printing.

I. Dyen.

1956   Language Distribution and Migration Theory. Language, 32:611-626.

1965   A Lexicostatistical Classification of the Austronesian Languages. International Journal of American Linguistics Memoir 19.

1967   The Position of the Malayopolynesian Languages of Formosa. Asian Perspectives 7.1-2:261-267.

1974   The Proto-Austronesian Enclitic Genitive Pronouns. Proceedings of the First International Conference on Comparative Austronesian Linguistics, vol. 2. Oceanic Linguistics 13:17-31.

1992   The Internal and External Classification of the Formosan languages. ISASRT 162-196.

S. Egerod.

1980   Atayal-English Dictionary. Scandinavian Institute of Asian Studies Monograph 35.

R. Ferrell.

1969   Taiwan Aboriginal Groups: Problems in Cultural and Linguistic classification. Taipei.

V. Fey.

1986   Amis-Dictionary. Taipei.

C. F. Hockett.

    1958   A Course in Modern Linguistics. New York.

            ISASRT - Papers for International Symposium on Austronesian Studies Relating to Taiwan, December 29-31. Organizing Committee: P. J-K. Li, C-H. Tsang, Y-K. Huang. Nankang, Taiwan.

J. C-H. Kuo.

    1986   Budai Dictionary. Taichung, Taiwan.

P. J-K. Li.

    1982   Male and Female Forms of Speech in the Atayalic Group. BIHP 53.2:265-304.

    1985   The Position of Atayal in the Austronesian Family. In A. Pawley and L. Carrington, eds. Austronesian Linguistics at the 15th Pacific Science Congress. Pacific Linguistics C-88. Canberra.

    1992   Formosan vs. Non-Formosan Features in Some Austronesian Languages in Taiwan. ISASRT 268-284. Appeared in this volume.

    1994   Some Plant Names in Formosan Languages. In A. K. Pawley and M. D. Ross, eds. Austronesian Terminologies: Continuity and Change, pp. 241-266. Pacific Lingustics C-27.

A. Meillet.

    1954   La Méthode Comparative en Linguistique Historique. Paris.

Y. Nihira.

    1983   A Bunun Vocabulary: a Language of Formosa. 2d. ed. Tokyo.

N. Ogawa and E. Asai.

    1935   Taiwan Takasagozoku Densetsu-shū (Myths and Traditions of the Formosan Native Tribes). Taihoku Imperial University. Tokyo.

F. Pecoraro.

    1977   Essai de Dictionnaire Taroko-Français. Paris.

S. Tsuchida.

> 1976   Reconstruction of Proto-Tsouic Phonology. Study of Languages d
>   Cultures of Asia and Africa Monograph 5. Tokyo.

T'-H. Tung.

> 1964   A Descriptive Study of the Tsou Language, Formosa. Taipei.

S. H. Wang.

> 1976   Amis Word List (Fataan). n.p. (Mimeo.).
>
>   WPL - Working papers in Linguistics. Department of Linguistics,
>   University of Hawaii-Manoa.

Abbreviations of Language Names

Agt - Agta. North Luzon, Philippines.

Akl - Aklanon. Panay, Philippines.

Ala - Alangan. Mindoro, Philippines.

AmC - Central Amis as reported in Fey 1986.

AmF - Fataan (= Vata?an) dialect of Amis (reported in Wang 1976) Taiwan.

Ami - Amis. Taiwan.

AmiC - Central Amis. Taiwan.

AmiN - Northern Amis. Taiwan.

AmiS - Southern Amis. Taiwan.

AmiSk - Sakizaya dialect of Amis. Taiwan.

AmiT - Tavalung dialect of Amis. Taiwan.

AmiV - Vata?an (= Fataan) dialect of Amis (by ST). Taiwan.

Atc - Atayalic subgroup.

Atta - Atta. Mindanao, Philippines.

Aty - Atayal. Taiwan.

AtyC - Cʔuli dialect of Atayal. Taiwan.

AtyMay - Mayhuman dialect of Atayal. Taiwan.

AtyMs - Maspazi? dialect of Atayal. Taiwan.

AtyMx - Mayrinax dialect of Atayal. Taiwan.

AtyMxm - Mayrinax dialect of Atayal, men's speech. Taiwan.

AtyMxw - Mayrinax dialect of Atayal, women's speech. Taiwan.

Aty[SE] - Atayal in Egerod 1980.

AtySk - Skikun dialect of Atayal. Taiwan.

AtySq - Squliq dialect of Atayal. Taiwan.

Bab - Babuza. Taiwan.

Ban - Bantik. North Sulawesi, Indonesia.

Bik - Bikol. South Luzon, Philippines.

Bkd - Binukid. Mindanao, Philippines.

BM - Bolaang-Mongondow. North Sulawesi. Indonesia.

BonG - Guinaang dialect of Bontok. North Luzon, Philippines.

Btk - Batak. Palawan, Philippines.

Bun - Bunun. Taiwan.

BunC - Central dialect of Bunun. Taiwan.

BunN - Northern dialect of Bunun. Taiwan.

BunS - Southern dialect of Bunun. Taiwan.

DgtC - Casiguran dialect of Dumagat. Luzon, Philippines.

Fav - Favorlang. Taiwan.

Gad - Gaddang. North Luzon, Philippines.

Grt - Gorontalo. North Sulawesi, Indonesia.

Hlg - Hiligaynon dialect of Bisayan. Philippines.

Hnn - Hanunoo. Mindoro, Philippines.

Hoa - Hoanya. Taiwan.

Ibg - Ibanag. North Luzon, Philippines.

Ibl - Inibaloy. North Luzon, Philippines.

IfgBt - Batad dialect of Ifugao. North Luzon, Philippines.

IfgBy - Baynihan dialect of Ifugao. North Luzon, Philippines.

IfgK - Kiangan dialect of Ifugao. North Luzon, Philippines.

Ilk - Ilokano. North Luzon, Philippines.

IltK - Kakidugen dialect of Ilongot. North Luzon, Philippines.

Isg - Isneg. North Luzon, Philippines.

Itb - Itbayat. Batanes Islands, Philippines.

Ivt - Ivatan. Batanes Islands, Philippines.

Kan - Kanakanabu. Taiwan.

Klg - Kalagan. Mindanao, Philippines.

KlnKl - Keleyqiq dialect of Kallahan. North Luzon, Philippines.

Knk - Kankanay. North Luzon, Philippines.

Kpm - Kapampangan. Central Luzon, Philippines.

Kvl - Kvalan. Taiwan.

Mal - Malay. Malay Peninsula, Sumatra and other areas of Indonesia.

Mar - Maranao. Mindanao, Philippines.

MbK - Cotabato Manobo. Mindanao, Philippines.

MbAD - Dibabawon Manobo. Mindanao, Philippines.

MbI - Ilianen Manobo. Mindano, Philippines.

MbS - Sarangani Manobo. Mindanao, Philippines.

MbKC - Kalamansig Cotabato Manobo. Mindanao, Philippines.

Msk - Mansaka. Mindanao, Philippines.

MTb - Tagabawa Manobo. Mindanao, Philippines.

Pai - Paiwan. Taiwan.

PaiQ - Southern Paiwan. Taiwan.

PaiTb - Tjubar dialect of Paiwan. Taiwan.

PaiW - Western Paiwan. Taiwan.

Pap - Papora. Taiwan.

Paz - Pazeh. Taiwan.

Png - Pangasinan. North Central Luzon, Philippines.

Puy - Puyuma. Taiwan.

PuyKs - Kasabakan dialect of Puyuma. Taiwan.

PuyKt - Katipul dialect of Puyuma. Taiwan.

PuyLik- Likavung dialect of Puyuma. Taiwan.

PuyPi - Pinaski dialect of Puyuma. Taiwan.

PuyPu - Puyuma dialect of Puyuma. Taiwan.

Ruk - Rukai. Taiwan.

RukBd - Budai dialect of Rukai. Taiwan.

RukMg - Maga dialect of Rukai. Taiwan.

RukMn - Mantauran dialect of Rukai. Taiwan.

RukTa - Tanan (Tainan, Dainan) dialect of Rukai. Taiwan.

RukTo - Tona dialect of Rukai. Taiwan.

RukTr - Taramakao (= Tanan) dialect of Rukai in Ogawa and Asai 1930.
Taiwan.

Sai - Saisiyat. Taiwan.

SaiTa - Taay dialect of Saisiyat. Taiwan.

SaiTu - Tungho dialect of Saisiyat. Taiwan.

Sar - Saaroa. Taiwan.

SblBt - Botolan dialect of Sambal. Central Luzon, Philippines.

Seb - Sebu dialect of Bisayan, Philippines.

Sed - Sediq. Taiwan.

SedRa - Raus dialect of Sediq. Taiwan.

SedSka - Skadang dialect of Sediq. Taiwan.

SedTd - Toda dialect of Sediq. Taiwan.

SedTk - Taroko dialect of Sediq as reported in Pecoraro 1977. Taiwan.

SedTn - Tongan dialect of Sediq. Taiwan.

SedTr - Truwan dialect of Sediq. Taiwan.

Sir - Siraya. Taiwan.

SL - Samar-Leyte dialect of Bisayan. Samar and Leyte Is., Philippines.

Sng - Sangirese. Sangir Is., Indonesia.

SngSi - Siau dialect of Sangirese. Sangir Is., Indonesia.

SubS - Sindangan dialect of Subanun. Mindanao, Philippines.

SubSc - Siocon dialect of Subanun. Mindanao, Philippines.

Tag - Tagalog. Central Luzon, Philippines.

Tao - Taokas. Taiwan.

Tau - Tausug. Sulu Is., Philippines.

Tbl - Tagabili. Mindanao, Philippines.

TbwA - Aborlan dialect of Tagbanua. Palawan, Philippines.

TbwK - Kalamian. Kalamian Is., Philippines.

Tha - Thao. Taiwan.

Tir - Tiruray. Mindanao, Philippines.

Tsc - Tsouic subgroup.

Tso - Tsou. Taiwan.

Tsw - Tonsawa. North Sulawesi, Indonesia.

Ttm - Tontemboan. North Sulawesi, Indonesia.

WBM - Western Bukidnon Manobo. Mindanao, Philippines.

Yam - Yami. Botel Tobago I., Republic of China.

The Position of the Austronesian Languages of Taiwan within the Austronesian Group

John Wolff

Cornell University

There is clearly much by way of phonological, morphological, syntactic, and lexical innovation that would support the proposition that the Austronesian languages of Taiwan are close to the Philippine languages, especially those of northern Luzon. This paper examines the commonalties among selected Philippine languages and Austronesian languages of Taiwan, evaluates these shared elements for status as inheritance from Proto-Austronesian or as shared innovations, and draws conclusions as to the implications of these facts for our understanding of the place of the Austronesian languages in Taiwan within the Austronesian family. The emphasis is on commonalties which are shared by the Philippine and Taiwan Austronesian languages, but which are lacking elsewhere, for it is these features which most strongly argue that the Philippine and Taiwan Austronesian languages are especially close.

Introduction

The Austronesian (AN) languages of Taiwan are thought to be aberrant and very much different from all the other AN languages. On the basis of this belief Blust, Dahl, and others have proposed that the AN languages outside of Taiwan are in one group and the AN languages of Taiwan are in one or more other groups (Dahl 73, Blust 77). This view seems to have been widely accepted, although not by everyone. Dyen (1965:269) does not

present this view and suggests that the AN languages of Taiwan are in a subgroup together with other western AN languages as opposed to the eastern AN languages, a view which is supported by Tsuchida (1976:13-14). There are several arguments which have been adduced for the conclusion that the AN languages of Taiwan are an early off-shoot (or that all the other AN languages have split off from the AN languages of Taiwan). These arguments are of three types: (1) phonological, (2) morphological, and (3) lexical. The phonological arguments center around the retention of the putative phonemes *C and *N which were thought to have been lost in all the AN languages outside of Taiwan (and also in some of the AN languages in Taiwan), and the retention of *s as a sibilant (my symbolization - this is usually written *S in the literature and was written *h by Dempwolff).[1] I discuss these in §1.ff., below. The morphological arguments have to do with a small number of morphological features in the pronouns which are thought to continue the Proto-AN (PAN) state of affairs (Blust 1977). I discuss these in §2.1. The lexical arguments have to do with the low number of cognates between the AN languages in Taiwan and outside of it. The bulk of this paper is aimed at refuting this agument (§3.1ff.). There are a couple of other developments which Blust (1992) believes support a theory that all the AN languages outside of Taiwan form a subgroup. We shall look at these arguments in detail in §2.2. The low amount of cognition between the AN languages of Taiwan and other AN languages is a chimera, due principally to the comparatively skimpy amount of data which are available for the AN

---

1   Dahl quotes other putative phonemes to support his belief that the AN languages other than the AN languages of Taiwan form a subgroup, but those are not well reconstructed, and until their existence can be proven with some certainty, they can have no place in the argument.

languages of Taiwan. As we shall see in §3.1ff, a rather haphazard search through the small amount of lexicographical information available on the AN languages of Taiwan reveals that the number of cognates between the AN languages of Taiwan and other AN languages is a great deal larger than originally thought and is quite comparable in amount to the number of cognates which have been found between the Oceanic languages and the AN languages outside of Oceania. This lays to rest any theory which proposes that the AN languages of Taiwan evolved in splendid isolation from the mainstream of Austronesian through the thousands of years since the times of Proto-Austronesian. We shall examine these data for clues as to whether they are innovations or retentions and test the thesis that the Philippine languages and the AN languages of Taiwan form a subgroup.

1. Phonology

There are three features of phonology which represent shared innovations not likely to have developed independently and which argue strongly that the languages which share them formed a subgroup. Since these three phonological features lead to conclusions as to subgrouping which are incompatible with the lexical and morphological arguments (§3ff), I draw some conclusions about dialectology of PAN (§4, below).

1.1 *C

It was believed that there were two PAN phonemes symbolized *C and *t. In a paper (Wolff:91) I argue that the distribution of the reflexes of *C and *t are complementary. In certain AN languages of Taiwan *t developed affricate allophones in two environments: in monosyllabic roots and in medial

and final position in roots with an iambic stress pattern on the first two syllables (that is, disyllabic roots with short vowels, and in trisyllabic roots with long vowels). This rule is carried through in all the languages which reflect this development, although there are lexical items in individual languages or dialects which show exceptions. In view of the variability of the accentuation of roots and the scope for analogical reformation of the accent pattern of the root, the exceptions to the rule in individual lexical items are explainable in terms of accentual variation or analogical reformation. The specific and idiosyncratic nature of the rule for the distribution of the allophones *C and *t argues for a one-time innovation in a single language. All of the AN languages of Taiwan for which we know the reflexes except Bunun, Amis, Kavalan, Siraya, Basay, Ketagalan, and Yami show the development of *C - that is, Saisiat, Atayal, Sedeq, Pazeh, Thao, Tsou, Kanakanavu, Saaroa, Rukai, Puyuma, and Paiwan share this innovation, the development of affricate allophones of *t which then became phonemicized, a development which may be symbolized *C.[2]

---

2   There are cases of forms which undergo this change in certain languages but the cognate in certain other languages which normally make this change fails to make this change - that is, for certain forms in which a change of *t to *C took place in most of the *C-innovating languages the change failed to take place in one or two of the C-innovating languages. The best explnation for such exceptions is as remnants of forms which failed to make the change - that is, one assumes a scenario whereby this development of affricates proceded root by root and failed to affect all roots. The exceptions are the following roots: *taqi "feces" *tasiq "sew", taŋiʃ "weep" tiŋas "food particles in teeth", where the Tsou, Kn and Sar reflexes show stop reflexes and *tebus "sugar cane" where a Rukai Budai and Mantauran dialects show an affricate where other languages all show stop reflexes. (The citations can be found in Tsuchida 76, pp.149 and 151.)

1.2 N

As I showed in my paper 1993a *N is in fact the phoneme *ñ. In all of the Austronesian languages outside of Taiwan and in Yami this phoneme developed idiosyncratically in a way that cannot have happened independently.[3] In initial position in short vowel roots *ñ became /1/, and in long-vowel roots it became ñ (and often subsequently n). In medial position in short vowel roots *ñ became /n/ and in long vowel roots became /ñ/ (and subsequently /n/ in many languages independently). In final position *ñ

---

3   The evidence that this change took place in all, and not just some, of the AN languages outside of Taiwan is slender. I have found only two forms in Oceanic languages which show reflexes of *ñ- in roots with a short penult for which there is a cognate in one of the AN languages of Taiwan:*ñiʃesaq "nit" and ñuwang "hole", and the reconstruction of both of these forms offers difficulties. The form for "nit" is attested to by the following forms: Tg lisaq, Cb lusaq, Pai -liseqes, Puy liñsa, Futuna liha "nit", but both of the forms from Taiwan show exceptions which need explanation. The word for "hole" is attested to by the following forms: Puy luwang "hole" Cb luwang "bilge" (in ships), Old Javanese luwang "hold" Futuna lua "hole, pit". Because there are many forms meaning "hole" with similar sounds, it is very likely that these resemblances have nothing to do with inheritance in this case, but are coincidental. Thus the evidence is very slender that any languages other than the western Austronesian languages shared this innovation of *ñ- to 1- before unstressed syllables. There is one form which provides counter-evidence (namely, indicates that the Oceanic languages did not share this innovation). That is the word for a kind of fish with poisonous dorsal spines *ñ epuq Tag lūpuq "black fish with poisonous spine", Malay ikan lepu Tongan nofu "fish with poisonous spine". On the basis of this form one could conclude that Oceanic languages reflect ñ- with ñ rather than 1-, but it may be that the Oceanic forms reflect a root with long vowel penult (which is reflected as /ñ/ and often subsequently /n/ everywhere in the extra-Taiwan languages) whereas the Philippine and Malay forms reflect a short vowel penult (in which case the Tagalog word is in fact a Malay borrowing, as many fish names are, and not an inheritance from PAN). In summary we cannot say for sure that all of the extra-Taiwan languages share the idiosyncratic innovations involved in the development of PAN *ñ. This change is reliably attested only for the western languages.

became /n/ (and became lost in some of these languages independently). The change in initial position and medial position is of a sort that cannot have taken place independently. This argues strongly that all the extra-Taiwan languages (including Yami) form a subgroup. On the other hand in the remaining languages of Taiwan except for Kavalan, Bunun and Kana-kanavu (Atayal-Sedik, Tsou, Saaroa, Saisiyat, Thao, Rukai, Paiwan, Amis, Pazeh, and Puyuma)[4] *ñ sifted to a palatalized lateral [t] in all positions and subseq-uently merged with other phonemes in some languages. This shift of *ñ to [t] may possibly be a shared innovation. However, it may also be inter-preted as a shift which spread after the split-up of the proto-language. A third possible interpretation is that the PAN phoneme was in fact *ɬ and that the changes subsequently in the extra-Taiwan languages were develop-ments from *t (as were the changes in Kavalan, Bunun, and Kanankanavu).

1.3. *s

PAN *s (usually symbolized *S) remained a spirant in all the AN languages of Taiwan except Puyuma, Saaroa, and Yami. In these languages and in all other Austronesian languages *s became *h and subsequently was lost independently in many of them.[5] Although the change of s to h is not unusual and may well have happened independently, one fact argues for the hypothesis that this is in reality an early change which was shared by all or most of the extra-Taiwan languages: there are no remnants anywhere over the entire spread of the AN languages outside of Taiwan of a failure to make this change. Although this change also took place in Puyuma and

---

4 I do not know what the developments of *ñ in the dead AN languages of Taiwan were.

5 *s was lost in Saaroa except in a few lexical items. Where *s was not lost in Saaroa, Tsuchida reconstructed a new phoneme, which he symbolized *S₆ (p.160).

Saaroa, it is almost certain that the change was independent in these two languages. In fact, both of these languages show a few remnants in which the *s was not lost. Further, Amis, which normally reflects *s with a spirant /s/ in a few forms reflects *s with /h/.

2.1 The PAN Pronouns and Subgrouping Hypotheses

Blust 1977 adduced materials from the pronoun system of the AN languages to show that AN languages should be grouped into one or more groups in Taiwan and a "Malayo-Polynesian" group which comprised all of the languages outside of the AN languages of Taiwan (and also Yami, which is clearly closely related to other languages of the Batanes Straits). His argument to support this grouping in a nutshell was that the extra-Taiwan AN languages innovated in creating the second singular genitive form *mu*, which is found widely distributed throughout Austronesia with the exception of the AN languages of Taiwan (but where it is absent in other AN languages, it was presumably lost). In other words, the creation of -*mu* is an innovation of a kind that could not have been made more than once, and when this innovation was made, the languages which participated must have been a single language. Two questions arise about this argument: first, was -*mu* in fact created subsequent to PAN times, or was it in the proto-language and subsequently lost? Second, if -*mu* is indeed a later creation, why can the creation not have taken place independently in many languages more than once? We may note that any one of these three possibilities could explain the forms and their distribution in the attested languages: (1) that -*mu* arose once as an innovation in an ancestor of the extra-Taiwan AN languages or (2) that -*mu* was a PAN form lost independently in many of the AN languages, including all of those of Taiwan or (3) that -*mu* was created independently

many times. Blust argues for the first hypothesis - that -*mu* was created after PAN times and that this happened only once. His reasoning is that there is a basis for believing that originally -*mu* was a second person plural pronoun, and since the shift to singular in meaning is of a nature that it cannot have happened independently, the creation of -*mu* was a one-time thing. Blust's argument depends crucially on the hypothesis that -*mu* was plural originally. However, he can provide no evidence to substantiate this hypothesis other than the following argument: in the PAN pronoun paradigm the genitive consisted of the last two phonemes of the nominative (as can be well substantiated on the basis of attested forms for the first person singular), and since the nominative of the second person plural was *kamu (or actually Blust believes *i-kamu), the genitive must have consisted of *mu. This argument does not constitute evidence, however, since a second person genitive plural shaped -*mu* is attested only in Puyuma. In Puyuma the pronominal paradigm is completely regular and clearly innovative in many respects, so that the -*mu* second person genitive plural is highly unlikely to be inherited and provides no evidence for a -*mu* genitive plural at an earlier stage. In short, it is nothing more than a hypothesis that the genitive in all persons consisted of the last two phonemes of the nominative, even if on the basis of attested forms one can reconstruct a genitive consisting of the last two phonemes of the nominative for some of the pronouns.

Thus there is no basis for the argument that the first hypothesis - that -*mu* arose once as an innovation in an ancestor of the extra-Taiwan AN languages, offers the only possible explanation for the absence of -*mu* singular in the AN languages of Taiwan. There are two other possible hypotheses to explain the fact that -*mu* singular does not appear in the AN languages of Taiwan. Both of these explanations are perfectly reasonable and

hardly require one to <u>assume</u> a certain paradigm for the proto-language not reconstructible on the basis of existing forms in the daughter languages. The first is that the form *-mu* was lost at a late point in the AN languages of Taiwan after the break-up of PAN. Since in fact these languages share a number of innovations, it is not an impossible scenario for the second person genitive singular to have gotten lost at a late point in these languages. The second possibility is that *-mu*, the second person gentitive singular, developed independently in various AN languages by the normal kind of process which Blust proposed for the third singular genitive (Blust 77:§3), whereby the third singular *ña* developed from earlier *niya quite independently in many languages by the process of loss of unstressed vowels. By this hypothesis *-mu* could have developed independently in many languages from the genitive former *m, which certainly can be reconstructed for PAN[6] plus the reconstructed second person pronoun singular root *su,[7] *s became *h* or disappeared entirely in AN languages outside of Taiwan, so that a change of *msu>*mhu>*mu*, is natural, and it is not unlikely that such a change could have occurred independently in widely separated regions.

To sum up: Blust's arguments do not substantiate the subgrouping of the extra-Taiwan languages: there are three possible hypotheses for the non-occurrence of *-mu* as second person singular genitive in the AN languages of

---

6   Blust rejects the reconstruction of *m- as a genitive as "ad hoc" (Blust 77:§6). However, genitives with *m-* are widespread, not only in the AN languages of Taiwan but also in other western AN languages (e.g. second plural genitive *mamu* in the Malay of the Srivijaya inscriptions). They are also found in Celebes. It probably is possible to find traces of this genitive elsewhere as well (and not just the *-mu* second singular genitive).

7   According to the traditions which have developed in the literature, this is transcribed *Su, but I am here following my own system as explained in the second chart of the appendix.

Taiwan, and a hypothesis which makes the creation of -*mu* a shared innova-
tion of the extra-Taiwan languages is not better substantiated than one which
states that it occurred independently or one that states that the innovation
was in the AN languages of Taiwan when they lost -*mu*, the second person
singular pronoun.

2.2 Other Phonological Arguments

Blust 1992 adduces further phonological arguments. These are (1) that
PAN *e before final s (his transcription) merges with *a (1992 §5.1.4). This
is based on two words, the word for "louse egg", which he reconstructs as
*lisequS, and the word for body louse, which he reconstructs as *tumeS. In
the case of the word for "louse egg" the occurring forms can be best
accounted for by a reconstruction *ñiqesaq (I write "s" for the phoneme
written in the literature as "S")- i.e. this form does not exemplify the rule.
As for the reconstruction *tumeS, Blust cites Kavalan /tuməs/ Amis/tumus/
Puyuma/Tumuh/ and Itbayaten/tomah/.[8] However, there is no reason to
assume that the Amis, Puyuma and Itbayaten forms reflect *tumes (or
tumeS). Puyuma -h reflects *-g not *-s, and it is not sure that Itbaytaten -h
reflects *-s; Amis /u/, Puyuma /u/, and It bayaten /a/ do not reflect *e. It
is circular to assume a specific development for PAN *eS and then say that
this assumed development proves a specific subgrouping. This leaves Kuvalan
/tuməs/. This is a language about which we have very little information and
about which the phonological history has been worked out in no more than
most general terms. It is certainly premature to say that this form provides

---

8   The data are such that one suspects that the forms in southern Taiwan and
    Batanes languages are not directly inherited (although indeed they must be from a
    form which in some language was inherited from PAN).

evidence for shape of the PAN form, especially since this word in neigh-boring languages can be shown to be of secondary spread. In short there is little evidence for reconstruction *tumes instead of *tumas or *tuma. (2) The second phonological development is the occurrence of a disyllabic in the form meaning "pandanus" whereas the AN languages of Taiwan show a trisyl-labic. Blust (1992 §5.1.6) quotes *paNuD₄aN. I would reconstruct this as *paŋudañ, and the PAN form probably was trisyllabic. However, the change of a trisyllabic to a dissylabic is a normal change which can happen quite independently again and again, it has in fact happened in other roots. It is hardly an innovation of the sort that must have been a common develop-ment. (3) The final argument adduced is that the form for "four" reflects initial *ʃ[9] in the AN languages of Taiwan, whereas the other languages reflect no initial *ʃ (Blust 1992 §5.1.5). The initial *ʃ in the AN languages of Taiwan must reflect a prefix of some sort, for otherwise initial *ʃ is retained in many of the AN languages outside of Taiwan. I would suggest that the initial *ʃ in this root in the AN languages of Taiwan is cognate with the morph *ʃa "one" and that the AN root is *pat. The initial vowel, which is found in many of the attested languages, is the product of the normal disyllabification process which is found with many monosyllabic roots, and which often has the form of a prefixed /ə/. E.g., Malay *empat* can be explained as the normal development of *pat (with the commonly occurring sporadic nasalation) and thus is perfectly cognate with Javanese *papat* and *pata* (as in *pata-ng-puluh* "forty") which are dissyllabified forms of *pat as well. Thus, the development of a prefixed form in the AN languages of Taiwan is

---

9 I write *ʃ for the PAN phoneme which others write with "s" and Dempwolff wrote "t'" (because I use the symbol *s for the PAN phoneme which others write with "S").

secondary. It could serve as evidence for subgrouping of the languages which show this innovation, although secondary spread is very possible in this case.

3. Commonalities between the Languages
of the Philippines and the AN Languages of Taiwan

So far we have argued that some of the AN languages of Taiwan show shared innovations in phonology and that all extra-Taiwan languages also show innovations in phonology (although we argued in §§2.1 and 2.2 above against some of the hypotheses of specific changes which were proposed). However, there is also an extensive amount of data which argues the other way around: namely, that the AN languages of Taiwan and the Philippine languages belong in a single group. The large number of commonalities between the AN languages of Taiwan and the Philippine languages suggests that the AN languages of Taiwan were by no means isolated as has often been thought. The low amount of cognition between the AN languages of Taiwan and other AN languages is an artifact of the comparatively skimpy amount of data which are available for the AN languages of Taiwan. As we shall see in the following sections, a rather haphazard search through the small amount of lexicographical information available on the AN languages of Taiwan reveals that the number of cognates between the AN languages of Taiwan and other AN languages is a great deal larger than originally thought and is quite comparable in amount to the number of cognates which have been found between the Oceanic languages and the AN languages outside of Oceania. These data make it impossible to believe that the AN languages of Taiwan evolved in splendid isolation from the mainstream of Austronesian through the thousands of years since the times of Proto-Austronesian. We

shall examine these data for clues as to whether they are innovations or retentions and test the thesis that the Philippine languages and the AN languages of Taiwan form a subgroup. There also are features of syntax and morphology which are shared by the AN languages of the Philippines and Taiwan. These commonalities could be due to inheritance from PAN, and indeed, some of these are to be attributed to inheritance. However, as we shall see in the following sections, it is likely that some of them are innovations which were shared between the AN languages of Taiwan and the languages of the Philippines. A good number of these commonalities (especially lexical items) were features which spread through borrowing, but some of these are not likely to have been borowings, and insofar as they can be shown to be neither spread by borrowing nor to be inheritances from PAN, they provide proof of a close relation betweem the AN languages of Taiwan and Philippine languages.

3.1 Lexical Commonalities between the AN Languages of Taiwan and the Philippine Languages

The number of forms which are found in one or more languages in he Philippines[10] and in one or more languages of Taiwan, but not outside of these two sets of languages, is appreciable. I can give some data based on the material which I have so far in the data base of citations from the literature

---

10　It would be foolish to restrict the concept of "Philippine languages" to the national area of the Philippines. Clearly the languages of adjacent regions of Indonesia, Malaysia and even of Brunei, have so much in common lexically, grammatically, and semantically with the languages of the Philippines that one should include them when one talks of the "Philippine languages". Any delimitation of the concept "Philippine languages" in terms of a given set of features would be arbitrary. I include data from Sangir-Talaud and N. Celebes as far west as Gorontalo and data from Kadazan in Sabah.

on AN languages. I am attempting to confine this data base only to forms which are of some antiquity, that is, they are assignable to PAN or at least to one of its first-order subgroups. I have so far inputted 1450 forms. This probably represents 90% of the total number of forms which I will be able to have in the data base in the end. Of these 1450 perhaps 500 will have to be discarded as being forms which have spread from language to language by processes of borrowing or which are in fact confined to a group which is more recent than a first-order subgroup. Of these 1450 some 1200 are found either in the AN languages of Taiwan or the Philippine languages (or in both). 470 of these are cited in the literature for the AN languages of Taiwan and 800 of these are attested in the literature for the Philippine languages. 420 of these occur both in the AN languages of Taiwan and in the Philippine languages. Of these 420 I have found 65 which are not attested elsewhere except in these languages, and therefore these are part of the 500 of the forms now in my 1450-form data base which will have to be removed as not belonging to PAN or to one of the first-order subgroups. This list of 65 forms has been expanded by other citations in the literature, chiefly Dyen '92 and '93 or by forms which I have found in my perusal of the few small dictionaries available on the AN languages of Taiwan. This brings the total to approximately 100 forms to consider. Dyen and Tsuchida (91) state that they have found more than 400 such forms, but do not list them. If we add to these 100 forms an additional 38 forms which are found in adjacent languages as well as in languages of the Philippines and Taiwan, we may come up with a list which indicates even more persuasively that there is a close connection between the AN languages of Taiwan and those to the south of them. We may now turn our attention to these forms.

3.1.1 Forms which Show Regular Correspondences between the Phillipine and the AN Languages of Taiwan

A large number of the forms which are found in the AN languages of Taiwan and the Philippine languages but not outside of these languages show regular corespondences. The reglar correspondences as I see them are shown in Table I of the appendix. This table does not give all of the outcomes in all environments, but it is sufficient to demonstrate the correspondences which turn up in most environments and enable one to distinguish the forms which reflect regular correspondences from those that do not. The following list gives these forms.[11] Each group of forms is introduced by an asterisked form which represents the form of the proto-language[12] which gave rise to the forms attested. This proto-language was not PAN, for these forms do not have cognates outside of the Philippines and Taiwan. The substantial number of forms which the Philippine languages and the AN languages of Taiwan share is a strong indication (though no proof) that these form a subgroup within the AN family.[13] Even if cognates for some of these forms should

---

11 For the cognate sets which are well attested I do not attempt to give an exhaustive list of all forms found in the sources. However, for forms which are poorly attested, I have attempted to list most, if not all, of the forms which can be found in published sources.

12 The transcription of the forms given follows that of the sources. For those languages where there is a dictionary cited, we use the transcription found in the dictionary. For Amis we use Fey's transcription. Glottal stops are represented by ? except in the Philippine languages where it is transcribed with q. The reconstructed forms follow the phonology which I proposed in Wolff 88 with the revisions of Wolff 91, 93a. A table of correspondences between my transcription of Proto-Austronesian reconstructions and those found normally in the literature is given in the second table of the appendix.

13 The number of shared forms is substantial and they have the semantic character of forms that tend to be inherited. In view of the very skimpy information available for the AN languages of Taiwan, this number of cognates should be viewed

turn up outside the AN languages of Taiwan and the Philippines languages, the number remaining will still be substantial. The following chart shows the language abbreviations:

Am = Amis	Il = Iloko	PAN=Proto-Austronesian	Rk=Rukai(Budai)
At=Atayal	Kn=Kanakanvu	PMin=Proto-Minahasa	SL=Samar-Leyte
Bk=Bikol	Kp=Kapangpangan	PSan=Proto-Sangir	Tg=Tagalog
Cb=Cebuano	Pa=Paiwan	Pu=Puyuma	Ts=Tsou

*amen[14] "our"   *Pa* amen *Tg* ámin *SL* ámun "our"

*bakãng "bow-legged" (D.93)[15]   *Pa* vakang "bow-legged" *Cb* bakang "bow-legged"

*balatung "kind of bean"   *Pu* valatung "red beans" *Cb* balatung "string bean (Vigna sesquipedialis)"

*banaw "lake, pond, wash(?)" (D.93)   *Pa* vanaw "take bath" *Cb* banaw "for liquids to spread"

*basaw "cold left overs (?)" (D.93)   *Am* fasaw "cooled off (rice, anger)" *Tg* bahaw "cold victuals"

---

as very high. It is speculative, but nevertheless reasonable, to suppose that dictionaries more thorough than the brief word lists or limited monographs which we possess for these languages would bring to light a much more extensive list of forms connectable with forms in AN languages outside of Taiwan, and a fair proportion of these would no doubt be exclusively shared with languages of the Philippines.

14   The Paiwan and the Philippine forms are not necessarily inherited from a form in the protolanguage. Independent analogical development is possible in the case of these forms.

15   A reference in parentheses after the reconstructed form refers to the article or monograph where the cognate set is discussed: D.93 refers to Dyen 93; I, II, III, and IV refer respectively to Blust 80, 83-83, 88, and 89; T refers to Tsuchida 76.

*batu "throw, cast"(D.93) *Bunun* ma-batu "throw" *Tg* bato "thow, cast"

*beñut "pull out" *Pa* veɬuts *Cb* bunut "pull up grass, etc."[16]

*biliñ "behind, last" At s-bil "leave behind, abandon" *Saisiat* may-biil
(T.141) "later" *Ts* ua-frihi "last" *Rk* us-birli "follow" *Pa* viliɬ
"behind" *Cb* bilin "leave behind"

*biqel "goiter" (T.170) *Sedik* biqir *Proto-At* *biqer *Ts* fʔiri *Bunun* bihiq *Pu*
vaħer *Il* biqél "goiter"

*bunguq "head, skull" *Ts* fnguu "head" *Am* fogoh[17] "head" *Bunun* bunguh
(T.133) "head" *Tg* bungoq "skull"

*buñbuñ "down, body *Pa* -al-vuɬavuɬ+an "down, small feathers" *Cb* bulbul
hair" (D.93) "pubic hair" himulbul "remove feathers"

*buqul "swelling, lump" *Sediq* bql+it "leg" *Pu* v-en-uħur "get a swelling,
bump" *Tg* buqol "ankle" Cb buqul-búqul
"anklebone or bulge in neck like adam's apple"

*buʃuɣ "enemy"(III.58) *Pa* vutu *Bontok* bósol "enemy" *Ifugao* bohol "be
angry at" *Il* ka-búsor *Pg* bosol "enemy" *Bk* maka-
búsog "fight with"

*dapa "sole"(D.93) *Bunun* dalapa, dalapaʔ (North) "sole" *Cb* lapa-lapá
"sole" *Alangan* (*Mindoro*) darapa "foot"

*dapañ "sole of foot" *At* dapal "sole of foot" *Saisiat* rapal "foot" *Ts* caphə
(T.154) "sole of foot" *Rk* dapalə "foot" *Pa* djapaɬ "hind leg
of animal, thigh" *Pu* dapal "foot" *Il, Pg,* and *Bk*

---

16 There is an Am form *foqnot* "remove from ground", which looks connected
with these forms. However the /q/ is not explainable, and also *ñ is normally
reflected as /d/in Am.

17 /-q/ is the expected reflex of *q in Amis, not /-h/. This word is perhaps not
inherited in Amis.

dapan "sole"

*daw "quotative" (D.93) *Bunun* daw "they say" *Tg* daw "they say"

*debdeb "chest" (III.86) *Pazeh* zebzeb *Tg* dibdib *Hanunoo* dubdub "chest"

*duma "other" (T.153) *At* roma *Saisiat* ʔa-rumaq *Ts* mo-cmo *Rk* ḍma-né *Pa* zuma *Am* roma *Pu* na-zuma "other" *Il* agdūma "differ" *Ibanag* dumurumã "unusual" *Manobo* duma *Tboli* dumuh *Tiruray* dumo-n "companion"

*ɣawus "extract" (T.161) *Kn* m-ari-a-raūsu "scoop out with ladle having holes" *Il* gaó "draw out, extract cooked food from pot"

*ɣeken "coiled base for cooking pots" (IV 517) *Pu* ɣeken "pad to put pots on" *Bk* gukón "bamboo shield to put pots on"[18]

*indi "no" *At* ini? *Pa* ini "no, not" *Pu* ini-yan "not exist" *Tg* hindiq[19] *Hiligaynon* indiq "no, not"

*jañukap "sole, palm" *Pu* dalukap "sole" *Ibaloi* calukap "sole" *Il* dakūlap *Hanonoo* dalūkap *Sambal* dawkap "palm of hand"

*ka- "prefix for adverbs referring to past time" *Pa* ka-tiaw "yesterday" *Tg* ka-hápon "yesterday" *Cb* ka-gabíqi "last night"

---

18 I would reconstruct PAN *deken based on the following forms: Tg dikin, Kelabit reken "coiled rattan base on which cooking pots are placed when hot" Yamdena réan "woven palm leaf base on which hot cooking pots are placed" Kei raun "basis". The correspondence comes from Blust (I.370). The innovation here is the substitution of ɣ- for d- as the initial (probably through contamination with a word of similar meaning with initial ɣ- as illustrated by Kankanay gíken, Bontoc gíken "support coil").

19 The correspondences are not completely regular here: Tg has an unexpected initial /h-/ and both Tagalog and Hiligaynon show final /-q/. These irregularities are due to the special intonation this word receives, and the forms are connected historically.

"kamet "grab" *Am* kamet "grasp, grab" *Tg* kamit "get" *Cb* kamut "hand"

*katawan "body" *Pa* kinacavacavan *Tg* katawan "body"

*ki- "verb prefix" *Pa* ki-langeda "listen"[20] *Pu* ki-ngeɤ "listen" Tg pa-ki-nig "listen"

*lapek "rotten, broken into bits" *Am* lpek "break apart" *Tg* lapok "rotten" *Cb* lāpuk "mud"

*lebleb "inside room" *Am* loflof "inner roon of house" *Tg* liblib "hidden, secret" *Cb* lublub "inside, hidden"

*liñung "calm, tranquil" (I.279) *Pa* ma-liɬung "be not moving (water)" *Hanunoo* linung "peace, quiet"

*lum "ripe" *Pa* ma-lum *Rk* ma-ləmə *Ilk* na-luqum *Kankanay* na-lqum "ripe"[21]

*lupas "spit out" (IV.388) *Pa* lupas "something solid spat out" *Gaddang* maluppa "spit"

*maduq"grain"(T.165) *Ts* mcuu "ear or top of rice-plant or millet" *Rk* po-mdhoo "bear fruit" *Paz* maduq "fruit" *Cb* māluq "pollen of corn" *Il* maro "a variety of awned rice"

*mūging "face (forehead?)" (D.9) *Sediq* muhing *Ivatan* muying *Pai* muding+an *Pu* muding+an "face" *Bunun* muing "eyebrow" *Thao* mudin *Pazeh* muzing "nose" *Ilk* mūging "forehead"

---

20 The prefix *ki-* is productive in Pa, Pu, and Tg (and perhaps elsewhere) and some of the meanings of the affix in all three languages are comparable, eg. Pa *ki-tja-qata* "ask for assistance" Tg *pa-ki-tūlong* "ask for help" Pu *ki-liksa* "look for nits."

21 There is a Malay form *masak lum* "overripe". If the second word of this is connected the Philippine and AN forms from Taiwan, then this is an inheritance, and not a shared innovation.

*ngabngab "bite off (like meat off bones)" (II.247, IV.416)[22]   Am gafgaf "eat grass (cow)" Tg ngabngab "bite meat off bones" Bontok ngabngab "eat corn off cob" Kan ngabngab "take bite from" Ifugaw ngab-ngab "bite off a mouthful"

*ngalay "saliva"(D.10)   Ts ngroi Rk ngree Pu ngalay Am galay Itbayatan naxay "saliva"

*nguynguy "wail" (II.256)[23]   Am galiwgiw "complaining" Cb danguynguy "wail" Il ngoyngoy "whimper"

*ña "already"   At la Pa ła Pu la Tg na Cb na "already"

*ñibuq "den, lair" (T.142)   At libū? "pig pen" Saisiat libu? Kn niivu Rk libūu Pa łivu "den of wild pigs" Pazeh libu-patakan "bamboo-fence around garden" Tg ulboq "pigpen"

*paliq "spleen" (T.171)   Kn páali "bile, gall-bladder" Tg pāliq Pg pāli Kp pali "spleen" Il pāli "pancreas"

*paqaq "chew (without eating)" (IV.425)   Am pa?pa? "chew" Cb paqpaq "bite or chew on something without eating it" Maranao papaq "bite, chew"

*pilek "eyelash" (D.14)   Rk s-kərpə "eyelid" Am sa-pelek Tg pilik "eyelash"

*pulu "handle of   Pa pulu "handle (as of hoe)" Cb pūlu "handle of

---

22   Blust (89a:154) connects Minangkabaw ngangap "snap at flies" with this root, which is certainly correct and indicates a PAN origin, at least of *ngab. However, I list this here as a commonality between the AN languages of Taiwan and the Philippine languages because there is a common semantic development which the Minangkabaw form does not show.

23   Blust (83-4:90) compared Malay ngoingoi with this, but it certainly is not a normal word (I have not found it listed in any Malay dictionary), and the resemblence with these forms is coincidental, or perhaps it is from a Malay dialect in contact with one of the Philippine languages. Further, oi is not a normal sound sequence in Malay and is confined to words containing sound symbolism.

instrument" (III.244)   bladed instruments" *Proto-Sangir, Mongondow* pulu "handle (as of axe)"

*punuq "head" (T.172)   *At* puqing "beginning" *Ts* pnuu "brain" *Rk* t-punú-a "forehead" *Pa* punuq "brain" *Am* pono? "brain" *Pu* punuh "brains" *Tg* púnuq "chief"[24]

*qadeng "mole" (D.19)   *Bunun* hadung *Bikol* árung "mole" Itbayaten qadeŋ "large mole"

*qaleb "knee, shin(?)"   *Bunun* (*Central*) qaa? *Magindanao* aleb *Maranao* leb "knee" *Bunun* (*South*) haab "shin"

*qamis "north" (T.160)   *At* qmis-an "winter" *Saisiat* ?ameʃ-an "winter" *Kn* ? amis-anə "year" *Pa, Am* ?amis "north" *Pu* hami "year" *Tg* amih-an "north" *Cb* amihan, *Il* ami-án, *Pg* ami-an, *Bk* amih-an "north wind"

*qañipúgu "whorl on   *At* (*C?uli?*) qalipugu? *Saisiat* ka?alipoza?an *Pa*
head" (II.6)   qulipapudu-an *Pu* haripudu *Cb* alimpúlu *Bk* alim-púpuro "hair whorl"

*qañup "hunt" (T.140)   *At* qālup *Ts* hup-a *Rk* u-lupu *Pa* qmalup *Am* ?adup *Pu* halup *Il* anup *Pg* anóp "hunt"

*qati "lower water levels   Kn ?umá?aci "dam up stream to catch fish" *Rk*
(to catch, fish etc.)"   u-cii "dam up side stream" *Bunun* ma-hciq "dam
(T.160)   up a side stream to catch fish" *Tg* k-āti *Bk* ati "low tide" *Pg* āti "decrease water in rivers"

*qebel "smoke" (T.172)   *Kn* ?əəvə *Rk* əbələ *Tiruray* kíbĭl "smoke"

*qetuɤ "sexual arousal"[25]   *Am* ?etol "sexual arousal, sexual passion" *Cb* utug

---

24   Blust (92:15) suggests that this form is connected with M1 pohon "base of tree", in which case it does not belong in this list.

25   Because of its meaning, this word is not likely to appear in dictionaries. For this

"erection of penis"

*qusaw "thirst" (II.440)   *Pa* qusaw *Am* so?aw *Pu* ħaw *Tg* uhaw *Cb* uhaw "thirsty" *Ivatan* ma-qwaw *Maranao* oao "thirsty"

*qutus "stalk, stem"   *Pa* qutsus "stalk, stem" *Cb* uhut "straw of rice or other grain plants"

*sangetej "arrive, up to"   *Pa* pangetjez "arrive" *Cb* hangtud "until"

*sapsap "feel by groping" (IV.593)   *Kn* mati-sapəsapə *Rk* o-spəspə *Am* sapsap *Tg* apuhap *SL* hapuhāp *Maranao* apap "feel around for, grope" *Cb* hapúhap *Bk* hapihap "stroke gently"

*sawsaw "wash, rinse" (iv.594)   *Am* sawsaw "wash (generic term)" *Tg* hawhaw "rinse clothes"

*sikam "sleeping mat" (T.160)   *Kn* sikāmə "mat woven from wild ginger leaves" *Pg* ikam-en "mat woven from buri leaves" *Il* ikam-en "sleeping mat"

*simaɣ "oil" (I.409)   Ts simro "fat, oil, salve" *Proto-Rk* sima?a "fat" *Pa* sima "raw animal fat" *Am* simal "oil, grease" *Pu* imaɣ "grease, fat" *Cb* himag "kind of oil for healing purposes"

*sinaw "wash" (T.146)[26]   *At* sminaw *Kn* mari-sināu "wash utensils" *Rk* usnisnoo "wash clothing" *Pa* smenaw "wash hands" *Pazeh* me-senaw "wash utensils" *Tg* hinaw "wash hands" *Cb* hunaw *and* hinaw "wash hands" *Bk* hanaw "wash hands or feet"

---

reason, there may be reflexes of this form in existence in many other languages, and it may in fact be inherited from PAN.

26   Some languages reflect *senaw instead of *sinaw.

*siwid "aslant" (IV.600)   *Am* siwit "slant, at an angle" *Tg* hiwid "out of line, at an angle"

*suɤsuɤ "thread a
    string"   *Am* solsol "string together, run drawstring through" *Cb* halughug "drawstring in the waist"

*sūɤut "pull, draw"
    (IV.606)   Am solot "pull out, draw sword, etc." *Tg* hūgut "unsheathe" *Cb* hūgut "pull rope to make tight" *Bontok* ólot "remove rice from panicle by pulling" *Han* hūgut "pull up (loincloth)"

*suqeyab "yawn" (III.334) *At* suyab *and* swab *Pa* mi-suaw *Am* sowaf *Cb* huyqab *Proto-Sangir* kioyab/koyab *Itbayaten* qahua *Ivatan* uab *Il* wab *Pg* wab *Kp* uyab *Kadazan* uvab *Mongondow* uwab *Gorontalo* huwabu "yawn"

*suqsuq "pour, drain
    out"   *Am* soso? "let drip out, drain" *Tg* huhuq "flowing off of grains of pieces of things" *Cb* huqhuq "empty container by turning upside down"

*ʃabay "together"   *Am* cafay *Tg* sabay "together"

*ʃakal "restraining device *Am* cakal "muzzle for ox, buffalo" *Bk* sakal "yoke"
    for animals" (IV.533)

*ʃakʃak "chop up"
    (IV.538)[27]   *Am* cakcak "cut into pieces, hoe ground to till it well" *Cb* saksak "chop into pieces to mix it into something"

*ʃāleng "resin producing *At* hayung *Saisiat* haəng *Ts* srongə *Rk* srəngə *Pa*

---

27   Blust (89a: 163) connects Malagasy *sasaka* "half" with this, but on semantic grounds this is not likely. He also suggests that the forms which Dempwolff listed under his reconstruction *t'akt'ak "stab, stick" may be connected, but there is no meaning connection between the forms here cited and the Javanese and Malay forms which Dempwolff cites to attest to *t'akt'ak. Otherwise, Dempwolff cites only a Tg form, which may indeed be conneced with these forms.

tree"(T.127)

taleng Am caleg "pine" *Bunun* sãang "torch" *Tg* sã hing *Cb* salong "resin" *Il* saleng "pine" *Bk* salong "resin" *Kadazan* hasong "dammar"

*ʃaliw "buy, sell, exchange"(IV.542)

Thao thaariw "buy, sell" *Am* caliw "borrow, lend" *Cb* saylu "go beyond something" *Ivatan* mapa-saliw "sell" *Ilonggot* taiw "buy" *Ata, Proto-Sangir* saliu "trade, exchange" *Mongondow* mo-taloy "buy"

*ʃãñaɤ "come out(sun)" (IV.532)

Am cahdal[28] "come out between showers (sun)" *Cb* sanag "bright (of countenance)" *Hiligaynon* sãnag "bright, shiney (as sun)"

*ʃapiñ "pad, patch"

Saaroa sapili? "patch" *Tg* sapin "padding"

*ʃaqʃaq "flatten bamboo and split it" (IV.544)

Pa taqtaq "split bamboo" *Cb* saqsaq "split bamboo"

*ʃegep "enter" (II.359)

Pa tedep "enter house" *Kanakanay* segep *Gorontalo* tolopo "enter, be inside"[29]

*ʃibu "urine" (T.128)

Saisiat kæ-hbu? *Ts* sifu *Il* isbu "urine"

*ʃiwʃiw "chick"

Am ciwciw *Tg* siwsiw *Proto-Minahasa* sisi "chick"

*ʃukud "walking stick" (III.317)

Pa tukuz-an *Am* cokor *Cb* sungkud *Maranao* songkod *Kadazan* sukud "staff, walking stick"

*ʃuñud "follow" (T.155)

Ts mou-su-suhcu "move forward step by step" *Bunun* mu-sunu? "next" *Tg* sunod "follow" *Cb* sunud "follow, next" *Bk* sunod "next"

---

28 The /h/ in the Am form is not explained. Perhaps the reconstruction should be *ʃaseñaɤ where *s is reflected by /h/in Amis, as is the case in other forms where *s is reflected by /h/instead of the more frequent /s/.

29 It is not entirely certain that this form is an innovation. Ml *serap* "seep in" may be cognate and Blust lists some possible Oceanic cognates. In that case, this form is a shared inheritance from PAN.

*ʃupʃup "suck" (IV.588) *Am* copcop *Maranao* sosop *Bk* supsup "suck"

*taɤa "wait" (T.232) *At* naga? *St* may-naa? *Ts* moo-tro *Pazeh* maxu-taxa?
Am mi-tala *Pu* mal-ta-talah *Il* tága *Kp* ma-naya
"wait"

*takid "adhere" (T.182) *Kn* maa-takici *Kp* takid "adhere"

*taktak "detach and fall Ts m-o?to?ə "clear land for swidden" *Pu* thakthak
of"[30] (T.220) "partly come off" *Cb* taktak "detach and fall off"
Kp taktak "cut grass with a bolo"

*taliuk "go around, *Am* taliyok "perimeter, go around" *Cb* taliyuk
circumference" (IV.621) "turn body around"

*tanglad "Andropogon *Pa* tsanglad "miscanthus grass" *Cb* tanglad
citratus" "Andropogon citratus"

*tapuʃ "finish"(IV.628) *Favorlang* tapus "finish" *Tg* tápus *Manobo* tapus
"the end, finish, complete"

*tasaw "open area" *Pa* tsasaw "outdoors, outside" *Tg* táhaw "clearing in
(IV.129) forest", tahaw "open to view" *Il* taáw "sea, ocean"
Pg taéw "middle of river, ocean, deep sea"

*tebteb "cut through the *Pu* tebteb "chop" *Tg* tibtib "sugarcane cuttings
base" used for planting" *Cb* tubtub "cut at the base
through something cylindrical"

*tenek "thorn"(III.73)[31] *Proto-Rk* cənəkə "thorn" *Tg* tinik "thorn" *Cb* tunuk
"thorn"

---

30 Tsuchida (220) connects Fijian *taa* "chop" with this root, in which case *tak is
PAN, but in reduplicated form with this meaning, it is clearly an innovation in
the AN languages of Taiwan and the Philippine languages.

31 Blust (88:32) connects Fijian *tono* "pierce, probe, poke", and if it is connected,
the development to the meaning "thorn" is an innovation in the AN languages of
Taiwan and the Philippine languages. However, the resemblance between the
Fijian form and the others is probably coincidental.

*téñuq "coconut milk" (T.215)
Kn tənəʔə "resin, sap" *Rk* tulu "resin, sap" *Cb* tunuq "coconut milk or milk of other fruit" *SL* tunu "thick material as laundry starch or coconut milk"

*tuku "prop, post" (IV.677)
Am toko "center beam of house or ship's mast" *Cb* tuku "prop up, post"

*tumek "pulverized" (IV.134)
Pu ma-thumek "crumbling, falling to pieces" *Kankanay* tumék "pounded, crumbling" *Bontok* tómek "crush (stone), crumble rice"

*tūqed "trustworthy, true"
Am tiʔer[32] "trust, trusworthy, faithful" *Cb* tūqud "true"

*uni *or* sawuni "later" (T.270)
At saoni "a little while ago, today" *Ts* osni "immediately, right way" *Pa* sawni "a short while" *Am* anu-sauni "after, later" *Cb* uny-aq *SL* uni-na "later" *Ifugao* qawni "wait! later!"

*uñay "something embedded" (IV.705)
Pa ułay "something broken off inside something else (thorn in flesh)" *Cb* unay "integral part of something or sliver embedded"

*-ūseɤ "nasal mucous"
Bunun ngusul *Itbayaten* ŋuhey *Samar-Leyte* mūhug "nasal mucus"[33]

*witwit "swing outward and back" (T.145)
Ts reu-vtivti "move and swing tail or ears (of animal)" *Pa* vari-vitjivitj "dart back and forth (as fish, wagging tail, etc)" *Tg* witwit "shake finger in scolding" *Bk* witiwit "swing ride in carnival"

---

32 The /i/ in Am is not explained. Cf. *qūsung and *ɤamut.

33 The non-corresponding initial consonants can be explained by the fact the these forms are probably petrified verb forms.

3.1.2 Shared Forms which are not Inherited but have Spread Secondarily

Some of the forms which are shared exclusively between the Philippine languages and the AN languages of Taiwan do not enter the sound correspondences of Table I of the appendix. These must have spread secondarily (after the times of the proto-language) through processes of borrowing. In addition there are other forms which are perfectly regular in correspondence, which I nevertheless do not believe to have been inherited from the proto-language. These are discussed individually.

The following forms refer to flora, fauna, goods and utensils. In some cases their sound correspondences are regular, but in those cases the meanings are such that they do not seem to be inherited but rather seem to have been affected by semantic changes which occur in contact situations.

*bakal "iron tool"(D.93) *Rk* bkalə "small knife" *Pa* vakaɬ "dagger" *Pu* vakal "single-edged knife"

*bangaw "noxious insect" (I.42) *Am* fagaw "bedbug" *Tg* bángaw *also* bangyaw "bottlefly"

*banuɣ "hawak" *Am* fanol "dove" *Cb* banug *Proto-Sangir* banegha "hawk"

*bañituk "gold" (T.143) *Kn* vanituku "money" *Pa* vaɬitjuq "any shiney metal object" *Tg, Ilk, Pg* balitok "gold"

*būqut "rabbit, squirrel" *At* bhut *St* ka-bhöt *Ts* vuūtu *Rk* būtu *Pa* vutj *Am* fohet *Pu* vut "squirrel" *Tg* būqot "rabbit" *Bk* boqot "mountain rat"

*dayap "kind of citrus" *Pa* djalayap "Citrus depressa" *Am* rarayap *Cb* dayap "kind of citrus"

*kamuɣaw "citrus" *Rk* (*Tana*) kamuraw *Pa* kamuraw "lemon, pomelo"

Am kamoraw "white bark fig, *or* pomelo" *Pu* kamuɤaw "pomelo" *Cb* kabūgaw "kind of citrus"

*qatipa "kind of turtle" (T.266)
At qsipa "soft-shelled turtle" *Ts* acipa "turtle" *Pa* sipa? "soft-shelled turtle" *Bunun* hansipa? "turtle" *Kp* antipa "kind of turtle" *Pg* ansipa "riverine turtle"

*ñateng "kind of vegetable" (T.150)
Ts natəngə *Rk* lcəngə *Pa* łatseng "vegetables" *Am* dateg "cooked vegetables" *Il* nateng "vegetables" *Bk* natong "taro leaves" *Proto-Sangir* nating "plant with edible leaves"

Some of the forms which have spread by borrowing through the AN languages of Taiwan and the Philippine languages are also found in languages outside these languages and usually originate from outside them.[34]

*bulaw "copper"
Am fodaw-an "copper, brass" *Pu* vulawan "brass" *Tg* būlaw "reddish colored" *Cb* bulāwan "gold" *Ngaju Dayak* bulaw "gold"

*kalaw "hornbill"
Pa kalakalaw "bird, sp." *Cb* kālaw *Subanon* kalau *Kayan* kalau *Murnaten* alakwe *Wolio* halo "hornbill"

*kudiʃ "skin disease" (I.232)
Pa kuris "itching sores which leave dark spots on body" *Am* kodic "skin disease" *Malay* kudis "scabies"

*pariuk "pan"
Pa paliuk *or* pariuk "pan" *Tg* palayuk "clay cooking pot" *Malay* periuk "pot"

*piray "sickness" (T.181)
Rk (*Manatauran*) ma-piłay "crippled" *Pa* piłay "lame" *Tg* pilay "sprain, limp" *Il* *Bk* pilay

---

34 I give only a few examples. Most certainly a large number of forms in this category could be turned up in a through-going search.

"crippled" *Malay* piray "kind of disease, tumor"

*tageɤang "chest cavity"   *Sedik* tekelang "ribs" *Ts* t?ornga "chest, breast" *Pa*

(T.227)   tagerang "trachea, larynx" *Pu* taheɤang "chest,

breast" *Itbayaten* tagrang *Pg* taglang, *Kp* tagyang

Miri tagreng "ribs"

There are other forms which are not of a semantic character that one would think extremely likely to have been spread by borrowing. They show irregularities in their reflexes and therefore should be taken to be connected by borrowing (or the resemblance in form and meaning is coincidental). I have noted the following forms which show resemblances but do not entirely conform (or do not at all conform) to the normal correspondences.

*alabuk "dust"   *Am* alafoh "dust blown by wind" *Tg* alabok "surface

dust, dust from the ground"

*bali "strong wind"   *Pa* vali *Pu* vari? "wind" *Tg* bāli-bāli "strong change-

(D.93)   able wind"

*dulem "dark" (D.93)   *At* yulung "cloud" (*Proto-Atayal.* *rulung) *Pa*

ve-lerem "cloudy day" *Pu* dema-durem "pitch dark"

Tg ma-dilim "dark"

*geqgeq "shake up"   *Pa* gesges "move, shake something" *Cb* luqguq

"shake up"

*ɤetɤet "cut through"   *Am* retret "cut through" *SL* getget "saw with

knife"

*kademel "thick"[35]   *Pa* keḍemel, kuḍemel *Pu* kezemer *SL* ma-dakmul

"thick"

---

35   Mahdi (75) connects Javanese *kandel* "thick" with this form, and if it is connected, this word is inherited from PAN. However, the /ḍ/ in Pa reflects *d, and the /d/ in Javanese reflects *j. The case is not completely clear, for Pa has

*lumaq "old"   *Pu* ruman *Tg* lumaq "old"

*saweliq "return"   *Am* coli *Tg* sauliq "return"

*ʃuled "enter"   *Am* coled "insert" *Cb* sulud "insert" *Maranao* soled
  (IV.583)[36]   "enter"

*[s,q]i[d,j]iɤ "leaning"   *Am* ʔirig "lean to one side" *Tg* hilig "inclined
                                 (mentally)" *Cb* hilig "inclined (physical position)"

*tañam "taste"   *At* tmalam *Saisiat* ʃan-talam *Ts* oo-thomə *Pazeh* mu-
                                 talam *Am* mi-tanam *Pu* tanam-un "taste" *Itbayaten*
                                 taxam-en "taste"

*teneb "submerge"   *Pu* tenep "submerge" *Il* táneb "submerge partly,
  (III.370)   immerse in part" ma-tneb "sink"

*tindeɤ "stand"   *Am* patireg "make stand" *Tg* tindig "stand" *Cb*
                                 tindug "stand"

3.1.3 Shared Forms which Show Sound Symbolism

Some of the forms which are shared exclusively by the AN languages of
Taiwan and the Philippine languages reflect sound symbolism. Some of these
words refer to sounds and others refer to smells. I believe that the words

---

three reflexes of *d (ḍ, dj, and z), and the conditions for their occurrence have
not been established. The reflex /ḍ/ occurs in only three other forms, whereas
/dj/ and /z/ occur as reflexes in many forms. Mahdi believes that the form
*kademel is a proof of the existence a contrast between PAN *d as opposed to
*ḍ, which Dempwolff reconstructed, but stronger evidence than this word has to
be provided to substantiate this contrast, and I believe this resemblance to be
coincidental.

36   Blust (89:166) connects roots meaning "plug up" in Roti and Sasak with these,
      but I do not think they are connected. The Am /-d/ corresponding to /-d/ in
      the Philippine languages also raises a problem. Either the Amis form is not direct-
      ly inherited or the resemblace between Amis and the Philippine languages is a
      coincidence.

which refer to smells are in fact formed by sound symbolism, as there are literally dozens of forms which refer to smells which are similar in form and found throughout Indonesia, the Philippines, and the AN languages of Taiwan. A few of these happen to correspond systematically, but most show irregular correspondences. The following list gives the forms which refer to sounds and smells which enter into regular correspondences and which are exclusive to the AN languages of Taiwan and the Philippine languages.

*angsu "smell of urine" *Am* ancoh, kagcoh "smell of urine"[37] *Cb* angsu "smell of urine"

*daɤing "make noise" (IV.142) *Ts* troe-crengi *Sar* taraa-saringi "moan" *Pa* zmaing "make noise, shout" *Bunun* dadaling "moan (when sick)" *Manobo* daging "noise, sound" me-zaging "loud, noisy" *Sangir* dahing "breathe loudly when asthmatic"

*kiskis" scrape off" (IV.297)[38] *Am* kiskis "scrape off surface" *Kuvalan* kiskis "shave" *Cb* kikhi "scrape off caking dirt"

*kuskus "scrape" (IV.814)[39] Kn kumakusûkusu "cut hair" *Saaroa* kumakuukusu "cut hair, shave" *Rk* (*Tanan* dialect) ukukuqu "cut hair" *Pa* kuskus "scraper" *Am* koskos "scratch body"

---

37 The final /-h/ in *Am* is not regular.

38 Blust (89a:145) quotes a Yamdena form in connection with this, but a large number of reconstructions could give rise to the Yamdena form, and there is not likely any direct connection.

39 There are roots with the shape *qisqis, *qiʃqiʃ, *kiskis, *kiʃkiʃ, *kuskus, and *kuʃkuʃ, all with the meaning "rub, scrape, shave", and the like. Clearly there is a process of sound symbolism at work here, but one or more of these roots may well have been inherited. *kuskus is actually a good candidate as an inherited form, as we have a derived form of this root meaning "(finger, toe) nail" as discussed in §3.2 below.

	Cb kalukhu "scrape something off"
*pakpak "clap"	*Am* pakpak *Tg* palakpak *Cb* pagakpak "clap"
*qangɤis "stench of fish, bats" (IV.27)	*Am* ʔaglis "stench of fish" *Il* angri "stench (fish, bats)"
*qangtud "smell bad"	*Am* ʔagtol "stink of something bad" *Pu* ɦangtur "smell bad" *Cb* angtud "smell bad"
*qiʃqiʃ "scrape, scrub"	*Ts* siʔsi "scrape, as bamboo knots holding bolo in both hands" *Pa* ki-qisaqis "rub oneself to relieve itching" *Am* ʔisʔis "shave" *Tg* is-is "scrub"

3.1.4 Shared Forms which also Appear in Languages Adjacent to the Philippine Languages

There is a list of forms which are found exclusively in the AN languages of Taiwan and in the Philippine languages and also in one or more neighboring languages. There are two possible interpretations for this sort of data: (1) that the form is actually of PAN origin and by chance does not show up in the available dictionaries or is in fact missing in most languages (in which case there must be a reason for it); (2) that there is a substrate language that covered the neighboring languages. Later developments (language death, language mixture, and other contact phenomena) created a rift between these languages and the AN languages of Taiwan and the Philippine languages. To be sure, explanation(1) is possible, for there are forms of PAN origin which are reflected in a few widely separated languages,[40] and an explanation that

---

40   For example *latuq "kind of edible seaweed", which is not found in many languages, but is attested in scattered distant languages, is very likely of PAN origin. The reason for its disappearance in many languages probably has to do with the fact this plant is not found everywhere, and communities which brought the language to areas where this seaweed did not occur lost the knowledge of it. Even if later the community speaking the language moved to an area where it

such scattered forms are inherited does not clash with the Stammbaum model of language history. Still in light of what detailed studies of historical events in language change reveal (studies of on-going change, close studies of contact phenomena, and studies which have focussed in the origins and mechanisms for language change) one must take it as more probable that in most cases explanation (2) is closer to an account of historical events. The following list gives some of the forms which have turned up in the data. First we give forms which have cognates which are attested for languages as far as the Lesser Sundas, Celebes, and/or Moluccas:

*bakeʃ "belt" *Am* fakes[41] "put a belt on" *Cb* bakus "belt" *Sawu* wake "belt"[42]

*bayu "pound with pestle" *Bunun (North and Central)* ma-baəu *Tg* bayo *Baree* mbaju *Solor* bayo "pound with pestle"

*bekut "hunched over" (IV.79,D93) *Pa* ma-bekuts "bent down (as from burden)" *Am* fkot "bend one's body as to hump over double" *Tg* bukot *Tiruray* bekut "hunchbacked" *Tboli* bekut "bend or fold" *Rembong* wengkut "lower head in fear, curl in sleep"

*buleg "single hill" *Pa* vuled "a single mountain" *Tg* burol "hill" *Cb*

---

did occur and went back to consuming it, the name was lost. An example of another widely represented type of development is *niniq "kind of reed (Donnax canniformis)" a word which occurs in the Batanes Straits and then does not reoccur until languages in eastern Indonesia and Oceania. In this case the plant was known to all or most of the AN speech communities, but the name was replaced by *banban in most languages which have been studied and survives only in isolated spots or in the outskirts of the area covered by *banban.

41　Amis /s/ does not normally correspond to Cb /s/ from PAN*ʃ, but in this and in several other forms we get Amis /s/ for PAN *ʃ for reasons which probably have to do with developments of PAN *ʃ in Amis dialects.

42　The Sawu form was suppplied by R. Blust (personal communcition).

(IV.106)   būlud "hill" *Bontok* bulud *Blaan* bulul *Taw Sug* buud *Sangir* bulhude? *Uma* bulu? *Baree* buju *Bugis* vulu *Kei* vuur "mountain"

*dakut "grasp in hand"   *Pa* djakuts "grab, take in hand" *Tg* dakot "handful, (II.75)   grasp in hand" *Chamorro* hakot "snatch, grab"[43] *Tetum* rakut "pull, grab, snatch away"

*iqeta "rice husk" (D.12)   *Saisiat* kæ-sæq "rice husk" *Rk* icaa "husked rice" *Pu* ñetha? "husk and bran" *Ilk* itta "rice husk" *Baree* ota "chaff" *Sawu* m-əda "unshelled grains in pounded rice" *Lio* həta "straw, chaff"

*kawayan "bamboo"   *Rk* kavadanə *Pa* kavayan *Pu* kawayan *Tg* kawayan (I.199)   *Cb* kawayan *Manobo* kewayan *Buru* kawan "bamboo *or* kind of bamboo"

*kañuskus "fingernail"   Pa kałuskusan "fingernail" *Cb* kalukhu "scrape out" (IV.314)   *Maranao, Sangir* kanuku *Mongondow* konuku "fingernail" *Tae, Makkasar* kanuku "nail, claw, hoof"

*mula "to plant"   *At* muya? *St* ma-mula? *Am* pa-loma *Ilk* mūla *Baree* mula *Bima* mura *Kambera* pa-mula "plant"

*pedek "wink" (IV.434)   *Am* mapa-pdek "blink" *Maranao* perek "wink" *Bare'e* pida "wink, blink"

*qaɤaw "snatch"   *Pu* ñaɤaw-i *Tg* āgaw *Cb* āgaw *Manobo* aɤew *Maranao* agaw *Mongondow* agow *Kadazan* aaw "seize, snatch" *Kelabit* aro "snatch"[44]

---

43   The Chamorro phonemes do not correspond to the others in the list, and this form may not be connected.

44   The Kelabit form was pointed out to be by R. Blust (personal communication).

*qañuwang "bovine"   *At* p-qanu-x "deer" *Kn* ʔi-ũangə "female deer" *Pa*
  (T.139)   ɬuwang "cattle" *Pazeh* nuang "carabao" *Bunun*
                      hanvang "deer, carabao" *Il* nuang "carabao" *Kulawi*
                      anoa "kind of wild bovine"

*qũsung "mushroom"   *At* qihung[45] *Ts* ungo *Cb* ũhung "white mushroom"
  (T.171)   *Manggarai* ung *Balinese* ong "mushroom"[46]

*saliñ "transfer" (III.327) *Pa* saliɬ "isolate, put off by itself" *Cb* halin
                      Hanunoo hãlin *Manobo* halin *Proto-Sangir* alin
                      "move, transfer" *Manggarai* aling "move, transfer,
                      change, go out from inside"

*ʃakat "rise, climb" *Am* cakat "arise, go up" *Bk* sakat "go up, climb"
  (IV.535)   *Maranao* sakat "step up, rise" *Ngadha* saka *Buru*
                      saka-h "ascend"[47]

*ʃekel "bent" (IV.552) *Am* ckel "bow head" *Bk* sukól "bow head" *Baree*
                      sengko "curved, bent", *Buru* seke-k "bend to form
                      an angle"

*tam "smack lips"(II.420) *Pa* tjalamtjam "smack lips" *Maranao* tantam "taste"
                      Chamorro tamtam "test, taste" *Buli* camcam "taste"

*timij "chin, jaw" (I.42) *Pa* tjimiz "chin, mandible, jaw" *Pu* timiz *Atta* simik
                      Isneg simid *Il* timid *Pg* timir "chin" *Roti* timi

---

45   The /i/ in At is not explained.

46   The Balinese and Manggarai forms were pointed out to me by R. Blust (personal
     communication).

47   The Ngadha and Buru forms may not be connected with *ʃakat, as there is no
     proof that these words come from a root with a final -t. There are also forms
     with reflexes from *ʃaka in various languages, and these forms may be cognate
     with those. Also the Bare'e and Buru forms in the next item on the list may not
     be connected for the same reason: namely, they give no clue as to the final
     consonant.

"chin, jaw"

*tuktuk "top, top of head" (II.434, D.25)   *Siraya* toucktouck "top, crown of head" *Tg* tuktuk "top, peak, summit, top of head" *Balinese* tuktuk, *Kei* tutu "top, tip, extremity"

The following forms have cognates in the AN languages of Taiwan and the Philippine languages and in languages of northern Borneo. These languages extend as far west as Iban, which is a Malay dialect, but like Brunei Malay has vocabulary from the Philippine languages. There are also two examples of forms with cognates in Moken, a language of the Mergui Archepelago of the Tenasserim of Burma. The small amount of data available in word lists of this language makes it look rather close to Malay (which certainly would not be surprising in view of its location). However, a fleeting examination of published texts in this language gives the impression of a language which in fact is not very closely related to Malay, and I have put the two examples of roots in common with Moken and Taiwan-Philippines in the following list. I concede that at this point it is quite speculative to claim that Moken shows a substratum of Philippine-Taiwan forms, but on the other hand the occurrence of a form in Moken and the Philippines-Taiwan is hardly grounds for concluding the form is inherited from PAN.

*betu "callus" (IV.95)   At bsuh "callus" *Pa* vetsu "callus" *Pu* vuthu "callosity, corn" *Cb* butu "blister" *Maranao* beto "callus, bunion" *Mongondow* boto "blister, callus" *Moken* beto? "blister"

*dañ "old"   *At* ra-ral *Cb* dãqan *Proto-Minahasa* daqan "old" *Kelabit* dadan "old (things)"

*ɣamut "root"   *Am* lamit[48] "root" *Tg* gamot "medicine" *Cb* gamut "root" *Kelabit* ramut "fibrous roots"[49]

*jañi "near" (T.157)   *At* dalih *Saisiat* ʔal-ʔalih-an *Kn* ará-cani *Rk* me-d-dali *Paẕeh* ʔalih "near" *Am* nani "come from a place" *SL* ha-ráni *Bk* ha-rani, *Sangir* dani "near" *Il* dan-dani "nearly, almost" *Katingan, Lawangan (Kalimantan)* dani "near"[50]

*lejep "submerge"(I.274) *Pa* ledep "dive, plunge" *Dumagat* ledep "swim under water" *Kayan* lejep "sumerged by rising flood"[51]

*putaq "foam"   *Kn* pucaʔ "foam" *Rk* puca "foam" *Pa* putsaq "foam"
(III 86,238)   *Kadazan* puta "foam, lather, froth" *Malagasy* fota-fota "saliva, spittle"[52]

*qálaj "wooden fence,   *Pu* ħaraz *Cb* álad *Proto-Minahasa, Manobo, Maranao*
enclosure" (II.3)   alad *Kadazan* tahad "fence" *Kelabit* alad "wall"

---

48 The /i/ of the final syllable in Am is irregular. Normally *u is reflected by /o/. Cf. *qúsung and *túqed.

49 The Kelabit cognates of *dañ and *ɣamut were supplied by Blust (personal communication).

50 The Dayak forms were pointed out to me by Blust (personal communication).

51 Blust connects Sundanese *lejep* "close the eyes" with this, but it is too distant semantically to justify a hypothesis of cognition.

52 The Malagasy cognate (if it is indeed a cognate) indicates that perhaps this is in fact inherited from PAN. Still, I think it is very possible that it is a form shared exclusively by the AN languages of Taiwan, the Philippine languages, and the neighboring languages. Kadazan normally reflects *-q with a final glottal stop. The most likely hypothesis is that this form spread by borrowing and had been borrowed into Malagasy before Malagasy was brought away from Borneo. Another possible hypothesis is that the Borneo and Malagasy forms resemble the others because of sound symbolism, as there is a number of forms with meaning "foam" and the like with similar sound.

*qañit "leather (I.40)[53]   *Ts* hici "leather" *Pa* qałits "skin" *Cb* ánit "leather" *Kelabit* anit "skin, bark"

*qateb "trap" (T.167)   St ʔæsəb *Ts* cəfə *Bunun* hatub "deadfall trap to catch rats" *Cb* átub "pitfall and other kinds of traps" *Bk* k-ateb "pig *or* rat trap" *Itbayaten* atob "crab trap" *Sangir* atabeʔ "crab trap" *Kelabit* ateb "deadfall trap"

*qatimela "flea" (I.23; IV.665)   *Ts* timro *Am* ʔatimela *Pu* hatimuraq *Hanunoo* timla *Il* timel "flea" *Bintulu* temela "louse" *Kelabit* gesimel "bedbug"

*qekung "owl" (IV.179)   *Am* ʔkong *Kelabit* ekung "owl"

*qupung "bunch, cluster" (II.438)   *Pa* qupung "swarm of honey bees" *Cb* upung "bunch of stalks, stems enough to hold in two hands" *Manobo* upũng "hold together in one hand" *Hanunoo* upũng "harvest with knife" *Iban* upong "bunch of fruit"

*qusenget "angry" (IV.712)   *Am* ʔosnget "expression of anger" *Il* unget "get angry" *Kayan* unget "anger"

*ʃuqagh "thorn" (I.404)   *Pu* suhaɤ *Bilaan, Tiruray* suʔal *Tboli* suwwar *Long Anom* suaʔ "thorn"

*tālaw "cowardly, ashamed" (I.413)   *Am* ma-talaw "afraid" *Cb* tālaw "be cowardly" *Isneg* na-tālaw "afraid" *Il* ag-tālaw "flee" *Maranao* talao "coward" *Kadazan* t-um-ahou "be a coward" *Moken* talau "ashamed, shy"

---

53 Blust (80:47) hypothesizes that *bañit (his transcription:baNiC) is connected with this and contains a prefix b-. In that case this root is PAN. This is a possibility, although no prefix *b- has been reconstructed for PAN or for a later subgroup.

*tebas "cut off" (III.357)[54] — *Pa* tjevas "cut, prune, clear (vegetation)" *Tg* tibaq "cut down banana for fruit" *Maranao* tebaq "cut banana tree" *Manobo* tevaq "cut banana stem" *Iban* teba? "cut small plants through" *Malagasy* tevy "cut trees to make clearing"

*teja "left-over" (III.359) — *Pa* tjeza "something saved, left over, remainder" *Tg* tira "leave over" *Iban* teda? "leavings left-over"[55]

*tingas "food caught in teeth" (T.149) — *Saisiat* ʃingaʃ "food particles in teeth" *Ts* ru-ngtosə "remove food particles from teeth" *Rk* mu-tingāsə "remove food particles from teeth" *Pa* tsingas *Bunun* cingas *Pu* thinga *Tg, Cb, Kp, Bk, SL, Proto-Minahasa Bare'e* tinga *Itbayaten* tiñah *Iban* tinga? "food caught in teeth"

*ulaw "dazed, not able to see properly" (II.435) — *Pa* ulaw "fail to see, recognize, know" *Kankanay, Il* ūlaw *Bontok* ólaw "dizzy, faint, muddled" *Kadazan* um-uhau "become mad" *Iban* ulau "dazed, not able to see properly"

3.2 Commonailities in the Syntax and Morphology between the AN Languages of Taiwan and the Philippine Languages

The grammatical similarlities between the AN languages of Taiwan and the Philippine languages are striking. In many cases one can literally translate

---

54 These forms certainly are connected in some way, but the final /s/ in Pa and the glottal stop in the Philippines are not regular correspondences. Also the same considerations in regards to Malagasy which were weighed in the case of *putaq (this section) must also be weighed here.

55 There is also a Javanese form *turah* which is similar enough in form and meaning to indicate a connection. In fact, however, three out of five phonemes in *turah* do not correspond, and the resemblance must be purely coincidental.

the sentences morpheme by morpheme from one language to the other, although the resemblances are not so close as for example within the Philippine languages themselves. In this section I will look at some features from At, Pu, and Pa and compare them with Tg and Cb in the Philippines. The choice of these languages has to do more with the materials which I have readily available to me at the time of the writing of this paper than that these languages are particularly suitable for such a comparison. They are as suitable as any other randomly chosen languages from the Philippines or Taiwan would be. The data discussed here only scrape the surface and only are confined to that which happened to attract my attention as I worked through the limited amount of textual material available. In short, this is merely an introduction to what could amount to a several-hundred-page listing of commonalities between the AN languages of Taiwan and the Philippine languages not found in other AN languages.

Of course not all such common grammatical features would provide evidence for a close relationship between the AN languages of Taiwan and the Philippine languages. A large number of them are inherited from PAN, and if they are not found outside of the AN languages of Taiwan and the Philippine languages, it is because they have been replaced, and in that case there often are traces in the languages which lost them. I discussed some of these features in previous articles (Wolff 73,76). In the following sections I discuss some of the features which have not been found elsewhere. Traces of these features cannot be adduced elsewhere. For this reason, they should be looked at as shared innovations or as innovations which were made independently.[56]

---

56 Reconstruction of the morphology and syntax of PAN - that is, an account of how the proto-forms developed to give rise to the attested forms, is very much

3.2.1 Morphology

Some of the affixes which are found exclusively in the AN languages of Taiwan and the Philippine languages are listed in §3.11 above. There are other morphological features which require detailed discussion. Until we have a good reconstruction of the PAN morphology it is clearly going to be impossible to state definitively which features are innovations and which are inherited. I list a few features here which I have failed to notice in AN languages outside of Taiwan and the Philippine languages and which therefore I presume not to have been present in the outside languages.

3.2.1.1 Durative (Progressive)Tense

The durative (progressive) form of the verb is formed in clearly related ways in Pu and Tg: the root is reduplicated and infixed with -*um*-. In Pu the initial vowel is always /a/ and *-um- has been changed to /em/.

Pu 1. Kemakeʀut ku za-vuʀasi haʀem /am-digging I sweet-potatoes

now/"I am digging sweet potatoes now."(205.49)[57]

---

in its infancy. For this reason, any judgment that this or that feature is in fact an innovation and not a retention is very much open to revision. Without a good account of how a given form came into being, it is impossible to say definitively that it does not contain traces of features which have gotten lost as active processes. Further, without an account of how the attested forms came into being, it is impossible to say whether or not a given parallel is in fact the product of a single event which spread over the speech area or if analogies for the creation of the form existed and that the form could well have come into being quite independently at different times in different places. In this paper, when there are close coincidences of form and meaning, I will assume that the innovation occurred only once and spread. This is certainly the case of a large portion of the forms described here, and if a certain percentage of them in fact developed independently, there still are enough in common to provide strong evidence that the AN languages of Taiwan and the Philippine languages form a subgroup.

57   References are to page and paragraph number in the sources quoted in the bibli-

Tg 1. Humúhukay ako nang mga kamóte ngayon/am-digging I some

plural sweet-potatoes now/"I am digging sweet potatoes now."

3.2.1.2 Future Tense

In At and Tg the future tense may be formed by reduplicating the initial consonant and the vowel following it. In At this vowel is reduced by automatically operating rules to schwa. In currently spoken Squliq At this future has gotten lost, but it is documented in texts recorded fifty years ago (Rau p. 61).

At1a. yat mmuwah ssiyuk bbiq niya? ita? laru? / not future-going-to

in return future-give by-them we particle / "He won't be going to

give us anything in return."

Tg 1a. Walá táyong mākukúha diyan/not we will-be-able-to-be-gotten

there/ "We won't get anything from him."

3.2.1.3 Nominal Formation with Reduplication Plus - *en* or - *an*

Reduplication of the first syllable plus *-en* or *-an* forms nouns in Pa, Tg, and Cb. In Tg this formation is dead, but in the other languages it is productive. In Pa the reduplication consists of the first consonant of the base plus/a/, in Tg it consists of the first consonant plus vowel of the base, and in Cb it consists of the first consonant plus vowel of the base and then the first consonant of the base is replaced by /1/. The noun with *-en* refers to the object of the action, and the noun with *-an* to the place of the action.

Pa: temkel "drink" tatekelen "drink" tatekelan "drinking place, cup"

vateq "wash" vavateqen "clothes to be washed" vavateqan "wash board"

---

ography for Paiwan, Puyuma, and Atayal (Egli 90, Tsuchida 80, and Rao 91). Cebuano and Tagalog examples are made up for the purposes of this paper.

Tg: bãsa "read" babasahin "reading matter" babasahan "place to read"

Cb: inum "drink" ilimnun "drink to have" ilimnan "place to drink, cup"

palit "buy" palalitun "things to buy" palalitan "place to buy"

3.2.1.4 *ka-* "co-"

There is a morpheme shaped *ka-* which occurs in words which refer to one of two people who are in a certain relationship. This occurs in Tg and Cb without further affixation. It also occurs together with the reflex of *maɣ-* in nouns or verbs which refer to two or more people who are in a relation.

Tg: sãma "go with" kasãma "companion" magkasama "two people who are companions"

Cb: ũban "go with" kaũban" companion" magkaũban"two people who are companions"

This morpheme also occurs in Pa, although in combination with different affixes.

Pa: matsiur "be together" katsiuri "let's go together"

The morpheme *ka-* in the following Pa forms may also be connected with this morpheme *ka-*"co."

Pa: tjengelay "love" markatjengelay "people who love each other."

3.2.1.5 *ma-*

There are two affixes shaped *ma-* in four of our languages: (1) *ma-* "adjective former" and (2) *ma-* potential passive verb form. The *ma-* adjective former is probably inherited from PAN, as it is also found in Oceanic languages. We list it here for reference only.

ma- "adjective former"

Pa: matsingki "clean (adjective)" cf. tsmingki "clean something"

Pu: ma-vutih "blind" ma-nadam "accustomed, tamed"

Tg: malínis "clean (adjective)" cf. maglínis "clean"

Cb: maqáyu "good(adjective)" cf. muqáyq "repair something (make it good)"

The potenital passive *ma-* seems to be a development within the AN languages of Taiwan and the Philippine languages. The potential passive most often has the meaning "for (so-and-so) to happen to (the subject) involuntarily."

ma- "potential passive"

Tg: mākita "be seen" matápos "be finished" mainit "become hot" malútoq "get cooked"

Cb: makitaq "be seen" mahuman "be finished" mainit "become hot" malútuq "get coooked"

Pa: mapatsun "be seen" maqatsuvung "be finished" masezam "get heated up" makesa "get cooked"

This affix is also productive with verbs which refer to psychological processes (the way one feels or the like) in the Philippins, but not in Pa, although the prefix *ka-* which is part of the paradigm of the *ma-* potential passive (§3.23 (d)below) is productive with forms of this meaning in Paiwan.[58]

Tg: matākut "be afraid" mahiyaq "be ashamed" magútom "feel hungry"

Cb: mahadluk "be afraid" maūlaw "be ashamed"

Pa: malekutj "be afraid" masiaq "be ashamed" matsua "feel hungry"

In At the prefix *ma-* (reduced to *m-*) is not productive.[59]

---

58   This assertion is based on Egli's statment p.107, §197, which gives a long list of bases which may potentially contain a *ka-* derivative prefix but cites only one example which actually contains *ka-*.

59   There are a great number of prefixes shaped *m-* in At, but only some come from an earlier direct passive potential form *ma-*. They are recognizable because

At: m-ʔuyay "be hungry" m-ngungu "be afraid" m-sayux "be ashamed"

3.2.1.6 *Pa*-Forming Words Meaning to go to a Certain Place

In Tg this affix is productive and comparable in meaning to the Pa. It occurs with many roots in Pa, but I do not know how productive it is. Cb also has *pa*- comparable in meaning to the Tg and Pa affix, but the formation is not productive in Cb.

Tg: palikud "go in back" pakánan "go to the right" palapit "getting near"

Pa: palikuz "go in back" panaval "go right" padut "go near" pazaya
"go inland"

Cb: palikud "go to the back" patúqu "go to the right" padúqul
"go near" pailaya "go inland"

3.2.2 Meanings of Specific Affixes

Some affixes which certainly are of PAN origin have developed special meanings which are found only in the AN languages of Taiwan and the Philippine languages.

3.2.2.1 *maka*- in a Potential Meaning

Four of our languages have a prefix shaped *maka*- which forms potential active verbs. The corresponding passives are formed with *ma*-

Tg: makapunta "can go somewhere" mapuntahan "can be gone to"

Cb: makaadtu "can go somewhere" maadtu "can be gone to"

Pa: makavaik "can go somewhere" mapatsun "can be seen"[60]

---

forms with this prefix as a direct passive potential, have a prefix shaped *k*-with other affixes (Rau §4.3.2 - cf. §3.23(d) below).

60   The sources unfortunately do not give paradigms. There is no reason that an active form *makapatsun* "can see" should not occur, but it is not given in our sources. Similarly, there is no reason that a passive form (*mavaik* or *mavaikan*) in paradigm with *makavaik* should not occur, but it is not listed.

3.2.2.2 Instrumental Passive Referring to Time

In Cb and Pa one of the meanings of the instrumental passive is a temporal orientation. I have not found this meaning of the instrumental passive in any language outside of the AN languages of Taiwan and the Philippine languages.

Pa: sisekezan a qadau/time-to-rest linker day/"day of rest" (Egli.263§474)

Cb: adlaw ng igpapahulay /day linker time-to-rest/"Day of rest."

3.2.3 Morphophonemics in the Affixation of Derivative Forms

The morphophonemic peculiarities involved in adding affixes to derived bases of our five languages are directly comparable with one another and must have a common origin.

(a) affixation of the causitive prefix *pa*-

The affix *pa*- is added agglutinatively without undergoing any morphophonemic alternation.

Pa:

pegatsel "itch (verb stem)" megatsel (= -um- + pegatsel) papegatsel (= pa- + pegatsel) "cause itching"

pigatsal "stand (verb stem)" migatsal (= -um- + pigatsal) papigatsal (= pa- + pigatsal) "cause something to stand"

pedjek "be aflame(verb stem)" medjek (= -um- + padjek) "burn" papadjek "set afire(= pa-+pedjek)

Tg:

pag-áral "study (verb stem)" mag-áral (= -um- + pag-áral) "study" papag-áral" cause to study"

pagpanibágo "change" (verb stem) magpani-bágo (= -um- + pagpanibágo) "change" papagpanibágo "make something change"

(b) the active potential affix *maka-* is added agglutinatively.

Tg: makapagdala "can bring" (maka- + pag + dala)

Pa: makapaseqeⅼing "can save someone"(maka + pa + seqeⅼing)

(c) the active potential affix *maka-* is analyzable as consisting of -*um*- plus *paka-*.

This analysis is possible in Pa and in Cb in a restricted way. Since the Cb is restricted, one can draw the conclusion that it represents something old and for this reason is cognate with the morphophonemics in Pa, which are totally regular.

Pa: makaqati "can do something" su pakaqati-n/by-you it-can-be-done/ "You can do it."

Cb: makakitaq "can see" magpakakitaq "can see (plural agents)"[61]

(d) The base to which the passive potential prefix *ma-* is added can be analyzed as containing *ka-*. The *ka-* is not present when *ma-* is present, but when other affixes are added to the same base, the *ka*-reappears.

Pa: marekutj "be afraid" ka-rekutju "be afraid (imperative)" ka-rekutjan "be afraio of it" (Egli.814) matjani "fall" ka-tjaniu "fall (imperative)" masiaq "be ashamed" si-ka-siaq "be ashamed on account of it"

Cb: mahadluk "be afraid" ka-hadluk "be afraid (imperative)" gi-ka-hadlûkan "be afraid of it" mahulug "fall" ka-hulug "fall (imperative)" maulaw "be ashamed" i-ka-ûlaw "be ashamed on a account of it"

Tg: matãkut "be afraid" pag-ka-tãkut "being afraid" k-in-a-tãtakûtan "be

---

61   This formation is found only in very lofty styles. The use of *mag-* plus a derived base to form a verb referring to plural agents is also found with verbs having a base with *paN-*: *mangⅼûya* "become weak" (=-*um*- plus *pangⅼuya*) *magpangⅼuya* "become weak (many things)"

afraid of it" mahúlog "fall" pagkahúlog "action of falling" mahiyaq
"be ashamed" i-ka-hiyaq[62] "be ashamed on account of it"

In At the affix *ka-* is productive, but *ma-* is confined to a restricted number
of roots.

At: mngungu? "be afraid" ini? k-ngungu "not be afraid" m-tala "be red"
k-tala "be red (dependent form)" (Rau §4.3.2)

3.2.4 Word Order

In all five of the languages which we examine, short words - that is,
short forms of pronouns and post-positive modifiers, come directly after the
first word of the predicate. A form which modifies the predicate (e.g. a
word referring to time, a negative, and the like) tends to stand before the
head of the predicate and attracts the short words. The following pair from
Pu illustrates how this feature works. The short-form pronoun *u* "you" comes
after the verb in Pu 2a, whereas it is moved immediately after the modifier
in Pu 2b:

---

62   The *ka-* of this prefix in Tg is part of the base and not part of the prefix, as
     has been often thought. The instrumental passive in Tg, as in Pa, Cb, and many
     other AN languages may refer to the reason on account of which the (genitive)
     feels a certain way or does a certain thing. With the instrumental passive affix
     the *ka-* appears in bases of verbs which have *ma-* in Pa and in languages of the
     Philippines. In Tg and Pa at least (probably in other languages as well), the
     instrumental passive is used with verbs which do not have *ma-*, and in most
     cases there is no prefix *ka-*. Thus *manibágo* (to the base *panibágo*) "feel unacus-
     tomed", *magtaka* (to the base *pagtaka*) "be surprised" *manghina* (to the base
     panghina) "become weak" have instrumental passives without *ka-* to form verbs of
     this sort: *ipanibágo* "feel unaccustomed on account of it", *ipagtaka* "feel surprised
     on account of it", *ipanghina* "weaken on acount of it." The following sentence
     exemplifies a Pa verb which uses this kind of instrumental passive:
        A nem a su sinivaik /specifier what? subject-marker by-you reason-
        on-account-of-which-went/"What did you go for?" (Egli.262)

Pu 2a. Kemeɤut u za-ʔeman?/dig you what/"What are you digging up?"
(201.38)

Pu 2b. hazi u ziya savuran zi malazam u la maɤngay/not you yet one-
month and know you already to-speak/ "You aren't here a month
yet, and you can already speak(Puyuma)."(300.G)

In Pa the process is totally analogous. The pronoun follows the active verb in
example 2a, but in example 2b it is attracted to the first word modifying the
predicate (and is reduced in form):

Pa 2a. Vaik aken. Vaik sun. Vaik itjen. "I go. You go. We (incl.) go."
(Egli.155)

Pa 2b. Ini ke vaik. Ini ana su vaik. Iɤu me kan. /not I go/not yet you
go/absolutely-not we eat/ "I won't go. You won't go. We abso-
lutely won't eat." (Egli.156)

In At the same process occurs. The pronoun follows the verb in example 2a,
but in example 2b it is attracted to the first word modifying the predicate
(and is reduced in form):

At 2a. Ngilis saku? 1ga? /cried I particle/

At 2b. si ku? ngilis/Indeed I cried/

Similarly, in Tg and Cb the form *ka* "you" comes immediately after the verb
in the first sentence, but in the second, where there are preposed modifiers
to the predicate, the *ka* comes immediately after them.

Tg 2a. Nakākapagtagālog ka na/can-speak-Tagalog you now/ "You can
speak Tagalog now."

Tg 2b. Hindi ka pa matagal dito, marúnong ka nang magtagālog. /not
you yet long here can you already speak-Tagalog/ "You haven't
been here long, you can already speak Tagalog."

Cb 2a. Makabinisayaq ka na /can-speak-visayan you now/ "You can speak

Visayan now."

Cb 2b. Walaq ka pa magdūgay dinhi mahibalu ka nang mubinisayaq/not

you yet long-time here know you already speak-Visayan/ "You

haven't been here long, (but) you can already speak Visayan."

In Pa, Tg, and Cb if the word which modifies and precedes the predicate is linked with a linking word, the pronoun which is attracted precedes the linker. In the following sentences the word for "I" is attracted to the first word of the predicate (the modifier) which is linked to the head with *a* in Pa, with *ng* in Tg and with *g* in Cb:

Pa3. Maya aken a qemtsi/not me linker kill/"Don't kill me." (Egli. 285)

Tg2. Hwag mo ako ng patayin/not by-you me linker kill/ "Don't kill

me."

Cb3. Ayaw ko g patya/not me linker kill/"Don't kill me."

3.2.5 Linking

All of the AN languages of Taiwan and the Philippine languages to my knowledge have linkers, and this is probably a feature inherited from PAN. There are some idiosyncratic features of the linker which are peculiar to specific languages and because of their peculiar nature, they may possibly be common developments. A case of this is the treatment of the preposed genitive forms in Tg, Cb, and Pa. Information on the syntax of Pa is not available in sufficient detail to enable me to state that the similarities of linkage are in fact of a common origin in all three languages, and so what I say here is a suggestion of an area of possible further investigation, not an uncontrovertable conclusion. The genitive has two important functions in the vast majority of the western AN languages: (1) it refers to the possessor and (2) it refers to the agent of a passive verb. In Tg and Cb the genitive may be

preposed or postposed. If it is preposed, it is linked to the noun which it modifies. In Pa the genitive of the third person may be preposed or post-posed. If the genitive of the third person is preposed, it is linked. For the first and second persons the genitive is always preposed, and may be a long form or a short form. If it is a long form (preceded by *ni* and in some cases differing in other ways formally from the short form), it is linked. If the genitive of the first and second persons is a short form, it is not linked. In Pa apart from the first and second persons there is no formal difference between the genitive in preposed constructions and in postposed construc-tions, whereas in Tg and Cb there is a difference. (The preposed genitive in Tg and Cb and other Philippine languages is a development of the dative form.)

The examples marked "4a" illustrate the preposed genitive (modifying the word for "child") and linked to the word it modifies. The examples marked "4b" illustrate a phrase with the same meaning with a post-posed genitive:

Pa 4a. Nua mamazangiłan a ałak /of-the chief linker child/"A child of the chief." ni aken a kama /belonging-to me linker father/ "My father."

Pa 4b. Ałak nua mamazangiłan /child of-the chief/ "A child of the chief." ku kama /my father/"My father."(Egli.295)

Tg 4a. Kay Maria ng anak /belong-to Maria linker child/ "Maria's child." āki ng anak /my linker child/"My child."

Tg 4b. Anak ni Maria /child of Maria/ "Maria's child." anak ku /child my/ "My child."

Cb 4a. Kang Maria ng anak /belong-to Maria linker child/ āku ng anak /my linker child/ "My child."

Cb 4b. Anak ni Maria /child of Maria/ "Maria's child." anak nākuq

/child my/ "My child."

4.Conclusions

In §1.1. we argued that the innovation represented by *C (the development of contrast between two allophones of *t) occurred only once and was shared exclusively by Saisiat, Atayal, Sedeq, Pazeh, Thao, Tsou, Kanakanavu, Saaroa, Rukai, Puyuma, and Paiwan (and probably some of the dead AN languages of Taiwan for which I do not have the reflexes). In §1.2 I argued that the development of initial *ñ to a lateral in roots with short penults and to n (or subsequently n) in roots with long penults was an innovation which could have occurred only once and is shared exclusively by the extra-Taiwan languages. Similarly in §1.3 I argued that the extra-Taiwan languages also shared the innovation of the change of *s to *h. If we are to believe that exclusively shared innovations are indicative of a subgroup we conclude that the family tree was one of the following four possibilities (A = extra Taiwan languages, B = languages of Taiwan which did not split *t, C = Taiwan languages which split *t into two phonemes *t and *C : (1) there were three primary subgrops A, B, and C; (2) B groups with A and then A splits off; (3) B groups with C and then C splits off; (4) Some B languages group with A and some with C and then A and C spit off. We do not know where the B languages go, and other possibilities exist in that some but not all of the B language could form a primary subgroup.
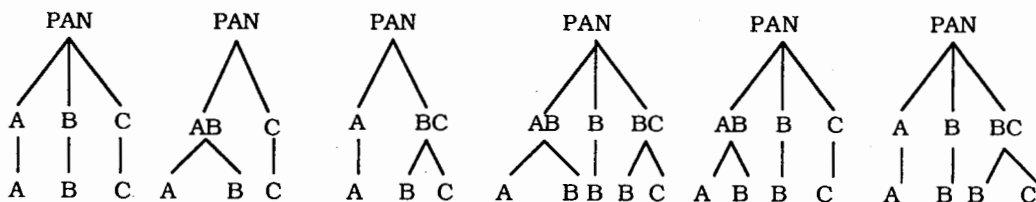
However, we need not take subgrouping to be a statement of historical fact. The changes which define A and those which define C may well have taken place during the time of the proto-language PAN and affected only part of the PAN dialect area: that is, the changes represented by the development of the new phoneme *C affected the northern or northwestern area, whereas the changes represented by A (the languages which changed *ñ to *1 initially and *s to h) affected the southern area. Only after the migrations or language movement which was to give rise to the eastern Indonesian and the Oceanic languages did the innovations develop which are detailed in §§3ff, - that is, the commonalities shared by languages of the Philippines and the AN languages of Taiwan. Some of these commonalities among the AN languages of Taiwan and the languages of the Philippines may well reflect inherited features of PAN which were subsequently lost elsewhere, but a good portion of them are innovations which developed during a period in which the ancestors of the AN languages of Taiwan and the Philippine languages were in close enough contact to be able to undergo shared developments. They may well have continued to form a single language or dialect chain during this period.

Another way to put this is that the contradictory evidence provided by the shared innovations in phonology on the one hand and the commonalities of development between the Philippines and Taiwan on the other can be explained by dialectal variation in the proto-language and by a period of common development between the Philippines and Taiwan after the eastern languages split off. This explanation is consistent with the anthropological and archeological view that the speakers of the AN languages originated somewhere in the north and west and spread southward and eastward. However, a study of shared innovations allow us to draw no specific conclusions as to

the homeland. This picture could as well be applied to a Philippine or even Philippine-Celebes homeland with the AN languages moving northward into Taiwan, as with a Taiwan homeland with little or no northward movement. In fact the details of shared innovations are even consistent with a PAN homeland and development on the mainland, west of the area in which AN languages are currently spoken.

Appendix

I. Table of Correspondences between the Philippine languages and the the Austronesian languages of Formosa[63]

PAN	At	Ts	Rk	Pa	Am	Pu	Tg	Cb	PMin	PSan
p	p	p	p	p	p	p	p	p	p	p
t	t,c	t,c	t,c	ts,tj	t	t,ṭ	t	t	t	t
k	k	?	k	k	k	k	k	k	k	k
q	q	?,#	#	q	?	ħ	?	?	?	#
b	b	f	b	v	f	v	b	b	b	b
d	r	c	ḍ	dj,z[64]	r	d,z	d,-1-	d,-1-	d,-r-	d
j	r	c	d	dj,z	r	d,z	d,-1-	d,-1-	d	d
g	k-,r	#	g	d	k-,-n -d	h-,d	k--1-d	k--1-d	k--r--d	k-,d
ɤ	g	r	?	#	l	ɤ	g	g	h	ɤ
m	m	m	m	m	m	m	m	m	m	m
n	n	n	n	n	n	n	n	n	n	n
ñ[65]	l	h	l	ɬ	d	l	l,n	l,n	l,n	l,n
ŋ[66]	ŋ	ŋ	ŋ	ŋ	ŋ	ŋ	ŋ	ŋ	ŋ	h
ʃ	h,-x	s	s	t	c	s	s	s	s	s
s	s	s	s	s	s	#	h	h	#	#
w	w	v	v	#	w	w	w	w	w	w
y	y	z				y	y	y	y	y

63 These are the correspondences which I believe to be correct, but I cannot at this point say that they are definitive. It is beyond the scope of this paper to discuss the environments which explain the exceptions. Also some of the infrequent correspondences are attested by only one example. Howerer, this table allows us to sift out forms which are regular in their correspondences from those which are not, which is the point of this paper.

64 The conditions for the occurrence of /dj/ or /z/ from *d and *j are unknown. In three case there is also a reflex /ḍ/ for *d. This is probably due to dialect mixture.

65 See Wolff 1993a for the reflexes of *ñ.

66 We transcribe ŋ with ng in this paper except in the case of Am, where this phoneme is written "g".

PAN	At	Ts	Rk	Pa	Am	Pu	Tg	Cb	PMin	PSan
l	l,#	r	l	l	l	r	l	l	l	l
nd,nj,ŋg							r	dl	nd	l
i	i	i	i	i	i	i	i	i	i	i
u	u	u	u	u	o	u	u	u	u	u
a	a	a	a	a	a	a	a	a	a	a
e	u	ə	ə	ə	e	e	i	u	ə	i,u,ə
ay	ay	oy	ai	ay	ay	ay	ay	ay	e	ay
ey	ay	oy	ai	ay	ay	ay	ay	ay	e	ay
aw	aw	o	aw	aw	o	aw	aw	aw	o	aw
iw	o(?)	iv		iw	iw	iw	iw	yu	i	uy(?)
uy	uy	uz	ui	uy	oy	uy	uy	uy	i	i

Transcriptions

I reconstruct a simpler phonology for PAN than is normally done (Wolff 1988, 1991, 1993a). The following chart shows the symbols for the consonants which I reconstruct compared for those commonly used to represent PAN phonemes:

Wolffs transcription	Traditional transcription	Wolffs transcription	Traditional transcription
p	p	ɣ	R
t	C,t[67]	m	m
k	k	n	n
none	T,c	ñ	ñ
q	q	ng	ng
b	b	ñ	N
none	d	l	l
d	D	none	r
j	Z	ʃ	s
none	z	s	S
g	j	w	w
none	g	y	y

67   PAN *C and *t are in complementary distribution, and *ñ and N do not contrast. Further, the phonemes symbolized *T, *c, *d, *r, *z, and *g are not reconstructible for PAN. Where I have listed forms with these letters, they refer forms which I do not believe can be reconstructed for PAN. These phonemes are discussed in the articles listed under my name in the bibliography.

REFERENCES

Blust, Robert A.

1973 The Origins of Bintulu ɗ and ɓ. BSOAS 36: 603-20.

1977 The Proto-Austronesian Pronouns and Austronesian Subgrouping: a Preliminary Report. Working Papers in Linguistics 9.2: 1-15, Honolulu, Department of Linguistics, University of Hawaii.

1980 Austronesian Etymologies I. Oceanic Linguistics 19.1: 1-181.

1983-4 Austronesian Etymologies II. Oceanic Linguistics 22-23: 29-149.

1984 The Austronesian Homeland: a Linguistic Perspective, Asian Perspective 26.1: 45-67.

1988 Austronesian Etymologies III. Oceanic Linguistics 25:1-123.

1989 Austronesian Etymologies IV. Oceanic Linguistics 28:111-180.

1992 The Position of the Formosan Languages: Method and Theory in Austronesian Comparative Linguistics. Papers presented to the International Symposium on Austronesian Studies Relating to Taiwan. Appeared in this volume.

Cauquelin, Josiane.

1991 The Puyuma Language. Bijdragen tot de Taal- Land- en Volkenkunde 147.1: 17-60.

1991 Dictionnaire Puyuma-Fran)ais. Paris Jakarta: École Française d'Extrême-Orient.

Dahl, Otto C.

1973 Proto-Austronesian. The Scandanavian Institute of Asian Studies, Monograph Series 25. Lund and London: Curzon Press.

1981 Early Phonetic and Phonemic Changes in Austronesian. The Insti-

tute of Comparative Research in Human Culture, Serie B: Skrifter LXIII. Oslo: Universitetsforlaget.

Dempwolff, O.

1934-1938   Vergleichende Lautlehre des austronesischen Wortschatzes. Zeitschrift für Eingeborenen-Sprachen, Supplements 15, 17, and 19. Berlin, Dietrich Reimer.

Duris, P. Antoine.

1970   *Dictionaire Amitsu-Francais.* Mission Catholique, Fuyuan, Hualien, Taiwan.

Dyen, Isidore.

1965a   Formosan Evidence for some New PAN Phonemes. Lingua 14:285-305.

1965b   *The Lexicostatistical Classification of the Austronesian Languages.* International Journal of American Linguistics. Memoir 19.

1990   Homomeric Lexical Classification. In Philip Baldi, ed. Trends in Linguistics: Studies and Monographs, 45: Linguistic Change and Reconstruction Methodology. Berlin: Mouton De Gruyter.

1991   Proto-Philippine as the Closest Relative of Proto-Formosan. In Ray Harlow, ed. VICAL 2: Papers from the Fifth International Conference on Austronesian Linguistics, Parts One & Two: 85-101. Auckland, Linguistic Society of New Zealand.

1992   The Internal and External Classification of the Formosan Languages. Paper presented to the International Symposium on Austronesian Studies Relating to Taiwan.

1993   Borrowing and Inheritance in Austronesian Linguistics. Paper circulated at the International Symposium on Austronesian Studies Relating to Taiwan. Appeared in this volume.

Egerod, Søren.

1980   Atayal-English Dictionary. The Scandanavian Institute of Asian Studies, Monograph Series 35. Lund and London: Curzon Press.

Egli, Hans.

1990   Paiwangrammatik. Wiesbaden: Otto Harrassowitz.

English, Leo James, C.Ss.R.

1986   Tagalog-English Dictionary. Manila, National Bookstore.

Ferrell, Raleigh.

1969   Taiwan Aboriginal Groups: Problems in Cultural and Linguistic Classification. Taipei, Institute of Ethnology, Academia Sinica.

1982   *Paiwan Dictionary*. Pacific Linguistics, Series C. No. 73. Department of Lingusitics, Australian National University, Canberra.

Fey, Virginia.

1986   Amis-Dictionary. Taipei: The Bible Society.

Ho, Dah-an.

1978   A Preliminary Comparative Study of Five Paiwan Dialects. Bulletin of the Institute of History and Philology, Academica Sinica 49.4:565-681.

Li, Paul Jen-kuei.

1977   The Internal Relationships of Rukai. Bulletin of the Institute of History and Philology, Academica Sinica 48.1:1-92.

1978   A comparative Vocabulary of Saisiyat Dialects. Bulletin of the Institute of History and Philo,ogy. Academica Sinica 49.2:133-199.

1980   The Phonological Rules of Atayal Dialects. Bulletin of the Institute of History and Phil,logy. Academica Sinica 51.2:349-405.

1981   Reconstruction of Proto-Atayalic Phonology. Bulletin of the Insti-

tute of History and Philology, Academica Sinica 52.2:235-301.

1985 The Position of Atayal in the Austronesian Family. Austronesian Linguistics at the 15th Pacific Science Congress, 257-280. Pacific Ling-uistics, C-88.

1988 A Comparative Study of Bunun Dialects. Bulletin of the Institute of History and Philology, Academica Sinica 59.2:479-508.

Mahdi, Waruno.

1988 Morphophonologische Besonderheiten und historische Phonologie des Malagasy. Veröffentlichungen des Seminars für Indonesische und Südseesprachen der Universität Hamburg. Band 20. Berlin, Hamburg: Dietrich Reimer Verlag.

Rau, Der-Hwa.

1992 A Grammar of Atayal PhD. Dissertation, Cornell University. Taipei: The Crane Publishing Co.

Sneddon, James N.

1978 Proto-Minahasan: Phonology, Morphology and Word List. Pacific Linguistics Series B. 54.

1984 Proto-Sangirese. Pacific Linguistics. Series B.

Ting, Pang-hsin.

1978 Reconstruction of Proto-Puyuma Phonology. Bulletin of the Institute of History and Philology, Academica Sinica, 49.3:321-392.

Travis, Ed.

1993 A Kei-English Dictionary. Summer Institute of Linguistics. Unpublished Manuscript.

Tsuchida, Shigeru.

1976 Reconstruction of Proto-Tsouic Phonology. Study of Language and Cultures of Asia and Africa, Monograph Series, No. 5. Tokyo

Gaikokugo Daigaku.

1980　Puyuma (Tamakalaw dialect) Vocabulary. Kuroshio no Minzoku, Bunka, Gengo, 183-307. Ed. by Kuroshio Bunka no Kai. Tokyo.

1983　Puyuma-English Index. Working Papers in Linguistics, University of Tokyo.

Vanoverbergh, Morice, C.I.C.M.

1956　Iloko-English Dictionary. Manila.

Wolff, John U.

1972　Cebuano Visayan Dictionary. Linguistic Society of the Philippines, Monograph No. 3 and Southeast Asian Program, Data Paper No. 84, Cornell University.

1975a　Proto-Austronesian *r and *d. Proceedings of the First International Conference on Comparative Austronesian Linguistics, Oceanic Linguistics 13:77-121.

1975b　Protoaustronesian Verbal Morphology. Essays Presented to Cecilio Lopez. Manila: 71-91.

1976　Verbal Morphology and Verbal Sentences in Proto-Austronesian. Michigan Papers on South and Southeast Asia No. 15. 1976b: 153-147.

1982　Proto-Austronesian *c, *z, *g, and *T. Papers from the Third International Conference on Austronesian Linguistics, ed. by Amran Halim et al. Vol 2: Tracking the Travelers. Pacific Linguistics C-75, Canberra:1-30.

1988　The PAN Consonant System, in D. McGinn, Ed., Studies in Austronesian Linguistics, Monographs on International Studies, SE Asia Series, No. 76. Athens, Ohio: 125-147.

1991　Was there a Contrast between PAN *C and *t? In Robert Blust,

Editor, Papers on Austronesian Languages and Ethnolinguistics in Honour of George Grace, 535-549 Pacific Linguistics, C-117.

1993a   The PAN Phonemes *ñ and *N. Oceanic Linguistics 32.1:45-61.

1993b   Proto-Austronesian Stress. In Jerold A. Edmondson and Kenneth J. Gregerson, eds. Tonality in Austronesian Linguistics, Oceanic Linguistics, Special Publications, number 24, pp. 1-15.

The Position of the Formosan Languages: Method and Theory in Austronesian Comparative Linguistics[*]

Robert Blust

University of Hawaii

Phonological, lexical and to some extent morphological comparison of the non-Formosan Austronesian languages yields a reconstructed language ("Proto-Malayo-Polynesian") which differs in certain significant respects from the reconstructed language which results when Formosan languages are included in the comparison ("Proto-Austronesian"). The purpose of this paper is to sketch out some of the more salient differences between Proto-Austronesian and Proto-Malayo-Polyesian, and to relate these differences to more general issues of method and theory in comparative linguistics. An attempt will be made to answer the question whether the Formosan aboriginal languages constitute a single primary branch or more than one primary branch of the Austronesian languages, and to relate this answer to the general problem of reconstruction in language families that have a binary tree structure.

Method and theory

If we define method as a set of constraints imposed upon the analysis of data, and a theory as a model of reality, then it can be shown that method-ological differences have been largely responsible for differing theories of the position of the Formosan aboriginal languages. In this paper I argue that

---

only unilateral inferences of change provide a secure basis for subgrouping, and that exclusively shared lexicon can support unilateral inferences of change only under certain well-defined conditions. Given these methodological constraints, three theoretical conclusions emerge: 1. the existence of a Formosan subgroup remains moot, 2. there is unambiguous evidence for a subgroup which includes all extra-Formosan languages, and 3. claims that have been made for a Formosan-Philippine subgroup are unsupported. The notion of "similarity" that has been invoked by some scholars as a basis for the formation of linguistic subgroups raises methodological issues which are strikingly parallelled in recent debates in the field of biological taxonomy.

The position of the Formosan languages

At least since Klaproth (1822) it has been generally recognized that the indigenous languages of Taiwan are Austronesian (AN). However, the subrelationships of these languages, both to one another and to their relatives outside Taiwan, remain controversial even today. Two fundamentally distinct views have been defended. In the first of these, which I will call the "Primary Branch Hypothesis", the Formosan languages are believed to form one or more primary branches of the entire AN family. In the second, which I will call the "Formosan-Philippine Hypothesis" these same languages are believed to subgroup immediately with the languages of the Philippines.[1]

---

1 I use the term 'hypothesis' to refer to a theory of limited scope. Hypotheses are thus intended to explain a relatively narrow and more derivative range of empirical observations in comparison with the broader and more fundamental explanatory aims of theories. In this sense a theory can be seen as an interlocking network of hypotheses. However, the terms are often used almost interchangeably in the literature on philosophy of science. For instance, Popper (1972:194),

1.1. The Primary Branch Hypothesis.

The earliest explicit statement of the view that the Formosan aboriginal languages constitute a primary division of the AN language family apparently is that of Haudricourt (1965:315), who recognized a tripartite structure: 1) Western Austronesian, which includes island Southeast Asia apart from Taiwan, together with Madagascar, parts of mainland Southeast Asia, Palauan and Chamorro in western Micronesia, and Yami of Botel Tobago Island (Lan Yü), 2) Northern Austronesian, which includes the Formosan languages, and 3) Eastern or Oceanic Austronesian, which includes the AN languages of Melanesia and Polynesia, together with most of the languages of Micronesia.

A similar view has been expressed by Dahl (1976:123ff), who, however, regards the non-Formosan languages as a unit which further divides into "Western" and "Melanesian" (= Oceanic) branches.

Blust (1977) proposed an AN family tree which recognized three primary branches in Taiwan (Atayalic, Tsouic, Paiwanic), and a single Malayo-Polynesian branch comprising all other AN languages. In later publications (as Blust 1982, and 1983/84a) the Formosan languages were treated for purposes of lexical reconstruction as constituting a single primary branch. This difference of treatment, however, did not reflect a commitment to the hypothesis that the Formosan languages constitute a subgroup. Rather, it reflected a view that no lexical reconstruction could safely be assigned to Proto-Austronesian (PAN) if its known distribution is confined to the

---

referring to 'Newton's theory' of gravity, continues "It was a hypothesis concerning the ultimate or essentialist causal explanation of gravity itself which he had in mind when he wrote in the *Scholium generale* at the end of the *Principia:* 'So far I have explained the phenomena ... by the force of gravity, but I have not yet ascertained *the cause of gravity itself* ... and I do not arbitrarily [or *ad hoc*] invent hypotheses.'"

Formosan languages, since the latter have been in close geographical proximity and hence in a potential borrowing relationship for perhaps six millenia.[2]

Most recently, Ross (1992) has defended the view that the first split within the AN language family produced one or more subgroups in Taiwan, and a single Malayo-Polynesian subgroup comprising all other AN languages.

1.2. The Formosan-Philippine Hypothesis.

So far as I know, the hypothesis of a Formosan-Philippine subgroup was first advanced by Dyen (1963:268ff), who based it on lexical distributions which appeared to be confined to languages in these two geographical regions. Tsuchida (1976:13-14) adopted a similar position for similar reasons, citing a number of proposed Formosan-Philippine lexical innovations in addition to those noted by Dyen.

More recently Dyen (1990) has returned to this matter, making use of what he calls "homomeries", defined as "Different sets of cognates distributed over exactly the same set of languages". By way of illustration he presents a sample set of 29 proposed cognate sets (said to be a selection from a total of 400 unique pairs) which are reportedly confined to the Formosan and Philippine languages. Dyen and Tsuchida (1991) reiterate this claim, presenting the same set of proposed innovations less three

---

2   The complete tree proposed in Blust (1977) and modified in subsequent publications has four major nodes, as follows (F= Formosan, MP= Malayo-Polynesian, W= Western, C= Central, E= Eastern, OC= Oceanic): 1. AN yields F + MP, 2. MP yields WMP + CEMP, 3. CEMP yields CMP + EMP, 4. EMP yields South Halmahera-West New Guinea + OC. The tripartite division of the Formosan languages was taken over largely from Dyen (1965b) and Ferrell (1969). Ferrell did not take a position regarding the wider relationships of Formosan languages; the position taken by Dyen in 1965 is discussed under 1.3.

(*$q_2$ali-W$_2$aDaŋ 'collar-bone', *Q$_2$uS$_{23}$aw 'thirsty', *[tT]imiD$_2$ 'chin') which presumably were eliminated because external cognates were identified between the writing of the two articles.[3]

Finally, Wolff (1991:536) argues on the basis of similarity in grammatical structure "and the portion of shared vocabulary in the basic items of the sort that tend to be old in a language (not just the 200-item Swadesh list, but a much broader vocabulary)" that the Formosan and Philippine languages constitute a single subgroup.

1.3 Hybrid hypotheses.

Several other proposals concerning the position of the Formosan languages are perhaps best characterized as "hybrid", since they contain features of both the Primary Branch Hypothesis and the Formosan-Philippine Hypothesis.

Dyen (1965a) found lexicostatistical evidence for an Atayalic Subfamily which appeared to form one of 40 primary branches of the AN language family. He did not consider material from any of the Tsouic languages, but grouped Amis, Paiwan, Bunun and Thao into the East Formosan Hesion, which he treated as one of seven primary branches of the Malayopolynesian Linkage, itself a primary branch of Austronesian. In effect, Dyen's position based on the lexicostatistical evidence contradicted his earlier (1963) position based on the evidence of exclusively shared cognates, since it assigned Atayalic and the other Formosan languages in his sample to different primary branches of Austronesian. Dyen (1990:212) acknowledges this contradiction,

---

3   PMP *timij or *timid 'chin, jaw', is proposed in Blust (1980), while PAN *qali-wadaŋ or *kali-wadaŋ 'clavicle, collarbone' and PAN *quSaw 'thirst' are proposed in Blust (1983/84b). For each of the first two etyma supporting evidence is cited from Formosan, Philippine and other AN languages.

and has attempted to resolve it by giving preference to the "homomeric" method. As a consequence he now views all of the aboriginal languages of Taiwan as constituting a single subgroup, called "Formosan", which is most closely related to the languages of the Philippines.

A somewhat different proposal was made by Harvey (1982), who accepted the basic framework of Blust (1977), but claimed that Amis subgroups with Malayo-Polynesian. This view is adopted by Reid (1982), and further elaborated in ways that need not detain us here.

To summarize, published views on the position of the Formosan languages relate to two issues: 1. ·.hether these languages form a subgroup, and 2. whether they (separately or jointly) subgroup immediately with the languages· of the Philippines, or are coordinate with the extra-Formosan languages as a whole. Each of these issues will be considered in turn. Before doing so, however, it will be useful to briefly consider the archaeological literature on Taiwan, particularly as this relates to the broader issues of determining the AN homeland and major population movements into other areas.

The archaeology of Taiwan in perspective

The archaeology of the AN world as a whole is magnificently presented in Bellwood (1978) and Bellwood (1985). Each of these publications is an exhaustive survey of the literature up to about one year prior to the publication date. As in any attempt to go beyond a dry recitation of disconnected observations, these surveys present a theory of culture history based upon the archaeological record.

Drawing largely on earlier work by K-C. Chang, Bellwood sketches the history of the Taiwan Neolithic (defined primarily by the presence of pottery)

as follows. The earliest Neolithic sites are concentrated on the western coast. Chang regarded them as manifestations of a single cultural tradition, which he called "Ta-p'en-k'eng." The oldest radiocarbon date associated with the Ta-p'en-k'eng culture is a date of about 4,300 B.C. obtained from the site of Pa-chia-ts'un, near T'ainan (Bellwood 1985:213). Around 2,500 B.C. there is evidence of two distinctive archaeological traditions in Taiwan which appear to be divergent continuations of the Ta-p'en-k'eng Culture. The first of these, which Chang called "Yüan-shan" is found in the north and east. The second, which Chang called "Lungshanoid" because it resembles the contemporaneous Lungshan of mainland China, is found in the south and west. It is Bellwood's belief that only the Yüan-shan culture of northern and eastern Taiwan has significance for the settlement of the Philippine and Indonesian islands to the south. Table 1 presents a summary of Neolithic sites in the AN world for which generally accepted radiocarbon dates have been obtained. For each major region covered (Taiwan, Philippines, Indonesia, Melanesia, Polynesia) I have chosen the earliest dates that are not considered questionable.[4]

Although it probably would be simplistic to read off the above dates (or some generalized form of them) as corresponding with the proto-languages in

---

4   Dates for island Southeast Asia are taken from Bellwood (1985), who associates the Neolithic assemblages of this region with populations ancestral to later AN-speaking groups based on widespread patterns of cultural continuity. In accordance with the usual practice of Pacific archaeologists I use Lapita pottery as a marker of archaeological cultures which probably were associated with speakers of AN languages in the insular Pacific. I follow Spriggs (1988) in dating the Lapita sites of the Pacific, including his rejection of dates produced at the Gakushuin Laboratory during the 1960's and 1970's. For the sake of uniformity (since Bellwood's dates are expressed in years B.C.) Spriggs' B.P. dates have been converted to B.C. by substracting 1,990 years.

TABLE 1

Radiocarbon dates associated with the Austronesian settlement of island Southeast Asia and the Pacific

AREA	LOCATION	SITE	DATE (B.C.)
Taiwan	T'ainan	Pa-chia-ts'un	4300
Philippines	n. Luzon	Musang Cave	post-3500
Philippines	Masbate	Bagumbayan	2000
Indonesia	Sangir	Leang Tuwo Mane'e	2500
Indonesia	Sabah	Madai	2000
Indonesia	s. Sulawesi	Ulu Leang	2500
Indonesia	e. Timor	(4 caves)	2500
Melanesia	Manus	Kohin Cave	1870
Melanesia	Mussau	ECA, Eloaua	1482
Melanesia	Siassi	KLK	1222
Melanesia	New Britain	Lolmo	1411
Melanesia	Nissan	DFF, Lebang Halika	1664
Melanesia	Fiji	Naigani 21/5	937
Polynesia	Tonga	Mangaia Mound	794
Polynesia	Samoa	Mulifanua	1075

a right-branching tree structure, the overall pattern surely cannot be without significance: in general as one moves south from Taiwan and east from Indonesia the earliest radiocarbon dates with likely AN associations show a steady decrease in age. The simplest culture-historical interpretation of this pattern, as Bellwood has stressed repeatedly, is that AN-speaking peoples settled Taiwan initally from the adjacent mainland of China (Yangzi delta to

Fujian) before moving on to the Philippines, Indonesia and the Pacific (cf. e.g. Bellwood 1991, Thiel 1984/85).

If the foregoing picture is accurate the aboriginal peoples of Taiwan have inhabited the island (perhaps initially with a pre-AN population of Negrito hunter-gatherers) for upwards of 6,000 years, or over 300 genera-tions at 20 years per generation. Under such circumstances ongoing cultural and linguistic borrowing over a period of millenia would be all but inevitable. Indeed, according to Ferrell (1969:13) "within Taiwan extensive inter-ethnic borrowings and reformulations were the pattern rather than the exception." During the ethnographic present the aboriginal peoples of Taiwan have been confined largely to the mountainous interior and the narrow eastern coast, and apart from the Yami have lacked seagoing boats. To conclude from the present situation that the native peoples of the island have always been so isolated would nonetheless be a serious mistake. According to Ferrell (1969:52) the only seagoing craft presently used by the Taiwan aborigines are large bamboo sailing rafts. However, sea-going canoes, possibly with outriggers, were reported for the Kavalan of Ilan early in the 19th century. Although Ferrell himself lays greater emphasis on the role of castaways in Taiwan as sources of borrowing between Formosan and other ethnic groups, it is clear that the possession of boats by the aborigines themselves would have provided opportunities for contact with other areas, particularly the northern Philippines. This theme is given archaeological credibility by Bellwood (1985:224), who observes that excavations on the island of Luzon "are well on the way to demonstrating a significant Taiwan-northern Philippine axis of Neolithic continuity."[5]

---

5   The former possession of boats by the ancestral aboriginal population throughout Taiwan is further confirmed by reflexes of *qabaŋ 'boat' in Kanakanabu, Saaroa,

The evidence for a Formosan subgroup

With a minor exception to be noted below, no phonological, grammatical or lexicostatistical evidence has been presented to date in support of a Formosan subgroup. The arguments for Proto-Formosan thus consist almost entirely of proposed lexical innovations. Most of these appear in Dyen (1963) and Tsuchida (1976). As a hedge against possible undetected borrowing or subgrouping error, Dyen (1963) required that all supporting evidence for a Formosan subgroup include proposed cognates in at least one Atayalic and one Paiwanic language, and this requirement is maintained in Tsuchida (1976).

Dyen (1963) drew his data from Ogawa and Asai (1935). While this source was a valuable pioneering study which contained the first comparative vocabularies of Formosan languages, it has been superseded by a wealth of new studies over the past two decades. For our purposes the most important of these studies (by language or language group) are: 1) Atayal (ATY): Egerod (1980); Atayal and Seediq (SED): Li (1980, 1981, 1982a); Seediq: Pecoraro (1977), 2) Saisiyat (SAI): Li (1978), 3) Pazeh (PAZ): Ferrell (1968), Tsuchida (1971),4) Tsou (TSOU): Tung (1964); Tsou, Kanakanabu (KNB) and Saaroa (SAR): Tsuchida (1976), 5) Kavalan (KAV): Li (1982b), Tsuchida (1971), 6) Thao (THAO): Li (1976, 1983), 7) Bunun (BUN): Jeng (1972), Li (1988), Tsuchida (1971), 8) Rukai (RUK): Li (1977), 9) Amis

---

Rukai and Siraya of Taiwan, and by Western Bukidnon Manobo and Tiruray of southern Mindanao in the Philippines (Blust 1973). To these we can now add Moken *kaban* 'boat', thus extending the evidence for a continuous cultural transmission of at least one type of boat in the AN world (including Taiwan) for a period of several millenia.

(AMIS): Fey (1986), 10) Puyuma (PUY): Ting (1978), Tsuchida (1980), 11) Paiwan (PAI): Ferrell (1982), Ho (1978). In addition to these publications, which are based on original fieldwork, Shigeru Tsuchida has collected and published a number of manuscripts composed during the Japanese occupation of Taiwan (most notably Tsuchida 1982, 1985, and Tsuchida, Yamada and Moriguchi 1991). The most poorly described languages that still are spoken are Pazeh, Thao and Kavalan.

In view of these new materials, and of ongoing comparative work in the AN family as a whole, it is clear that many of the 37 proposed Formosa-only lexical innovations in Dyen (1963) must be discarded. To avoid selection bias I will consider each comparison in turn, marking it with 'NO' if it is rejected (together with the reason for rejection), and 'YES' if it is accepted as a plausible candidate for Proto-Formosan. In general I do not attempt to distinguish dialect differences noted by Dyen, nor to cite more than the minimum material necessary to make my point. In addition, due to typographical limitations I use S for a voiceless palatal fricative, L for a voiceless alveolar lateral, G for a voiced velar fricative, and D for a voiced interdental fricative. The comparisons are numbered 1.1. to 1.37., as in the original source:

1.1. ATY *pali?*, BUN *pani?* 'feather'. NO. Li (1981) reconstructs Proto-Atayal (PATY) *palid 'wing'. Extra-Formosan reflexes of PAN *paNid/paNij 'wing' are found in the Philippines, Sulawesi, the Moluccas, and a number of Oceanic languages.

1.2. ATY *raho?*, RUK *madau* 'big'. NO. Li (1977) reconstructs Proto-Rukai (PR) *maDaw 'big'. Even if the initial syllable of this form is a fossilized prefix, as Dyen evidently assumed, the correspondence ATY -h- : RUK zero is irregular.

1.3. ATY *ramu*, PAI *jamoq*, PUY *?ada.moq*, PAZ *?damo?* 'blood' NO.

Subgroup reconstructions include PATY *damu? 'blood'(Li 1981:279), Proto-Paiwan (PPAI) *damuq 'blood'(Ho 1978:634), Proto-Puyuma (PPUY) *damuk 'blood' (Ting 1978:361), and Ferrell (1968:84) gives Pazeh-Kahabu *damu?*. The Atayal and Pazeh forms point to a prototype with final vowel (or non-contrastive glottal stop), the Paiwan form to *damuq and the Puyuma form to *damuk. The resemblance between these forms is too strong to deny a historical connection, but the irregularities point to borrowing. This compari-son is instructive, since it involves an item of basic vocabulary. If basic vocabulary has been borrowed between languages which are geographically as widely separated as the modern Atayal dialects are from Puyuma and Paiwan it follows: 1. that non-basic vocabulary has almost certainly been borrowed between these languages or others which are as widely separated, and 2. that the present geographical locations of the Formosan languages cannot be taken as a necessary indication of their locations in the past, particularly prior to the large-scale Chinese immigration of the 17th century.

1.4. ATY *juluŋ*, SED *ruluŋ*, RUK (Maga) *kroroŋo*, SAR *loloŋa* 'cloud'. No. Subgroup reconstructions include PATY *ruluŋ (Li 1981:264), reflecting earlier *luNuŋ, and PR *koroŋo (Li 1977:46) 'cloud'. These forms do not exhibit recurrent sound correspondences, and cannot be considered cognate. The expected SAR reflex of *luNuŋ would be **luLuŋə (Tsuchida 1976: 305ff), a form that disagrees in two respects with *lu:luŋa* 'cloud', cited by Tsuchida (1971:1).

1.5. SED *belebil*, PAI *venilevil*, RUK *bilbil* 'pull'. NO. I am unable to find the SED form in Li(1980, 1981), or Pecoraro (1977). For PAI Ferrell (1982) gives *viLviL* 'pull towards oneself', and for PR Li (1977) reconstructs *bilibili. The SED form cannot be compared with the others under any known formula, and although the forms in Paiwan and Rukai probably are

cognate this distribution could easily be a product of borrowing.

1.6. ATY *meqwas, peqwas,* SED *uwes,* Thao *maqa-qūjiS,* BUN *ka-xūdas* 'sing'. NO. Li (1980:399) gives the Squliq Atayal stem form *pqwas.* Even if this is segmented (apparently arbitrarily) as *p-qwas* the correspondences with Thao and Bunun are irregular.

1.7. ATY *hema-li,* SED *he:ma,* PAI *sema,* PUY *sima,* AMIS *sima* 'tongue'. NO. Likely CMP cognates include Roti *ma,* Atoni *ma-k* and Sika *ma-ŋ* 'tongue' (Lampung *emah* 'tongue' appears to be a chance resemblance).

1.8. ATY *kawas,* THAO *kā.waS,* PAZ *qawas* 'year'. NO. Li (1981:297) reconstructs PATY **kawas* 'year', and Ferrell (1968:94) gives PAZ kawas 'year'. All forms appear to be cognate, but the languages in question are geographically contiguous or nearly so, and borrowing cannot safely be ruled out.

1.9. ATY *hera, so:-hesa,* RUK *ko-eja,* AMIS *ina-tsira* 'yesterday'. NO. Reconstructions for subgroup ancestors include PATY **cu-hig'a?* (Li 1981:297) and PR **koDa* 'yesterday' (Li 1977:81). In addition Fey (1986) gives AMIS *inacila* 'recent past time, yesterday.' None of these forms show regular correspondences.

1.10. ATY *mipusal,* RUK *maposal,* BUN *mapusan,* KNB *mapusanu,* THAO *mapuSaD* 'twenty'. YES. Reconstructions for subgroup ancestors include PATY **ma-pusal* (Li 1981:295) and PR **ma-posa-lə* 'twenty' (Li 1977:78). In addition, regularly corresponding forms are found in Tsou, Kanakanabu, Saaroa, Thao, and Bunun. This item must be considered one of the strongest pieces of lexical evidence yet presented for a Formosan subgroup. However, the reconstruction of PF **ma-puSaN* 'twenty' raises some still unanswered questions. A decimal system of counting can be reconstructed for PAN. In PMP this system persisted, and multiples of ten were expressed by preposing

the multiplier to the multiplicand with an intervening ligature *ŋa (hence PMP *duha ŋa puluq 'twenty', etc.). Since a similar method of expressing multiples of ten is found in Ketagalan, Pazeh, Papora, Favorlang, Amis, Paiwan, and perhaps Kavalan (Ferrell 1969:407-18), the question naturally arises whether these Formosan-extra-Formosan agreements in method of expressing multiples of ten are: (1) a product of convergence, (2) a product of borrowing, or (3) a retention from a common ancestor. If (1) it would be instructive to know what factors might have favored convergent development; if (2) we would have further evidence of borrowing which connects languages at the geographical extremes of the island; if (3) we would be compelled (a) to reconstruct two methods of expressing multiples of ten for PF (even though no attested language uses more than one), or (b) to conclude that reflexes of *mapuSaN have acquired their known distribution through diffusion at a time antedating many of the sound changes that distinguish the modern languages.

1.11. ATY *juŋai*, PUY *uŋai*, AMIS *oʔŋai* 'monkey'. NO. Li (1981:288) reconstructs PATY *ruŋay 'monkey'. (Ting 1978:351) cites *uŋay* 'monkey' for the Kasabakan and Katipul dialects of Puyuma, but does not reconstruct a PPUY etymon. In any event the Atayal forms cannot be related to those in Puyuma or Paiwan because of problems with the initial consonant. This may be loan distribution.

1.12. ATY *qaom*, PAI *ʔaːm*, PUY *harum*, BUN *qalom*, SAR *ʔarumu* 'ant eater'. NO. Many external cognates in Borneo, including Kiput *arəm*, Batu Belah *am*, Kayan *hem* and Ma'anyan *ayɛm* 'pangolin'.

1.13. SED *walo*, BUN *wanno*, KNB *anu*, SAR *ʔaloʔo* 'bee'. YES. Cf. Tsuchida (1976:147) for further supporting evidence.

1.14. ATY *rotok*, SAI *ɣotok*, PAI *lotok*, RUK *lutuku*, TSOU *jutuka*

'rabbit'. YES. I am unable to find the Atayal item in Egerod (1980) or Li (1981), but the following forms are confirmed in the more recent literature: SAI, Taai dialect *Lotok* (Li 1978:153), PAI *lutjuk* (Ferrell 1982), PR *Lotoko (Li 1977:65), TSOU *eutūka* (Tung 1964), all pointing to PF *lutuk 'rabbit'.

1.15. ATY *bǝtakan*, BUN *batakan*, TSOU *puʔtsoknu* 'bamboo sp.'. NO. Li (1981:278) reconstructs PATY *batakan 'bamboo'. The Tsou forms cited by Dyen (1963) do not appear to exhibit regular correspondences with this form, and although the Bunun words can be considered cognate, a hypothesis of borrowing between Bunun and Atayal-Seediq cannot safely be ruled out.

1.16. ATY *hajuŋ*, SED *haroŋ*, PAI *taliŋ*, AMIS *tsalɨŋ* BUN *saaŋ*, TSOU *seoŋɨ*, SAR *(h)alɨŋɨ* 'pine'. NO. External cognates include Isneg *tālaŋ*, Bontok *sáleŋ*, Ifugaw *hāloŋ* 'pitch pine', Tagalog *sāhiŋ* 'white sticky resin', Bikol *sáloŋ*, Cebuano *sáluŋ* 'resin', Bolaang Mongondow *taloŋ* 'tree from which pitch is obtained', and Bisaya (Brunei) *saloŋ* 'pitch, resin', all pointing to *saleŋ.

1.17. ATY kai, SED *ka:li*, PAI kai, SAR kari 'language'. NO. Likely external cognates include Ifugaw *kali* 'speech, language, word; to speak, to say', and Western Bukidnon Manobo kaGi 'say, speak, talk'.

1.18. SED *bulbul*, PAI *vulvul*, RUK *bulbul*, PUY *vulvul*, BUN *bun-bun*, KNB *ta-bunɨbunɨ* 'banana'. NO. Reconstructions for subgroup ancestors include PATY *blǝbul (Li 1981:278), PR *bǝlǝbǝlǝ, PPUY *bǝlbǝl (Ting 1978: 345), and PT *vǝǝvéLǝ (Tsuchida 1976:212) 'banana'. Together with BUN *bunbun* and PAI *veLveL* 'banana' these forms appear to justify a PF *beNbeN 'banana'. However, the Atayalic word (attested only in Seediq, at the southeastern corner of Atayalic territory) evidently is irregular, and is best explained as a loan. Variant forms such as Siraya *bulbil*, Favorlang *bilpil*

and Thao *fiDfiD* 'banana' may also point to borrowing.

1.19. SED *wassau*, PAI *asao*, RUK *vasau* 'leaf'. YES. Although neither Ferrell (1969) nor Li (1977) cites a Rukai form resembling the Taramakau Rukai word *vasau* which Dyen (1963) obtained from Ogawa and Asai, the SED and PAI forms point to PF *waSaw 'leaf'.

1.20. SED *buŋa*, PUY *boŋa*, RUK *voŋa* 'sweet potato'. NO. The sweet potato is not native to Southeast Asia. This item assumes some importance, then, in showing that a borrowed cultigen can acquire cognate names which provide prima facie evidence for a reconstruction. It is, of course, possible that *buŋa* was a PF term with some other meaning, but no reconstruction is currently attainable.

1.21. SED *sepog*, THAO *Smu.piL*, BUN *maSipul*, PAI *semopo* 'count'. NO. Only the Thao and Bunun terms appear to be cognate, and borrowing cannot safely be ruled out as an explanation of this distribution.

1.22. SED *baki*, AMIS *faki* 'grandfather'. NO. Likely external cognates include Maloh *baki?*, and probably Timugon Murut aki, Delang *aŋki*, Iban aki 'grandfather'.

1.23. SED idas, PAI *qilas*, PUY *helas*, PAZ *ʔilas* 'moon'. NO. PAZ *ʔiLas* (Ferrell 1968:81) and PAI *qiLas* (Ferrell 1982) point to *qiNaS. SED *idas* cannot reflect such a form, and must be either a product of convergence or, as appears more likely, a loan (in this connection cf. Favorlang *idas* in Ferrell 1969). The Paiwan and Pazeh forms could be taken to justify a PF reconstruction, but the evidence of borrowing in (at least) Seediq, together with the competing form *bulaN strongly suggests that this is a loan distribution.

1.24. SED *dalih*, RUK *me-dali*, SAR *ma-sa:li* 'near'. NO. Reconstructed subgroup ancestors include PATY *dalih (Li 1981:288), and PR *maʔadidali(?)

(Li 1977:61). Tsuchida (1976:157) cites external cognates from the Philippines, and these can be supplemented by e.g. Lawangan, Katingan (Borneo) *dani*, Kambera (Sumba) *ma-reni*, Helong (Timor) *dani* 'near'.

1.25. SED *ukka*, BUN *ukka*, TSOU *uk?a*, SAR *uka?a* 'not exist'. YES. Reconstructed subgroup ancestors include PATY *?uka?* 'not exist' (Li 1981: 289) and Proto-Southern Formosan *uk₂a 'there is none' (Tsuchida 1976: 176), and to these we may add e.g. BUN uka 'don't have; without' (Jeng 1972). However, as noted by Tsuchida (1976:203) irregularities are found in a number of apparent cognates.

1.26. SED hoda, PAI sola, THAO *?ûlda?*, AMIS *sorla*, KAV *suɤona?*, PUY *?orla?* 'snow'. YES. A convincing case is made by Tsuchida (1976:250), who reconstructs PF *S₂uRəLa 'snow'.

1.27. SED *broa*, BUN *bilva* 'thunder'. NO. The distribution of these forms does not permit borrowing to be eliminated as a plausible explanation of the agreement.

1.28. SED *rebu*, TSOU *sifu*, SAR *i:bu* 'urine'. NO. SED *rebu* does not appear to be cognate with the Tsouic forms. The latter, and such Philippine forms as Ilokano *isbū*, Ifugaw *ihbū*, and apparently Palawan Batak *si?bu* 'urine' can be assigned to *sibu.

1.29. ATY *royiq*, SED *daoriq*, PAZ *?daurik* 'eye'. NO. Both the limited geographical range of these forms, and the irregularity in the final consonant correspondence strongly suggest borrowing as the explanation for this distribution.

1.30. ATY *timu*, SED *timo*, RUK *timus* 'salt'. NO. External cognates include Maranao *timos*, Western Bukidnon Manobo *timus* 'salt', Samal *timus* 'salt from sea spray'.

1.31. SED *balay*, THAO *mu?bala?bā.laj* 'right (correct)'. NO. Li (1981:

291) reconstructs PATY *balay 'right (correct)'. If the Thao form is not a chance resemblance, it could easily be a product of borrowing.

1.32. ATY *siyik*, THAO *riSi?* 'liver'. NO. The correspondences between these forms are irregular.

1.33. ATY *giqas*, PAZ *hijaS* 'new'. NO. Egerod (1980:132) gives ATY *geqas*, and Ferrell (1968:92) gives PAZ *xias*, both meaning 'new'. The correspondences between these forms are irregular.

1.34. ATY *mastamaq*, PUY *matama?* 'rotten'. NO. Probably a chance resemblance. Tsuchida (1980:234, 1983:53) gives PUY *ma-TemeH* 'rotten, of sweet potatoes'. This points to earlier *Cemeq, a form that would regularly yield an Atayal reflex with /c/-.

1.35. ATY *utas*, SED *uttas*, PUY *?as* 'penis'. NO. Li (1981:289) reconstructs PATY *?utas 'penis', but a Puyuma cognate is highly doubtful. Ting (1978:361) refrains from reconstructing a Proto-Puyuma etymon, although most of the dialects he cites contain a reflex of PAN *qutiN. The dialect which he calls simply 'Puyuma' has *iTas*, but this isolated form does not show regular correspondences with the Atayalic material.

1.36. ATY *ruma*, SED *doma*, PAI *dzoma* 'some'. NO. This is a very dubious comparison. Egerod (1980:578) gives ATY *ruma?* 'some, sometime, some time ago, sometimes, from time to time'. The Paiwan form apparently is *zuma* 'other (person, thing)', a form which has external cognates such as Kalamian Tagbanwa *duma?* 'other' and Western Bukidnon Manobo *duma* 'companion; another; to accompany'.

1.37. SED *s-em-inaq*, PAI *semenaq* 'wash', KNB *ma-tsina* 'wash, wash body', TSOU *mamtsi:no*, SAR *ma:sinu* 'wash body'. NO. Again, this is a very dubious comparison. Pecoraro (1977) gives only SED *sinao, s-m-inao* 'wash (vessels, surfaces, solid objects)', while for Paiwan Ho (1978:658) gives *s-əm-*

naw 'wash utensils' for each of the five dialects he compares, and Ferrell
(1982) gives *s-em-enaw* 'wash tools'. Each of these has external cognates,
some of which were noted in another context by Dyen himself (1963:269).

To summarize, of the 37 Formosan lexical innovations proposed by Dyen
(1963:265-66) no more than six (16.2%) appear to be justified. Two of these
are names of animals (bee, rabbit), one names a feature of the natural envi-
ronment (snow) which clearly would have been lost by AN speakers as they
moved south from Taiwan, and a fourth is an alternative method for forming
higher numerals.

The cognate sets assembled by Tsuchida (1976:313-20) reflect a better
control of the relevant sound correspondences than those just considered.
Although Tsuchida reconstructed some 77 Proto-Southern-Formosan etyma,
he proposed only 20 Proto-Formosan forms. Eliminating duplications of
Dyen's list (snow, honeybee), simple and affixed forms of the same stem
(drink), and PFN etyma which are cross-referenced to PAN forms that they
continue (sew, eel), we are left with 15 additional proposed Formosan lexical
innovations. From these we must also subtract at least one and perhaps two
items (mortar, bile/gall-bladder) which have external cognates.[6]

At most, then, Tsuchida (1976) offers 14 new lexical innovations in

---

6   Tsuchida no doubt wished to include *Lusuŋ 'mortar' (p. 128), and *paQ₂əju (p.
     224) 'bile, gall-bladder' because they differ in certain respects from their extra-
     Formosan cognates. Dempwolff (1938) reconstructed *lesuŋ 'mortar', but reflexes
     of *esuŋ, *li(ŋ)suŋ and *lusuŋ are also widely distributed outside Taiwan, and
     other variants appear in individual languages, as with Roti *nesu-k* 'mortar' (Blust
     1980:69). Nonetheless all Formosan forms cited by Tsuchida do appear to reflect
     a variant that to date has not been found outside Taiwan. The second item is
     cognate with reflexes of *qapeju in many of the languages of the Philippines and
     Indonesia. Its value as evidence for a Formosan subgroup rests on a determination
     of the direction of metathesis in the first two consonants.

support of a Formosan subgroup. Dyen (1991:92) claims that Formosan is "supported by over 97 sets," but as has been shown, no more than 20 (6 + 14) convincing cognate sets have been proposed in print.

In addition to this lexical evidence Tsuchida (1976:13, fn. 8) has noted that Formosan and non-Formosan cognates which contain a reflex of *S in medial or final position often differ in the position of the two consonants (e.g. 'head hair', with reflexes of *bukeS in Taiwan, but of *buSek elsewhere). He concludes that "These phonological phenomena, together with lexical innovations, indicate that the Formosan languages as a whole form a single subgroup." However, as observed in Blust (1981a:207), *S metathesis can serve with equal plausibility as evidence for a non-Formosan (Malayo-Polynesian) subgroup. Consider Table 2:

TABLE 2
*S metathesis and the Formosan/Malayo-Polynesian language boundary

	F	MP	PAN		
I	-s-	-h-	-S-	1. *iSiq	urine
				2. *luSeq	tears
				3. *quSuŋ	mushroom
				4. *CuSuR	to string
				5. *SuSuq	empty out
II	-s	-h/0	-S	6. *paRiS	stingray
				7. *CaliS	rope
				8. *tebuS	sugarcane
				9. *CumeS	body louse
				10. *tuqaS	old; mature
III	-s	-h-	(?)	11. *bukeS/buSek	head hair
				12. *liseqeS/liseSeq	nit
				13. *tapeS/taSep	winnow
				14. *CaqiS/CaSiq	sew
				15. *tuduS/tuSud	knee

Table 2 schematizes *S metathesis in AN languages. Reflexes in Groups I and II are unambiguous in indicating the position of *S, since Formosan and extra-Formosan witnesses generally agree (e.g. PAI /isiq/, TAG /ihi?/ 'urine', AMIS /tfos/, TAG /tubô/ 'sugarcane', /tubuh-ān/ 'sugarcane plantation'). By contrast, Formosan reflexes of Group III point to final *-S, and extra-Formosan reflexes to PMP *-h-. For these items any PAN reconstruction that we choose forces us to accept certain irregularities. If we follow Tsuchida (1976) and posit PAN forms such as *buSek the Formosan reflexes of Group III must be seen as innovations, but we have no explanation for the lack of metathesis in the Formosan reflexes of Group I. If instead we posit PAN forms such as *bukeS we imply that the extra-Formosan reflexes of Group III are innovations, but then we have no explanation for the lack of metathesis in the extra-Formosan reflexes of Group II. Appeals to conditioning evidently will not work, since instances of *q are involved both in *S metathesis and in exceptions to it.[7]

It is not my purpose to quarrel with what appears to be an interesting, if still somewhat limited *prima facie* case for the Formosan hypothesis.

---

7   Some extra-Formosan reflexes of *tuqaS appear to show metathesis of the last two consonants, as with Lawangan, Taboyan, Dusun Malang *tuha?* 'old' (Dahl 1981: 49), and Itbayaten toha 'maturity, idea of maturedness'. Here the innovative form seems clearly to be PMP *tuhaq, thus supporting the view that where Formosan and extra-Formosan languages disagree in the placement of the reflex of *S it is the latter, not the former, which are innovative. There is some evidence that *q and *S in non-final syllables also metathesized, as with PAN *qaSelu > TAG *hālo*, BKL *há?lo*, HAN *hál?u*, CEB *hal?ū/alhū*, SML *hallu* 'pestle'; PAN *quSaNap > WBM *hun?ap*, TBL *unaf* (without /k/-) 'fish scale'. However, since some non-Formosan languages agree with Formosan witnesses in pointing to *q- (e.g. LOU kuna- 'skin, peeling, bark, scales', REN *?una* 'outer shell, as of turtle') the metathesis of *q and *S in at least the latter item appears to have post-dated the break-up of PMP.

Rather, I wish to stress the difficulty of interpreting the available evidence. As I have attempted to show, the published lexical support for a Formosan subgroup presently consists of no more than 20 plausible candidates. This is not a great deal of evidence. Moreover, in a limited geographical area which has been inhabited by speakers of AN languages for perhaps 6,000 years the task of distinguishing exclusively shared lexical innovations which reflect common inheritance from exclusively shared lexical innovations which reflect diffusion can present formidable difficulties. As already observed, there are some phonological indications that items of basic vocabulary (blood, moon) have diffused between Paiwanic and Atayalic languages. Given these cases it is impossible to rule out similar borrowing which may have left no phonological trace because the phonemes composing the loanword had the same reflexes in donor and recipient languages at the time of borrowing. This situation would have obtained more commonly in earlier periods, when fewer phonological innovations distinguished the languages, but would still have been possible in later periods for lexical items in which the component phonemes did not involve distinctive reflexes (e.g. Atayalic reflexes of *pat_1aS_1 'write', cited by Tsuchida 1976:151, could easily have been borrowed from Bunun).

The evidence for a Formosan-Philippine subgroup

Given the longstanding geographical proximity of the Formosan languages the Formosan hypothesis is essentially irrelevant to higher-level reconstruction, since in any case basic methodological controls require us to treat these languages as if they form a single branch. The matter is far different with the Formosan-Philippine hypothesis. If the Formosan-Philip-

pine hypothesis should be confirmed by further research many lexical reconstructions that are now labeled 'PAN' would need to be reassigned to a lower-order proto-language. With such major reconstructional and cultural-historical implications at stake the published evidence in support of the Formosan-Philippine group should be carefully scrutinized.

Dyen (1990:224) claims to have assembled some 400 "unique lexical pairs" connecting the Formosan and Philippine languages. Dyen and Tsuchida (1991:89) reiterate this claim, and further propose that the fundamental division within the AN family separates the Oceanic languages from all others (labeled as 'Indo-Formosan' + 'East Indonesian').

Before examining this claim it will be worthwhile to take a moment to question the intended meaning of the expression "unique lexical pair". According to Dyen (1990:212) "Different sets of cognates distributed over exactly the same set of languages are said to be homomerous and to be homomerously distributed. All of a collection of interhomomerous cognate sets constitute a homomery." From the wording of this passage one could easily be led to believe that what is meant by "exactly the same set of languages" is an identical collection of individual languages (e.g. all cognate sets that are uniquely confined to say, Paiwan and Tagalog). In other words, the unmarked interpretation of "set" is of an inductively determined category. But it is clear from e.g. Dyen and Tsuchida (1991:89) that "exactly the same set of languages" refers rather to a concatenation of assumed or predetermined subgroups such as "Formosan" and "Philippine". Since the validity of "Formosan" is open to question, and since Dyen and Tsuchida make no attempt at all to justify such proposed subgroups as "Phillippine" or "Oceanic", it is clear that the expression "exactly the same set of languages" must be understood very loosely. To return to our hypothetical

example, both a cognate set A, which is unique to, say, Paiwan and Tagalog, and another cognate set B, which is unique to, say, Atayal, Amis, Ilokano and Maranao could be said to have a distribution over "exactly the same set of languages", despite the obvious contradiction if we understand "set" from the standpoint of inductively constituted categories.

Keeping such problems of definition in mind I will now consider the evidence offered in support of a Formosan-Philippine subgroup. As noted above, Dyen claims to have assembled some 400 "unique lexical pairs" connecting the Formosan and Philippine languages. The published evidence, however, is far less extensive. Dyen (1990:226ff) lists 29 proposed lexical innovations, all having "meanings associated with body parts". In Dyen and Tsuchida (1991:92ff) the same list is repeated, less three examples which have been dropped (*$q_2$aliW_2aDaŋ 'collar-bone', *Q_2uS_{23}aw 'thirst', *[tT]imiD_2 'chin'). The 26 surviving candidates are evaluated below in the same way that the "Formosan-only" candidates were evaluated above.

1. *buŋuH_1: KNB *na-vuɲu*, SAR *vuɲuʔu*, TSOU *fŋuu*, AMIS *vuɲuh*, BUN (north and central) *buɲu*, SIR *voŋo* 'head', TAG *buŋóʔ* 'skull'. NO. The Formosan forms clearly are cognate (though their distribution may be a product of diffusion). It is not at all clear, however, that TAG *buŋoʔ* is connected with these. Not only are the meanings different (cp. BKL *buŋoʔ* 'skeleton, skull'), but the correspondences for the final consonant are made "regular" through the use of an *ad hoc* device: the Tagalog reflex of *-H_1 can be said to be /ʔ/ simply because it is unique.

2. *buqel: SED *bql-it* 'leg', SAI *boʔol* 'bone', WBM *buqel* 'knee', TAG *buʔól* 'ankle', BIK, HLG bu?ól 'heel, heel-bone'. NO. The segmentation of SED *bql-it* is arbitrary, and the vowels of the Saisiyat form cannot regularly reflect those of *buqel (cf. Li 1978:140, and examples).

3. *[dD₃Z]aLukap: PUY *daLukap*, IBL *calukap* 'sole (foot)', SBL *dawkap* 'palm (hand)'. YES. This set appears to be confined to Puyuma (Tsuchida 1980:223) and to various languages of northern Luzon (Reid 1971:114, McFarland 1977:34).

4. *D₃apaN: ATY *rapal* 'sole, footprint', PUY *dapal*, TSOU *capha*, PAZ *da-dapal* 'sole (foot)', BIK, ILK *dapan* 'sole (foot)'. YES. However, as noted in Blust (1980:65), a doublet *Da(m)pa 'palm of the hand, sole of the foot' is supported by apparent cognates in Cebuano Bisayan (Philippines) and Sasak (western Indonesia).

5. *kamut: PAZ *ka-kamut*, MSK *kamut* 'finger'. YES.

6. *L₂ewS₁e[q₁₄Q₂]: PAI *Luseq*, RUKBd *Léese*, SAR *Leeʔe*, TSOU *rsəə*, DGTC *ləwaq* 'tears'. NO. External cognates include OJ *eluh* and BIMA *olu* 'tears'.

7. *Li[S₁₃Xx₂]iʔ: PUY *ma-Liʔiʔ*, SAR *paLi-a-Lii* 'pregnant', TAG *mag-lihi*, KAP *ag-li* 'to conceive'. NO. External cognates include KEL *malih* 'pregnant', *alih* 'pregnancy' (assumed to be a back-formation from *ma-lih).

8. *liliq: ATY *lilyeq*, FAV *lili* 'armpit', ILK *qak-lili* 'carry under the armpit'. NO. The segmentation of ILK *qak-lili* is arbitrary.

9. *mujiŋ: SED *muhiŋ*, PAI, PUY *mudiŋ-an*, IVT *muyiŋ* 'face', PAZ *muziŋ* 'nose', ILK *mu:giŋ* 'forehead'. NO. Although the Ilokano form is ambiguous for medial *j or *R, the Ivatan form can point only to the latter. Further confirmation of *muRiŋ in the northern Philippines is seen in PGS *moliŋ* 'forehead' (as against *bajaq > *bagá* 'tell, say', *pija > *pigá* 'how much?', etc.).

10. *ŋaLay: PUY *ŋaLay*, RUKBd *ŋaLai*, SAR *ŋaLiʔi*, TSOU *ŋoi*, ITB *ŋaxay* 'saliva', AMIS *ŋaLay* 'to drool'. YES.

11. *ŋiwa[tC], ŋawi[tC]: KAV *mŋait*, ILK *ŋi:wat* 'hare-lipped'. NO.

Probably a convergent resemblance.

12. *ŋu[S$_{1236}$x$_1$]eR: BUN (northern and central) *ŋusul*, ITB *ŋuhey, ŋuhuy* 'nasal mucus', IVT *ŋuhey* 'catarrh, cold'. YES.

13. *pikpik: AMIS *sa-pikpik*, TAU *pikpik*, MbAD *p-ag-ikpik* 'wing'. NO. Most convincingly treated as a convergent semantic innovation, from *pikpik 'sound of patting or tapping', with reflexes extending into western Indonesia (Blust 1983/84b:95-96). For an obvious parallel cf. PUY *pakpak*, DGT *pakpak*, WBM *pakpak* 'wing', from *pakpak 'clap the hands, beat the wings'.

14. *piLek: RUKBd *sa-keLepe* 'eyelid', AMIS *sa-peLek*, WBM *pilek* 'eyelash'. NO. Next to reflexes of earlier *pilek 'eyelashes', Reid (1971:77) lists forms in Palawan Batak, Mamanwa, Ilianen Manobo, and Aborlan Tagbanwa that point instead to *pidek or *pirek. Fey (1986) gives AMIS *plek*, *pdek* 'go off and on; blink', *saplek, sapdek* 'eyelash'; the second of these variants can be compared with BAR *pindo* 'close the eyes'.

15. *pil[ae]y: PAI *ma-pilay* 'crippled, lame', RUKBd *ma-pilai* 'lame', TAG, ILK, BKL *pi:lay* 'crippled'. NO. Ferrell (1982) gives PAI *(ma-)piLay* 'lame, crippled', and Li (1977:58) reconstructs PR *mapilay 'lame', both pointing to earlier *piNay. The Philippine forms can more plausibly be compared with IBAN *pilai* 'paralyzed, paralyze', or (in the case of Tagalog) with MAL *pirai* 'rheumatism, rheumatic or gouty stiffness of the joints'.

16. *piŋi?: KAV *piŋi*, SUBS *si-piŋi* 'cheek'. NO. Sindangan Subanun *si-piŋi* appears to be cognate with other Philippine forms that reflect *pis(e)ŋi (ITB *pisñi*, TAG *pisŋi*, MSK *pisŋi*, SML *pisŋi*, perhaps GTL *putoŋi* 'cheek', with unexplained *i > /u/ in the initial syllable). Since Kavalan reflects *s as /s/, and permits various medial clusters (Li 1982b) there is no obvious way in which it could be cognate with reflexes of PPH *piseŋi*. Similar, but non-corresponding forms meaning 'cheek' appear in other AN languages, as with

BUN (Ishbukun) *pi?iŋ*, KEL *piŋit*, and KAY *piŋah*.

17. *puju?: PAI *quli-pa-pudu-an* 'crown (head), whorl (head), cowlick', PUY *HaLi-pudu-an* 'hair whorl', BKL *qarim-pu:ru*, SL *qalim-pu-pu:ru* 'cowlick, hair whorl'. YES. However, other two-syllable stems affixed with the *qali/ kali-prefix occur in various languages of western Indonesia in the meanings 'hair whorl; whirlpool, whirlwind' and the like (Blust 1983). Although none of these are known to be cognate with *puju?, one may be regarded as reflecting a doublet: BKL *ali-pūros* 'whirlwind, cyclone, tornado', TAE' *talim-puru?* 'whirlwind'.

18. *pu[nñ]u[q₁Q₂]: PAI, AMIS *punuq*, PUY *punuH*, SAR *punu?u*, TSOU *pnuu* 'brain', RUKBd *punup* 'forehead', THAO *pu:nuq*, PAZ *punu?* 'head', TAG *puno?*, BKL *puqun*, KAP *pun* 'chief, head of a group', TAG *pu:noq*, KAP *pu:nuq*, CEB *punu:q-an*, MAR *sapi-puun* 'beginning, source, origin'. NO. None of the Philippine forms cited here can be said to be "associated with body parts." Rather, all refer to 1. a social leader, or 2. the meaning 'beginning, source, origin'. As argued in Blust (1991:122), the whole extra-Formosan set almost certainly reflects *puqun 'beginning, source, origin; base of a tree'.

19. *[qQ]a[dDZ]eŋ: BUN (southern) *haduŋ*, BKL *qa:ruŋ* 'mole', ITB *qadeŋ* 'a large mole'. YES.

20. *qaLeb: BUN (northern and central) *qaa?*, MGD *aleb*, MAR *leb* 'knee', BUN (southern) *haab* 'shin'. NO. External cognates include KEL *aleb* 'knee' (Blust 1970:137), KEN *lep* 'knee' and Dalat Melanau *teb-eleb* 'thigh' (the latter two, like Maranao, reflecting a variant with penultimate *e).

21. *[q₁Q₂]añiC: ATY *qali-n* 'bark', PAI *qalic* 'skin', SAR *?alici*, TSOU *hici* 'leather', TAG *qa:nit* 'scalp, skin', BKL *qa:nit* 'leather, hide'. NO. External cognates include KEL *anit* 'skin, bark', and Tebeduh *añit* 'skin'

(Blust 1970:137).

22. *rukap: KAV *ruqap* 'palm, sole', YAMI *rukap* 'palm (hand)'. YES.

23. (a) *[sθ]ebuʔ: SAI *kä-hbuʔ* 'urine', *h-om-buʔ* 'urinate'.

23. (b) *q₄i-[sθ]ebuʔ: KNB *iivuʔ*, SAR *ʔivu*, TSOU *sifu*, ILK *qisbu*, BTK *siqbu* 'urine'. YES.

24. *[tT]eLek: KNB *téekeʔ* 'ear', SBLBt *teek* 'ear; deaf', PGS *telek* 'deaf'. NO. Tsuchida (1976:297ff) suggested that Kanakanabu reflects Proto-Southern Tsouic (and PAN) *l as zero when a pre-Kanakanabu *n follows, but as /l/ elsewhere. Dyen and Tsuchida (1986) instead propose to replace the earlier two-way contrast of *l and *N with a three-way contrast between *l, *L and *N by splitting the old *l into two distinct phonemes. The recon-struction of *L in this form is best seen as an *ad hoc* device, since the only other instance of *L > zero that Dyen and Tsuchida (1986) cite in Kanakanabu is seen in *ZaLan > KNB *caáneʔ* 'road, path', where the original condition proposed by Tsuchida (1976) can explain the loss of the medial consonant.

25. *[tT]uk[tT]uk: SIR (toucktouck), ITB *tuktuk* 'top, crown (of head)'. NO. External cognates include BAL *tuktuk* 'top, tip, extremity', and KEI *tutu* 'top, peak, extremity, upper part; south' (Blust 1983/84b:136). As evidence that the Siraya and Itbayaten forms do not exemplify a semantic innovation, cp. ITB (Yamada 1976:334) *toktok* 'peak, top, tip, topmost, summit, acme'.

26. *CiŋaS₁₃: ATY, SED *siŋas*, PAI *ciŋas*, PUY *Tiŋaʔ*, RUKBd *muaciŋase*, SAR *li-u-tiŋa-a*, KAV *tiŋas*, PAZ *siŋas-en*, TAG *tiŋa* 'interdental food parti-cles'. NO. External cognates include IBAN *tiŋaʔ* 'shred, splinter, esp. food stuck between teeth' (Blust 1970:124), and BAR *tiŋa* 'food particles caught between the teeth'.

In summary, only nine of the 26 proposed Formosan-Philippine lexical

innovations in Dyen and Tsuchida (1991) can be considered convincing. Of these, two (nos. 4 and 17) must be qualified, since similar (but non-cognate) forms occur outside the proposed subgroup, and four others (nos. 3, 10, 12 and 22) are confined to the northern Philippines and Taiwan. Perhaps even more damaging to the Formosan-Philippine hypothesis, three of these four items (all but no. 3) are confined to the extreme northern Philippines and Taiwan, or even to Botel Tobago Island and Taiwan (22).

It is important to emphasize that my decisions regarding the above comparisons are not motivated by negativism or special considerations of any kind. On the contrary, for the past three years I have been engaged in compiling a new comparative dictionary of the AN languages, a task that --- although far from complete at this time ---has already required me to make many thousands of decisions about the causes of lexical similarity in AN languages. To one familiar only with textbook descriptions of how to apply the comparative method the matter may seem to be cut-and-dried, but in practice it is not. I have used material from about 120 regular sources and a smaller number of occasional sources to reconstruct nine different proto-languages which I feel are supported by the comparative evidence: 1. Proto-Austronesian (PAN), 2. Proto-Malayo-Polynesian (PMP), 3. Proto-Western Malayo-Polynesian (PWMP), 3a. Proto-Philippines (PPH), 4. Proto-Central-Eastern Malayo-Polynesian (PCEMP), 5. Proto-Central-Malayo-Polynesian (PCMP), 6. Proto-Eastern Malayo-Polynesian (PEMP), 7. Proto-South Halma-hera-West New Guinea (PSHWNG), and 8. Proto-Oceanic (POC). For each of the regular sources I have systematically searched narrow bands of the lexicon (e.g. reflexes of PAN *i, *hi, *qi, *Si, then reflexes of *iV-, *hiV-, *qiV-, *SiV-, then reflexes of *ib-, *hib-, *qib-, *Sib-, etc.). The results do not show a sharp binary split into obvious cognates and obvious non-

cognates. Rather, they show a continuum from a modest percentage of obvious cognates at one end to a large percentage of obvious non-cognates at the other. The analytical problem for the comparativist is what to do with the small percentage of material that falls between these extremes.

As one feature of the dictionary I have kept an ancillary file of "Rejects" which includes material of two kinds: 1. probable loans, either of native AN forms, or of extraneous forms that have been spread most commonly through the medium of Malay, and 2. weakly attested resemblances that are in my judgement best treated as convergent innovations. To date the dictionary contains about 830 entries that have been checked thoroughly against all sources (plus some 3,000 others which have not). In generating these 830 entries I have also generated 337 rejects. In assigning candidates to the "Reject" category I have consistenly adhered to the following methodological principles. PRINCIPLE 1: No form which is found only in geographically contiguous members of different subgroups will be attributed to the parent language of those subgroups, PRINCIPLE 2: No comparison which depends crucially on an ad hoc morphological analysis (what Geraghty 1983:2 calls the "benign slash") will be allowed, PRINCIPLE 3: Comparisons which have no members that are well-matched in meaning are not allowed, PRINCIPLE 4: Comparisons which are attested in only two witnesses for a given proto-language are suspect, and are generally rejected.

To illustrate the methodological controls exercised under Principle 1, even though I reject Dyen's view that the languages of Sabah subgroup with those of the Philippines, I do not accept a southern Philippine-Sabahan cognate set as evidence for a Proto-Philippine or Proto-Western Malayo-Polynesian etymon, since borrowing offers an equally plausible explanation of the facts (and is in some cases demonstrable). The matter would, of course, be

different for a cognate set confined to Sabah and the *northern* Philippines, since then geographical contiguity would not play a role. For much the same reason I do not accept a cognate set which is confined to southern Taiwan and the northern Philippines as evidence for a PAN etymon unless the phonological development of the form involves such regular and fundamental differences that borrowing appears highly unlikely.

To illustrate the methodological controls exercised under Principle 2, I reject Dempwolff's (1938) *ibay 'nausea', since the supporting evidence (TAG *ibay* 'giddiness or intoxication caused by overchewing betel nut', MLG *ivi-vy* 'nausea in the stomach') can be maintained only through use of a "benign slash."

The methodological controls exercised under Principles 3 and 4 are probably best illustrated together, since a decision under Principle 4 often depends on one made under Principle 3. Among comparisons which are superficially appealing, but which have been assigned to my "Reject" file on the grounds that they are best regarded as chance resemblances, are the following: 1. TAG *palá* 'so, so then', MGG pala 'so that, in order to', 2. SIM *ituŋ* 'grandmother' (Lekon dialect), OJ *ituŋ* 'great-great grandchild', 3. KB *ugus* 'rub against something', TB *ugus* 'rub, scrub oneself', SA'A *usu, usu-usu* 'rub, daub, wipe, grate', 4. NgD *babaŋ* '(used only in combination) flow strongly (tears)', NGA *bhabha* 'ooze, flow slowly out (as palm sap from a cut trunk)', 5. ATY *ima?* 'who; anyone', KOI *ima* 'the one who, which', 6. KNB *ta-ʔəvəŋ-a*, SAR *ta-ʔuvuŋ-a* 'pigpen, pigsty', ILK *óboŋ* 'pigsty, pigpen; a small enclosure for swine, situated at some distance from the house', PGS *óobŋ* 'nest (of bird); be nesting'. Space does not permit an extended discussion, but it should be obvious from this sample that my decisions regarding the evidence offered to date for Formosan, and Formosan-Philippine subgroups is

broadly consistent with the principles I have followed in compiling the comparative dictionary. Others may criticize me for being methodologically too strict, but I would prefer to err on the side of caution than on the side of speculation.

The evidence for Malayo-Polynesian

What, then, is the evidence for an extra-Formosan (= Malayo-Polynesian) subgroup of the AN languages? Dyen (1990:224) asserts that my classification has been "presented with little if any validation." Elsewhere Dyen and Tsuchida (1991:86) point out that "the Formosan and Philippine languages resemble each other more than either does the Oceanic languages," and describe my decision not to use this resemblance for subgrouping as a "fault." Each of these remarks calls for a reply, and each for a different reason.

Dyen's claim that my classification has not been supported by evidence is surprising. In his references he cites none of the relevant literature, and Dyen and Tsuchida (1991) cite only Blust (1978). In fact, Blust (1977) presents evidence for one of the most important Malayo-Polynesian innovations summarized below (shift of *-mu '2pl. actor/possessor' to *-mu '2sg. actor/ possessor'), and support for other parts of the classification is given in Blust (1978, 1982, and 1983/84a).

While the statement in Dyen (1990) simply misrepresents my classification, the statement in Dyen and Tsuchida (1991) expresses a fundamentally flawed idea about the nature of linguistic subgrouping, one which is periodically repeated by scholars who pay lip service to the comparative method, but who suffer from what might be described as a failure of nerve when

faced with its results. The theoretical issue in linguistics can perhaps be stated most succinctly by comparison with another field of investigation, biological taxonomy. By any commonsense classification a dolphin (or whale) is to be included with fish. The gross anatomical form supports this classification entirely: no one would think to argue that a dolphin and a kangaroo, a dolphin and a giraffe, or a dolphin and a human being "resemble each other" more than do a dolphin and a shark, or a dolphin and a tuna fish. But, as we all know, the classification of dolphins with fish is flawed. It is flawed precisely because it fails to assign a *weighting* to shared features, a weighting which would give far greater importance to the inconspicuous lungs, hair, mammary glands, and skeletal anatomy of the limbs which dolphins share with other mammals as a result of common ancestry than to the superficial but far more conspicuous features of body form which dolphins share with fish as a result of convergent adaptation to a similar environment. Likewise, but for quite different reasons, the lungfish is not a fish, but rather a highly conservative member of the genetic unit which includes amphibians, reptiles, birds and mammals. [8]

---

8   My target here is not a straw man. Qualified comparativists should, in principle, know what is relevant and what is not relevant to the determination of linguistic subgroups. But in reality even experienced comparativists sometimes appear to forget basic principles when faced with the pitfalls of actual data. Wolff (1991:535-36), for example, maintains that "if the Philippine languages were more closely related to Malay, Polynesian, Moluccan etc. than they are to Puyuma and Paiwan it would be a unique situation, for languages like Kapampangan, Ilocano and even Tagalog are so similar in structure to Paiwan and Puyuma that one could almost translate from them into the latter on a morpheme-by-morpheme basis, and the portion of shared vocabulary in the basic items of the sort that tend to be old in a language (not just the 200-item Swadesh list, but a much broader vocabulary) is more like the portion shared by, say, French and Rumanian, than by French and German." Setting aside questions about specific factual claims, it is clear that Wolff's objections to the Formosan : Malayo-Polynesian dichotomy could be gener

In general there are four isolable causes of "similarity" (implicitly meaning "special similarity") between genetically related languages, and each of these has its parallel in biological taxonomy. These are: (1) convergence, (2) common retentions, (3) parallel development, or drift, and (4) exclusively shared history. (1) is most clearly seen in linguistic areas (India, the Balkans, etc.), and is the explanation for the similarity of body form in dolphins and fish; (2) is seen when widely separated languages preserve features of their common proto-language which have generally been lost (e.g. the retention of Proto-Indo-European aspirates and accent in Greek and Sanskrit); (3) is seen when languages that diverge from a common ancestor undergo similar changes independently (e.g. the development of pluralizing ablaut in English and High German); (4) is seen when a language undergoes innovations which are inherited (perhaps with further modifications) by its descendants. Only the last of these causes of similarity is relevant to genetic subgrouping.

In a charming and insightful essay on problems of biological taxonomy, Gould (1984) draws attention to the clash between traditional approaches to classification, cladistics (tree diagrams based on 'shared derived characters' or 'synapomorphies'), and phenetics (tree diagrams based on overall similarity alone). Proponents of cladistics reject 'shared primitive characters' (traits found in the ancestor of the whole group) as the basis for subgrouping organisms, allowing only 'shared derived characters' as evidence. Proponents

---

alized to any two neighboring languages or language groups on opposite sides of a major subgroup boundary, particularly where at least one of the subgroups is widely dispersed. For example, both structurally and lexically there is greater "similarity" between Persian and neighboring Armenian than between Persian and the genetically closer but geographically more distant Bengali. While this type of similarity may usefully be mentioned in typological classifications, it is of no value to subgrouping.

of phenetics, on the other hand, care only about "large suites of characters, all expressed numerically and processed by computer (364)." Traditional approaches appear to be characterizable as a largely pre-theoretical mixture of phenetics and cladistics. In Gould's words (363) "no debate in evolutionary biology has been more intense during the past decade than the challenge raised by cladistics against traditional schemes of classification." This intensity is readily comprehensible if we consider only that the traditional concept of 'ape' is regarded as cladistically unsound, since gorillas, chimpanzees and humans form a group apart from orangutans, or more distantly related primates. Historical linguists need not read far to recognize in this biological debate the functional equivalents of exclusively shared innovations (= shared derived characters), retentions (= shared primitive characters), genetic classification (= cladistics) and typological classification (= phenetics). As linguists we can feel a certain satisfaction and justifiable pride in knowing that essentially the same conceptual problems that have only recently been sharply posed and clarified in biological taxonomy were worked out in basic outline by Indo-European scholars well over a century ago. At the same time, as pointed out by Grace (1985), the lessons of biological taxonomy provide a useful backdrop to the parallel debates regarding classification in linguistics: in each case the adversaries fall into two camps divided by the importance given to overall similarity as against the often far less conspicuous evidence of exclusively shared innovations.

To summarize, what is relevant to subgrouping languages (like organisms) is not "similarity", but clear evidence of exclusively shared ancestry. The only unambiguous evidence which may be used for this purpose is that set of differences between languages which permit a unilateral inference of change (that is, where we can be virtually certain which feature is innovated and

which retained). The fact that this type of feature may sometimes appear descriptively insignificant in comparison with broad structural agreements that are less relevant to subgrouping should not surprise us. After all, beneath the obvious fish-like appearance of the dolphin one must look deeper to find the hair, mammary glands, lungs and skeletal anatomy which reveal its true genetic lineage.

5.1. Unilateral inferences of change.

When cognate morphemes are identical in form and meaning we need make no inference about change. However, where a difference exists we must infer the direction of change in accordance with general linguistic principles. One of the basic principles advocated by the Neogrammarians is the regularity of sound change, and in accordance with this principle we are required to rule out unconditioned phonemic splits as a possible type of sound change. This requirement forces us to conclude that where a language, A, has two phonemes corresponding to a single phoneme in another language, B, we must either state phonological conditions for a split in A, or conclude that a phonemic merger has taken place in B. This and other principles governing likely directions of sound change, likely directions of morphological change and the like lie behind all of the inferences in this section.

Many of the innovations which have been proposed for a Malayo-Polynesian subgroup have already been discussed in print. For this reason and to conserve space I will simply refer to published evidence where it exists, with only minimal discussion.

5.1.1. Shift of PAN *S to PMP *h.

Most Austronesianists have maintained that PAN *S has no sibilant reflexes outside Taiwan. The principal exception to this view was expressed

in Blust (1969, 1974b). However, I now agree with Dahl (1981), Zorc (1983) and others that there is little convincing evidence for the survival of *S as a sibilant in the North Sarawak languages, and write PMP *h as the continuation of PAN *S.[9]

Since *h > /s/ is unattested in natural languages, while the shift of a sibilant to /h/ is commonplace, the only serious issue here is whether the change in extra-Formosan languages is a product of one, or of multiple historical events. Given the number of extra-Formosan languages (estimated at over 900) the probability is relatively low that this change was independent in more than a few cases (since otherwise we would expect it to be equally common in Formosan languages).

5.1.2. Merger of *t and *C.

As noted by Dahl (1976:125) the distinction of *t and *C, first recognized by Ogawa and Asai (1935) as *t₁, *t₂, is not made outside Taiwan. The subgrouping value of this merger will be discussed in the following section.

5.1.3. Merger of *n and *N.

Like the distinction of *t and *C, the distinction of *n and *N, first recognized by Ogawa and Asai (1935) as *n₁, *n₂, is not made outside Taiwan. Although both this and the preceding merger are found in some Formosan languages, Dahl attributes these shared innovations to parallel development.[10] It might be objected that if Formosan and extra-Formosan

---

9   In departure from my original simplifying assumption (Blust 1969), I now believe that the voiced aspirates of Bario Kelabit have multiple sources. For example, all pre-Proto-North Sarawak voiced obstruents became voiced aspirates after *e, and when following another obstruent in a reduplicated monosyllable.

10 Only Bunun (Tsuchida 1976:305ff) and Kavalan (Li 1982b:487-88) show complete loss of both contrasts. Since there is little if any other evidence that these two languages form a subgroup, it would seem to follow that the mergers which they share are a product of independent changes. In addition, Amis shows

languages have undergone these two mergers independently, parallel change might also lie behind the shared innovations in various extra-Formosan languages. This possibility cannot be excluded, but once again, given the number of extra-Formosan languages in which the mergers have taken place the probability is very low that both changes were independent in more than a few cases. This in itself would imply the existence of a very large subgroup or subgroups outside Taiwan, even if it were not accepted that *all* extra-Formosan languages have shared an exclusive common ancestor.

Since the subgrouping value of the foregoing set of innovations depends crucially on our assumption about the direction of change, this assumption requires critical scrutiny. If the distinction of *t and *C and of *n and *N should turn out to be not a retention, but rather an innovation, it would count as evidence for a Formosan subgroup, not an extra-Formosan subgroup. Wolff (1991) has proposed just such a reversal of the generally accepted view, maintaining that the distinctions in question have arisen in Formosan languages under suprasegmental conditioning that was previously overlooked. I will not dwell on Wolff's analysis here apart from noting that there are numerous exceptions to the conditions he proposes. The whole matter is treated thoroughly by Ross (1992), who finds that the *t/C and *n/N distinctions should be maintained for PAN.

5.1.4. Merger of *a and *e before *-h.

PAN *-S became PMP *-h, which was retained in Itbayaten, but otherwise disappeared in absolute final position throughout the MP group.[11] There

---

loss of the *t/C contrast (but not the *n/N contrast), and Kanakanabu shows loss of the *n/N contrast (but not the *t/C contrast).

11   It has been known at least since Dyen (1953) that a "thematic" /h/ appears in final position under suffixation in Tagalog and some other languages of the central Philippines. However, where Formosan evidence can be used to check the

are several PAN etyma which end in *-eS, and in at least two of these *e clearly merged with *a in all extra-Formosan languages. Dempwolff (1938) reconstructed *tuma 'clothes louse', but Saisiyat /somaeh/, Kavalan /tumes/ 'body louse', Amis (Kiwit) /tumus/ 'head louse' (Ogawa 1979), Itbayaten/ tomah/ 'small lice-like insects on clothes', point to PAN *CumeS, PMP *tumah, with an unambiguous merger of PAN *e and *a before PMP *h. This inconspicuous and superficially unimportant merger of PAN *a and *e in what might be characterized as a phonological "micro-environment" turns out to be highly significant to an understanding of the historical dispersal of the Austronesian languages. Since unique examples are easily dismissed it is fortunate that a second example of the same merger can be cited. Based on evidence from Tagalog, Toba Batak, Malay, and Sa'a-Ulawa Dempwolff (1938) posited *buRa 'spray water from the mouth', but Formosan evidence points instead to *buReS: Kavalan /buRes/ 'spray water from the mouth', Paiwan /bures/ 'a water squirter of bamboo (child's toy)', /ma-bures/ 'be sprayed or spewed upon'. Kavalan /buRes/ points to *buReS, a form which would have yielded PMP *buRah, and hence forms without -h in all extra-Formosan languages except Itbayaten. Although Paiwan /bures/ is irregular, and may be a loan from another language in Taiwan, the subgrouping issue is unaffected. In these two examples, then, we see a parallelism which is not

historical source of this segment it is often found to be secondary. Zorc (1981) has presented evidence that PMP *-h (written *-H) is reflected as glottal stop in some of the languages of the Philippines and western Indonesia. While there is an undeniable correlation between *-h and -? in some of his best witnesses (as Iban), this correlation is imperfect. Moreover, in some critical examples as *baRiw > Iban /bari?/ "musty, 'gone off', as rice" final glottal stop must be secondary. As a general statement based on clear evidence of regular change, then, it seems fair to say that PMP *-h disappeared in absolute final position in all languages but Itbayaten.

likely to be a product of chance, and which involves a change which could only have been in the direction *-eS > *-ah, not the reverse.

Two other etymologies which may illustrate the same change are more problematic. As noted in Table 2, Formosan and extra-Formosan languages sometimes differ with regard to the order of *S and a stop consonant in an adjacent syllable, as with *bukeS (Formosan) vs. *buhek (extra-Formosan) 'head hair'. Although the great majority of extra-Formosan languages reflect *buhek, some of the languages of the central and southern Moluccas instead reflect a form *buka: Selaru /huka-re/, Yamdena /buke/, /buka-r/, Asilulu /huʔa/, Paulohi /hua/ 'head hair'. While the similarity of these forms to reflexes of PAN *bukeS may be a product of chance, given the above observations about the development of PAN *CumeS 'clothes louse' and *buReS 'spray water from the mouth' we cannot rule out the possibility that PAN *bukeS became PMP *buhek/bukah, with reflexes of the latter variant surviving only in eastern Indonesia. A final example of PAN *-eS > PMP *-ah probably is seen in the word for 'nit, egg of a louse'. Formosan languages as a group reflect *liseqeS (SAI /liʔSiS/ and PUY /liHsa/ show metathesis of *s and *q), while extra-Formosan languages sometimes reflect *lisehaq, and sometimes a form which lacks *-q, arguably *liseqah. Examples of the first of these variants are ITB /lisaha/ 'nit, egg of a louse or other parasitic insect', WBM /lisehaʔ/ 'eggs of the insect parasites of humans or animals', BM /litaʔ/ 'nit, egg of a louse' and KEL /liaʔ/ 'nit, egg of a louse' (< *liqas, Met. from *lisaq). Examples of the second variant are KB /lisa/, SND /lisa/ 'nit', OJ /liŋsa/ 'nit, egg of a louse'. If we reconstruct PAN *liseqeS and assume a change to PMP *liseheq by *S metathesis, we have no way to explain 1. why some extra-Formosan languages appear to have irregularly lost *q, and 2. why all extra-Formosan languages appear to

reflect last-syllable *a rather than *e. Although the merger of *a and *e before final *q is common to many extra-Formosan languages, Toba Batak /lisa/, Kwaio /lita/ 'nit', next to Toba Batak /tano/, Kwaio /ano/ 'earth' (< *taneq) are incompatible with a PMP reconstruction *liseheq. Moreover, Iban /linsa?/ 'eggs of the louse' points to a final consonant other than *q. The conclusion appears to be unavoidable that PAN *a and *e also merged before the final consonant in the PMP word for 'nit, egg of a louse', whether the form of that word was *liseqah, *lisehaq, or both.[12]

5.1.5. PAN *Sepat to PMP *epat 'four'.

Dahl (1981:47) maintains that extra-Formosan languages which should refelct PAN *S_1əpat$_1$ 'four', *S_2apui 'fire' and *H_2uɤat$_2$ 'vein, tendon' with /h/- have instead lost the initial consonant. However, there is little reason to reconstruct an initial consonant in the last item, and the word for 'fire' appears with initial /h/ in Itbayaten, Binukid, and a number of the Manobo languages (Reid 1971:82). The word for 'four' is another matter. Although at least four Philippine languages (Binukid, Ata, Tigwa Manobo, Western Bukidnon Manobo) reflect this form with an initial /h/, the same languages also reflect PAN *enem 'six' with an initial /h/: BKD /haqɨpat/ 'four', 'four', /haqɨnɨm/ 'six', ATA /hopqat/ 'four', /honqom/ 'six', MbT /hɨpqat/

---

12 A change from PAN *liseqeS to PMP *liseqah, *lisehaq would, of course, parallel the development noted in PAN *tuqaS to PMP *tuqah, *tuhaq (fn. 7), and like it would provide evidence for an extra-Formosan innovation in regard to *S-metathesis. It is perhaps worth noting that in a number of the languages of Borneo the contrast of /a/ and /ə/ is neutralized (as /a/) before a final /h/ from any source. This suggests either a universal tendency for [a]/ [ə] neutralization to occur in this environment, or at the very least a tendency for the same kind of change to repeat itself in the history of AN languages. A similar condition explains SAI /somaeh/ 'body louse', with a last-syllable vowel that Li (1978: 141) assigned erroneously to *a rather than *e (*CumeS >someh > somah > /somaeh/).

'four', /hinqim/ 'six', WBM /hiqipat/ 'four', /hiqinim/ 'six' (Reid 1971:85, 132). The same relations hold for Dampelas, of the Tomini-Tolitoli group in Sulawesi, where we find /hapat(A)/ 'four', /honoŋ/ 'six' (Himmelmann 1990:15). In all of these cases it is clear that the forms in question reflect reduplications in which the reduplicating vowel is stereotypically *a (Blust 1974a:135ff, Yamada 1991). What is interesting for our purposes, however, is the initial consonant.

If we appeal to this material as evidence for PMP *hepat we must similarly reconstruct *henem, despite the absence of Formosan evidence for an initial consonant in the latter form. There would seem to be two ways out of this dilemma: 1. assume that PAN *Sepat became PMP *hepat, and that the form *ha-hepat served as a model for the reanalysis of PAN *a-enem as *ha-henem, 2. assume that PAN *Sepat became PMP *epat, and that the attested extra-Formosan reflexes of PAN *Sepat and *enem have acquired an initial /h/ from some other, still undetermined source. The first alternative forces us to conclude that all languages which irregularly lost the initial consonant in the word for 'four' also irregularly lost the historically secondary initial consonant in the word for 'six'. While not implausible, this hypothesis assumes an extra-Formosan innovation PAN *enem > PMP *henem. If instead we choose the second alternative we posit a different extra-Formosan innovation, PAN *Sepat > PMP *epat. In either case there appears to be evidence for a period of exclusively shared history in which the Formosan languages did not participate. In the case of *Sepat > *epat the languages encompassed in the extra-Formosan subgroup include at least those of the central Philippines and Malay, while in the case of *enem > *henem the implied subgroup would include at least various languages of the southern Philippines and the Tomini-Tolitoli languages of Sulawesi.

5.1.6. PAN *Si- to PMP *i 'prefix of the instrumental focus'.

Wolff (1973) reconstructed PAN *i- 'nonpast instrumental passive'. However, the Formosan evidence which was critical to allowing this reconstruction points not to *i-, but rather to *Si-: ATY /s/-, BUN /is/-, PAI /si/- 'marker of the instrumental focus'. What clearly led Wolff to the shape of his reconstruction was the shape of reflexes in Central Philippine languages, since none of these reflect the initial consonant: TAG, BKL, AKL, HLG, CEB /i/- 'marker of the instrumental/benefactive focus'. As with the case of PAN *Sepat > PMP *epat, the direction of change in this form is also unambiguous: PAN *Si- became PMP *i- 'marker of the instrumental focus'.

5.1.7. PAN *Sipes to PMP *ipes 'cockroach'.

Dempwolff (1938) reconstructed *ipes 'cockroach', but the Formosan evidence points instead to *Sipes: SED /supuh/ (< Assim.), SAI /hipih/, KAV /sipes/ 'cockroach'. Diagnostic witnesses for *S- in extra-Formosan languages invariably point to a form with initial vowel: ITB /ipes/, TAG /ipis/, HAN /ipus/, CEB /ipus/, WBM /ipes/ 'cockroach'.

The foregoing three examples (*Sepat, *Si-, *Sipes) all manifest the same tendency: loss of *S-. Since PAN *S became PMP *h more generally, it appears likely that a general weakening of *S was underway in the immediate ancestor of the extra-Formosan languages at the time that it began to diversify and give rise to the MP group. Not only were all examples of PAN *S altered to *h, but some were already beginning to disappear entirely in initial position.

5.1.8. PAN *Siwa to PMP *siwa 'nine'.

Somewhat different from the preceding changes which affected PAN *S was a change of the initial consonant in the word for 'nine'. PMP *siwa

'nine' is among the most stable words in AN languages, and is widely reflected from the southern Philippines to Madagascar, and in many Oceanic languages as far east as Polynesia. Wherever the initial consonant of this form has been preserved it reflects *s (or is ambiguous for *c/s). Contrary to the massive extra-Formosan evidence supporting *siwa, the reflexes in Taiwan point instead to *Siwa. The difference in initial consonant between Formosan and extra-Formosan reflexes of the word for 'nine' is very clear, but what was the direction of change?

The influence of successive numerals upon one another in serial counting is a well-known phenomenon, often cited in textbook discussions of the causes of phonological irregularity (e.g. Latin /quattuor/ 'four', /quinque/ 'five', where /pinque/ is anticipated, but was modified through the influence of the labiovelar onset in the word for 'four'). PAN had a decimal system of counting which can be reconstructed as: 1. *esa/isa, 2. *duSa, 3. *telu, 4. *Sepat, 5. *lima, 6. *enem, 7. *pitu, 8. *walu, 9. *siwa, and 10. *puluq. What must be understood, however, is that *puluq did not signal 'ten' as a specific numeral, but rather 'a group of ten' as a unit in counting by tens. The specific numeral 'ten' was signaled by a proclitic form of *esa joined with *puluq. There is solid comparative evidence that the linking of this clitic with the base for 'group of ten' was mediated by a numeral ligature *ŋa in the common ancestor of the extra-Formosan languages, hence PMP *sa-ŋa-puluq 'ten'; this may have been the case in PAN as well, but the evidence for PAN *ŋa 'numeral ligature' is yet to be collected. In any case it is clear that the PAN word for 'ten' was morphologically complex, and began with *s: *sa-(a)-puluq. In serial counting PAN *Siwa would have immediately preceded *sa-(a)-puluq. PAN *S probably was an alveolar fricative, and *s a palatal fricative. An anticipation of the palatal initial of 'ten' would have trig-

gered a change in the initial of 'nine' from *S to *s, whereas no motive is evident for a change in the opposite direction (from a hypothetical PAN *siwa to *Siwa in a hypothetical "Proto-Formosan").

The claim that the PAN word for 'nine' was *Siwa rather than *siwa, and that a change took place in the immediate common ancestor of the extra-Formosan languages not only is supported by observations about the contaminating influence of successive numerals in human languages generally, but is also supported by observations about the assimilatory behavior of *s and *S in successive syllables of other Austronesian morphemes. In Paiwan *S normally became /s/, while *s became /t/. Yet the reflex of *Sasaq 'whet, sharpen' is not the expected **sataq, but rather /tataq/, a form that points to earlier *sasaq, modified from *Sasaq by anticipation of the quality of the medial sibilant. Similarly, in Saisiyat *S normally became /S/, *C became /s/ and *s became /h/ (Li 1978:140). In light of these regular changes we would expect PAN *CaliS 'rope' and the very problematic PAN word for 'nit' discussed earlier to yield SAI **saLiS and a form with /h/ and /S/, yet what we actually find is SAI /SaLiS/ 'to tie', /S-in-aLiS/ 'rope', and SAI /LiʔSiS/. In both cases *s has been assimilated to *S or the reverse (depending upon the order of consonants in the PAN word for 'nit'), in the former case after *C > /s/, in the latter before *s > /h/. These three examples from two widely divergent AN languages in Taiwan show clearly that the type of anticipatory assimilation hypothesized for the change of PAN *Siwa to PMP *siwa is well-supported by the more general historical phonology of AN languages.

5.1.9. Reduction of PAN *paŋudaN to PMP *paŋdan.

Dempwolff (1938) reconstructed *panDan 'pandanus', but material presented by Tsuchida (1976:208) points clearly to a PAN trisyllable, which he writes *paŋuD4aN. Although the velar nasal is preserved in several Philip-

pine languages (e.g. ILK, CEB paŋdan), there is no evidence for retention of the medial vowel. The reconstruction of PMP *paŋdan is problematic, since apart from CVCCVC reduplications and instances of preconsonantal *R, heterorganic consonant clusters are not allowed in PMP. However, if we reconstruct PMP *paŋudan we are left without an explanation for the absence of a medial vowel in widespread reflexes throughout the Philippines, western Indonesia, eastern Indonesia and Oceania.

5.1.10. Change of *biRbiR to *bibiR 'lip'.

Dempwolff reconstructed *bibiR 'lip', but the Formosan evidence points instead to *biRbiR: SIR /vigbig/, AMIS /filfil/, PUY /viRviR/, PAI /birbir/ 'lip'. All diagnostic witness outside Taiwan support a form without the medial cluster: ISG /bibig/, KAN /bibil/, ILK /bibir/, TAG /bibig/, HAN /bibig/, AKL /bibig/, HLG /bibig/, WBM /bivig/, TB /bibir/, PS *bibiR, BM /bibig/, MAK /bibere/ 'lip'. The direction of this sporadic change can be debated, but given the general direction of regular change in CVCCVC reduplications (often reduced to CVCVC by loss of the first consonant in the cluster) it certainly appears more likely that *biRbiR was original.

5.1.11. Change of *-mu from plural to singular.

Perhaps the single most significant piece of evidence for the Malayo-Polynesian hypothesis is the shift of PAN *-mu '2pl. agent/possessor' to PMP *-mu '2sg. agent/possessor'.

As noted in Blust (1977), the direction of this change is unambiguous. Dahl (1976:122) reconstructed a general pattern of PAN pronouns in which the agent/possessor set is identical to the last syllable of the corresponding subject pronoun: 1sg. *aku:ku, 2sg. *iSu:Su, etc. The one apparent exception to this pattern is *ia:a '3sg.'. To account for this and other superficially dissimilar observations Blust (1977) proposed that all agent/possessor

pronouns were optionally preceded by *ni 'marker of agency and possession'. Omitting the preposed person markers *i and *si, which are irrelevant to the discussion, the entire PMP set can be reconstructed as follows: 1sg. *aku:(ni) ku, 2sg. *ihu/kahu:(ni)hu/(ni)mu, 3sg. *ia:(ni)a, 1pl. incl. *(k)ita:(ni)ta, 1pl. excl. *(k)ami:(ni)mi, 2pl. kamu:(ni)mu, 3pl. ida:(ni)da. Since there is no controversy about the reconstruction of PAN *kamu as a plural subject pronoun there can be no controversy that PAN *-mu was also plural, as it continues to be in all recorded Formosan reflexes. The widespread reflexes of *-mu '2sg.' outside Taiwan must, then, be an innovation.[13]

The mere shift of a plural pronoun to a singular would have little value as subgrouping evidence, since such shifts are common strategies in natural languages for deriving polite singulars (which may, in turn, eventually be bleached of their special character). What is significant about the shift of *-mu to a singular is that only the agent/possessor form was affected, bringing about a new paradigmatic asymmetry in the relationship between long form and short form pronouns, from PAN *kaSu : Su '2sg.' to PMP *kahu : mu '2sg.'. Although the form *-Su evidently was retained as PMP *-hu, reflexes of *-mu are far more widely distributed, extending from Yami of Botel Tobago Island through the Philippines and Indonesia and into the insular Pacific as far east as Fijian and Gilbertese. Since the difference between Formosan and extra-Formosan languages in this case involves a demonstrable innovation in the latter, and since this innovation is of an arbi-trary type which is not likely to be repeated as a parallel development, it carries great weight as subgrouping evidence. The alternative would be to

---

13 I have simplified the presentation in Blust (1977) by omitting details that are irrelevant in the present context. The reflexes of *-mu '2pl. agent/possessor' in most Formosan languages are embedded in longer strings of morphemes.

argue that *-mu (but not *kamu) shifted to a singular pronoun on multiple occasions in the history of the AN languages, affecting all known extra-Formosan reflexes, but no Formosan reflexes.

5.1.12. Change of *iten to *aten.

As also noted in Blust (1977), a set of absolute possessive pronouns, including at least *aken 'mine', *iten 'ours (incl.)' and *amen 'ours (excl.)' must be reconstructed for PAN. Since these forms have a clear morphological relationship to *aku 'I/me', *k-ita 'we/us (incl.)' and *k-ami 'we/us (excl.)', there can be little doubt that in the comparison PAI itjen 'we, us, our' : YAMI y-aten 'ours', TAG ātin 'our, ours (incl.)', CEB n-tun 'our' it is the Philippine languages which have changed. Since no non-Formosan reflexes of PAN *iten have yet been found outside the Philippines it is impossible to determine whether the change of *iten to *aten took place in a language ancestral to all non-Formosan AN languages or in a lower-order proto-language such as Proto-Philippines.

5.2. Bilateral inferences of change.

Dahl (1976) has drawn attention to a number of other features in which Formosan and extra-Formosan languages present an interesting contrast. Features found only in Formosan languages are: 1. reflexes of voiced obstruents symbolized as $*d_1$, $*d_2$, and $*d_3$; features found only in extra-Formosan languages are: 2. contrast of plain and prenasalized medial stops (e.g. PMP *putiq 'white', *punti 'banana'), 3. use of homorganic nasal substitution in verbal morphology (as in Malay *pukul* 'hitting', *me-mukul* 'to hit'), 4. prefixes *paŋ- 'agentive', *maŋ- 'active transitive', *paR- 'prefix of deverbal nouns', and the like which are widely distributed in the Philippines and western Indonesia, but are unreported in Taiwan.

As with the use of "homomeries", the direction of most of these changes is ambiguous: the innovation could be in the Formosan languages as a group, or in the extra-Formosan languages as a group. They do not, therefore, support a unilateral inference of change. However, the use of homorganic nasal substitution in verbal morphology is almost certainly an innovation. The process is fully active in most languages of the Philippines and western Indonesia (including Malagasy), and in Palauan and Chamorro of western Micronesia. Although it is absent in eastern Indonesia and Oceania, traces of its former existence in various Oceanic languages were noted by Dempwolff (1938), and further examples have since been found. If this morphological process had been present in PAN one would expect similar traces to be found in Formosan languages, but very few have been reported.[14]

Finally, some indisputably cognate lexical items show one phoneme in Formosan languages but a different phoneme in extra-Formosan languages. An example is Dempwolff's *balija "batten used in loom weaving". While all extra-Formosan languages which contain an unambiguous reflex point to *l as the second consonant of this form, all Formosan languages which contain an unambiguous reflex point instead to *R: *baRija. There is nothing about such forms which provides any indication of the direction of change, but given the evidence for a MP subgroup which has been considered in this

---

14   Examples that have been found include *kayaw > PUY -ngayaw '(of many people) to go headhunting' (Tsuchida 1980), and Saisiyat /manakiS/ 'climb, ascend', possibly connected with Malay /daki/, /men-daki/ 'climb uphill'. Such cases, however, are rare and show little evidence of former systematic relationships. By contrast, some Oceanic languages, as Roviana, retain functioning reflexes of PAN *-in- 'infix forming deverbal nouns', or PAN *-an 'suffix of location', and a number of languages contain multiple instances of fossilized *ma- 'prefix of stative verbs'.

section and the lack of convincing evidence to date for a Formosan subgroup, one is clearly on stronger ground in treating the shape in Formosan languages as original, hence PAN *baRija, PMP *balija "batten used in loom weaving".

A comparison of the Primary Branch and Formosan-Philippine Hypotheses

Both the PB and the FP hypotheses have been proposed to explain particular sets of observations. In evaluating the relative merits of these competing views, it is necessary to consider whether the same observations can be explained as plausibly in other ways. Since I have defended the PB hypothesis, I will leave it to others to find alternative explanations for the observations which I have cited in support of a Malayo-Polynesian subgroup. My obligation here is rather to show that the observational basis for the FP hypothesis does not lead ineluctably to the conclusion reached by Dyen (1990), Dyen and Tsuchida (1991) and Wolff (1991).

6.1. Shared structural similarity.

Wolff (1991) in particular, has emphasized the close structural similarity of many Formosan and Philippine languages. Dyen and Tsuchida (1991:86) are less explicit, noting simply that "the Formosan and Philippine languages resemble each other more than either does the Oceanic languages." It goes without saying that the systems of verbal "focus" which have (at least in part) given rise to these observations could as easily be a product of shared retentions as of common innovations. This interpretation is further strengthened by (1) the appearance of similar systems of verbal focus, or (2) of

elements of structure which suggest the former presence of such focus systems in languages which are neither Formosan nor members of the Philippine subgroup. In the first category are languages such as Chamorro, many of the languages of Sabah, Malagasy, and perhaps the Batak languages of northern Sumatra. In the second category are languages such as Kelabit-Lun Dayeh and Bintulu of northern Sarawak, Old Javanese, and various of the languages of central and southern Sulawesi. To my mind the simplest explanation of this entire range of observations is that PAN had a system of verbal focus marked by affixes *Si-, *-um-, *-in-, *-an, *-en and *-i (among others). The basic structural properties of this system were retained in PMP, but various other affixes, including *paR-, *maR-, *paŋ- and *maŋ- probably were innovated during this period. This morphologically expanded system of verbal focus was retained in Proto-Western Malayo-Polynesian, but underwent fundamental changes in Proto-Central-Eastern Malayo-Polynesian.

6.2. Exclusively shared vocabulary.

The arguments for the FP hypothesis in Dyen (1963, 1990) and Dyen and Tsuchida (1991) are supported entirely by exclusively shared vocabulary, or "homomeries." As already noted, only 9 of the 23 candidates for FP homomeries presented in Dyen and Tsuchida (1991) bear close scrutiny, and four of them have a distribution which can be explained as easily by borrowing as by common innovation. One possible response to this criticism would be to propose more FP homomeries, since there undoubtedly are many other cognate distributions which are restricted to these two areas. The theoretically important question is whether the addition of proposed FP homomeries in itself would make any difference to the FP hypothesis. Since the homomeric method does not distinguish innovations from retentions, and

therefore cannot permit a unilateral inference of change, at least three considerations (apart from borrowing) are relevant to evaluating a subgrouping claim based on this approach: 1. how much greater is the homomeric corpus linking two proposed genetic units than the competing homomeric corpus linking either of these units with some third unit?, 2. how much comparative material is available for each of the proposed genetic units compared?, 3. are there statistically significant differences in the rates of lexical replacement for any of the populations of languages compared?

6.2.1. Relative sizes of homomeric corpora.

Dyen and Tsuchida (1991:89) admit that the number of unique cognate pairs linking the Philippine and "West Indonesian" languages ("more than a thousand") is larger than that linking the Philippine and Formosan languages ("about 400") by a factor of roughly two and one half. They attempt to reconcile this discrepancy with the FP hypothesis by proposing an old dialect chain which included all three groups, but in so doing abandon strict adherence to a family tree model of subgrouping relationships. In the interest of providing what I hope is a more objective indication of the relative sizes of homomeric corpora linking Philippine languages with Formosan languages on the one hand with West Indonesian languages on the other, I have scanned two completed sections of my Austronesian Comparative Dictionary (*i, with 201 reconstructions, and *w, with 64) and recorded the number of cognate sets that are uniquely represented in any of the pairs which arise from comparing Formosan (F), Philippine (P), West Indonesian (W), East Indonesian (E), and Oceanic (O). The results are displayed in Figure 1:

FIGURE 1

Relative sizes of homomeric corpora linking Formosan, Philippine, West Indonesian, East Indonesian and Oceanic languages

F-P	F-W	F-E	F-O	P-W	P-E	P-O	W-E	W-O	E-O
9	3	4	0	52	2	2	21	5	8

In general, the use of exclusively shared vocabulary for subgrouping is trustworthy only where the quantity shared by one pair of genetic units greatly exceeds that shared by other pairs of genetic units. The reason that there is a need for caution in using such methods is clear: where innovations are not distinguished from retentions two languages or language groups which share more retentions will be erroneously assigned to a subgroup to which they do not belong. However, it can be shown that the danger of confusing lexical innovations with lexical retentions is greatest where the languages compared fall into either of two primary branches of the language family as a whole, since in such a case there is no external control for making the innovation/retention distinction. Where higher-level branchings have already been established by unilateral inferences of change lower-level branchings may more confidently be established by homomeric methods, since in such a case external controls for making the innovation/retention distinction do exist.

It is important to recognize that the corpus represented in Figure 1 is not selected for any particular subgrouping purpose, but includes all data found in the relevant dictionary files. What immediately stands out is the number of Philippine-West Indonesian unique pairs, which exceed the number of Formosan-Philippine unique pairs by a factor of nearly six. If the material from these two files proves to be representative of the dictionary as a whole, it will lend far stronger support than Dyen and Tsuchida have indi-

cated to a homomeric argument for a Philippine-West Indonesian subgroup, at the expense of the FP hypothesis.

6.2.2. Differences in the amount of available material.

To some extent the results in 6.2.1. correlate with the amount of material available for comparison. Time and space will not allow me to enter into details here, but lexicographically the Philippines is one of the best-represented regions in the AN world. The number of Formosan languages for which descriptions are available, and the amount of lexical material in these descriptions is far less than is the case for the languages of western Indonesia. However it is done, it seems clear to me that some type of weighting must be given to these differences in the quantity of material available for comparison before reaching any subgrouping conclusions based on simple numerical values.

The W-E figure of 21 is surprising (compare the much lower P-E figure), and may indicate undetected borrowing from Malay, Makasarese or other WMP languages in various of the languages of the Lesser Sundas. If so, some of the etymologies that have been assigned to PMP based on cognate distributions that include WMP and CMP witnesses may be invalid.

Figure 1 confirms one of the central contentions of Dyen and Tsuchida (1991), that unique pairs confined to the Formosan and the Oceanic languages are extremely rare. However, it must be kept in mind that although nearly half of all AN languages are Oceanic, large parts of the Oceanic region are represented by little more than standardized comparative vocabularies of a few hundred words.[15]

15　Consider Vanuatu, where according to Tryon (1976:87ff) some 102 indigenous languages (exclusive of Polynesian Outliers) are spoken. Dictionaries or extensive word-lists are available for perhaps five of these (Mota, Lonwolwol, Southeast Ambrym, Kwamera, Lenakel). The Southeast Ambrym dictionary contains less

6.2.3. Differences in retention percentage.

Blust (1981b) documents a pattern of statistically significant differences in mean retention percentages of basic vocabulary for major AN subgroups. By comparing the recorded basic vocabularies of 226 attested languages with the reconstructed basic vocabulary of Proto-Malayo-Polynesian the following mean retention percentages were obtained: 1. Western Malayo-Polynesian: 40.5, 2. Central Malayo-Polynesian: 38.9, 3. South Halmahera-West New Guinea: 25.6, 4. Oceanic: 23.6.[16] For reasons explained in the original paper, retention percentages could not be calculated for the Formosan languages. Assuming that these results generalize to the vocabulary as a whole, and assigning an arbitrary value for the Formosan languages, it is clear that the number of uniquely shared retentions (in this case from PAN, not PMP) should be lower for Formosan and Oceanic than for Formosan and WMP or CMP. From the standpoint of a homomeric argument this difference of *individual* retention percentages is less important than the content of retained vocabulary in the Oceanic languages as a whole (hence in POC), but inferences about the latter are crucially dependent on adequate descriptive sources, and these, as already noted, are sorely lacking for much of the Oceanic region.

---

than 1,500 entries, and the Lonwolwol, Kwamera and Lenakel dictionaries less than 2,500. This leaves only one dictionary (Mota, with over 6,000 entries) that is comparable to typical dictionaries for the languages of the Philippines and western Indonesia. Much the same weakness of coverage is typical for the central and western Solomons, and for virtually all of western Melanesia.

16 Blust (1981b) was a conference presentation which has never been published. It made use of a data base that included only 55 languages, but the original data base has continued to expand over the past decade.

6.3. Stochastic network models.

Even if neither of the variables noted in 6.2.2. and 6.2.3. were relevant to determining differences in the size of homomeric corpora, there is a general theoretical explanation for the rarity of Formosan-Oceanic homomeries which follows in a straightforward manner from the structure of the family tree that I have proposed and defended in various publications (cf. fn. 2).

In such a right-branching binary tree structure information which enters at the top may be uniquely retained in just two branches, but the probability that that this will be the case is not the same for all pairs of branches, or dyads. The probabilities for information retention in given homomeric dyads are described under the general heading of "stochastic network models", or the stochastic analysis of "branching processes" (Guttorp 1991, Kelly 1979). Inspection will show that the dyad least likely to uniquely retain information from the top in a right-branching binary tree structure contains the leftmost branch and either of the branches from the rightmost node (hence Formosan-SHWNG or Formosan-OC). This follows from a basic asymmetry in right-branching tree structures: information continuity is terminated only through extinction in a right branch.[17]

To illustrate, a Formosan-WMP homomery requires only one extinction (lexical replacement) at any point along the branch leading from MP to CEMP, since this historical event will be reflected in all lower-order proto-languages (CEMP, CMP, EMP, SHWNG, OC), leaving a retained cognate set only in Formosan and WMP. A Formosan-CMP homomery, on the other

---

17   The same, of course, applies *mutatis mutandis* to a left-branching structure. More generally, then, in any unilaterally branching structure information continuity is terminated only through extinction in an outer branch.

hand, requires two extinctions: one in WMP, and another at any point along the branch leading from CEMP to EMP. Finally, a Formosan-SHWNG or Formosan-OC homomery requires three extinctions: one in WMP, a second in CMP and a third in either SHWNG or OC. Not only are these extinctions mutually independent, but in each case they must take place shortly after branching from a major node to avoid further magnifying the differences between the probabilities of occurrence of different homomeric dyads. It follows that the homomeric dyads F + WMP, F + CMP and F + SHWNG/F + OC occupy positions of inherently decreasing likelihood, and that differences in size of homomeric corpora that distinguish them may reflect little or nothing more than this purely ahistorical factor.[18]

Conclusion

In this paper I have compared two theories of the position of the Formosan aboriginal languages, called the Primary Branch Hypothesis and the Formosan-Philippine Hypothesis. Each of these theories is justified by a different method, the first by the comparative method of linguistics, and the second by the homomeric method of Dyen (1990) and Dyen and Tsuchida (1991). Although scientific theories are deductive in application they may be,

---

18 For much the same reason a homomeric argument for the *lowest* node in a right-branching structure (in the present case, EMP) carries an inherently lower risk of error, since the interpretation that such exclusively shared vocabulary consists entirely of retentions would require us to assume independent extinctions in each of the higher nodes of the tree. In the case of Formosan-Philippines it is probable that borrowing has also contributed to differences in size of homomeric corpora, while in the case of SHWNG-OC this would appear to be far less likely. For further discussion of theoretical issues arising from the use of exclusively shared vocabulary for subgrouping, cf. Blust (1978:218ff).

and often are, inductive in construction. In an important sense, then, the theories compared here can be seen as *results* of methodological differences.

It must not be forgotten that the essential feature of the comparative method as it relates to subgrouping is the distinction between *innovations* and *retentions*. The homomeric method (like lexicostatistics, and traditional biological taxonomy) ignores this critical distinction. As a result factors which give rise to differences in retention rate, or to stochastic differences in the composition of homomeric sets, may mistakenly be regarded as reflections of history. These are errors which can, and must be avoided if we are to distinguish the true lineage of languages from deceptive "special similarity" produced by convergence or common retentions. Appearances notwithstanding, neither the dolphin nor the lungfish is a fish, and historical linguists who are concerned with method would do well to pursue the work of subgrouping with such non-linguistic examples in mind.

REFERENCES

Bellwood, Peter. 1978. Man's Conquest of the Pacific. Auckland, Collins.

_____. 1985. Prehistory of the Indo-Malaysian Archipelago. Sydney, Academic Press.

_____. 1991. The Austronesian Dispersal and the Origin of Languages. Scientific American 265.1:88-93.

Blust, Robert. 1969. Some New Proto-Austronesian Trisyllables. Oceanic Linguistics 8:85-104.

_____. 1970. Proto-Austronesian Addenda. Oceanic Linguistics 9:104-62.

_____. 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. Honolulu, Department of Linguistics, University of Hawaii.

_____. 1974a. The Proto-Austronesian Word for Two: a Second Look. Oceanic Linguistics 13:123-61.

_____. 1974b. The Proto-North Sarawak Vowel Deletion Hypothesis. Unpublished Ph.D. dissertation, xiii + 319pp. Honolulu, Department of Linguistics, University of Hawaii.

_____. 1977. The Proto-Austronesian Pronouns and Austronesian Subgrouping: A Preliminary Report. Working Papers in Linguistics 9.2:1-15. Honolulu, Department of Linguistics, University of Hawaii.

_____. 1978. Eastern Malayo-Polynesian: A Subgrouping Argument. In S.A. Wurm and Lois Carrington, eds., Proceedings of the Second International Conference on Austronesian Linguistics, Fascicle 1:181-234. Pacific Linguistics C61. Canberra, Department of Linguistics, Research School of

Pacific Studies, The Australian National University.

_____. 1980. Austronesian Etymologies. Oceanic Linguistics 19:1-181.

_____. 1981a. Review of Shigeru Tsuchida, Reconstruction of Proto-Tsouic Phonology. Language 57:205-11.

_____. 1981b. Variation in Retention Rate among Austronesian Languages. Paper presented at the Third International Conference on Austronesian Linguistics, Den Pasar, Bali, January, 1981.

_____. 1982. The Linguistic Value of the Wallace Line. Bijdragen tot de Taal-, Land- en Volkenkunde 138:231-50.

_____. 1983. A Linguistic Key to the Early Austronesian Spirit World. Keynote address to the Third Eastern Conference on Austronesian Linguistics, Ohio University, Athens, Ohio, May 6-7, 1983. Ms. 101pp.

_____. 1983/84a. More on the Position of the Languages of Eastern Indonesia. Oceanic Linguistics 22-23:1-28.

_____. 1983/84b. Austronesian Etymologies - II. Oceanic Linguistics 22-23:29-149.

_____. 1991. The Greater Central Philippines Hypothesis. Oceanic Linguistics 30:73-129.

Dahl, Otto Christian. 1976 [1973]. 2nd ed. Proto-Austronesian. Scandinavian Institute of Asian Studies Monograph Series, No. 15. London, Curzon Press.

_____. 1981. Early Phonetic and Phonemic Changes in Austronesian. Oslo, Institute for Comparative Research in Human Culture.

Dempwolff, Otto. 1938. Vergleichende Lautlehre des Austronesischen Wortschatzes. Zeitschrift für Eingeborenen-Sprachen. Vol. 3. Austronesisches Wörterverzeichnis, Supplement 19. Berlin, Dietrich Reimer.

Dyen, Isidore. 1953. The Proto-Malayo-Polynesian Laryngeals. Baltimore,

Linguistic Society of America.

_____. 1963. The Position of the Malayopolynesian Languages of Formosa. Asian Perspectives 7.1-2:261-71.

_____. 1965a. A Lexicostatistical Classification of the Austronesian Languages. IJAL Memoir 19. Baltimore, Waverly Press.

_____. 1965b. Formosan Evidence for Some New Proto-Austronesian Phonemes. Lingua 14:285-305.

_____. 1990. Homomeric Lexical Classification. In Philip Baldi, ed., Linguistic change and reconstruction methodology:211-30. Trends in Linguistics Studies and Monographs, No. 45. Berlin, Mouton de Gruyter.

_____, and Shigeru Tsuchida. 1986. Proto-Austronesian Laterals and Nasals. In Paul Geraghty, Lois Carrington, and S.A. Wurm, eds., FOCAL II: Papers from the Fourth International Conference on Austronesian Linguistics:23-31. Pacific Linguistics C94. Canberra, Department of Linguistics, Research School of Pacific Studies, The Australian National University.

_____. 1991. Proto-Philippine as the Closest Relative of Proto-Formosan. In Ray Harlow, ed., VICAL 2: Papers from the Fifth International Conference on Austronesian Linguistics, Parts One & Two:85-101. Auckland, Linguistic Society of New Zealand.

Egerod, Søren. 1980. Atayal-English Dictionary. 2 vols. Scandinavian Institute of Asian Studies Monograph Series, No. 35. London and Malmö, Curzon Press.

Ferrell, Raleigh. 1968. The Pazeh-Kahabu Language. BDAA 31/32:73-97.

_____. 1969. Taiwan Aboriginal Groups: Problems in Cultural and Linguistic Classification. Institute of Ethnology, Academia Sinica Monograph No. 17. Taipei.

_____. 1982. Paiwan Dictionary. Pacific Linguistics C73. Canberra, Department of Linguistics, Research School of Pacific Studies, The Australian National University.

Fey, Virginia. 1986. Amis Dictionary. Taipei.

Geraghty, Paul A. 1983. The History of the Fijian Languages. Oceanic Linguistics Special Publication No. 19. Honolulu, University of Hawaii Press.

Gould, Stephen Jay. 1984. What, If Anything, Is a Zebra? In Stephen Jay Gould, Hen's Teeth and Horse's Toes: 355-65. New York, W.W. Norton & Company.

Grace, George W. 1985. Oceanic Subgrouping: Retrospect and Prospect. In Andrew Pawley and Lois Carrington, eds., Austronesian Linguistics at the 15th Pacific Science Congress:1-18. Pacific Linguistics C88. Canberra, Department of Linguistics, Research School of Pacific Studies, The Australian National University.

Guttorp, Peter. 1991. Statistical Inference for Branching Processes. New York, John Wiley & Sons, Inc.

Halim, Amran, Lois Carrington and S.A. Wurm, eds., Papers from the Third International Conference on Austronesian Linguistics, Vol. 2: Tracking the Travellers. Pacific Linguistics C75. Canberra, Department of Linguistics, Research School of Pacific Studies, The Australian National University.

Harvey, Mark. 1982. Subgroups in Austronesian. In Halim, et al.:47-99.

Haudricourt, André G. 1965. Problems of Austronesian Comparative Philology. Lingua 14:315-29. (Originally published in French in the _Bulletin de la Société de Linguistique de Paris_ in 1964).

Himmelmann, Nikolaus P., ed. 1990. Sourcebook on Tomini-Tolitoli

Languages. Typescript, 336pp. Department of Linguistics, University of Köln.

Ho, Dah-an. 1978. A Preliminary Comparative Study of five Paiwan Dialects. BIHP 49/4:565-681.

Jeng, Heng-hsiung. 1972. A Bunun-English Dictionary. Mimeographed. 147pp. Taipei.

Kelly, F.P. 1979. Reversibility and Stochastic Networks. New York, John Wiley & Sons.

Klaproth, J.H. 1822. Sur la Langue des Indigénes de l'Île de Formose. Asia Polyglotta. Paris.

Li, Paul Jen-kuei. 1976. Thao Phonology. BIHP 47/2:219-44.

_____. 1977. The Internal Relationships of Rukai. BIHP 48/1:1-92.

_____. 1978. A Comparative Vocabulary of Saisiyat Dialects. BIHP 49/2: 133-99.

_____. 1980. The Phonological Rules of Atayal Dialects. BIHP 51/2:349-405.

_____. 1981. Reconstruction of Proto-Atayalic Phonology. BIHP 52/2:235-301.

_____. 1982a. Male and Female forms of Speech in the Atayalic Group. BIHP 53/2:265-304.

_____. 1982b. Kavalan Phonology: Synchronic and Diachronic. In Rainer Carle, et al., eds., GAVA': Studies in Austronesian Languages and Cultures Dedicated to Hans Kähler: 479-95. Veröffentlichungen des Seminars für Indonesische und Südseesprachen der Universität Hamburg, Band 17. Berlin, Dietrich Reimer.

_____. 1983. Notes on Thao Dialects. BDAA 43:48-50.

_____. 1988. A Comparative Study of Bunun Dialects. BIHP 59/2:479-

508.

McFarland, Curtis D. 1977. Northern Philippine Linguistic Geography. Study of Languages & Cultures of Asia and Africa monograph series, No. 9. Tokyo, Institute for the Study of Languages and Cultures of Asia and Africa.

Ogawa, Naoyoshi. 1944. Indoneshia-Go ni Okeru Taiwan Takasago-Go no Ichi (The place of the Formosan Languages in the Indonesian Languages). Pp. 451-502 in Taiheiyô-Ken (Pacific Areas). Tokyo.

_____. 1979 [1934]. Amis (Kiwit dialect) Vocabulary. Compiled by Raleigh Ferrell. Typescript, 100pp. (= English Translation of Ami Goshû. Taipei, Taiwan Government General).

_____, and Erin Asai. 1935. Myths and Traditions of the Formosan Native Tribes. Taipei, Taihoku Imperial University.

Pecoraro, Ferdinando. 1977. Dictionnaire Taroko-Français. Cahier d'Archipel 7. Paris.

Popper, Karl. 1972. Objective Knowledge: An Evolutionary Approach. Oxford, Clarendon Press.

Reid, Lawrence A. 1982. The Demise of Proto-Philippines. In Halim, et al.: 201-16.

_____, ed. 1971. Philippine Minor Languages: Word Lists and Phonologies. Oceanic Linguistics Special Publication No. 4. Honolulu, University of Hawaii Press.

Ross, Malcolm. 1992. The Sound of Proto-Austronesian: An Outsider's View of the Formosan Evidence. Oceanic Linguistics 31:23-64.

Spriggs, Matthew. 1988. Dating Lapita: Another View. In M. Spriggs, ed., Lapita Design, Form and Composition: Proceedings of the Lapita Design Workshop:6-27. Occasional Papers in Prehistory 19. Canberra, Dept. of

Prehistory, Research School of Pacific Studies, The Australian National University.

Thiel, Barbara. 1984/85. Austronesian Origins and Expansion: the Philippine Archaeological Data. Asian Perspectives 26.1:119-29.

Ting, Pang-hsin. 1978. Reconstruction of Proto-Puyuma Phonology. BIHP 49/3:321-92.

Tryon, D.T. 1976. New Hebrides Languages: An Internal Classification. Pacific Linguistics C50. Canberra, Department of Linguistics, Research School of Pacific Studies, The Australian National University.

Tsuchida, Shigeru. 1971. List of Words of Formosan Languages. Typescript, 40pp.

_____. 1976. Reconstruction of Proto-Tsouic Phonology. Study of Languages and Cultures of Asia and Africa Monograph Series, No. 5. Tokyo, Institute for the Study of Languages and Cultures of Asia and Africa.

_____. 1980. Puyuma (Tamalakaw dialect) Vocabulary --- with Grammatical Notes and Texts ---. In Kuroshio Bunka no Kai, ed., Kuroshio no Minzoku, Bunka, Gengo:183-307. Tokyo, Kadokawa Shoten.

_____. 1982. A Comparative Vocabulary of Austronesian Languages of Sinicized Ethnic Groups in Taiwan, Part I: West Taiwan. Tokyo, Memoirs of the Faculty of Letters, University of Tokyo, No. 7.

_____. 1985. Kulon: Yet Another Austronesian Language in Taiwan? BIE 60:1-59.

_____, Yukihiro Yamada and Tsunekazu Moriguchi. 1991. Linguistic Materials of the Formosan Sinicized Populations, I: Siraya and Basai.

Tung, T'ung-ho. 1964. A Descriptive Study of the Tsou Language, Formosa. Institute of History and Philology, Academia Sinica Special Publications,

No. 48. Taipei.

Wolff, John U. 1973. Verbal Inflection in Proto-Austronesian. In Andrew Gonzalez, ed., Parangal kay Cecilio Lopez: Essays in Honor of Cecilio Lopez on His Seventy-fifth Birthday. Philippine Journal of Linguistics Special Monograph Issue No. 4:71-91. Quezon City, Linguistic Society of the Philippines.

_____. 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:535-49. Pacific Linguistics C117. Canberra, Department of Linguistics, Research School of Pacific Studies, The Australian National University.

Yamada, Yukihiro. 1976. A Preliminary Dictionary of Itbayaten. Manuscript, 400 + 4pp. Kochi, Japan.

_____. 1991. The Numeral Systems of the Formosan and the Philippine Languages. Department of Foreign Languages, Himeji Dokkyo University, Publication No. 4:119-35.

Zorc, R. David. 1981. Where, O Where, Have the Laryngeals Gone? Austronesian Laryngeals Reexamined. In Amran Halim, Lois Carrington and S.A. Wurm, eds., Papers from the Third International Conference on Austronesian Linguistics, vol. 2:111-44. Pacific Linguistics C75. Canberra, Department of Linguistics, Research School of Pacific Studies, The Australian National University.

_____. 1983. Proto-Austronesian Accent Revisited. Philippine Journal of Linguistics 14.1:1-24.

Formosan vs. Non-Formosan Features
in Some Austronesian Languages in Taiwan

Paul Jen-kuei Li

Institute of History and Philology, Academia Sinica

In this paper I shall try to sort out linguistic differences between typical Formosan and non-Formosan (or extra-Formosan) features, including lexical, phonological and morphological ones. I shall then argue that not all Austronesian languages in Taiwan can be treated as truly "Formosan," since they contain some linguistic features which are typically non-Formosan. These languages include Ketagalan, Qauqaut, Siraya and Kavalan. The type and amount of non-Formosan features as found in each of these languages vary. While Qauqaut resembles the Oceanic type of language, the others are similar to Philippine and Indonesian languages to a certain extent.

1. Typical Phonological Features in
Formosan Languages

Typical Formosan languages such as Atayal, Tsou and Rukai distinguish between *t and *C, *n and *N. Only a few Formosan languages do not make these distinctions, including Bunun, Amis, Kavalan, Siraya and Ketagalan (including Basay and Trobiawan). No Austronesian language outside Formosa makes these distinctions. If these distinctions are retentions of PAN features, as most comparative Austronesianists have assumed, then most Formosan languages still retain these two phonemic contrasts from PAN, since these

contrasts have merged in only a few Formosan languages and in all extra-Formosan languages, i.e. the Malayo-Polynesian languages outside Taiwan (hereafter abbreviated as MP). Conversely, if these phonemic distinctions are common innovations in most of the Formosan languages, then the MP and a few Austronesian languages in Taiwan have not shared the innovations.

Another phonological feature in typical Formosan languages is the retention of PAN *S, which has become h or lost in non-Formosan languages. A few Formosan languages that have lost *S include Saaroa and Puyuma. In fact, it is retained in Kanakanavu, a language closely related to Saaroa. In addition to the retention of *S, Tsuchida (1976:13) noted that in several cognate sets "a medial *S is metathesized in Formosan and consequently appears in final position," e.g., PAN *buSek > PFN *bukeS 'hair,' PAN *CaSiq > PFN *Caqis 'sew,' PAN *lit'eSeq > PFN *lit'eqeS 'nit.' However, Formosan languages may have retained the original order and the metathesis of *S may have taken place in extra-Formosan languages, as Blust (1981, 1993) has argued.

Still another phonological feature retained only in Formosan languages is PAN *q, which is retained as an uvular stop [q] in Atayal, Sediq, Bunun, Paiwan and Thao, but as a pharyngealized stop in Amis. It is reflected as [k] in the Philippines, Moken, Watubela, various languages of the Admiralty Islands and elsewhere in Melanesia, as [h] in Malay, Chamic, Acehnese, Nias, Sundanese, and Lakalai, and as [gh] in Muna (Blust, p.c.), or lost in the other extra-Formosan languages.

Consonant clustering of nasals and stops such as /mp, nt, ŋk/ in the word-medial position do not occur or occur only rarely in Formosan languages. A few attested cases for this type of clustering can be explained as due to the loss of an intervening vowel (Tsuchida, p.c.), e.g., Tsou oŋko

'name, reflection,' or borrowing, e.g., paŋka 'table,' paŋki 'a non-Tsou personal name'(Tung 1964:507). Similarly, Proto-Rukai *nakua > Maga ŋkua '1st pers. sg. accusative.' There are only a few Formosan forms that contain /nt/ or /nS/, not due to the loss of an intervening vowel, e.g., PAN *bi(N)tuqan > Bun bintuqan, Sai bintoeʔaen 'star,' cf. also Paz bintul 'star;' BunIs bantas 'leg;' PAN *binSeq > BunTkb binsiq, BunIs binsah.

Nasal accretion is a phonological process widespread in western Austronesian languages such as the Philippine and Indonesian languages. In these languages the stem-initial voiced and/or voiceless stops are lost when they are immediately preceded by homorganic nasals. For example, in Yami,

palas-en,	malas	'tell a lie'
tovis-an,	manovis	'answer'
tawag-an,	manawag	'call'
kekeR-en,	maŋekeR	'plant other crops'

Such a whole series of phonological process of nasal accretion is not found in any Formosan language.

2. Typical Lexical Features in Formosan Languages

Dyen (1963) gave a list of 37 cognate sets that are found only in Formosan languages. Tsuchida (1976:13, 22) removed some of them from the list as he found cognates in extra-Formosan languages and he also added a few others. Based on Dyen and Tsuchida, the "Formosan only" cognate sets are the following. I have eliminated some of them because they do not show regular sound correspondences and they are most likely borrowed from each other. I have also up-dated the data.[1]

---

1   Most Formosan data for the extant languages are based on my field notes and manuscripts; data for the extinct languages are based on Tsuchida (1982, 1985)

1) *damuq > Ata[2] ramu-ux, Pai djamuq, Paz damu? 'blood,' Bun damuq, Tha samuq 'dew'

2) *ruluŋ > Ata yuluŋ, Sed ruluŋ, Sar luuluŋ-a (<A l/ɬ), Paz ruruŋ (<A r/l) 'cloud'

3) *bilbil > Sed blebil, Ruk wa-bilibili, Pai v-in-iljvilj 'pull'

4) *kawaS > Ata kawas, Sed kawas, Tha kawaS, Paz kawas 'year'

5) *t'iRa > Ata cu-hisa (-s- irreg.), Sed c-higa?, Tha tiɬa?, Sai ka-hiLa?, Paz nu-ka-zixa?, Ami na-ciLa?, Tao ni-diha, Pap sera, Sir icha (s- expected) 'yesterday'

6) *ma-puSaN > Ata ma-pusal, Sed m-pusal, Tso m-pusk-u, Kan ma-pusan,

---

and Tsuchida et al (1991). A number of Formosan cognates illustrated in this paper were originally identified by Tsuchida (1976), but subsequently updated by myself by adding a few more examples and/or giving more accurate transcriptions. The letter "e" stands for the phonetic symbol ə, and "S" for ʃ.

2   Abbreviations for Formosan language names are: Ata, Atayal; Sed, Sediq; Tso, Tsou; Kan, Kanakanavu; Sar, Saaroa; Ruk, Rukai; Bun, Bunun; Pai, Paiwan; Puy, Puyuma; Tha, Thao; Sai, Saisiyat; Paz, Pazeh; Kav, Kavalan; Ami, Amis; Tao, Taokas; Fav, Favorlang; Bab, Babuza; Pap, Papora; Hoa, Hoanya; Sir, Siraya; Ket, Ketagalan; Bas, Basay, Tro, Trobiawan A major dialect is chosen for each language unless specified otherwise (see Li 1990). Dialect names under each language are also abbreviated: AtaSq, Squliq dialect of Atayal; AtaSk, Skikun dialect of Atayal; SedTn, Toŋan dialect of Sed; SedTd, Toda dialect of Sed; SedTr, Truwan dialect of Sediq; SedIn, Inago dialect of Sediq; RukTa, Tanan dialect of Rukai; RukMg, Maga dialect of Rukai; RukTo, Tona dialect of Rukai; RukMn, Mantauran dialect of Rukai; BunTkb, Takbanuð dialect of Bunun; BunIs, Ishbukun dialect of Bunun; PaiSt, Stimul dialect of Paiwan; PaiTb, Tjubar dialect of Paiwan; PuyKl, Katipul dialect of Puyuma; PuyLp, Lower Pinlang dialect of Puyuma; PazKh, Kahabu dialect of Pazeh; AmFa, Fataqan dialect of Amis; AmiFr, Farangaw dialect of Amis.

Abbreviations for extra-Formosan language names include: Yam, Yami; Mal, Malay; Tag, Tagalog; Skt, Sanskrit.

Abbreviations for proto-language names are : PHF, Proto-Hesperonesian-Formosan; PFN, Proto-Formosan; PMP, Proto-Malayo-Polynesian; PNF, Proto-Northern-Formosan; PSF, Proto-Southern-Formosan.

Sar ma-puɬe, Ruk ma-pusale, Bun ma-pusan, Tha ma-puSað 'twenty'

7) *waNuH₁ > Sed walu?, Kan aanu , Sar aɬu?u, Bun va-vanu, Pai alju 'bee,' AtaSq walu? 'bee-hive,' Ruk valu 'honey,' Puy walu 'sugar,' Sai walu? 'honeybee, sugar,' Paz walu? 'honeybee, sugar,' Kav wanu? 'honeybee, sugar,' Ami waðu? 'honeybee, sugar'

8) *Lituk > Tso rtuk-a, Kan lituk-a, Sar lituk-a, Ruk Lutuku (<A u/i), Pai Lutjuk (<A u/i), Sai Lutuk (<A u/i), Sir routock (<A u/i) 'rabbit'

9) *batakan > Ata batakan, Sed btakan, Bun takan 'bamboo'

10) *belbel > Sed blebul, Tso fxûfxû 'wild banana,' Kan ta-venevene 'raw banana,' Sar ta-veɬeveɬe, Ruk belebele, Bun bunbun, Pai veljevelj, Puy belbel, Tha fiðfið, Paz beɬebel, Bab bilpil, Pap bibul, Hoa bulbul, Sir bulbil 'banana'

11) *waSaw > Sed wasaw, RukBu vasaw, Pai asaw 'leaf'

12) *buŋa > Ata buŋa?, Sed buŋa?, Puy buŋa 'sweet potato.' This is probably a Formosan innovation, as the sweet potato is of South American origin (Blust, p.c.).

13) *SupeR > SedIn s-m-epug, Tso s-m-uprû, Kan s-um-a-sepere (<A e/u), Bun (ma-sipul (-i- irreg.)), Pai s-em-upu (<A u/e), Tha S-um-upiɬ, Sai SepeL (<A e/u) 'count'

14) *baki > Sed baki, Sai baki?, Kav vaqi? (-q- irreg.), Ami baki?, Hoa vaki, Bas baki 'grandfather'

15) *qi[DN]aS > Pai qiljas, Sai ?ilaS, Paz ?ilas, Tao idar, Bab idass 'moon'

16) *uka > AtaSk ?uka?, Sed ?uka?, Tso uk?a, Sar uka?-a, Bun ?uka?, Tha ?uka?, Sai ?uka?, Tao ua 'not exist'

17) *SuReNa > Ata hula-qiy, Tso ruxo, Kan erena (<A -e/u), Sar uruɬa (<A -u/e), Ruk u?ula (<A u/e), Pai sulja, Puy urla, Tha ?uɬða?, Sai hae-hoelae?, Paz ha-hela? (<A h/x), Kav suRna?, Ami suLða?, Sir ougla

'snow'

18) *mi-maH₁ > Sed m-imah, Tso m-imo, Kan m-iima, Sar m-ima 'drink'

19) *kuS₁a > AtaSq m-usa?, Sed m-usa?, Tso uso(us-a), Kan m-u-a-kusa, Sar m-aɬu-kua, Puy mu-kuwa, Tha m-uSa?, Paz mu-kusa?, Bab m-usa, Hoa usa 'go'

20) *Cabu > Ata c-um-abu?, Sed l-m-abu? (<A l/c), Tso c-m-ofu, Sar c-um-avu, Ruk wa-cabu, Pai c-em-avu, PuyKl T-em-avu, Kav t-m-avu?, Ami mi-tabu? 'wrap,' Kan cavu-cavu-a 'millet or rice cake wrapped in ginger leaves with pork inside'

21) *pataS₁ > Ata matiq/matas(patas), SedTd matas, patas 'letter,' Tso t-m-opsû (<M), RukTa wa-?acase, Bun ma-patas, Tha mataS, pataS-an 'book' 'write,' Kan tapase (<M) 'pattern, design,' Sar taa-tapa-a (<M) 'pattern, design,' Sai pataS 'tatoo'

22) *ReH₁ap > Ata ga-ghap, SedIn gehap. Tso xio-rapo, Kan raape, Sar erape (<M), Ruk eape 'seeds of grains'

23) *puNi > Kan ta-puni-a, Sar ma-puɬi, Ruk ma-?uli, Tha ma-puði?, Sir ma-poule 'white'

24) *t'ima > Ata ?-ima?, Sed ?-ima?, Bun cima?, Pai tima, Puy ?-ima-nay, Tha tima?, Paz ?-ima?, Ami cima? 'who'

25) *baLi > Pai vaLi, Puy baLi, Tha fari?, Sai baLi?, Paz bari?, Kav vaLi?, Ami baLi?, Tao bari, Bab barri, Pap bari, Sir vare, Bas BACI 'wind,' Hoa ma-bali 'to blow'

Based on my own study, the following cognate sets can be added to the "Formosan only" list:

26) *DakeS > Ata rakus, Tso c?osû, Kan càkese, Ruk Dakese, Bun dakus, Pai Dakus, Puy Dakes, Tha SakiS (<A S-/s-), Sai rakeS, Paz dakes, Kav zaqes, Ami rakes 'camphor laurel'

27) *ki > Ata ki?, BunTb ki, Sai ki 'and,' Paz ki 'because'

28) *Capah > RukMn capa, Bun tapa?, Tha θapa? 'barn, granary,' SedTr sapah 'house'

29) *qaRiDaŋ > Ata qagiraŋ, Kan ?aricaŋe, Sar ?arisaŋe, Bun qalidaŋ, Pai qaRizaŋ, Paz xaidaŋ (<M) 'beans, peas,' Tso recŋɨ (<A i/e) 'pigeon peas'

30) *SiSiN > Ata sisil-iq, Sed sisil, Kan sisini, Sar iini, PaiTb sisilj, Sai SiSil, Paz sisil, Kav sisin 'omen bird, Alcippe morrisonia'

31) *qag'iS > Ata qais/qagis, Tso es-a, Kan ?es-ane, Ruk sagi-agise, Bun qais, Sai ?aezis (<A s/S) 'boundary,' Sar ais-a 'middle, between'

32) *tunu > Ata tunu?, Tha tunu?, Sai tunu?, Paz tunu? 'brains,' Sir tunu 'head'

33) *ma[dD]as > Ata maras, Sed madas, Bun madas, Sai maraS, Kav mazas̩ 'bring'

34) *qemu > Ata qumu-li?/qumu? (<A u/e), SedTd ?emu?, Bun qamu (-a- irreg.), Tha qmuu? 'rice cake,' Pai q-in-emu, Puy tinua-?umu 'flour,' Tso t-umu 'to powder'

35) *lawa > AtaSq m-lawa?, Sed m-lawa?, Tha ma-la-lawa?, Paz ma-lawa? 'call, bark'

36) *teRakuk > Tso troo?u-a, Kan tarikuuk-a, Sar turukuuk-a, Ruk tarukuku, Bun tulkuk, Puy terkuk, Kav teLaquq, Ami tuLakuk, Pap tokkoa, Hoa tokkoa, Sir tahuk-a (-g- expected for -h-), Bas tarahok (-h- irreg.), Tro toqqoq, Ket torako 'chicken'

37) *NaNaLi > Ata lalai?, Kan nane-nanali, Sar lalali, Ruk lalaLi, SaiTh lalai?, PazKh lalay? 'cicada'

38) *minaŋ > Ata minaŋ 'to clear a forest,' Bun minaŋ 'to sow'

39) *lukNaw > Ata akl-i?, Sed rkel-ic, Tso r?uxo, Kan ukunau, Sar lukulu,

Ruk Likulaw, Bun (huknav), PaiSt Likulaw, Puy Likulaw, Tha rukðaw, Sai Luklaw, Kav Luqenaw, Ami Lukðaw 'leopard'

40) *baLit' > Ruk wa-baLiθi, Pai ma-pa-vaLit, Puy mu-baLis, AmiFr ma-vaLic 'change'

41) *qatabaŋ > Kan ʔa-ta-tavaŋe, Sar ʔa-ta-tavaŋe, Ruk a-ta-tabaŋe, Bun qata-baŋ, Pai qa-tja-tjavaŋ 'cockroach'

42) *qaSip > Ruk wa-asipi, Puy ʔaip, Ami mi-qasip 'count'

43) *CaReb > Sar careve, Ruk wa-caebe, Kav taRev 'box,' t-m-uuv 'cover,' Tso p-a-crofû 'sleeping mat,' m-a-crofû 'cover with coverlet,' Kan c-um-a-caruvu 'cover with coverlet,' Pai c-m-auv 'cover someone or something'

44) *puNaS > Ata pulas, Sed pulas, Sai pulaS, Kav pulas, Ami puðas 'dandruff'

45) *qaNian > Ata qalian, Kan ca-ʔania, Bun qaniʔan 'daytime'

46) *qaNaH > Sar ała?a (ʔ- expected), Pai qalja, Puy ʔala, Sai ʔaelaʔ, Ami qaðaʔ 'enemy'

47) *d'awiN > Tso covxi, Kan ara-caini, Sar ma-saił-a, Ruk daili, Puy a-dawil 'far'

48) *ʔupʔup > Puy ʔupʔup, Tha ʔupʔup, AmiFr ʔupʔup 'frog'

49) *maDuq > Kan macuʔu, Sar masuʔu, Ruk maDu, Paz maduʔ 'fruit'

50) *baNiw > Puy baliw, AmiFr faniw, Sir bariu 'type of edible fungus'

51) *Du(k)Duk > Tso cucʔu, Sar suusuku, Bun duduk, Tha suksuk, Paz dukuduk 'ginger'

52) *SiDi > Bun sidiʔ, Pai siði, Tha sisiʔ (<A s-s/S-s), Sai Siriʔ, Kav siziʔ, Ami siðiʔ 'goat,' Puy siDiʔ 'sheep'

53) *qaCapi > Tso copi, Ruk ka-ca-caapi, Pai (qanapi) (-n- irreg.), Sai ka-ʔsapi, Paz (ʔatapiʔ) (-t- irreg.) 'goby'

54) *CebeR > Tso c-m-ûfrû, Kan c-um-evere, RukMn u-ceve?e, Paz me-zebex (<A z/s) 'grow'

55) *quDaS > Ata quras, SedTd qudas, Kan ?usase (<A -s/c), Sar ?usae, Ruk uDase, Bun qudas, Pai quDas, Tha (qutaS) (-t- irreg.), PazKh ?udas, Fav oras 'gray hair'

56) *qubiS > Tso fsi-fsi, Ruk ubisi, Pai quvis, Puy ?ubi 'pubic hair'

57) *NubeR > Ata lubug, SedTr lubug, Ruk leber (<A e/u), Puy luber 'Jew's harp'

58) *Lekep > Tso rû?pû, Kan lekepe, Sar e-lekepe, Puy Lekep 'hawk,' Ruk Le-Leke?e 'Asiatic sparrow hawk'

59) *taNa > Kan t-um-a-timana, Sar t-um-imała, Bun tan?a (-?- irreg.), Tha t-m-aða?, Paz t-um-ala?, Tao t-im-ala?, Hoa t-um-aala 'hear'

60) *keRiw > Bun kaliv, Puy keriw, Tha kłiu?, Sai ka-kLiw, Paz kixiw (<A i/e), Kav qeRiw, Ami keLiw 'hemp plant'

61) *paqet > Tso po?te, Bun paqut, Tha pa-paqut 'horsefly'

62) *SiLaw > Tso siro (srov-a), Ruk ma-siLaw, Bas s-om-elaw 'hungry'

63) *NukiS > Tao x?isi, Kan nukisi, Sar łukii, Bun nukis 'husk, chaffs, bran'

64) *taRuqan > Tso trova, Kan taru?ane, Sar taruan-ane, Ruk tauvanaane, Bun taluqan, Puy taru?an, Tha tałuqan, Paz taxwan, AmiFr taLuqan 'hut,' Sai taLoe?aen 'house'

65) *SeŋeR > Kan s-um-a-seŋere, Sar s-um-a-seŋere, RukMg u-sŋ^^, Bun ma-sa-sŋul, Sai SeŋeL, Ami mu-seŋeL 'immerse'

66) *CuŋuN > Tso c-m-uŋxu, Kan ma-cuŋunu, Sar c-um-a-cuŋułu, Ruk wa-cuŋulu, Pai c-m-uŋul, Puy ma-Ta-Tuŋul 'join, link'

67) *puDeR > Tso pcûrû, Sar pesere (<A e/u), Ruk peDee (<A e/u), Pai puDu (<A u/e) 'kidney'

68) *ku(ŋ)kuŋ > Ruk wa-kuŋukuŋu, Kav k-m-ukuŋ, Ami mu-kuŋkuŋ 'knock'

69) *Dalam > Puy ma-laDam (<M l/D), Sai ralam 'know'

70) *pika > Ata ma-pika?, Sed m-pika?, Tso pi?o 'lame'

71) *wiNi > Sed wili?, Ruk vili, Bun vini, PaiTb vilji, Puy wili, Paz wili?, Ami wiði? 'water leech,' Tha vivi? (<A v/ð 'mountain leech,' Sai wili? 'mountain leech'

72) *Ritu > AtaSq gitu?, Sed gitu?, Tso rtuu, Kan riitu, Sar ritu, BunIs litu?, Sai Litu? 'loquat,' Pai itju 'fruit of the hardwood tree'

73) *DaRaCu > Tso trocu, Kan carace, Sar vara?e, Ruk daacu (d- irreg.), Pai tacu, PuyKl raraTu 'body louse'

74) *temuy > Kan ?ukula-tumulu 'full,' ?i-tumulu, Sar kula-tumułu 'full,' ma-tumułu 'much,' Tso mo-tmuzu 'take many,' o-tmuzu 'eat much,' Bun ma-tmuð 'full,' Puy ma-temuy 'full'

75) *DaRa > Ata raga?/raa?, Sed dara?, Bun dala?, Tha łała? (<A ł-/s-), Sai raLa?, Paz daxa? 'maple tree (Liquidamar formosana (Hance))'

76) *NuqeS > Ata luqi?/luqus, Sed luqi?, Tso xûsû (<A û/u), Kan ne?ese (<A e/u), Sar łee?e (<M)(<A e/u), Ruk luusu (<A u/e), Bun nuqus, Pai ljuqes, Sai loe?es, Ami ðuqus 'marrow'

77) *Cebuŋ > Sar taru-cuvuŋ-a, taru-a-cuvuŋu, RukMn ?ii-civuŋu (<A i/e), Pai se-cevuŋ, Puy mar-Tebuŋ, Sai sa-sebuŋ 'meet,' Tso tro-cfuŋ-a 'confluence of rivers,' Kan cuvuŋ-unu 'confluence of rivers'

78) *baSaR > Ata basag, Sed basaw, Sai basaL, Tao basau 'millet'

79) *LikeS > Tha rikiS, Ami Likes, Bas ries, Sir rikig 'mosquito'

80) *pug'ek > Tso pucku, Kan peleke (<A -e/u), Sar pełeke (<A -e/u), Pai pudek, Puy pudek, Sai puzuk (<A -u/e) 'navel'

81) *pug'a > Ata puga?, Sed puga?, Tha puða?, Paz puza?, Ami puna?́, Tao puza, na-puda, Pap puda 'navel'

82) *Rubu > AtaSq ?ubu?, RukTa rubu, PaiTb ruvu, Puy rubu, Paz xubu

'nest'

83) *aray > Tso roi 'scoop net,' si-roi, Ruk aLay, Pai aray, PuyKl aray 'fish net,' Sar mi-a-arai 'catch fish with a scoop-net at the waterfalls'

84) *ini > Ata ?ini?, Sed ?ini?, Ruk ini, Bun ni-tu?, ni?, Pai ini, ?in-?ini, Paz ?ini? 'not,' Sai ka-ini? (don't)

85) *Sanaq > Ata sanaq, Tso snoo, Kan sana?e, Sar sana?e, Ruk sana, Pai sanaq, Kav sani?, Ami sanaq, Fav channa 'otter'

86) *RiNaS > Ata gila-quŋ, Sed gla-quŋ, Bun linas, Tha ɬiðaS-al, Sai LilaS-an, Paz xilas-an 'pheasant'

87) *qeluD > Sed ?eruc, Kan ?uucu, Sar ?ulusu, Ruk ?uLuDu, Bun qau?, Pai qeLuz, Tha qruus, Sai kae-?Lur, Paz ?urut (<A u/e) 'pillar'

88) *tanaq > Ata tana? (-? irregular), Tso tnoo, Kan tana?e, Sar tana?e, Ruk tana, PaiTb tjanaq, Puy tana?, Tha ta-tanaq, Sai tane? (e irrregular), Paz tana?, Kav tani?, Ami tanaq 'plant sp., Aralia decaisneana Hance'

89) *Remg'a > Sed gmeya? (<M), Tha ɬimða?, Sai Lemza? 'plant sp., Imperata cylindrica, cogan grass'

90) *Naya[dD] > Ata laya?, Tso xzocû, Kan nalace, Ruk laLaDe, Bun naða?, PaiTb ljayaz, Puy layaD, Sai layar, Kav layas, Ami ðayas 'plant sp., Ebulus formosana'

91) *Sina > Ata sina?, RukBu lasia-sina, Sai Sina? 'plant sp., Erechtites S.P.'

92) *biNuaq > Tso fkuo (<D k/x), Kan vinua?e, Sar viɬua (-?e missing), Ruk lubu (<M, <A), Pai viljuaq (Oreopanax formosana) 'plant sp., hibicus taiwanensis'

93) *Di(Ne)Let' > Tso cûrsû, Kan ne-cere (<A e/i), Sar ?aɬi-seer-a (<A e/i), Ruk DilLe, Kav zines (loss of L is unexplained) 'plant sp., Lager-stroemia subcostata Koehne in Engl'

94) *maqaw > Ata maqaw, Bun maqav, Tha maqaw, Sai ma?aw 'plant sp.,

Litsea cubeba (Lour.)'

95) *baŋaS > Tso fŋosû, Kan vaŋase, Sar vaŋae, Ruk baŋase, Pai vaŋas, Sai baŋaS, Kav vaŋas, Ami faŋas 'plant sp., Melia azedarach'

96) *put'ek > AtaMx puhuk, RukMg lŋu-ps^k^, Bun pus-l-uk, PaiTb putek, Sai ka-poehik, Kav puseq 'plant sp., Oxalis repens Thunb'

97) *LiNuk > Ata iluk, RukTa Liluk, Sai Liluk, Kav Rinuk 'plant sp., Rubus parvifolius Linn; Rubus taiwanianus Matsum'

98) *Samaq > Kan sama?e, Sar sama?e (s- irreg.), Ruk sama, Bun samaq, PaiTb samaq, PuyLp amaR, Tha Samaq, Paz sama?, Kav sami?, Ami samaq 'plant sp., Sonchus oleraceus Linn,' Bas sama 'mustard plant'

99) *bacaR > AtaSk bacax, SedTn basaw, RukMg bcaa, Sai basaL 'plant sp., Panicum miliaceum (Linnaeus),' Tao basau 'millet'

100) *basay > Ata basay 'type of vine,' Tso fsoi, RukMg bsee, BunIs basað 'plant sp., pueraria hirsuta (Matsum)'

101) *qaRlu > Ata qaglu?, SedTr glu?, AmiFa qarɬo? 'plant sp., Phragmites longivalvis (Steud.)'

102) *Si > Ata s-um-i, Tso mo-si 'put'

103) *Rami[CS] > Tso rmisi, Kan ramisi, Sar ramii, Ruk amici, Bun lamis, Puy rami, Tha ɬamiθ, Ami Lamit 'root'

104) *kuSkuS > Sai k-um-uSkuS, Ami mu-kuskus 'shave,' Kan k-um-a-kusukusu 'cut hair,' Sar k-um-a-kuukusu 'cut hair,' RukMn u-kuku?u 'cut hair,' Pai k-m-uskus 'scratch, remove hair from pig'

105) *damay > Kan camai, Sar camai, Ruk damay, Pai djamay 'side dish'

106) *Semet > Ruk a-sm^t-ne, PaiTb ma-semutj (<A u/e), SaiTh pane-Semet-en 'smothered,' Sar aa-semet-a 'half asleep'

107) *baNaySan > Sed brisan, PuyLp vaLaysan (-s- irreg.), Kav vLaysan, Ami baLaysan 'sorghum'

108) *kaka[tC]u > AtaSk kkasu?, Bun kakatu?, Paz kakasu?, Ami kakatu? 'spider'

109) *tiRpeS > Sai t-um-epeS, Paz tixipes, Kav t-m-iRpes 'spit'

110) *lapiC > AtaSq yapit, Sed rapic, Sai Lapis, 'flying squirrel,' PuyKl LapiT, Sar ?a-lapic-a 'bat'

111) *NawaR > Tso rvorû, Kan laale (<A -l/x), Sar laare, Ruk ravar (<A -r/L), Bun haval, Pai Lava, Tha ravað (+ expected), Kav LawaL (<A L/R) 'flying squirrel'

112) *RabaR > Ata gabag-an, Sed rbag-an, Bun ta-labal, Sai LabaL-an, PazKh abax-an 'summer'

113) *CaliH > Ata cai?, SedIn sari?, Tso u-cri, Bun tai?, Tha łari? (<A ł/θ), Ami taLi? 'taro'

114) *Deme[lR] > Tso o-cmûrû, Kan m-aki-cemere, Sar m-aki-semere, Ruk maku-DemeLe, Pai ku-DemeL, PuyKl ke-zemeL 'thick'

115) *waLay > Ata wayay, Sed waray, Kan alai, Sar alali, Ruk valay, Pai aLay, Puy waLay, Sai waLay, Kav waRay, AmiFr waLay 'thread,' Paz waray 'hemp yarn for weaving'

116) *waNiS > Tso xisi, Kan anisi, Sar ałii, Ruk valisi, Pai aljis, Puy wali, Sir walig 'tooth,' Bun vanis 'tusk, wild pig,' Tha waðiS 'wild pig,' Sai waliS 'tusk,' Paz walis 'tusk,' Ami waðis 'tusk'

117) *qaCaŋ > Ata qalaŋ (<A l/c), Sed ?alaŋ (<A l/c), q-n-alaŋ 'fence,' Sai ?aesaŋ 'village,' Sar ?acaŋe 'fence'

118) *qaNib > Tso xifi, Kan ?anivi, Sar ?ałivi, Ruk alibi, PaiTb qaljiw, Tha qaθif (<A θ/ð) 'wall, roof, slate slab'

A few animal names turn out to be cognates shared by only the Formosan languages in the Austronesian family, e.g., *Cumay 'bear,' *likuNaw 'leopard,' *RiNaS 'pheasant,' *SiSiN 'bird sp., Gray-eyed Nun

Babbler,' although these animals are also found outside Taiwan. Some plants which are native or endemic to Taiwan are rarely or not found outside Taiwan at all, e.g., *DakeS 'camphor laurel,' *NayaD 'Ebulus formosana,' *tanaq 'Aralia decaisneana Hance,' and several other plant species.

Listed above are some 118 cognates shared by Formosan languages. Not all of them can be reconstructed as "Proto-Formosan" if such an Austronesian subgroup ever existed at all. Depending on how one subgroups Formosan languages, many of the cognates will have to be relegated to a lower level subgroup. I have excluded a number of cognates which are shared by only languages that are geographically contiguous. Since all these languages are located on the island of Taiwan, chances of mutual borrowings are very high. Consequently even though there are a large number of cognates shared by these Formosan languages, this does not necessarily prove that they all constitute a subgroup in the Austronesian family at a higher or lower level. It may simply indicate that these languages have been spoken in the same geographical area for centuries.

3. Typical Morphological Features in Formosan Languages

The prefixes such as man-, maŋ- and mag- are common in the Philippine languages, but totally lacking in most Formosan languages. What we do find in most Formosan languages is only the prefix ma-. In Mayrinax, the conservative dialect of Atayal, a few forms contain the prefix mag-, e.g., mag-baytunux 'pretty,' mag-lakaam 'to go head-hunting,' but with quite different functions from that in Philippine languages. Similarly, Mayrinax also has a few forms containing the prefix man-, e.g., man-caqrug 'to stand up,' man-

cahuu? 'straight,' again with different functions from those in Philippine languages. The focus system in Philippine languages are generally manifested in verb inflections, including -um-, -en and -an. Such a neat focus system is not found in some Formosan languages.

Moreover, the suffix -aken, which is common and widespread among languages in Indonesia and Malaysia, is not found in any Formosan language.

On the other hand, the atemporal Undergoer imperative marker -u attested in several Formosan languages, including Saaroa, Paiwan and Puyuma, is not found in the Malayo-Polynesian languages outside Taiwan (see Ross, in this volume).

The infix -um- is totally lacking in Rukai, which has wa- instead.

Reflexes of the suffix *-en are absent in both Tsou and Kavalan. The suffix -an is also absent in Tsou. Both suffixes may have been merged as *-ã > -a in Tsou.

4. Austronesian Languages in Taiwan with Non-Formosan Features

Instead of sharing the typical linguistic features with most other Formosan languages, a few Austronesian languages in Taiwan contain certain features that are found only in extra-Formosan languages.

4.1. Ketagalan

Ketagalan is a sinicized language formerly spoken in the northernmost part of Taiwan, in and around Taipei. More linguistic data recently became available for Basay and Trobiawan, two divergent varieties of the Ketagalan language. Asai collected his field data for the language in 1936, which was

not published until 1991 in Tsuchida et al.

4.1.1. Ketagalan with Non-Formosan Features

Phonologically, Ketagalan contains a few forms with the extra-Formosan consonant clustering (/mp, nt, ŋk/) in the word-medial position, e.g.,

Trobiawan /mampaita/ 'Come to eat!' (Asai [ ], Text 4)

Trob. /maqsomnta/ 'to check' (Asai [ ], Text 4)

Trob. /pusanta/ '?' (Asai [ ], Text 4)

Trob. /qmantito 'to eat' (Asai [ ], Text 4)

Basay /pantiti/, Trob. /vantiti/ 'doll' (Tsuchida et al 1991:218)

Basay /baŋka/, Trob. /vaŋka/ 'canoe' (Tsuchida et al 1991:209)

Basay /maneŋke/ 'thick, fat' (Tsuchida et al 1991:213)

In addition, Basay has some other combinations of nasals and stops, which are extra-Formosan in nature, such as

/pantʃawan/ 'bowl' (Tsuchida et al 1991:218)

/palaŋpaŋ/ 'pot, water jar' (Tsuchida et al 1991:217)

/tatoŋpet/ 'stumble' (Tsuchida et al 1991:226)

/tuluŋtuŋ/ 'drum' (Tsuchida et al 1991:228)

Lexical evidence for extra-Formosan features in Ketagalan is more conspicuous. First of all, the cognate *baŋka? 'canoe' has so far been attested only in Malayo-Polynesian languages outside Taiwan, yet it is found also in Ketagalan, as listed above. Compare the following Formosan and Malayo-Polynesian (MP) lexical forms:

	Formosan[3]	MP	Basay	Trobiawan
1. 'banana'	*belbel (10)	*punti	puti	puti
2. 'count'	*SupeR (13)	*bilaŋ	bilaŋ	vilaŋ

3 See citation forms as numbered in the parentheses following the reconstructed forms.

3. 'canoe'   *qaCu, *qabaŋ   *baŋka   baŋka   vaŋka

All these lexical forms seem to indicate that Ketagalan is extra-Formosan rather than Formosan, since they are not apt to be borrowed from languages outside Taiwan. These are common words in daily use: "canoe," "banana," and "count." To borrow these words from languages outside Taiwan would seem rather unlikely if we consider the fact that Ketagalan was spoken in the northernmost part of Taiwan, far away from other extra-Formosan languages.

Moreover, Trobiawan, a variety of Ketagalan, uses the form imu 'your (sg.)' as in tama-imu 'your father' (Asai [ ], Text 6), rather than (i)su as in most other Formosan languages.

Given below are more Ketagalan lexical forms which resemble Philippine languages, although not all the sound correspondences are regular:

4. Basay surab, Phil. Ata. MbT. silab, TbwA. surib 'burn' (Reid 1971:56)

5. Basay lapuŋ, Phil. libiŋ, Ata. loboŋ 'bury' (Reid 1971:56)

6. Basay kapowa, Kavalan kpua?, Phil. Agta kap^s, ItgB 'kapas, KlaG 'kapos 'cotton' (Reid 1971:64)

Based on such linguistic evidence in morphology, phonology and lexicon as presented above, we can infer that the Ketagalan tribe may have settled in Taiwan later than most, if not all, other Formosan tribes. However, the Ketagalan people must have settled in Taiwan long enough (about 2,000 years)[4] so as to undergo the change: dropping the /n/ before /t/ in the form *punti > puti 'banana.' All the available data show that Ketagalan retains only a few forms with consonant clustering of a nasal immediately followed by a stop, as illustrated above.

Aside from Ketagalan, no Formosan language has a cognate form for

---

4   The archaeological dating in the area is 1,800-800B.P.

*baŋka? 'canoe,' *punti 'banana' or *manuk 'bird.' Only Kavalan, a sinicized Formosan language formerly spoken in the northeastern coast of Taiwan, has the following three forms similar to Ketagalan:

karavaw, qavaw 'carabao'

vilaŋ 'count'

RayaR 'sail,' cf. Basay rayar 'sail'

I conjecture that Kavalan may have borrowed these three words from Ketagalan as they were geographically adjacent to each other and in close contact for centuries. See Section 4.4 below.

According to reports by archaeologists, so far no excavation in Taiwan has turned up any remains for carabao. Coupled with linguistic evidence, we may conclude that carabao was introduced to Taiwan rather late in its history, perhaps only a few hundred years ago during the Dutch period (1624-1662).

4.1.2. Ketagalan with Formosan Features

Ketagalan also shares some features with the majority of Formosan languages. Phonologically, it retains PAN *S as s in the following few forms.

1. PAN *buSuk > Bas busukke, Tro vusuk 'drunk'

2. PAN *Sepat > Bas sepat 'four'

3. PAN *daqiS > Bas laise 'face'

4. PHF *Sikam > Bas sikkam 'mat'

5. PMP *t'iwa, PFN *Siwa > Bas siwa 'nine'

6. PHF *qamiS > amis 'north, west'

7. PHF *paliSi > Tro m-Lisi-na 'taboo'

The above are all the Ketagalan forms that contain reflex for PAN or PHF *S.

In addition, the following Ketagalan lexical forms are similar to the

other Formosan languages:

8. PFN *baLi > Tro vatsji 'wind'

9. PFN *Silaw > Bas s-om-elaw 'hungry'

10. PSF *Samaq > Bas sama 'mustard plant,' Tro sama 'vegetable'

11. PFN *teRakuk > Bas tarahok, Tro toqqoq 'chicken'

12. PFN *Cumay > Bas tomay 'bear'

13. PSF *qatimuLa > Bas timula 'flea'

14. PSF *Lidam > Bas lilam 'tongue'

The last 5 items may have been borrowed from the other Formosan languages.

In short, Ketagalan contains both Formosan and non-Formosan features. Not all of them can easily be explained by borrowing. Different hypotheses can be advanced to explain this case. One is that Ketagalan was spoken by a non-Formosan tribe, which has been settled in northern Taiwan long enough to acquire all its Formosan features. An alternative hypothesis will be that Ketagalan was a genuine Formosan language, but have borrowed fairly extensively from extra-Formosan languages due to its contact with the outsiders, perhaps the Philippine peoples.[5] The main difficulty with such a hypothesis is that Ketagalan was geographically more distant to the outsiders than all the other Formosan languages, which contain no such features at all.

4.2. Qauqaut

The Qauqaut tribe was situated in the northeastern coast near Su'au.

---

5 The northern parts of Taiwan were occupied by the Spaniards for the short period 1626-1642 until they were expelled by the Dutch. As Tsuchida (p.c.) suggested, the Spaniards might have brought a number of Filipino sailors with them. If so, the Ketagalan people had a chance to have some contact with Philippine languages.

The only language data available for Qauqaut is the 10 numerals listed in katakana by Namikoshi (1924:76-77), who probably recorded them around the end of the 19th century (1885-1900). He marked the other lexical items as "unknown" especially for Qauqaut. The old Chinese documents for the Kavalan areas noted that Qauqaut was linguistically and culturally distinct from all the other Formosan natives, and that there was no intermarriage between Qauqaut and the other tribes (Chen 1840:229). Finally Qauqaut became extinct at the turn of the century.

According to the oral traditions, Qauqaut originally settled in the Taroko area in Hualien in the east coast. Around 1690, Qauqaut had disputes with "their own people, Atayal" (more appropriately Sediq), so they migrated northward to the coastal area near Su'au (see Namikoshi 1924:92). Tsuchida's (1983) atlas indicated Qauqaut as Basay with a question mark. However, if we compare the 10 numerals in Qauqaut with Atayal, Sediq or Basay, they do not look like any of the languages at all:

Gloss	Katakana	Phonetic[6]		PAN	Atayal	Sediq	Basay
one	イス	is	<	*et'?a	qutux	kiŋal	ca
two	ヅ-ス	zus	<	*DuSa?	?usayiŋ	daha	lusa
three	ド-ル	dor	<	*teluH	tugal	teru	cu
four	ソ-プ	sop	<	*Sepat	sapaat	sepat	sepat
five	リ-ム	rim	<	*lima	?imagal	rima	cima

---

6   Every Japanese syllable ends with a vowel. If we should follow Namikoshi faith-fully, most of the Qauqaut numerals would end with the vowel [u], i.e. isu 'one,' zusu 'two,' doru 'three,' etc. Apparently Namikoshi's transcription in katakana was influenced by his native language, Japanese. So I left out the final vowel [u]. The numeral 'seven' would only be pi, as based on Namikoshi's transcription. I conjecture that he simply missed the final [t], which is difficult for an untrained Japanese field worker like Namikoshi to hear.

six	ン-	en	<	*enem	mamatuu?	mteru	aɲem
seven	ピ-	pit	<	*pitu	mapitu	mipitu	pitu
eight	アール	ar	<	*walu	mamaspat	msepat	wasu
nine	シエ	siw	<	*Siwa	mamaqisu?	mŋari	siwa
ten	トル	tor		*puluq	magalpug	maxal	labatan

Apparently Qauqaut has lost the second vowel and the final consonant in the forms derived from PAN. This is typical of many Oceanic languages including nuclear Micronesian languages like Trukese (Dyen 1949), Ponapean and Kusaiean, Gedaged and some other languages of western Melanesia, the languages of the Admiralties, many languages of Vanuatu (New Hebrides) and New Caledonia, Kei in the Moluccas and various other languages of eastern Indonesia (Blust, p.c.). We may infer that Qauqaut may have migrated to Taiwan from the South or East Pacifc fairly recently, perhaps only a few hundred years ago.

Perhaps the greatest difficulty in not treating Qauqaut as a Formosan language is that Qauqaut reflex for PAN *S̲ is s̲, as in the words for 'two' and 'four,' as both Blust and Tsuchida pointed out at the Symposium. Tsuchida (p.c.) suspects that those Qauqaut words might have been collected from an old speaker, who had only a vague memory of his own language and was able to recall only the first part of those words. Unfortunately there is no other source for the extinct language.

Incidentally, it would be difficult to explain *t̲ derived as d̲, as in *teluH̲ > dor 'three' if there was no mistake in Namikoshi's transcription.

4.3. Siraya

Siraya was formerly spoken in the southwestern plains of Taiwan. It seems to consist of the following three main subgroups: Siraya (or Sideia),

Taivoan and Makatao; see Tsuchida et al 1991 for language data, and Tsuchida 1991 and Li 1992 for discussion. Siraya was the first Formosan language to adopt a writing system devised by the Dutch missionaries in the 17th century and may have subsequently become a lingua franca in southern Taiwan. It must have been very influential for a certain period of time before it finally became extinct by the mid-19th century. Some of its vocabulary, for instance, was borrowed by several Formosan languages in the same areas. Such vocabulary is typically non-Formosan. For example,

1. *t'urat > Sir soulat 'book,' s-m-oulat, Bas s-um-ulat, Kan mari-sunate, Sar s-um-a-sulate, RukMg u-slatɨ, RukTo wa-solate, RukMn o-solate 'to write,' Pai sunat 'paper'

2. *qabaŋ > Sir avaŋ, Kan ʔavaŋe, Sar ʔavaŋe, RukMg avaŋɨ, RukTo avaŋe 'boat, canoe;' cf. Babuyan and Yami avaŋ 'canoe'

3. *puLut > Sir porot, Kan pulutu, Sar pa-pulut-an, Pai puljutj-an 'cotton'

4. *baLituk > Sir vannitock, Hoa manituk, Kan vantuku, Sar valituku, Pai vaLitjuq 'money'

More typical corresponding Formosan forms for the above are:

*pataS 'to write' (see #21 in Section 2)

*qaCu > Ata qacuʔ, Sed asu, Bun qatuʔ 'boat'

*pila > Ata, Sed pilaʔ, Hoa pira, Bas pila 'money,' Sai pa-pilaʔ,

       AmiSa piða? 'paper money,' Paz pila? 'silver'

Also compare the following forms in the sinicized Formosan languages all in the western plains in Taiwan (Tsuchida 1982):

5. Sir vagat, Paz baxat 'cucumber'

6. Sir moula, Tao mura, Bab mura, Hoa mula 'face'

Sander Adelaar listed the following Siraya forms as "apparent loanwords" although he did not always indicate their sources in his unpublished

manuscripts:

7. sacyt, sahkit 'ardent,' pa-sacyt-an 'struggle, war'

8. ma-voulas 'sad'

9. likough 'revert,' sa-likough 'convert'; cf. rikour 'back'

10. voussouk 'dirty'

11. tamsi, mitamsi 'lose its savor'

12. da-dila 'tongue,' d-m-ira 'lick'

13. tabe '(greeting)' < Skt

14. pingang 'plate, saucer' < Mal pingan < Persian pingan

15. caiou 'firewood' < Mal kayu, Yami kayu 'wood, tree'

16. kamau-en 'will, desire' < Mal ke-mau-an

17. raja, r-m-aja 'great,' karaj'-en ki ryh 'joy' < Mal (arch.) raya 'great'

Siraya may have originally borrowed these lexical forms from some extra-Formosan languages such as Philippine and Malay languages, as suggested by Adelaar. Alternatively, Siraya could conceivably be an extra-Formosan language itself. We have detected not only some lexical evidence as listed above, but also some phonological and syntactic evidence for this alternative explanation. Phonologically, Siraya manifests certain features that are not typically Formosan, such as the lack of distinction between PAN *t and *C (even though it separates *n and *N) and its reflex of *S as h (g, gh or ch, depending on the sources); see Li 1993 and Adelaar (manuscripts) for citations. Nevertheless, Siraya also contains certain Formosan features, including the distinction between *n and *N, and its reflex of *-S as a velar fricative in the material of van der Vlis, e.g., *paRiS > pagig 'stingray.'

As for syntactic evidence, we have to wait for the final results of Adelaar's study, who is working on the Siraya texts as preserved in the Gospel of St. Matthew (originally Gravius 1661, 1662, ed. by Asai 1939 and

Campbell 1888 respectively). Siraya requires further study in determining its linguistic position.

While the occurrence of *S in one Siraya form (18) is the same as the other Formosan languages, yet in another form (19) it is similar to extra-Formosan. In still another (20) Formosan languages have different manifestations for the order of the consonants:

	PMP	PFN	Siraya
18. 'hair'	*buSek	*bukeS	voukig
19. 'sew'	*CaSiq	*CaqiS	t-m-ahy
20. 'knee'	*tuSud	*tuduS	tourouh Cf. PuyLp tuzu,

Ami mi-pi-turus 'kneel'

*tuSud Cf. Kav tusuz, Ami tusuð 'knee'

Again we have conflicting evidence for treating Siraya as Formosan or extra-Formosan.

4.4. Kavalan

Kavalan is one of the three Austronesian languages in Taiwan that show no contrast between *t and *C, or between *n and *N. Could Kavalan also be linguistically an extra-Formosan language, but geographically located in Taiwan, just like Ketagalan? Our linguistic evidence for such a claim is weaker than that for Ketagalan. Nevertheless, Kavalan does have some peculiar features of its own, not shared by any other typical Formosan language. For instance, it has no reflex for PAN *-en, as mentioned in Section 3 above.

Kavalan does not seem to be genetically close to any other Formosan language. Lexically it looks closer to Amis, perhaps due to mutual borrowing by close contact with each other; see Li 1990. In more recent studies (Li 1991, 1992) I have found some phonological and lexical similarities between

Kavalan and Ketagalan. Phonologically, they both have (1) merged *t and *C, *n and *N; (2) merged *d, *D, *d' and *Z; (3) lost *q, *H and *?; i.e., merged as zero; (4) retained voiced and voiceless distinctions in obstruents, as in all other western Austronesian languages. Conversely, they have these phonological differences: (1) while Ketagalan merged *R with *d, *D, *d' and *Z, Kavalan kept them distinct; (2) while Ketagalan merged *t' and *l as c [ts], Kavalan kept them separate; (3) *k split into k and q (adjacent to back vowels) in Kavalan, but not in Ketagalan; (4) *a split into a and i (adjacent to *q) in Kavalan, but not in Ketagalan. See Li 1982, 1991, 1992.

Lexical similarities between Kavalan and Ketagalan (represented by Basay and Trobiawan, see Tsuchida et al 1991) include the following forms (Li 1992):

	Basay	Trobiawan	Kavalan
1. carabao	kalabaw	kLavaw	qavaw
2. tooth	baŋcaw		vaŋRaw
3. eyelash	kulupu		qLupu
4. claw	kanuukus		qnuqus
5. side dish	tabun		tamun
6. sweet potato	hawpit		qawpiR
7. fire	namaD	zamal	zamaR
8. boat		vawa	vawa
9. sail	rayar		RayaR
10. fish	vaute	vaut	vaut
11. duck	kulaba	kuLava	kLava
12. year	tasaw,tatasaw		ta-tasaw
13. morning	rabe-rabe		ta-Rav-Ravi
14. cotton	kapowa		kəua

15. count	bilaŋ	vilaŋ	vilaŋ
16. wine	raaq		Raaq
17. pepper	rirum	siri	siLi

As Tsuchida (1991) pointed out, Kavalan may have borrowed some of the forms, such as qavaw, vilaŋ, Raaq and RayaR, from Philippine languages because (1) they show close resemblance to the forms in Philippine languages, and (2) they are not regular derivations from proto-forms. Let's compare Kavalan and Tagalog forms below:

> Kav qavaw, Tag kalabaw 'carabao'
>
> Kav vilaŋ, Tag bilaŋ 'count'
>
> Kav Raaq, Tag alak 'wine'
>
> Kav RayaR, Tag layag 'sail'

It seems more likely, however, that Kavalan may have borrowed these forms from Ketagalan since they are were geographically much closer to each other than to the Philippines in the last three and half centuries or so. Tsuchida has also shown a dozen lexical forms which Kavalan may have borrowed from Spanish (or Portugese) via Philippine languages: vyavas 'guava,' mais 'corn,' vnina 'banana,' siLi 'pepper,' tvaku 'tobacco,' vaka 'cow,' kvayu 'horse,' pRasku 'bottle,' plumu 'lead (metal),' paskua 'new year,' u 'or,' savvun 'soap.'[7] We may raise the question: Why was Kavalan, in the northeast coast of Taiwan, the only language to borrow so heavily from Philippine languages?

In short, Kavalan has quite a few typically non-Formosan features, including syntactic (Li 1978), phonological (Li 1982) and lexical (see above). It is possible that Kavalan was originally spoken in the Philippines and then migrated to Taiwan, probably earlier than Ketagalan (2000B.P.), but later

---

7　All the Kavalan forms are listed in my orthographic system.

than most other Formosan tribes.[8]

5. Linguistic Implications for the Migration History of the Formosan Peoples to Taiwan

Judging from the linguistic evidence as presented above, we may infer that not all Formosan peoples migrated to Taiwan at the same time. Instead, they might have come to Taiwan in successive waves over a period of six millenia. The latest immigrants were probably the Yami people on the Botel Tobago. Yami is clearly closest to Itbayaten in the Batanic group of languages; see Tsuchida et al (1987, 1989). Its date can be tentatively set at 400B.P. Qauqaut may have arrived in Taiwan only a few hundred years ago, perhaps no more than 1000B.P. Ketagalan may have landed in northern Taiwan around 2000B.P. The earliest immigrants to Taiwan include most of the major Formosan ethnic groups today, i.e. Atayal, Tsou, Rukai, Paiwan, Saisiyat, Pazeh, Thao, etc. Based on the carbon 14 dating as reported in the archaeological excavations, the earliest settlement on Taiwan may be set at 6,000B.P.

---

8   Siraya and Amis have retained the contrast between *n and *N, though not between *t and *C. Siraya also has some typically non-Formosan features (see Section 4.3 above). All these indicate that they may have been in Taiwan for a longer period of time than Kavalan, though not as long as most typically Formosan languages such as Atayal and Tsou.

REFERENCES

Adelaar, Sander

[　] Siraya Historical Phonology. Unpublished manuscripts.

Asai, Erin

1939 Gravius's Formulary of Christianity in the Siraya Language of Formosa, fascimile edition of the orginal of 1662. In Erin Asai, ed., Memoirs of the Faculty of Literature and Politics, Taihoku Imperial University, Vol.4, No.1.

[　] Kavalan Texts. Unpublished fieldnotes.

[　] Trobiawan Texts. Unpublished fieldnotes.

Bellwood, Peter

1991 The Austronesian Dispersal and the Origin of Languages. Scientific American (July):88-93.

Blust, Robert

1977 The Proto-Austronesian Pronouns and Austronesian Subgrouping: a Preliminary Report. University of Hawaii Working Papers in Linguistics 9.2:1-15.

1981 Review of Shigeru Tsuchida, Reconstruction of Proto-Tsouic Phonology. Language 57:205-211.

1985 The Austronesian Homeland: a Linguistic Perspective. Asian Perspective 26.1:45-67.

1993 *S Metathesis and the Formosan/Malayo-Polynesian Language Boundary. In Øyvind Dahl, ed. Language--A Doorway between Human Cultures, 178-183. Oslo: Norvus Forlag.

Campbell, Rev. William

1888   The Gospel of St. Matthew in Formosan (Sinkang Dialect), with corresponding versions in Dutch and English, edited from Gravius's edition of 1661. London: Trubner and Co.

1896   The Articles of Christian Instruction in Favorlang-Formosan, Dutch and English. London: Trubner and Co.

Cauqelin, Josiane

1991   Dictionnaire Puyuma-Francais. Paris: Ecole Francais d'Extreme Orient.

Chen, Shu-jun

1840   Records for the Kavalan Area (in Chinese).

Dyen, Isidore

1949   On the History of the Trukese Vowels. Language 25:420-436.

1963   The Position of the Malayopolynesian Languages of Formosa. Asian Perspectives 7.1-2:261-271.

Ferrell, Raleigh

1982   Paiwan Dictionary. Pacific Linguistics C-73. Canberra: The Australian National University.

Fey, Virginia

1986   Amis Dictionary. The Bible Society, Taiwan.

Li, Paul Jen-kuei

1978   The Case-Marking Systems of the Four Less Known Formosan Languages. Pacific Linguistics C-61:569-615.

1982   Kavalan Phonology: Synchronic and Diachronic. GAVA': Studies in Austronesian Languages and Cultures Dedicated to Hans Kahler, 17:479-495. Berlin.

1990   Classification of Formosan Languages: Lexical Evidence. BIHP

61.4:813-848.

1991   Classification of the Sinicized Tribes in Northern Taiwan and Their Linguistic Evidence (in Chinese). The Taiwan Folkways 41.4:197-214.

1992   Classification and Relationships of the Sinicized Tribes in Taiwan (in Chinese). The Taiwan Folkways 42.1:211-238.

1993   New Data on Three Extinct Formosan Languages. BIHP 63.2:301-323.

1994   Some Plant Names in Formosan Languages. In A.K. Pawley and M.D. Ross, eds Austronesian Terminologies: Continuity and Change, 241-266. Pacific Linguistics, C-127.

Namikoshi, Shigeyuki

1924   Native Affairs of Kavalan before Japanese Occupation (in Japanese). Taiwan Jiho 55.4:66-92.

Murakami, Naojiro

1933   Sinkang Manuscripts. Memoirs of the Faculty of Literature and Politics, Taihoku Imperial University.

Reid, Lawrence

1971   Philippine Minor Languages: Word Lists and Phonologies. Oceanic Linguistics Special Publication No.8. Honolulu.

Tsuchida, Shigeru

1976   Reconstruction of Proto-Tsouic Phonology. Tokyo: Study of Languages & Cultures of Asia & Africa.

1982   A Comparative Vocabulary of Austronesian Languages of Sinicized Ethnic Groups in Taiwan. Part I: West Taiwan. Memoirs of the Faculty of Letters, University of Tokyo, No.7.

1985   Kulon: Yet Another Austronesian Language in Taiwan? Bulletin

of the Institute of Ethnology 60:1-59.

1991   Miscellanies of Languages of Sinicized Ethnic Groups in Formosa (in Japanese). Papers in Linguistics, University of Tokyo, 12:146-179.

Tsuchida, Shigeru, Yukihiro Yamada and Tsunekazu Moriguchi

1987   Lists of Selected Words of Batanic Languages. University of Tokyo.

1991   Linguistic Materials of the Formosan Sinicized Populations I: Siraya and Basai. University of Tokyo.

Tung, T'ung-ho

1964   A Descriptive Study of the Tsou Language, Formosa. Institute of History & Philology, Academia Sinica, Special Publications No.48.

A Grammatical Subgrouping of Formosan Languages[1]

Stanley Starosta

University of Hawaii

The usual practice in the reconstruction of Proto-Austronesian verbal morphology has been to reconstruct anything found in any Formosan language plus one extra-Formosan language. This paper in contrast hypothesizes that a large number of the features of modern Austronesian verbal morphology are innovations. The use of shared morphological innovations produces a subgrouping picture which equates Proto-Formosan with Proto-Austronesian, interprets morphological evidence for the putative Rukai-Tsouic subgroup as retentions rather than shared innovations, identifies almost all extra-Formosan languages as a relatively low-level subgroup of the Austronesian family, and for the first time posits Chamorro to be a mid-level offshoot of the Formosan family tree.

The Plan of the Paper

1. The Languages

In this paper, I will try to supplement previous attempts at lexically based reconstructions of Formosan languages by applying the comparative method of subgrouping to grammatical rather than lexical properties[2] of twelve languages: Amis, Atayal, Bunun, Chamorro, Kanakanavu, Paiwan,

---

1 I am grateful to Lawrence Reid and John Wolff for helpful comments on the original version of this paper. Neither should be blamed for what I have done with their comments.

2 As a precedent for this approach, see Pawley 1966.

Tanan Rukai, Saisiyat, Saaroa, Seediq, Tsou, and Yami. Ten of these are geographically and genetically Formosan languages. Yami is not considered to belong genetically to the Formosan language family in the narrow sense, but is included to provide a broader basis for the reconstruction. Chamorro on the other hand is a mystery. Starosta and Pagotto (1990) found that it was not Micronesian in terms of its grammatical properties, but that it could also not be clearly grouped with Philippine languages, as has often been assumed. Instead, it turned out that all of the features that had been previously put forward to support grouping Chamorro with the languages of the Philippines could equally well be taken to support a closer genetic relationship with the aboriginal languages of Taiwan. One of the purposes of this paper then is to take that proposal seriously, and to find out what happens when the comparative method is applied to the morphological properties of a group of Austronesian languages which includes most of the aboriginal languages of Taiwan, plus Yami and Chamorro.

2. The Data

The data for this study are taken from available published sources, theses and dissertations, and from some of my own notes from the field research I have been conducting off and on since 1964. I would especially like to acknowledge the importance to my study of work by T'ung-ho Tung, Paul Li, and Shigeru Tsuchida, and by a newer member of the club, Lillian Huang, and also to note the recent contributions made by two M.A. students from Tsing Hua University, Arlene Ho (Ho 1990) and Mei-li Yeh (Yeh 1991). I should state at the beginning that my survey of these resources has been necessarily rather superficial, and constitutes more of a "first pass" than a definitive study.

3. The Theory

Grammatical reconstruction cannot be done meaningfully outside the framework of a formal, explicit, and tightly constrained linguistic theory. Without such a theory, reconstruction is hardly more than a party game. Anyone can reconstruct anything he or she likes, and there is no objective way of evaluating the alternatives. A reconstruction formulated without a theory might for example posit a proto-language with five different distinctive kinds of *w*, six *s*'s, and six different *q*'s (Tsuchida 1976:126, 144, 160, 163). This may be a convenient way of filing correspondences, but it surely cannot be taken seriously as a claim about any kind of reality. The same criticism could be made for a reconstruction of a language with word classes and syntactic patterns that do not now exist in any known language.

The framework I employ in my own reconstruction is formal, explicit, and constrained. It is a version of dependency grammar called lexicase (cf. Starosta 1988a), and it has been refined and tested against almost fifty languages. Because it is constrained, it limits the choices of possible reconstructions for a given set of data, and to the extent that it correctly circumscribes the class of possible languages, it will also select the correct proto-language.

Finally, I think it is necessary for any grammatical reconstruction to provide a plausible abductive mechanism to move from one posited stage to the next (cf. Starosta 1991a:509 and Starosta, Pawley, and Reid 1982:157-158, hereafter referred to as "SPQR"). That is, it is not enough to claim, say, that a passive construction in one stage of a language somehow became a noun phrase at the next. Rather, it must be possible to show a scenario in which the children of one generation could plausibly reinterpret a passive construction in their parents' language as a nominal one in their own.

Linguistic Subgrouping

1. Linguistic and Archaeological Reconstruction of Prehistory

The prehistory of the ancestor of the modern aboriginal languages of Taiwan is especially significant because it is arguably also the prehistory of all the Austronesian languages of the Pacific (cf. Reid 1982). Linguists have a contribution to make to the resolution of this question because linguistic reconstruction techniques allow us to use information available in modern languages to make inferences about the prehistory of language communities. There are of course limitations to this approach. For example, as in archaeology, tracing the spread of a particular language family or a particular pottery technology does not necessarily allow us to associate that language family or that pottery technology with a particular group of people. Nevertheless, when linguistic and archaeological techniques are combined and found to reinforce each other, the likelihood of their shared account of prehistoric distributions and migrations being accurate is increased dramatically. An excellent example of this kind of cooperation in the area of Oceanic languages is Andrew Pawley and Roger Green's article, *Dating the dispersal of the Oceanic languages* (Pawley and Green 1973). I take it that the reason for assembling this particular group of scholars at this symposium is to attempt to achieve similar results in the area of the prehistory of the Formosan languages, and ultimately of the Austronesian language family as a whole.

2. The Comparative Method

(1) Similarities versus relatedness

Human beings have been fascinated for centuries by resemblances

between languages. However, although there has been a recent upsurge in sophisticated techniques for counting and quantifying such similarities and sprouting trees out of them computationally (cf. Ho 1983, Li 1990), such counts actually do little more than assign a number to a subjective impression (cf. Starosta 1990). The reason for the minimal utility of this approach is that there can be many explanations for the existence of similarities between two languages. First of all, there can be chance similarities between languages. One of my favorite examples of this is the Japanese folk etymology for the English word *kennel: ken* "dog" + *neru* "sleep". Secondly, the words in two languages may be similar because words are cultural artifacts, and cultural artifacts can be passed from culture to culture. In linguistics, the euphemistic term for this is "borrowing". As an example, counting similar words would tell us that English and Japanese or Chinese and Japanese are quite close to each other in some sense, because Japanese has borrowed heavily from both languages.

(2) Genetic relationship and subgrouping

Information on the transmission of cultural artifacts such as words is useful and interesting, and does tell us something about the history or prehistory of a particular group of speakers in the sense that it allows us to document a direct or indirect contact between two linguistic traditions. However, there is another kind of relationship between two linguistic traditions that is different from this kind: if two linguistic traditions can be shown to have evolved in a continuous line from a single common tradition, called a PROTO-LANGUAGE, then we can say that they are GENETICALLY RELATED. The century-old COMPARATIVE METHOD of linguistic reconstruction is a technique for establishing such a relationship. Once again, practitioners of this method have recognized that simply counting similarities is of little use in determining

genetic relationship. Thus they have developed a technique for distinguishing words with a common origin, cognates, from noise (loan words and accidentally similar forms) by requiring such word sets to exhibit regular sound correspondences.

More important for the paper I am presenting today is a technique employed within the comparative method for establishing the degree of genetic relatedness. If two or more languages within a language family (say, A, B, and C in Figure 1) all descend from the same common ancestor D, and if that ancestor existed at some time later than the progenitor of the entire family E, then these languages are said to constitute a SUBGROUP, and to be more closely related to each other than any of them is to some other member F of the same family which does not descend from the same immediate ancestor:

Figure 1
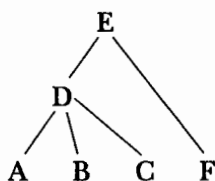
3. Subgrouping and Prehistory

(1) Family trees

The analogy between subgrouping diagrams and human kinship systems should be obvious, and is reflected in the conventional name for such diagrams in the linguistic literature: FAMILY TREES. The problem in constructing such a tree is not just the one of identifying cognate elements of the languages in the first place, but rather in showing that the common source of some of these items existed at some time later than the ancestor of

the entire family. Thus for example in constructing the tree above in which A and C are closer to each other than either is to F, it is necessary to demonstrate that the cognate words in these languages derive from some word that existed in a reconstructed language D which is itself a descendent of E. The criterion for establishing this is called SHARED INNOVATION. All descendants of E will have undergone changes in evolving into their modern forms. If we can show that languages A, B, and C underwent exactly the same changes, then the null assumption is that these changes happened only once, in the process of evolving from E to D, rather than many times independently. The more such shared innovations we can find, the stronger evidence we have for establishing the existence of D as distinct from E.

(2) Shared innovations versus feature-counting

There are two reasons why I am spending so much time on the question of subgrouping. The first point is that a family tree is a claim about prehistory. The most logical inference we can draw from the fact that existing languages can be shown to be related in accordance with such family tree patterns is that at some time prior to the present, there existed a group of speakers which shared the words and structures which we reconstruct for E, that this community divided into two different communities, that certain changes happened to one group which resulted in the configuration corresponding to D, and that subsequent to that point, the group of speakers corresponding to D split up into three groups which eventually gave rise to the modern languages A, B, and C. The paper I am presenting today uses grammatical properties of modern Formosan languages to establish such a family tree, and is thus a claim about actual events in the prehistory of the languages described.

The second point I want to emphasize is 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.

(3) Morphological versus lexical reconstruction

The paper I am presenting today attempts to make a contribution to the task of reconstructing the prehistory of the Formosan languages by providing an independently established subgrouping of Formosan languages, one based on shared innovations in grammar rather than phonology. To the extent that it turns out to be isomorphous to a subgrouping based on phonological innovations, the two reconstructions will reinforce each other. Correspondingly, in the few areas in this paper in which I have had to assume the diffusion of grammatical features across language boundaries, I hope that evidence from lexically based subgrouping will be of assistance in supporting or rejecting these proposals.

(4) Diffusion

Finally, I want to concede that languages do not evolve only by faulty transmission across generations or by simple borrowing of lexical items. Rather, it is also possible for linguistic features to "diffuse" across language boundaries, even when languages are related only distantly or even not at all. In fact, the difficulty in distinguishing the diffusion of phonological features from internal evolution seems to be one of the primary reasons for the lack of agreement among scholars of Formosan languages about the form of the common ancestor. Distinguishing between these two different kinds of change is inherently much more difficult than for example separating loans from cognates.

A Morphological Subgrouping of the
Formosan Languages

1. The Family Tree

My work so far suggests the following grammatically based subgrouping of the languages under study:

Figure 2

FO : Proto-Formosan

F1

Rukai

F2

Tsou

F3

Saaroa

Chamorro   F4

F5   Kanakanavu

F6   F7

F8   Bunun   Atayalic   Saisiyat

F9   F10   Seediq   Atayal

Yami   F11   Paiwan   Amis

Tagalog   Ilokano
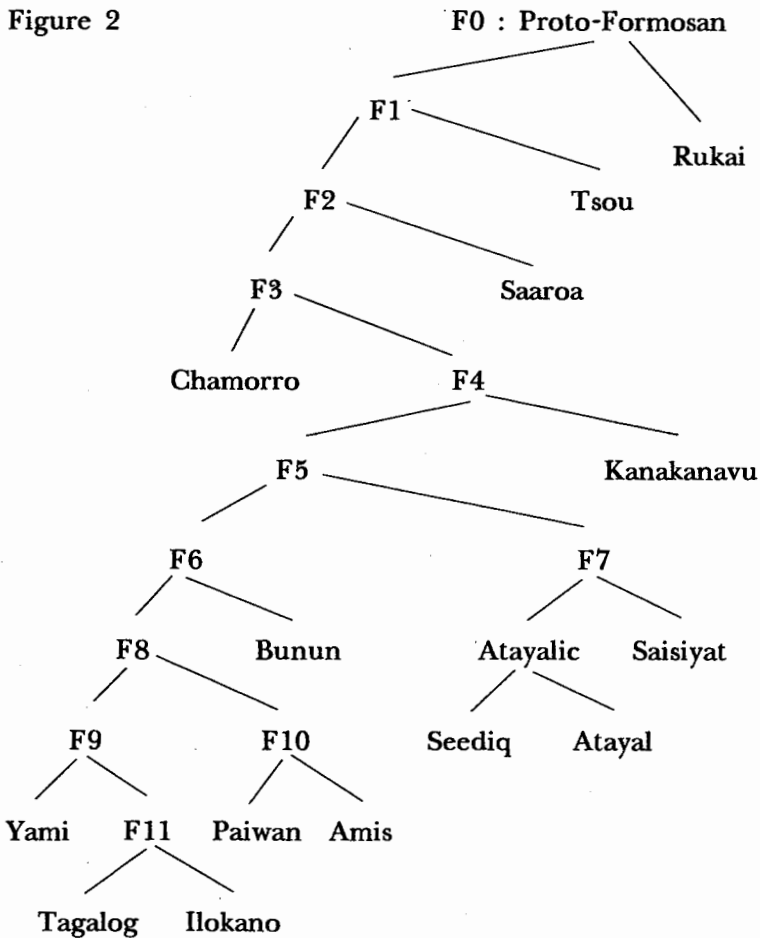
In accordance with the comparative method, each of the nodes in the tree is associated with one or more shared innovations that apply to all the

languages below it.[3] As will be seen in the discussion below, the reconstruction is not without problems. One such problem is that the evidence for several nodes is rather weak and negative. Another is that the forms which can be expected to appear in the daughter languages based on this tree have not always been found. This is partly no doubt a result of the inadequacy of my data, especially for Yami, Bunun, and Saisiyat. Such predictions have an obvious heuristic value, since they can be confirmed or disconfirmed based on further field research. To the extent that they cannot be confirmed or explained in some other way, they count against this reconstruction, and any new alternative which does not have the same or comparable failings will be preferred to this one.

There are interesting similarities and differences between the tree constructed from grammatical data and previously proposed family trees based on lexical reconstruction. First of all, it is completely incompatible with Blust's picture of Proto-Austronesian splitting up into two primary divisions, Formosan and Malayo-Polynesian. From a morphological point of view (though not from a lexical one, surely), Blust's tree would appear to be the result of a naive assumption that all the popular Austronesian morphology is uniformly distributed throughout all the daughter languages and thus can be reconstructed all the way to the top of the tree. My study will attempt to refute any such assumption, and replace it with one much closer to Reid's view (Reid 1982:213), in which Blust's Malayo-Polynesian

---

3   Several colleagues have dismissed this whole hypothesis by saying that all the innovations I propose are actually just retentions, and that 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.

(my F9) subgroups with Amis (my F8) and branches off much farther down the genetic tree.

The northern Formosan subgroup proposed by Li (Li 1985) is supported, though weakly, as F7, and node F6 matches Paiwanic, previously often considered a wastebasket. Regarding the controversial question as to whether Rukai subgroups with Paiwanic or with Tsouic (cf. Ho 1983), the answer provided by this analysis is: neither. Instead, Rukai and the Tsouic languages do not form a subgroup at all. Rather, they are successive chips off the old block, and the features that they have in common are shared retentions, not shared innovations that would justify establishing a subgroup.

2. Implications for Prehistory

As mentioned above, this family tree is a claim about the prehistory of the languages under study. Thus it claims that there was a single community speaking a common language, and that this community divided into two parts, proto-Rukai and F1. The community that spoke F1 subsequently divided into proto-Tsou and F2, etc. Linguistic techniques cannot establish exactly where these reconstructed languages were spoken, but if we assume as few movements as possible and movements which are as short as possible consistent with the modern distribution, it is possible to construct a hypothetical picture of possible prehistoric events. In order to get a feel for the kind of reality that the family tree matches, I have constructed the map shown as Figure 3. The locus of F0, Proto-Formosan, was fixed on the southwest coast after experimenting with various choices to see which one gave the fewest and shortest migrations and the fewest crossings of the central mountain range.[4]

---

4   I walked across it once myself, and it is not something I would like to try with my youngest child in my arms and all my household possessions on my back.
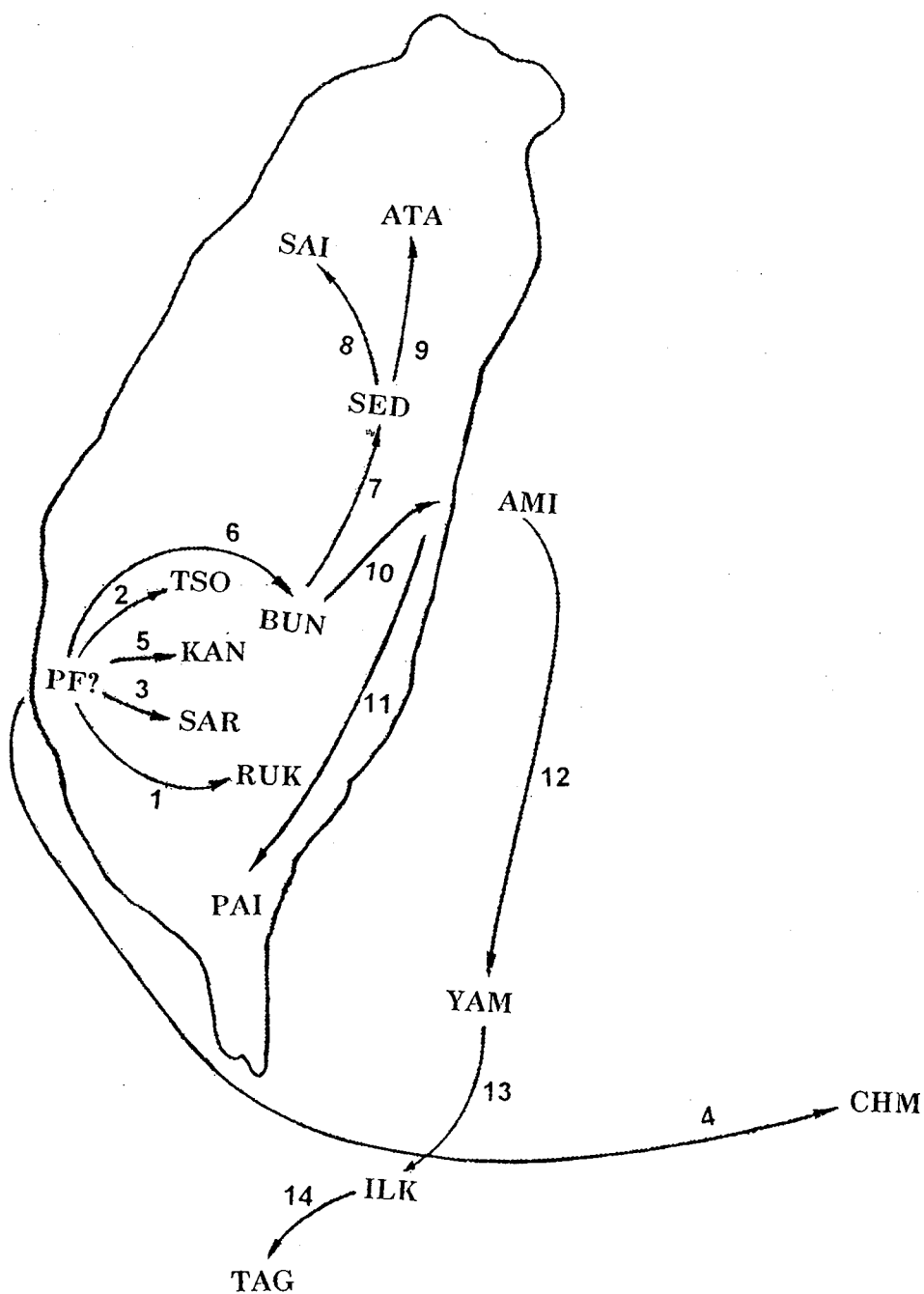
A possible scenario for prehistoric Formosan migrations; map based on Tsuchida 1976:xxx.

A narrative description of the history pictured here might go as follows: Proto-Formosan was spoken on the southwest coast of Taiwan. One group of speakers, the proto-Rukais, moved inland up the Lower Tamshui River valley toward the modern Rukai area [1]. F0 evolved in situ into F1, and then a second group, the proto-Tsous, left for Ali Shan [2]. F1 evolved into F2 and then the proto-Saaroas headed east [3]. F2 evolved in situ into F3, and then the proto-Chamorros headed out to sea [4]. Somehow they rounded Ouluan Pi and headed out into the Pacific, arriving after many harrowing adventures in the Marianas. Meanwhile, F3 was evolving into F4, upon which the proto-Kanakanavus moved into a position between the Tsous and Saaroas [5]. F4 evolved into F5, Proto-Paiwanic, in connection with a migration up the Choshui River valley into the central part of the island [6]. From the new location, another subgroup headed north, evolving into the proto-Northern Formosans. While this group evolved into Proto-Atayalic in the area now occupied by the Seediqs, the proto-Saisiyats moved northwest [8]. Then another group headed north from the Seediq region [9] to found Atayal.

Meanwhile, F6 (Proto-Paiwanic) had been evolving from F5 somewhere in the interior. One group crossed the remaining mountains through Taroko Gorge [10] and arrived in the Promised Land, the Taitung rift valley. Those disinclined to make the arduous journey remained behind, evolving into the Bunun. Once on the coast, the F8 speakers reencountered or reinvented ocean voyaging, and group F9 moved out to Botel Tobago to take advantage of the good flying fish fishing [12]. From there it was only a matter of time until they discovered the Philippines [13,14], and subsequent southward movements were the beginnings of the great Malayopolynesian migration into the Pacific.[5] Some of those left behind on the coast moved down the valley

---

5   According to Lawrence Reid (p.c.), "There seems to be archaeological evidence

toward the southern end of the island [11] and, after exchanging some morphology with the earlier Rukai settlers they met along the way, evolved into the Paiwans [11].

Locating F0 at modern-day Tainan is a convenience which makes it possible to draw a neater map. An alternative view has been proposed by Paul Li (Li 1979, Li 1993), who believes that Nant'ou County, closer to the geographical center of Taiwan, was the center of dispersal for the Formosan natives. Li notes (p.c.) that this is the most diversified linguistic area, where we find the Atayalic, Bunun, Tsou, Thao, Pazeh, and other languages spoken, and asserts 1) that aside from the Atayalic group which spread northward, most Formosan tribes spread southward from Nant'ou, including Bunun and Tsou, and 2) that Rukai, Paiwan and Amis also dispersed to the south, although they may not have started from Nant'ou. My account then contradicts Li's view in several fundamental respects in the area of initial dispersal point and the earliest migrations, though the two accounts are consistent in most other ways. If my genetic tree is correct, Li's scenario would require more walking than a southwest coast staging area would, and my account also has the advantage of providing a beachhead for a sea landing, whereas Li's view seems to require an airborne invasion into the interior. The linguistic evidence I offer here from morphology then counts as counterevidence to

for placing PAn speakers on the west coast, if they are associated with the carriers of the Tapenk'eng culture, the sites of which have been found down that coast, according to Bellwood. This is the earliest Neolithic culture found in Taiwan and almost certainly represents a continuation of mainland Neolithic cultures. Similarly, the earliest Philippine Neolithic assemblages appear to be associated with east coast Yuan-shan sites, which are probably somewhat more north than where you would have F8 speakers....". This of course supports the idea of a west coast entry point for Proto-Formosan, and an east coast exit point for Malayo-Polynesian.

Li's linguistic evidence from the lexicon.

It should be emphasized that some aspects of such a map cannot really be established based on linguistic evidence alone. In support of Li's claim, he refers not only to linguistic evidence, but also anthropological evidence, oral traditions, and Dutch documents citing Mabuchi 1954. However, Dutch records can of course only be used to document recent historical migrations, while oral traditions are just that, traditions, subject to great distortions in short time spans. The migrations I am positing are thousands of years old, so these latter two kinds of evidence have little or no bearing on them. Archaeological evidence on the other hand is of far greater potential significance in providing an acceptable answer to this question.

What I take as well established is that there was a language corresponding to F0, that it was spoken somewhere, probably though not necessarily in Taiwan, and that there were successive physical divisions within the original community speaking the language, most probably as a result of migrations away from the original site of F0 and toward the locations of the respective modern languages.

3. Innovations and Evolution

Each of the subgroups in the family tree proposed here is established in accordance with one or more shared innovations. The tree diagram below is the same one presented above, but it is broken into successive chunks, with abbreviatory annotations at each node[6] referring to some of the grammatical features and innovations associated with it. In the following sections, I will discuss some of the grammatical innovations that characterize each stage of the evolution of the language, and will be paying particular attention to the

---

6   A key to these abbreviations is given in section VI.

evolution of particular morphemes, especially the fabled Four Foci: AF *mu-*,
OF *-en*, LF *-an*, and IF *Si-* .[7]

In discussing each stage and the changes that happened at each node in
the tree, I will speak for convenience as if I were talking about attested
facts. However, it probably doesn't need to be emphasized for a group such
as this that I am talking about a hypothesis, not facts. Such a hypothesis is
only valid until it is replaced by another one which is founded in an equally
constrained and explicit theory and covers a comparable range of observa-
tions more economically.

Proto-Formosan clause and NP structure were strongly right-branching,
and clauses were commonly introduced by a root auxiliary verb which
"attracted" bound pronouns. A "ligature" preposition *ka* or *a* introduced
clauses following a negative auxiliary verb, and probably other complement
clauses as well. PF was an ergative language, so that the Patient in every
clause was marked with the Nominative case, with one other case to mark
non-subjects, including transitive agents, locatives, antipassive "objects", and
adnominal adjuncts.[8] There was a relator noun of the shape *ku* which inter-
acted with pronouns and determiners at various points in the evolution of
the family. It was the base to which genitive bound pronouns attached to

---

7   AF= "Actor focus", OF = "Object focus" (also sometimes GF "goal focus"), LF =
    "Locative focus", and IF = "Accessory (Instrumental, Reason, Benefactive, Referen-
    tial) Focus".

8   For three recent discussions of ergativity in Formosan languages (in the broader
    sense), see Ho 1990, Zeitoun 1992, and Huang 1994. Ho and Huang cite morpho-
    logical, syntactic, and semantic evidence to demonstrate that Yami and Atayal
    respectively are ergative, while Zeitoun draws on Chomskyan misconceptions about
    transitivity and ergativity plus some of her own personal confusion (Zeitoun 1992:
    30-33) in concluding that Tsou is not ergative but rather a "focusing" language,
    whatever that may be.

F0: Proto-Formosan
RGTV, XLRY, BP
CPRP *ka-*, *a-*
DET *i*[+dfnt], *a*[-dfnt]
PR *ku-* PSN
N>N DSTR R
V>N -*an*
V>N [_N *ta-*V-*an*] "place of Ving"
N>N inhabitant -*ana*
N>N *na-* LCTN
V>V NTRN RLS *m-*
V>V MPRT NTRN -*a*
V>V MPRT NTRN -*i*
BP GEN NOM ACTR
S > F /V__V
V>V MPRF C > C*a*C
N>V MOTN V
V>V *ka-*NCHO
V>V CAUS *pa-*
V>V STTV *m-*
*m*S > *m*
V>V OF PRFC *ni-*; -*in-*?

F1

Rukai
NOM DET DMNS *k*V-
N>N PAST *na-*
V>N IF *sa-*
N>V *si-* STTV
P>V CPRP NGTV *ka*
V>V FUTR *ay-*
V>V MOTN *u-*
V>V [_V *wa-*V] PAST
V>V NFNT [_{V2} C-*u*-aX] / V₁ __
V>N PAST, PRFC, FUTR
V>V OF PRFC -*in-*   diff.?
PRSL   diff.?

form nominative pronouns and determiners in Rukai and Amis and nominative and oblique pronouns in Kanakanabu, and it may be related to the Atayal nominative relator noun *qu?* and the *-u* of imperative and subjunctive OF forms above F5 in the tree.

There were two non-nominative determiners, a definite *i* and an indefinite *a* (cf. Starosta 1992). There was a process of full-word reduplication for deriving distributive nouns. A suffix *-an* derived various kinds of nouns from verbs and a circumfix *ta-..-an* derived "place of Ving" nouns from verbs. *-ana* derived names for inhabitants of a place from place nouns, and *na*-prefixing produced "former N" nouns.

PF had verbal derivation and limited inflectional affixation. The inflectional affixes were (1) a realis prefix *m-* which appeared on a limited subset of intransitive verbs and was omitted in irrealis (future and negative) contexts, (2) *-a*, an intransitive imperative suffix possibly derived through the capture of the indefinite non-nominative determiner *a*, (3) *-i*, an intransitive imperative suffix on three-argument verbs, possibly derived through the capture of the definite non-nominative determiner *i*,[9] (4) an infinitival morpheme *-u-*, and (5) an imperative morpheme *-u-*, the latter both infixed after C in a C*a*-initial verb. Stem-final semivowels *-y* or *-w* commonly alternated with homorganic voiced fricatives when followed by a vowel-initial suffix or bound pronoun.

F0 was an ergative language (cf. SPQR) in which a bound pronoun (Nominative for intransitive verbs and Genitive for transitive verbs) marked the actor in a verbal clause, but there was no third person Nominative

---

9   Essentially the same capture process that produced these two intransitive imperative suffixes later derived the transitive suffixes *-a* and *-i* at the next stage down the tree; cf. Starosta 1992.

bound pronoun. There was a process of initial *Ca-* reduplication for deriving imperfective verbs. The prefix *u-*, etymologically related to the verb *ua* "go", derived motion verbs "go to N" from place nouns, a *ka-* prefix derived inchoative verbs from non-stative verbs, and causative verbs were derived by prefixing *pa-* to non-stative verbs.[10] A stative prefix *m-* derived intransitive *ma*-initial stative verbs from *ka*-initial inchoatives, and *mu*-initial potential verbs from *ku*-initial middle verbs. In this process as well as in *m*-inflected realis forms, a following stem-initial *k-* or *p-* were lost after *m-*.

The prefix *ni-* derived perfective OF verbs from transitive stems, and zero nominalization applying to the output derived perfective OF nouns, "thing affected by V". This process established the analogical pattern for the later reverse derivation of well-known Philippine-type verbal focus morphology by means of the verbalization mechanism described in SPQR.

As Rukai evolved from Proto-Formosan, a relator noun of the shape *k*V fused with the original **i* and **a* determiners and demonstrative pronouns *iDa* and *ini* in the formation of the nominative determiners and demonstratives. The same form fused with possessive pronouns to produce the set of *ku*-initial free topic/predicate pronouns (cf. Starosta 1992). The remote demonstrative pronoun *na* fused with following nouns to form "former N" nouns. The prefix *sa*-attached to verbs to produce a noun meaning "implement for Ving", and *si-* attached to nouns to produce stative verbs meaning "have N". Note though that these two processes could also be a result of diffusion: the identical two processes are found in Amis, a language which is much farther down the genetic tree but in fairly close geographical proximity. The Rukai distinction between common and personal nouns is also

---

10   "Since *ka-* causatives are also found in Austroasiatic languages, I would place them at F0." (Lawrence Reid, p.c.)

suspicious, since 1) it does not appear in the genetically closest "Tsouic" languages (Tsou, Saaroa, and Kanakanabu), but does appear in Amis, and 2) it seems to be based more on animacy than "personhood" in the Amis or Philippine sense.

The Rukai negative pronouns *kay* and *kaDu(a)* may be innovations resulting from the reanalysis of a sequence of an earlier negative auxiliary verb followed by the PF ligature *ka* plus the determiner *i* or *iDa* of the following NP. The future prefix *ay-* may have been derived by a similar process from the ligature *a*.

The *u*-initial motion verb derivation pattern of PF was extended to apply to verb roots as well as nouns, and the verb *ua* "go" fused with following complements to produce the past tense derivational prefix *wa-/-aw* (cf. English "<u>went</u> bad", "<u>went</u> wrong", "<u>went</u> sour"), and the same verb is presumably the source of the infinitival infix -*u*- in *Ca*-initial verbs (cf. English "try to <u>go</u> buy a car") and the imperative infix -<u>u</u>- (cf. English "Go get me a beer!").

Rukai extended the earlier zero-derivation strategy for deriving nouns from *ni*-prefixed perfective verbs to aspect-derived verbs in general, including verbs prefixed with *wa-* (intransitive past) and *ay-* (future).

Several infixation processes applied in the evolution of Rukai: infinitival -*u*- only appears as an infix, past tense *wa-* may appear as -*a*-, and the perfective *ni-* may appear as -*in*-. The appearance of -*u*- is conditioned by an initial *Ca*- sequence, *ni-* and -*in*- are usually in free variation,[11] and the infixation of -*a*- is morphologically conditioned. Although the -*in*- infix is ubiquitous in Formosan and Philippine languages, I hesitate to attribute it to

---

11   But see Li 1973:206 for semantic differences in some pairs of infixed versus prefixed forms.

PF because it does not appear in either Tsou or Saaroa, the next two branches down the tree from Rukai in my reconstruction. This suggests that this particular form might have been innovated at the F3 level and diffused back to Rukai from the south or east, reinforced by the presence of the internally innovated -*a*- and -*u*- formations.

1. F1, F2

F1
V>V OF -*a*
V>V OF -*i*
V>V LF -*i*
V>V NTRN -*m*-

F2
V>V OF MPRT -*u*
V>V MPRF *a*-
V>V AF -*um*-
V>V AF PRFC *ni*-
V>V LF -*a*-*na*

Tsou
XLRY SPCT
XLRY MPRT
XLRY ADV

F3

Saaroa
V>V SF *saa*-
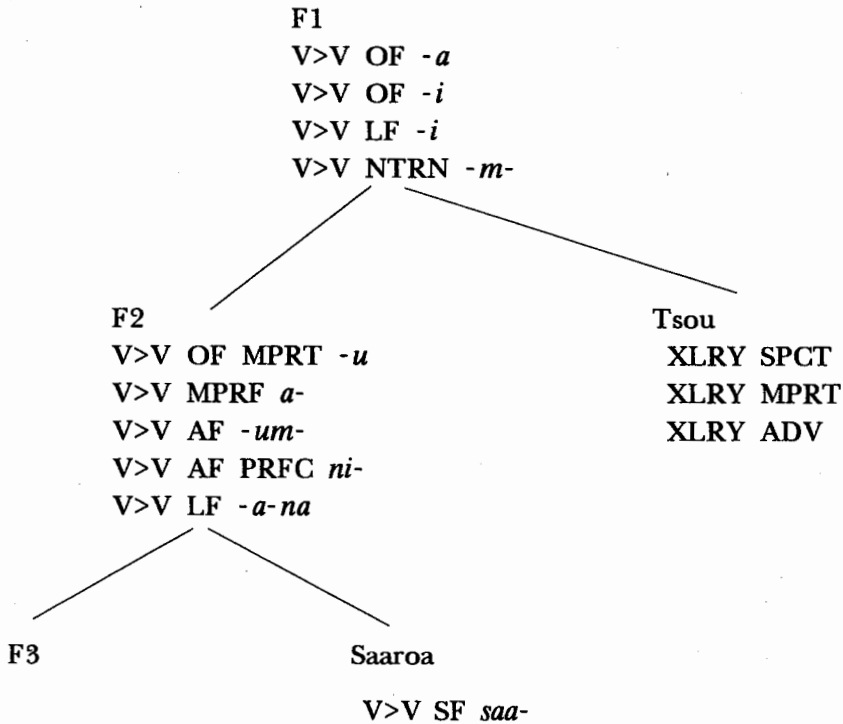
One important innovation at the F1 stage was the development of an OF suffix -*a* and an OF and LF suffix -*i* by capturing the following non-nominative indefinite determiner *a* and definite determiner *i*. The mechanism was presumably an extension of the same one that produced the intransitive imperative suffixes -*a* and -*i* in F0 (cf. Starosta 1992). By analogy with the *ni*-/-*in*- perfective alternation, the realis prefix *m*- (but not the stative m-) developed an infixed alternative form -*m*-. The expression of "adverbial"

concepts such as degree by auxiliary verbs also began at this stage.

Modern Tsou is very much like the structure reconstructed for F1, except for the elaboration of the inherited auxiliary verb system to more aspect, negative, and "adverbial" functions (cf. Starosta 1985).[12] For example, a new way of expressing imperatives using a future auxiliary verb *te* plus a second person bound pronoun displaced the earlier *-u-*, *-a*, and *-i* intransitive imperative inflections. A major consequence of this development was the replacement of the redundant future *ay-* and perfective *ni-/-in-* derivation processes and the loss of the expression of the realis-irrealis distinction by means of an m-/-m- ~ 0 alternation. These auxiliary verbs took over several functions formerly performed by affixes. The result of this latter change was the reinterpretation of *m-/-m-* as a derivational prefix marking intransitivity.

The non-future aspectual auxiliary verbs also appeared in *m*-initial intransitive and *m*-less transitive versions, an extension of the earlier aspectual use of the verb *m-u ~ u* "go", perhaps analogously to the way "go" is used for similar purposes in English. Non-aspectual auxiliary verbs also developed transitive versions through suffixing of the transitive suffix *-a*, and interpretation conventions for bound pronouns produced what appears to be a requirement of transitivity congruence between an auxiliary verb and its dependent verb (cf. Starosta 1991b).

A further determiner-capture processes resulted in an innovation at the F2 stage: an OF imperative *-u* was added to the verbal morphology at this point, possibly through the capture of a nominative determiner *u* by a process similar to the ones that produced the *-a* and *-i* transitive suffixes. The possible beginning of a distinct locative focus LF category in verbs

---

12 The resulting syntactic pattern is strikingly similar to the use of auxiliary verbs in Polynesian languages such as Samoan.

appears at this stage too, also by capture: a locative demonstrative pronoun *na* "there", which may appear adjacent to a locational *a*-suffixed transitive verb, is represented by Tsuchida as an optional suffix on the verb (Tsuchida 1976:71-75,77). That analysis may be premature, however, since Chamorro shows no evidence of *-an* in a LF function. This same *na* may form the beginning of a separate Locative case form in pronouns: Tsuchida shows a separate column of "Oblique pronouns" composed of "na + independent pronoun" (Tsuchida 1976:68).

A small semantic shift in the inherited future noun prefix *a-* and/or the future verb prefix *(a)y*-resulted in a new *a*-prefixed imperfective form which thereby came into competition with the earlier imperfective *Ca*-reduplication process. The realis prefix *m-* aquired a phonologically conditioned variant *-um-* by analogy with the perfective forms *ni-/-in-* and the aspectual use of the verb *mu ~ u* "go".[13] Finally, the transitive perfective affix *ni-* was extended to intransitives as well, with *ni-* preceding realis *m-* when they coocurred.

The primary innovation in the Saaroa language was a *saa*-prefixed "special focus" category (cf. Tsuchida 1976:69,77-78). As Tsuchida notes, it is grammatically "Goal Focus" (OF here), but its origin and distinctive function are unclear. It may derive from an old subjunctive transitive verb form retained in subordinate clauses, including those introduced by the complementizer preposition *sa*. It is tempting to relate this to the *sa*-prefixed implement nominalization in Rukai and Amis, but the evidence so far does not

---

13 This process was probably independent of the development of *-m-* in Tsou, which never shows up as *-um-*. If it turns out to be possible to account for this using Tsuchida's vowel syncope rules, the innovation will need to be posited only once, at the F1 level.

support this connection.

2. F3

```
                    F3
                    V>N OF NTNT -en
                    V>V PRFC -in- ?
                    V>N AF mu-
                    LIG na
                         /\
                        /  \
                       /    \
          Chamorro              F4
          V>N ABIL -on
          RF -i, BF -idzi
          DET GEN nu i > ni
          DET PRSL si    diff.
```
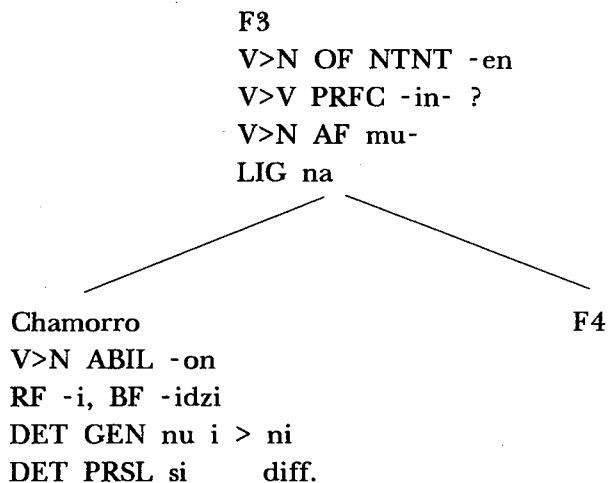
F3 is the stage at which a new verbal future OF suffix *-en* (*-n* in Kanakanabu, *-on* in Chamorro) was innovated, since reflexes are not found in Rukai or Tsou but appear in all the languages below this point. According to SPQR, this suffix came into existence by analogy with a nominalizing suffix **-en* which marked derived intent OF nouns, i.e. V >--> [N V-*en*] "the thing to be V-d". Unfortunately, I have not found such a nominalizing use of this form in either Saaroa or Chamorro,[14] raising the possibility that the verbal suffix came from some as yet unidentified source, and later got into nouns the same way that Rukai perfective *ni-* and past *wa-* did: by zero nominalization of morphologically complex verb forms.

As mentioned earlier, deverbal OF nouns marked by *ni-/-in-* and referring to past affect, i.e. V >--> [N *ni*-V] "thing which has been affected by

---

14 The Chamorro counterpart of *-en* is *-on*. All the examples of this form provided by Topping are forms with glosses like "capable of Ving", which fit the desired semantics fairly well. They are verbs instead of nouns, while the only nominalizing use of *-on* he provides (Topping 1973:181) looks AF rather than OF: guasa'on "can be sharpened" or "sharpener".

Ving", were formed from perfective verbs by zero derivation already at F0. F3 is the point at which the analogical pattern created by this process is first used in reverse, resulting in the creation of deverbal AF nouns marked by *mu-/-um-* and referring to the performer of an action, i.e. V >- -> [N *mu*-V] "one who does V". This formation is common in Chamorro (cf. Topping 1973:102), and I have at least one example from Bunun. It is ubiquitous in Philippine languages such as Tagalog and Ilokano.

In NP structure, F3 innovated a "ligature" element *na*, related to a PF non-proximate demonstrative noun **na*, which was used to mark an NP as an equational modifier of the preceding noun. This is the progenitor of the Tagalog "linker" *na/-ng*.

If Chamorro is, contrary to common sense, really a Formosan language (cf. Starosta and Pagotto 1990), then it was F3 which was waving from the shore when the proto-Chamorros set off on their fateful voyage.[15] According to this scenario, Chamorro's strikingly Filipinic features, previously used as evidence for a Philippine connection (cf. Topping 1973:3) actually on closer inspection give Chamorro a place in Formosan prehistory. The apparent absence of reflexes of LF **-an* and Accessory Focus **iSi*, embarrassing to a Philippine origin hypothesis, are accounted for by a geographical separation

---

15   "Nobody knows for sure where the first Chamorros came from. It is safe to assume that the original Chamorros belonged to the large group of Philippine peoples known generally as Malayo-Polynesian, but there is no certain evidence to tell us where the first inhabitants of the Mariana Islands came from. Laura Thompson has suggested that the first inhabitants of the Mariana Islands were the descendants of seafaring folk who migrated westward from Asia to the Philippines to the Western Carolines and, finally, to the Marianas (Thompson 1947). According to Spoehr (1954:38), there is good evidence . . . that the first settlers of the Marianas arrived somewhere around 1527 B.C. ± 200 years, or approximately 3,500 years ago!" (Topping 1973:2)

prior to the innovations which produced these forms. The "capable of being Vd" meaning of the Chamorro -on counterpart of OF *-en is accounted for here in terms of a slight semantic shift from an "intended affect" meaning. There is no garden variety verbal OF -en in Chamorro because when the Proto-Chamorros left Taiwan, it had not yet evolved.

The Chamorro Referential Focus Marker -*i* and Benefactive Focus Marker -*iyi* derive directly from the -i Locative Focus suffix which appeared at F1. The original meaning of "V to NP" has expanded to include "V for NP", but it is striking that close analogues of the semivowel-fricative alternation posited all the way back to F0 also appear in Chamorro. (The following examples are given in a revised phonemic transcription.)

Tsou

ahoɣ	"begin" (intransitive) (cf. Tung 1964:179)
ahoza	"begin" (transitive)
sifkow	"flay" (intransitive) (cf. Tung 1964:179)
sifkova	"flay" (transitive)
yansow	"breathe" (intransitive)
yansovi	"breathe" (transitive)
sɯʔno	"get angry" (intransitive) (cf. Tung 1964:530)
sɯʔnova	"get angry" (transitive)

Chamorro (cf. Topping 1973:191, 250)

taytaɣ	"read"
taytadzi	"read to"
hanaw	"go"
hanagwi	"go for"
kanta	"sing"
kantadzi	"sing to/for"

One feature of Chamorro which might at first glance seem to justify plugging the language into the Formosan family tree farther down is the existence of a genitive determiner *ni*, which I have reconstructed as an Eastern

Formosan innovation. However, this is synchronically in alternation with *nu i*, and thus must be an independent ongoing Chamorro-internal innovation. It would be nice to relate the *nu* to the demonstrative *na*, but I don't know if this is plausible in term of the correspondences. The personal determiner *si* on the other hand I would regard as a loan from a Philippine language.

Some of the evidence against a Formosan homeland for Chamorro within the parameters of this paper is the absence of reflexes of the verbal deriva-tional affixes imperative *-u* and motion *u-*, plus the lack of a set of *-a* or *-an-*suffixed locative pronouns. Since Chamorro is a well-described language, these lacunae are worrisome, since they cannot be readily dismissed as accidental gaps in the data.

3. F4

```
        F4
        V>V  LF  -an
        V>V  OF  -a >> -en  in main clauses
        V>V  OF  CM RRLS  -a-u



F5                          Kanakanavu
                            V>V  "SF"  -ai
                            V>V  AF  m-
                            PR  LCV  -a
                            LIG  na ??
                            V>N  AF  mu-/-um-  ??
                            V>V  LF  -i      ??
                            V>V  OF  -i      ??
```
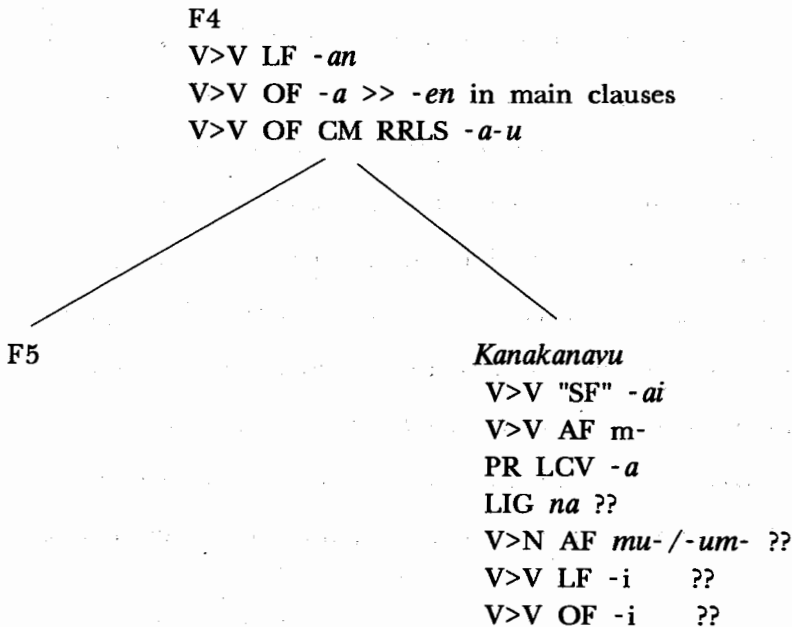
In F4, the association of a locative demonstrative noun *na* with a transi-tive perfective *a*-suffixed verb, which may have been only juxtaposition at F3, achieves affixhood via essentially the same capture process postulated in

the evolution of *-a* and *-i* suffixes; the resulting *-a-na* alternates with *-an* due to the word-final vowel loss occurring at this time, and *-an* is reanalyzed as the base form, with the final *-a* reinterpreted as an echo vowel. The new LF *-an* is now in competition with the earlier *-i*, and *-an* takes on the LF function in main clauses, with *-i* relegated to subordinate clauses. Analogously, the new OF *-en* innovated at F3 takes over the main-clause OF function, relegating OF *-a* to subordinate clauses. In subordinate clauses, an irrealis OF form *-au* arose, possibly through the capture of a Nominative determiner **u* by such an OF transitive *-a* verb.

The Kanakanabu intransitive realis *m-* prefix became lexicalized, and no longer alternated with its absence. The result was the loss of the realis function of this derivational process, leaving only its function as a marker of intransitive verbs (though not of all intransitive verbs). The "special focus" *-ai* appearing in Kanakanavu (cf. Tsuchida 1976:49-51) appears to be a kind of antipassive, and the inherited *k*-initial determiners are replaced by *s*-initial forms, though locative *na* is retained. *na* may also be the source of the Oblique *-(n)a* pronoun paradigm. The *na* precedes oblique pronouns in Saaroa, but if this was a demonstrative, it could either have preceded or followed the pronoun, allowing for both forms. On the other hand, no trace of this can be assigned to the intervening F3 node, suggesting that this may be a separate innovation. At any rate, it is a robust feature, carrying on down into Atayal and Amis. Other than these points, Kanakanavu is basically F4.

There are however several features which according to my tree should be present, but which have not yet turned up. They include a ligature *na*, the nominalizing use of the AF affixes *mu-/-um-*, and reflexes of the OF and LF *-i* innovated at F1. Since my data are limited to a single 23-page section of

Tsuchida (1976) plus a short paper on pronouns by Kuang Mei, at least some of these missing elements could turn out to be accidental gaps.

4. F5: Interior Formosan

F5: *Interior Formosan*
V>V OF -*an*
V>N LF -*an*
V>N OF FUTR -*en*
*ni-*C-*um-* > C-*um-in-*

F6: *Paiwanic*   *Northern Formosan*

F5 *Interior Formosan* is the common ancestor posited here for the Paiwanic and Northern Formosan protolanguages. One of its innovations must have been the extension of the Locative Focus to certain classes of OF verbs, since OF -*an* verbs exist in both Northern Formosan and Paiwanic languages, and well on down the tree to Tagalog (cf. Foley 1976:139). This innovation came about via a semantic shift which equated an action having a superficial effect with an action localized at a point. As Egli puts it, "While we are paying attention to the end product--letters, words--in writing, the Paiwan have in mind the surface on which the tattooing or writing is done. Thus even where symbols, letters, etc. are concerned, the Paiwan must use -*an: vetsik-an*. In terms of its meaning, therefore, *v/en/tsik* "write" should rather be glossed as "write on"" (Egli 1990:237, my translation). Similar examples could have been selected for example from Atayal verbs such as *pima* "wash".

Based on the existence of nouns of the form *Ca*-V-*an* "place of Ving" in Paiwan (Egli 1991:130) and V-*an* "place where someone Vd" in Atayal (cf. Huang 1990:34-35) and Bunun, we may conclude that locative nouns were derived from locative focus verbs by zero derivation at stage F5. However,

the zero derivation process itself goes all the way back to F0, and could in principle apply to any verb stem, so we will probably find evidence for the existence of this process going all the way back to the innovation of LF -*an* verbs in F4, and perhaps separate and independent instances of this innovation below this point in the tree. Identical considerations apply to derivation of nouns of the form V-*en* "thing to be Vd" in Paiwan (Egli 1990:127), Atayal (Huang 1993:36), and Bunun. A related innovation at this stage was the replacement of *ni-C-um-* derived perfective AF nouns by secondarily infixed C-*um-in-* forms.

5. F7: Northern Formosan

F7: *Northern Formosan*
S > F /V___V ??

Atayalic
XLRY for FUTR
XLRY for MPRF
XLRY for PRFC
NR PSN *na_*?
V>N OF NTNT -*en* ??
V>V MOTN *u-* ??

Saisiyat
DET GEN *ni* diff.?
DET PRSL diff.?
LIG *a* diff.?
V>N AF *mu-* ??
V>N OF PRFC *ni-* ??
V>V OF CM -*au* ??
V>V PRFC *ni-* ??

Seediq
V>V OF CM -*au* ?

Atayal
V>V OF PAST, partial affect -*an*
V>N AF *mu-/-um-* ??
V>N PRFC OF *ni-/-in-* ??
V>V PRFC OF *ni-* ??
V>V MOTN *u-* ??
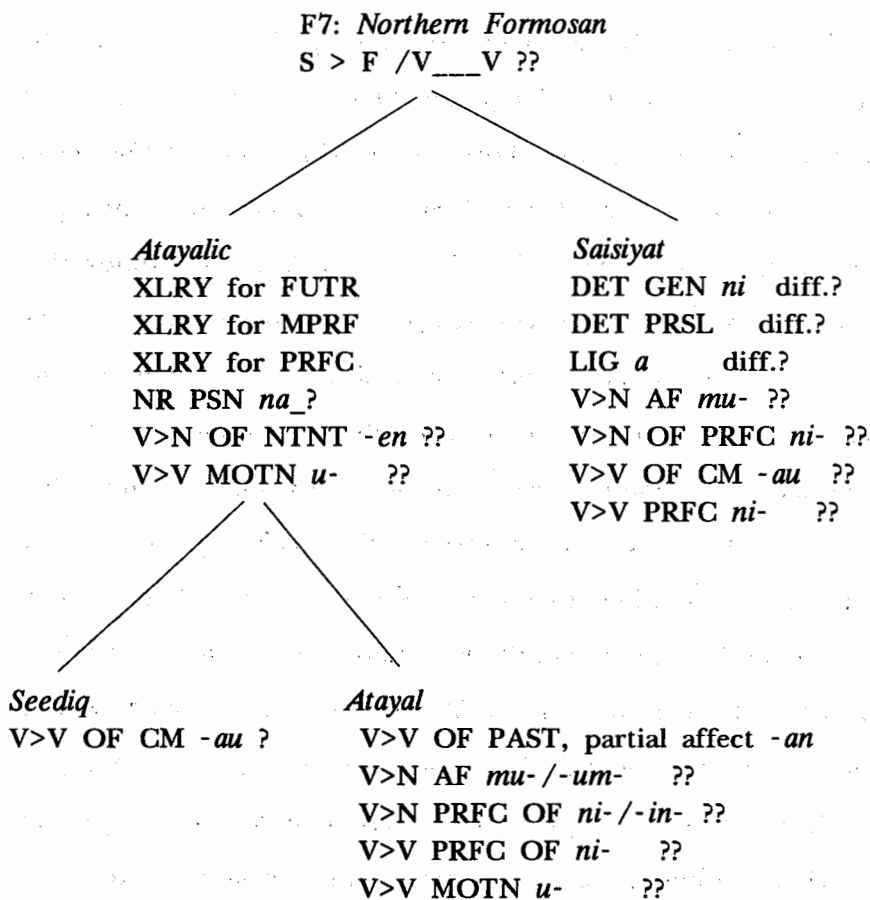
I have so far found no shared positive grammatical innovations to justify linking Saisiyat with the Atayalic subgroup, as Li has proposed (Li 1985), and in fact grammatically Saisiyat looks far more Paiwanic than Atayalic. There is a bit of negative evidence, however, for putting these languages together: the three Northern Formosan languages all seem to lack reflexes of the semivowel/voiced interdental fricative alternation present in PF. If Saisiyat does indeed subgroup with Atayalic, then the three innovations I have identified for this language so far would have to have diffused from other languages: the *n*V Genitive determiner, a personal-common noun distinction, and a ligature *a* all seem to derive from Eastern Formosan, though at present Saisiyat is off in the northwest, not in contact with any of these eastern languages.

What I have not found is reflexes of the prefixes *mu-* or *ni-* in their nominalizing functions. I have also not found the motion verb prefix *u-*, the subordinating OF suffix *-au*, or the prefix counterpart *ni-* of the verbal prefective infix *-in-*.

One grammatical innovation that supports regarding Seediq and Atayal as a subgroup, Atayalic, is the introduction of a new system of aspect-marking auxiliary verbs. This system is similar to the one innovated in Tsou, and is also based on the earlier use of *m-o* "go" as an auxiliary verb. Missing so far are examples of the denominalizing use of *-en*. In contrast to most of the other Formosan languages, the Atayalic languages hardly use determiners at all, though the *qu*? and *na*? are retentions from earlier forms attested in languages higher up the tree. This group innovated a genitive relator noun *na*? from an earlier non-proximate demonstrative noun. It should also be mentioned that this subgroup underwent some rather radical phonological reductions, which sometimes obscure the presence of the morphology being

tracked here.

Atayal has further extended the semantic domain of OF *-an* to mark partial effect (cf. Huang 1993:31-32) and past direct OF (cf. Huang 1993: 67), in which *-an* replaces *-en* (*-un* or *-on* in Atayal) in past tense functions. Atayal and Seediq are distinguished from the other Formosan languages in my data by what I expect them to have, based on my subgrouping hypothesis, but have yet to find: Seediq lacks an OF *-au*, and Atayal, like Saisiyat, lacks a motion verb prefix *u-*, a *mu-* nominalizer, and a prefix version of the perfective infix *-in-*.

6. F6: Paiwanic

<pre>
        F6: *Paiwanic*
        V>V MOTN *m-u* lost?
        V>V OF CM *-au* ??
                /\
               / \
              /   \
             /   \
            /   \
           /   \
    F8   *Bunun*
                       V>V OF, IF *is-*
                       LIG *na·*   ??
</pre>
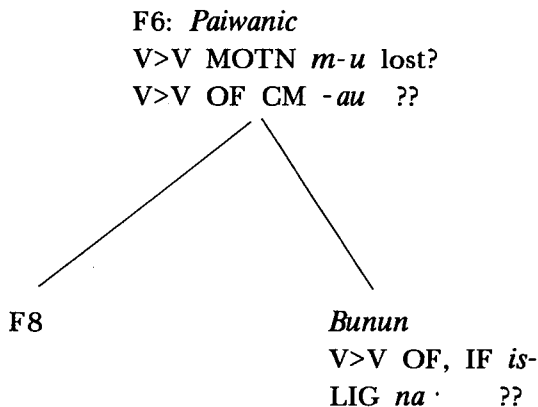
At this point there is almost no grammatical evidence for a Paiwanic subgroup other than the lack of a *u*-prefixed motion verb class, something that is also true of Saisiyat and Atayal for example, and the absence of the OF use of *-au*. I am not yet prepared to give up this subgroup until I get better data on Bunun, however. Bunun seems to have innovated a verbal OF and IF prefix *is-*, but otherwise is distinguished by what it should have but so far doesn't seem to: the *na* ligature.

7. Eastern Formosan

F8: *Eastern Formosan*
DET GEN *n*V-
DET OBL *t*V-
DET PRSL/CMMN
V>V AF *paka-*
LIG *a*
V>V MOTN *u-* ??

F9: Extra-Formosan   F10
                         V>V PRFC *na*   diff.
                         V>N AF *mu-/-um-* ??

Paiwan   *Amis*
V>N OF PRFC *ni-* ??   TRNS marks PRFC
V>V PRFC *ni-*   ??   V>N OF PRFC *-in-* ??
V>N *Ca-..-an* LF   V>V PRFC *ni-/-in-* ??
                                 DET NOM *k*V
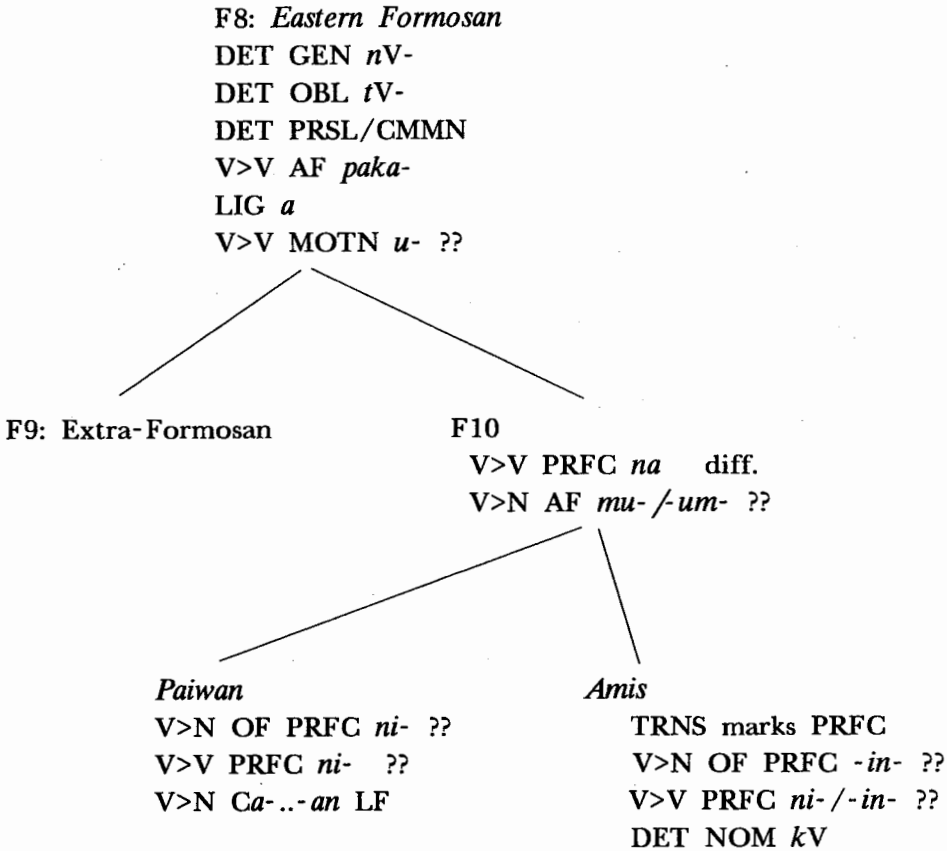
The most salient constituting grammatical properties of Eastern Formosan are two developments in the determiner system: the evolution of an *n*-initial Genitive case form category which is distinct from Locative, and of what may be referred to as an Oblique case form, marked by *t-* or *s-*, whose chief function was to mark the downgraded indefinite notional "objects" of non-OF constructions. Coinciding with this development was the innovation of a clear grammatically definable Philippine-style common-personal distinction in the nouns. (From this point on, Philippinists will find the scenery more familiar.) Eastern Formosan also innovated a verbal derivational prefix *paka-*, which however may turn out to be related to the

causative *pa?-* in Tsou and *pak-* in Saisiyat. Some languages of this group also show a ligature element *a*, but it is not yet clear whether this is a retention from the PF indefinite determiner **a* or perhaps derived from the inherited negative ligature *ka*. There are possible instances of PF **mu-* used as a motion-verb prefix, but it is hard to separate them from other uses of the prefix *mu-*.

Eastern Formosan divides up into two groups, the landlubbers (F10) and the seafarers (F9). As mentioned earlier, my Eastern Formosan matches Reid's *Amis-Extra-Formosan* except for the inclusion of Paiwan, and F9 is his *Extra-Formosan* (Reid 1982).

F10 is not a strongly supported subgroup. It is characterized only by two innovations: a new perfective auxiliary verb-cum-complementizer *na*,[16] and the loss of the nominalizing function of **mu-/-um-*. Subsequent research could well result in separating the two into independent branches of Eastern Formosan, or in supporting Reid's Amis-Extra-Formosan. One possible piece of evidence in support of this alternative is the existence of an apparent *ma-* derivational prefix in Amis, analyzed by Chen (Chen 1987:80-82) as a "Nominative instrument" formation. Grammatically, the structures in which this form occurs are antipassives requiring indefinite notional "objects", and this pattern can be matched by *mang-*prefixed forms in Yami, Ilokano, and Tagalog. The "instrumental" aspect of the Amis construction is also matched by the instrumental *pang-* in Ilokano and Tagalog.

---

16 Lawrence Reid (p.c.) suggests that this *na* implies an earlier *mina-* sequence, which must have been present in F8, if not earlier, since it was retained in Extra-Formosan. Two points against this hypothesis are that as far as I know, neither Paiwan, Amis, nor Yami exhibits the sequence *m-in-a-* in any other forms, and *na* occurs productively in Amis before noun predicates, where it would not have occurred if it derived from a sequence of two verbal prefixes.

Within F10, Paiwan seems to have taken over all the infix reflexes of perfective *ni-/-in-*, while Amis lost all of the infixes and all but a few of the prefix reflexes of this morpheme. I have only a few *ni-* nominalizations in my Amis data, and no verbal perfectives at all. In place of this morphological aspect encoding, perfectivity in Amis is marked by using a transitive in place of an intransitive verb form. In fact, this even works for notional intransitives. Thus the grammatically intransitive *-um-* form of *rakat* "walk", *rəmmakat*, is glossed as "is walking", while the transitive form, *rakatən*, is glossed as "walked". Grammatically, the latter form is a transitive impersonal and could be literally glossed as "He walked it to X". This accords well with Hopper and Thompson's correlation between transitivity and perfectivity (Hopper and Thompson 1980), and is probably a result of an accelerating tendency to use transitive forms to encode events perceived as telic.

The case-marking system of Amis is very neat and clear-cut, with Nominative determiners marked by *k-*, genitive by *n-*, Oblique by *t-*, and topics by the absence of these consonants. The *k-*marked forms have some similarity to Rukai forms, but since such forms fade out in coming down the genetic tree, the similarity could be accidental. The *i-*prefixed locative determiners and nouns too are ultimately related to Rukai and Paiwan forms, but the Paiwan *i* determiners are infrequent archaisms (cf. Starosta 1992), while the Rukai forms are remote in family relationship and perform quite a different function within the Rukai case-marking system. Instead, the *i-* in the Amis forms may be an innovation extended from the use of the Amis verb *i* "go to" in a non-root case-marking function.

Paiwan LF *-an* nominalizations all seem to require a cooccurring *Ca-*reduplication (cf. Egli 1990:130), which suggests that this may be an independent innovation rather than a retention of the earlier LF *-an* nominalization.

8. Extra-Formosan

F9: *Extra-Formosan*
V>V TF *ka-*
V>V IF *sa-* > *i-*
DET NOM *k*V-   ??

Yami
V>V NFNT -*um*-
LIG *na* ??
V>V AF *paka-* ??
V>V PRFC -*in*- ??
V>N AF *mu-/-um-* ??
V>N LF -*an*   ??
V>N OF PRFC *ni-/-in-* ??

Philippines, etc.

Tagalog   *Ilokano* etc.
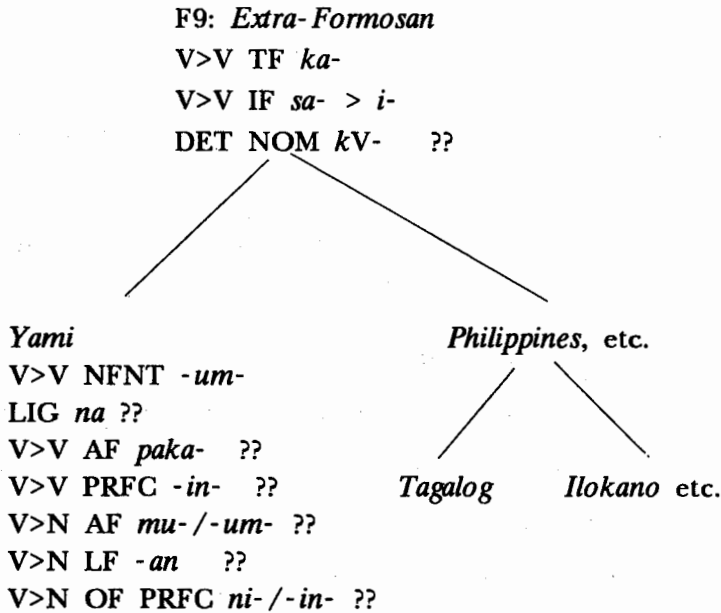
The remaining language in this study, Yami, does not subgroup grammat-
ically with any of the other Formosan languages, supporting the generally
accepted conclusion that it belongs rather with the Batanic languages of the
Philippines. It is distinguished from the other languages of the study by the
use of AF -*um*- in infinitival clauses, the existence of a "time focus" form *ka-*
similar to the Tagalog impersonal "recent perfective" focus (cf. Schachter and
Otanes 1972:371-375), and by a replacement of the IF form *sa-* by i-.[17]

Once again, although Yami is almost certainly a product of Formosan-
internal linguistic evolution, it is striking for the things that it does <u>not</u>

17   There is some question whether the *S of *Si-* was actually lost at this point:
      "A few Philippine languages still show *hi* rather than *?i-* as the IF prefix. Zorc
      (1977:134) cites Samar-Leyte, Waray and Northern Samar *hi-* as forming part of
      the IF potential affix forms (*nahi-*, *mahi-*, etc.), and Tausug *hipag-* as the IF
      dependent, durative form." (Starosta, Pawley, and Reid 1982:165) However, if
      Zorc's analysis is correct, then it may be that the *S was preserved only inside of
      these complex affixes but not initially.

appear to have: the *na* ligature,[18] *paka-*, verbal perfective infix *-in-*, and the denominal uses of *mu-/-um-, ni-/-in-*, and V>N LF *-an*. I suspect that most or all of these will turn up, since they do occur farther down the tree. If they don't, it suggests the possibility that more of the innovations recon-structed in this paper were diffusions than I had originally thought, and that the strip of Pacific Ocean between Taiwan and Orchid Island was a fairly effective impermeable membrane for some of the diffusions.

Conclusion

This paper is certainly not the final answer to the Formosan subgrouping problem, but just one more contribution to the ongoing saga. It is based on a limited database, and there is much more data in my own files which has not yet "come on line", while new studies of Formosan languages are under way right now in Taiwan. At the very least, I hope this paper will have a heuristic function, suggesting the kinds of data we need to look for in order to confirm or disprove the hypothesis. I myself will be looking especially closely at the pronoun systems of the Formosan languages, and at data from Bunun and the Sinicized plains languages in order to carry on from this point.

---

18   "For F8, I think you have to have a phonologically conditioned alternation between *na* and *a* ligatures to account for the ligature alternations in Philippine languages. Some languages maintain the alternation, some have lost it and retain only one of the forms usually *a* (Ivatan, Yami, Inibaloi, etc.). The same thing could account for the fact that some of your languages show *a*, others *na*." (Lawrence Reid, p.c.)

Key to Annotations

ABIL	abilitative
ACTR	actor; roughly, the interior of the action
AF	actor focus
a[-dfnt]	non-nominative determiner *a
BP	bound pronouns, typically clitics
C > CaC	reduplication of initial consonant, insertion of a C-u-a infix -u- before vowel a
CAUS	causative
CM	sentential complement
CPRP	complementizer preposition; "ligature"
DET	determiner
diff.	diffusion from a neighboring language
DMNS	demonstrative
DSTR	distributive
FUTR	future tense
GEN	genitive
i[+dfnt]	non-nominative determiner *i
IF	Instrumental Focus
LCTN	location
LF	locative focus
LIG	ligature
MOTN	motion verb
MPRF	imperfective
MPRT	imperative
NOM	Nominative
NTRN	intransitive
mS > m	m-k > m, m-p > m / #___
N>N	noun-to-noun derivation
N>V	noun-to-verb derivation
NCHO	inchoative
NFNT	infinitive
NGTV	negative auxiliary verb
NOM kV-	nominative determiners of the shape kV
NR	relator noun
OF	objective focus (= "goal focus")
P>V	preposition-to-verb derivation

PR	free pronouns
PSN	possessive nouns and pronouns
PRFC	perfective
PRSL	personal noun class
R	reduplication
RGTV	ergative
RLS	realis
RRLS	irrealis
S# ~ F /V__V	final semivowel alternates with intervocalic fricative
SPCT	aspect
STTV	stative
V>N	verb-to-noun derivation
V>V	verb-to-verb derivation
XLRY	root auxiliary vers

REFERENCES

Blust, Robert A.

1977   The Proto-Austronesian Pronouns and Austronesian Subgrouping: a Preliminary Report. University of Hawaii Working Papers in Linguistics 9.2:1-15.

Chen, Teresa M.

1982   Verbal Constructions and Verbal Classification in Nataoran-Amis. Ph.D. Dissertation, University of Hawaii. Pacific Linguistics Series C, No. 85.

Dahl, Otto C.

1973   Proto-Austronesian. Scandanavian Institute of Asian Studies Monograph Series no. 15. Lund: Studentlitteratur.

Egerod, Søren.

1966   Word Order and Word Classes in Atayal. Language 42:346-369.

Egli, Hans.

    1989   Mirimirngan--Die Mythen und Märchen der Paiwan.

    1990   Paiwangrammatik. Wiesbaden: Otto Harrassowitz.

Foley, William.

    1976   Comparative Syntax in Austronesian. Ph. D. Dissertation, University of California, Berkeley.

    1991   The Lexical Basis of Argument Structure in Western Austronesian Languages. Symposium on Grammatical Relations in Austronesian, Sixth International Conference on Austronesian Linguistics, Pacific Science Congress, University of Hawaii, Honolulu, May 20-24.

Ho, Arlene Y.L.

    1990   Yami Structure: a Descriptive Study of the Yami Language. M.A. Thesis, National Tsing Hua University, Taiwan.

Ho, Dah-an.

    1983   The Linguistic Position of Rukai Among Formosan Languages [in Chinese]. Bulletin of the Institute of History and Philology Academia Sinica, 54.1:121-168.

Hopper, Paul J., and Sandra A. Thompson.

    1980   Transitivity in Grammar and Discourse. Language 56.2:251-299.

Huang, Lillian M.

    1991   The Semantics of *s-* in Atayal. Studies in English Literature and Linguistics 17:37-50. Taipei: National Taiwan Normal University.

    1993   A Study of Atayal syntax. Taipei: The Crane Publishing Company.

    1994   Ergativity in Atayal. Oceanic Linguistics 33:1 129-143.

Li, Paul Jen-Kuei.

1973   Rukai Structure. Institute of History and Philology, Academia Sinica Special Publication No.64. Taipei.

1979   Linguistic Evidence for the Homeland of the Formosan Native Tribes [in Chinese]. Dalu Zaji 59.1:1-14.

1985   The position of Atayal in the Austronesian family. In Andrew Pawley and Lois Carrington (eds), Austronesian Linguistics at the 15th Pacific Science Congress. Pacific Linguistics, C-88, pp.257-280.

1990   Classification of Formosan Languages: Lexical Evidence. Bulletin of the Institute of History and Philology 61.4:813-848, Academia Sinica.

1993   Language Distribution and Migration of the Formosan Natives [in Chinese], First International Conference on Languages in Taiwan, Taipei, March 27-28. In Tsao, Feng-fu and Tsai, Mei-huei (eds), Papers from the First International Symposium on Languages in Taiwan, 1-16(1995). Taipei: The Crane.

Mabuchi, Tooichi.

1953-1954   Migration and Distribution of the Formosan Aborigines (Part I) [in Japanese]. Minzokugaku Kenkyu 18.1-2:123-154, 18.4:23-72.

Mei, Kuang

1982   Pronouns and verb inflection in Kanakanavu. Tsing Hua Journal of Chinese Studies, New Series 14.1-2:207-231.

Pawley, Andrew

1966   Polynesian Languages: a Subgrouping Based on Shared Innovations in Morphology, Journal of the Polynesian Society 75:39-64.

Pawley, Andrew, and Roger Green.

    1973   Dating the Dispersal of the Oceanic Languages. Oceanic linguistics 12.1-2:1-67

Reid, Lawrence A.

    1982   The Demise of Proto-Philippines. In Amran Halim, Lois Carrington, and S.A. Wurm (eds), Papers from the Third International Conference on Austronesian Linguistics, vol. 2: Tracking the Travelers. Pacific Linguistics C-75, pp. 210-216.

Starosta, Stanley.

    1985   Verbal Inflection versus Deverbal Nominalization in PAN: the Evidence from Tsou. In Andrew Pawley and Lois Carrington (eds), Austronesian Linguistics at the 15th Pacific Science Congress. Pacific Linguistics, C-88, pp. 281-312.

    1988a The Case for Lexicase: an Outline of Lexicase Grammatical Theory. London: Pinter Publishers.

    1988b A Grammatical Typology of Formosan Languages. Fang-kuei Li memorial volume. Bulletin of the Institute of History and Philology, Academia Sinica, 59.2:541-576. Taipei: Institute of History and Philology, Academia Sinica. (published in 1991).

    1990   Subgrouping by Lexical Similarity. Bulletin of the Institute of History and Philology, Academia Sinica, 61.4:836-840.

    1991a The Great AUX Cataclysm: Diachronic Justification for a Synchronic Analysis. In B. Lakshmi Bai and B. Ramakrishna Reddy (eds), Studies in Dravidian and General Linguistics: a Festschrift for Bh. Krishnamurti. Department of Linguistics, Osmania University, Hyderabad, pp. 505-515. Unabridged preliminary version in University of Hawaii Working Papers in

Linguistics 17.2:95-114 (published in 1987).

1991b Clitic Pronoun Reference and Ergativity in Four Western Austronesian Languages. Sixth International Conference on Austronesian Linguistics, Pacific Science Congress, University of Hawaii, Honolulu, May 20-24.

1992 The Case-marking System of Proto-Formosan, Plenary session, Third International Symposium on Language and Linguistics: Pan-asiatic Linguistics. Chulalongkorn University, Bangkok, Thailand, 10 January; to appear in volume IV of the proceedings.

Starosta, Stanley, and Louise Pagotto.

1989 The Grammatical Genealogy of Chamorro. In Ray Harlow and Robin Hooper (eds), VICAL 2: Papers from the Fifth International Conference on Austronesian Linguistics. Te Reo monographs. Auckland: Linguistic Society of New Zealand.

Starosta, Stanley, Andrew Pawley, and Lawrence A. Reid.

1982 The Evolution of Focus in Austronesian. In Amran Halim, Lois Carrington, and S.A. Wurm (eds), Papers from the Third International Conference on Austronesian Linguistics, vol. 2: Tracking the travelers; Pacific Linguistics C-75, pp. 145-170; referred to in this paper as 'SPQR'.

Topping, Donald M.

1973 Chamorro Reference Grammar. PALI Language Texts: Micronesia. Honolulu: University of Hawaii Press.

Tsuchida, Shigeru.

1976 Reconstruction of Proto-Tsouic Phonology. Monograph Series no.5, Study of Languages and Cultures of Asia and Africa. Institute for Study of Languages and Cultures of Asia and Africa.

Tokyo Gaikokugo Daigaku.

Tung, T'ung-Ho.

1964 A Descriptive Study of the Tsou Language, Formosa. Institute of History and Philology Special Publication no.48. Taipei: Academia Sinica.

Wolff, John U.

1973 Verbal Inflection in Proto-Austronesian, in Parangal Kay Cecilio Lopez (Essays in honors of Cecilio Lopez on his Seventy-fifth Birthday), ed. by Andrew B. Gonzales. Philippine Journal of Linguistics Special Monograph Issue no.4. Quezon City: Linguistic Society of the Philippines, 71-94.

1979 Verbal Morphology and Verbal Sentences in Proto-Austronesian. Austronesian Studies: Papers from the Second Eastern Conference on Austronesian Languages, ed. by Paz Buenaventura Naylor, 153-168. (Michigan papers on South and Southeast Asia no. 15) Ann Arbor: Center for South and Southeast Asian Studies, University of Michigan.

Yeh, Mei-li.

1991 Saisiyat Structure. M.A. thesis, Institute of Linguistics, National Tsing Hua University, Hsinchu, Taiwan.

Zeitoun, Elizabeth

1992 A Syntactic and Semantic Study of Tsou Focus System. M.A. Thesis. Hsinchu, Taiwan, R.O.C.: Institute of Linguistics, National Tsing Hua University.

Reconstructing Proto-Austronesian Verbal Morphology: Evidence from Taiwan

Malcolm D. Ross

Australian National University

In this paper the verbal morphology of Proto-Austronesian is reconstructed in order to provide a baseline for future research on the history of Austronesian verbal morphology. Particular attention is paid to data from the languages of Taiwan because of the special position they occupy in the Austronesian language family.

1. Introduction[1]

Because the aboriginal languages of Taiwan probably represent one or more primary Austronesian subgroups, any reconstruction of Proto Austronesian (PAN) must take very considerable account of them. The subgrouping hypothesis assumed here is that all Austronesian languages outside Taiwan belong to a single subgroup, dubbed Malayo-Polynesian (Blust 1977). I have explained my preference for this hypothesis in Ross (1992). The hypothesis entails the assumption that PAN was spoken on Taiwan (Blust 1985, Bellwood 1985) or at a neighbouring location, and that the languages of the

---

various Formosan[2] groups probably have a continuous five-thousand year history on the island. The higher-level relationships between Formosan groups need further research. It is possible that they evolved by gradual dialect differentiation from the ancestral dialect network from which Proto Malayo-Polynesian broke away, but geographical and social separations have obviously occurred at various points in their history.[3]

In this paper I seek to reconstruct the verbal morphology of PAN, using evidence from Formosan languages to go beyond the work of Wolff (1973, 1980), Dahl (1976, 1978), Pawley and Reid (1980), Starosta, Pawley and Reid (1982), and Starosta (1992) and to reconstruct a fuller version of the verbal system. There are features of verbal morphology found in Formosan languages (and having scattered cognates in the languages of the Philippines and western Indonesia) which have received little or no attention in the reconstruction of PAN morphology. My reason for including them in recon-struction is not simply a pedantic concern with completeness, but rather that their inclusion

(a) helps us to understand the historical development of the verbal morphology of PAN itself;

(b) potentially allows us to identify shared innovations in the verbal morphology of various groups of Austronesian languages, and therefore to reconstruct with greater certainty the early prehistory of Austrone-sian dispersion.

---

2   I retain conventional usage among Austronesian linguists by speaking of Taiwan's Austronesian languages as 'Formosan', since the term 'Taiwanese' is used of a Chinese dialect.

3   These matters are discussed in Ross (1994).

2. Reconstruction

For the sake of clarity, I first present my reconstruction of PAN verbal morphology, then give supporting evidence for it where this has not already been provided by other scholars, and discuss its implications for the history of PAN. Finally I look briefly at its implications for post-PAN developments.

2.1. Morphosyntax.

In an attempt to avoid the confusions implicit in terms like "topic" and "subject" in the description of western Austronesian morphosyntactic systems, I will adopt Foley and Van Valin's term "pivot" to refer to what others have termed the "topic" in Formosan and Philippine languages. "The pivot of a syntactic construction is the NP [noun phrase] which is crucially involved in it; i.e., it is the NP around which the construction is built" (Foley and Van Valin 1984: 110). If there is a morphologically encoded relationship between one noun phrase and the verb phrase, then that noun phrase is the pivot.

The pivot in English is the subject. However, the term "subject" sits uncomfortably in describing PAN and many of its Formosan and Philippine daughter languages because the pivot in these languages behaves rather differently from the English subject. The PAN pivot choice system apparently had the following characteristics:

(a) the unmarked choice for pivot was the undergoer (UG) (in English it is the Actor (AC));

(b) the pivot was always definite (in English it is usually, but not always, definite);[4]

---

4   Himmelmann (1991) points out that in Tagalog the pivot is obligatorily referential, but not obligatorily definite. This means in practice that the pivot is almost, but not quite, always definite. The same may have been true of PAN.

(c) PAN had three "voices", which respectively allowed the UG, the AC, and the Location (LC) or Beneficiary to be pivot; it is just possible that PAN had a fourth voice allowing the Instrument (IN) to be pivot (English has two voices, active and passive, which respectively allow the AC and UG to be pivot; the passive of ditransitive verbs also allows Beneficiary as pivot);

(d) in a PAN relative clause the (deleted) noun phrase coreferential with its head noun *had* to be its pivot (English has no such requirement).

The effects of (a), (b) and (c) can be seen in (1), drawn from a text in the Formosan language Paiwan.[5] Paiwan preserves the putative PAN system reasonably well. In English, the action sequence is best translated with a sequence of active, i.e. AC pivot, verbs ("loosened... saw... crushed... ate"), but in Paiwan the normal choice is a sequence of UG pivot clauses (under-lined). The passage is semi-literally translatable into English as "That monkey, the stones were loosened (by him) the water became muddy, the crabs were seen (by him), and (they) were crushed (by him) and (they) were eaten (by him).'

(1) Paiwan

a	zu'	a	ti	sa	ḍaiḍail	cəkaɬən	a	zu'	a	qaciɬai,
a	zua	a	ti	sa	ḍaiḍail	cəkaɬ -ən	a	zua	a	qaciɬai
PV	that	LIG	PV	RESPECT	monkey	loosen-UP	PV	that	LIG	stone

maɬimək	a
ma -ɬimək	a
PASS-mud	PV

---

zaɬum, pacunan   a   zuʼ   a   gaŋ,   qucəqucən   sa

zaɬum pacun-an   a   zua   a   gaŋ,   R-quc-ən   sa

water   see   -UP   PV   that   LIG   crab   DUR-crush-UP   and.then

kani   aya.

kan-i   aya

eat-UP.AT   thus

     "That Mr Monkey, he loosened some stones, the water became muddy,

he saw the crabs, and crushed and ate them.' (Paiwan 44)[6]

---

6   Sources which consist only of the language name and a number refer to published texts for which I have done a computer-aided interlinearisation. The texts are from the following sources: Saisiyat, Tsuchida (1964); Atayal, Egerod (1969); Seediq, Asai (1953); Pazeh, Ferrell (1968), Thao, Li et al. (1956); Paiwan, Egli (1990); and Puyuma, Tsuchida (1980).

Abbreviations used in interlinear glosses are as follows:

1,2,3   first, second, third person

1EP   first person plural exclusive

1IP   first person plural inclusive

AP   actor pivot marking on verb

AT   atemporal verbal suffix

CJ   conjunction

CM   construction marker

CS   causative prefix on verb

D   disjunctive pronoun

DUR   durative aspect

GEN   genitive CM or pronoun

IJ   interjection

IP   instrument pivot marking on verb

LIG   ligature joining head and attribute

LP   locative pivot marking on verb

NEG   negative auxiliary

NOM   nominaliser

NPV   non-pivot CM or pronoun

P plural pronoun

PASS   passive prefix on verb

PF   perfective aspect

The semi-literal English translation reads poorly because a major function of the English passive (= UG pivot structure) is to suppress the AC, a function not shared by the Paiwan UG pivot structure.

There were a number of circumstances where something other than UG was the normal pivot choice. It follows from (b) above that if the UG was indefinite and the AC definite, then the clause had an AC pivot. It follows from (d) above that if the deleted noun in a PAN relative clause was AC, then the clause had an AC pivot verb. Similarly, when a clause is juxtaposed or subordinated to an immediately preceding one, then the pivot of the first clause often serves as the pivot of the second and determines whether the latter is AC or UG or something else.[7]

The characteristics of the PAN pivot choice system listed above may reasonably be reconstructed for it because they are reflected in Formosan languages which largely retain PAN verbal morphology (Saisiyat, Atayal, Seediq, Paiwan, Puyuma, Siraya, Bunun),[8] in Tagalog and other Philippine languages discussed in the literature,[9] and in Malay/Indonesian and other

---

PJ   projective mood

PV   pivot CM or pronoun

R reduplication

RECIP   reciprocal

S   singular pronoun

UP   undergoer pivot marking on verb

VOC   vocative

7   See Levinsohn (1991) for discussion of this characteristic in Malay.

8   For Saisiyat, Atayal, Seediq, Paiwan and Puyuma, this assertion is based on analysis of the texts in Tsuchida (1964), Egerod (1969), Asai (1953), Egli (1990) and Tsuchida (1980) respectively; for Siraya, on notes made by K. A. Adelaar on his analysis of the seventeenth-century text of Matthew's Gospel; and for Bunun on examples and commentary in Jeng (1977).

9   See Naylor (1975), Hopper (1979b), Wouk (1986) on Tagalog, Cooreman, Fox and Givón (1988) on Chamorro and Tagalog.

western Indo-Malaysian languages.[10]

The basic constituents of PAN pivot structure were that:

(a) the role of the pivot (AC, UG, LC) was marked by an affix on the verb;

(b) the pivot received special marking: a pivot noun phrase was introduced by a special pivot-marking morpheme, and a pivot pronoun has a special form.

Thus in (1), *cəkaɬ-ən* 'loosen-UP', *pacun-an* 'see-UP', *qucə-quc-ən* 'IMPF-crush-UP' and *kan-i* 'eat-UP.AT' each have a suffix (*-ən* 'UG pivot', *-an* 'UG pivot unaffected by action',[11] or *-i* 'UG pivot/subjunctive verb') indicating that the pivot of the verb has the UG role, and UG noun phrases are marked with the Paiwan pivot-marking morpheme (PV) *a*. The last two verbs in (1) have no explicit pivot, as this is understood to be the pivot of the previous verb, *a gaŋ* 'the crabs'.

The (artificial) examples in (2) illustrate the basic affixes for the four pivot structures in Paiwan. The four examples contain respectively the suffix *-ən* 'UG pivot', the infix *<əm>* 'AC pivot' the suffix *-an* 'LC pivot' and the prefix *si-* 'IN pivot' In each case the pivot noun phrase introduced by *a*

---

10 The fact that *di* clauses, i.e. UG pivot clauses, are the unmarked choice for transitive verbs in action sequences in Malay/Indonesian narratives has received attention from Hopper (1977, 1979a, 1979b, 1983), Verhaar (1978, 1988), McCune (1979), and Levinsohn (1991). Hopper (1979b) makess a similar observation for Old Javanese, Wouk (1986) for Toba Batak, Brewis and Levinsohn (1991) for Timugon Murut. Cumming (1986, 1988) shows how Malay structure is shifting away from the earlier pattern.

11 The suffix *-an* usually marks LC or BN pivot in Paiwan, but in certain verbs where the UG is not affected by the action, in this case the 'thing seen', it marks UG pivot. It is not difficult to see that this is derived from an earlier stage whereby the thing seen was construed as a LC (*where* the AC looked) rather than an UG.

assumes the role indicated by the verbal affix.

(2) Paiwan

a. təkəl-ən   a   vaua

   drink-UP PV wine

   'the wine will be drunk' ('s/he/they will drink the wine')

b. t<əm>kəl   a   qała

   <AP>drink PV stranger

   'the stranger will drink (something)'

c. təkəl-an   a   kakəsan

   drink-LP PV kitchen

   'the kitchen will be drunk in' ('s/he/they will drink it/them in the

   kitchen')

d. si-təkəl   a   kupu

   IP-drink PV cup

   'the cup will be drunk with' ('s/he/they will drink it/them from a

   cup')

In natural discourse, verbs in Formosan and Philippine languages often have no noun phrase accompanying them (like the last two verbs in (1)) or only one (like the other verbs in (1)). Verbal clauses with two full noun phrases are rare, and clauses with more than two are almost never found. In many Formosan and Philippine languages the relationship of each noun phrase to the verb is indicated by a "construction marker" (CM) preceding the noun phrase. The CMs of most languages make a three-way distinction between the pivot (PV), the genitive (GEN), and other non-pivot (NPV) noun phrases. In the examples in (2) the CM a marks the pivot. The examples in (3) illustrate the other CMs in Paiwan.

(3) Paiwan

    a. təkəl-ən nua qała   a   vaua

       drink-UP GEN stranger PV wine

       'the wine will be drunk by a/the stranger' ('a/the stranger will drink the wine')

    b. t<əm>kəl a   qała   tua vaua

       <AP>drink PV stranger NPV wine

       'the stranger will drink wine'

    c. təkəl-an nua qała   a   kakəsan

       drink-LP GEN stranger PV kitchen

       'the kitchen will be drunk in by a/the stranger' ('a/the stranger will drink it/them in the kitchen')

    d. təkəl-an a   kakəsan tua   vaua

       drink-LP PV kitchen NPV wine

       'the kitchen will have wine drunk in it' ('someone will drink wine in the kitchen')

Where, as in (3a) and (3c), the AC is mentioned with a verb whose pivot is not AC, the genitive CM *nua* is used. I use the term "genitive' here because the same CM is used to mark the possessor: *ałak nua qała* 'the stranger's child'. Where, as in (3b) and (3d), a noun phrase occurs which is neither pivot nor AC, then it is marked with the non-pivot CM *tua* (if it is Loc, the preposition *i* is preposed: *i tua kakəsan* 'in the kitchen').

The three-way distinction among noun phrases marked by pivot, genitive and non-pivot CMs seems to have existed in PAN. As in the examples in (3), PAN noun phrases followed the verb (unless one was topicalised to clause-initial position). Starosta, Pawley and Reid (1982: 149) argue that the genitive-marked AC normally followed the head verb of its clause, "since

otherwise it could be interpreted as Genitive attribute of the noun preceding it."

There has been considerable discussion in the literature as to whether Formosan and Philippine languages with the kind of structure described here should be regarded as "ergative" languages.[12] I avoid using the term "ergative" here because it has been used in differing ways by various scholars. In the classical definition proposed by Dixon (1979), there is a distinction between accusative and ergative languages. An accusative language is one in which, like English, the single core noun phrase of an intransitive clause (Dixon's S) and the AC noun phrase of a basic transitive clause (Dixon's A) receive the same morphosyntactic treatment. In an ergative language like the Australian language Dyirbal, however, it is the UG noun phrase of a basic transitive clause (Dixon's O) which receives the same morphosyntactic treatment as Dixon's S. In the terms used here, in an accusative language the pivot of a basic transitive clause is AC, and in an ergative language it is UG. In both cases, it is assumed that the pivot of a basic transitive clause receives the same morphosyntactic treatment as the pivot of an intransitive clause.

The difficulty of applying these criteria to a language like Paiwan is evident. In English we take an active (AC pivot) clause like *The stranger is drinking the wine* to be a basic transitive clause, and regard its corresponding passive (UG pivot) *The wine is being drunk by the stranger* as derived because it is morphosyntactically more complex. Hence the pivot of the basic transitive clause is AC, and English is by definition an accusative language.

---

12 See, for example, Gerdts (1988) on Ilokano, De Guzman (1988) on Kapampangan, Cooreman, Fox and Givón (1988) on Chamorro and Tagalog, Shibatani (1988) on Philippine languages, and Himmelmann (1991) on Tagalog. Cartier (1976), Hopper (1988) and Verhaar (1988) raise similar issues with regard to Indonesian.

Furthermore, the morphosyntactic treatment of the English AC pivot matches that of Dixon's S in *The stranger is crying*. In Paiwan, however, there is no obvious difference in morphosyntactic complexity between the UG pivot and AC pivot structures of (2a) and (2b); that is, there are no morphosyntactic grounds for treating one structure as basic and the other as derived. One approach is to say is that the AC pivot receives the same morphosyntactic treatment as Dixon's S, in that intransitive verbs in Paiwan are marked in the same way as AC pivot verbs: *q< əm> au ŋ a qala* 'the stranger is crying'. On this analysis Paiwan is morphologically accusative. An alternative approach (Gibson and Starosta 1990) says that, since the AC pivot verb has the same morphology as the intransitive, the AC pivot verb is intransitive and the UG pivot verb is the canonical transitive.[13] On this analysis Paiwan is morphologically ergative. Both analyses seem to make much weaker assertions than we make when we say that English is an accusative language.[14]

By a more recent definition (Cooreman, Fox and Givón 1988) a language is said to be accusative if AC pivot clauses are the normal choice for transitive event clauses in a narrative, and ergative if UG pivot clauses are the normal choice.[15] By this criterion, Paiwan is "discoursally" ergative.

There are yet other criteria for ergativity which I will not discuss here.

---

13   Under this approach, the noun phrase marked as NPV of an AC pivot verb is an oblique. This analysis seems to leave pivots other than AC and UG somewhat out on a limb.

14   Although Formosan languages are morphologically accusative, the same is not true of Philippine languages. Some Tagalog intransitive verbs have the same morphology as AC pivot transitives, and others have the same morphology as UG pivot transitives. Hence we would have to say that Tagalog is morphologically a 'split-S' language.

15   See Cumming and Wouk (1987) for a critique of this formulation and a re-examination of 'discourse ergativity' in Tagalog and Malay/Indonesian.

The very fact that Paiwan could be described as morphologically accusative but discoursally ergative - and the same was apparently true of PAN - is sufficient to show the difficulties of using the term.

In this section I have described the morphosyntactic system of Paiwan. A similar system occurs in Atayal, Seediq, Saaroa, Bunun, Puyuma, and Siraya, and the systems of other Formosan and Philippine languages resemble it quite closely. Hence it appears to be the morphosyntactic system of PAN. However, there is a danger implicit in this assumption. Morphosyntactic systems (as opposed to the morphemes themselves) can readily be transferred from one language to a neighbour as the result of calquing by bilingual speakers. The Bel group of Austronesian languages on the north coast of Papua New Guinea, for example, has moved a long way towards the adoption of the morphosyntactic system of neighbouring Papuan languages, acquiring a system of clause-chaining and a distinction between sentence-medial and sentence-final verbs which is unknown elsewhere in Austronesia. However, although the system is quite Papuan, the morphemes of which it is composed are Austronesian forms which have undergone a change of function as speakers have increasingly calqued on the system of the Papuan language which they also speak (Ross 1987). So we cannot ignore the possibility that the morphosyntactic system we are attributing to PAN may have been acquired by a number of daughter-languages through calquing. In order to be reasonably sure that it was the system of PAN itself, we need to be able to reconstruct the morphemes of which the PAN system was composed, and it is to this task which I now turn.

Table 1: PAN pivot, mood and aspect morphemes

Key:

√   verb root

<X> X is infixed, normally after the root-initial consonant

-X   X is suffixed to the root

R   reduplication: Ca-√, where C is a consonant identical to the

        root-initial consonant

Bracketed (*√-a) represents an alternant to be discussed below.

	Actor	Undergoer	Location
INDICATIVE			
Neutral	<um>√	√-ən	√-an
	*k<um>áRaw	*kaRãw-ən	*kaRãw-an
	*k<um>aRãC	*kaRaC-ə́n	*kaRaC-án
Perfective	<umin>√	<in>√	<in>√-an
	*k<um><in>áRaw	*k<in>áRaw	*k<in>aRãw-an
	*k<um><in>aRãC	*k<in>aRãC	*k<in>aRãC-an
Durative	<um>R-√	R-√-ən	R-√-an
	*k<um>a-kãRaw	*ka-kaRãw-ən	*ka-kaRãw-an
	*k<um>a-kaRãC	*ka-kaRaC-ə́n	*ka-kaRaC-án
NON-INDICATIVE			
Atemporal	√	√-u (√-a)	√-i
	*kãRaw	*kaRãw-u	*kaRãw-i
	*kaRãC	*kaRaC-ú	*kaRaC-í
Projective	<um>√-a	√-aw	√-ay
	*k<um>aRãw-a	*kaRãw-aw	*kaRãw-ay
	*k<um>aRaC-á	*kaRaC-áw	*kaRaC-áy

2.2. Pivot, Mood and Aspect Morphology.

The reconstructions in this section are based on material from fifteen Formosan languages and various extra-Formosan languages. The verbal forms representing the various combinations of pivot, mood and aspect in the fifteen Formosan languages are set out schematically in the appendix to this paper, together with the sources consulted to ascertain these forms.

Table 1 sets out the pivot, mood and aspect morphemes of PAN in schematic form, together with their applications to two PAN roots, *kāRaw 'scratch' and *kaRāC 'bite'. These root-and-morpheme combinations are intended only to illustrate the structure of PAN verbal forms: there is no guarantee that these forms all occurred, as many verbs in Formosan languages have defective paradigms. These roots represent the two PAN stress types. PAN *kāRaw is a paroxytone root, i.e. a root with penultimate stress, *kaRāC an oxytone, i.e. with final stress. Infixes do not cause stress-shift, but the suffixes were probably all what Zorc (1978: 92) calls "same-accent" suffixes, that is, after suffix-addition a paroxytone remains a paroxytone, and an oxytone remains an oxytone. The outworking of this is that stress shifts one syllable to the right. On Zorc's Philippine evidence, *-ən, *-án, *-a, and *-i are all same-accent suffixes (Zorc 1977: 64), and Tsou confirms this for *-a and *-i (Ross 1992). To date no reflexes of *-aw, *-ay or *-u have been found in languages which are criterial for reconstructing stress, and I have simply assumed that the pattern covers all affixes in the paradigm (I suggest below that there is an ancient paradigmatic relationship between the atemporal and projective forms, so this assumption is not unmotivated).

There appear to have been four major formal classes of verb in PAN:

(a) Those like *kāRaw and *kaRāC above, which took AC pivot infixa-

tion of *<um> into the root;

(b) A small class of verbs whose AC pivot (and sometimes other pivot) forms had no affixes;

(c) Verbs whose root began with *pa- and whose AC pivot forms began with *ma-, derived historically from *<um> + *pa-, e.g. AC pivot neutral *maCāy 'die', AC pivot atemporal *paCāy 'die'. Many of these verbs are complex roots formed with the prefix *pa- 'causative'.

(d) Verbs similar to those in (3), but whose root began with *ka- and whose AC pivot forms began with *ma-, derived historically from *<um> + *ka-. Many of these verbs are complex roots formed with the prefix *ka- (perhaps 'inchoative').

The three pivot categories in Table 1 require little further explanation. The AC pivot infix is reconstructed *<um>: this is the form reflected most widely in Formosan and Philippine languages. Amis, Puyuma and Paiwan have <əm>,[16] and Thao <m>, but this probably results from the fact that the AC pivot infix was always unstressed in PAN. Atayal has <m>, the result of regular vowel deletion rules (Egerod 1965, Li 1980).

The LC pivot also seems to have served as a Benefactive pivot in PAN, as it does in a number of daughter languages. In other words, with semantically appropriate verbs, a human LC pivot was interpreted as Benefactive, as in these examples:

(4) Paiwan

uri   ku-su-pavay-an   tua   kakudan

FUTURE GEN:1S-PV:1S-give-LP NPV power

'I will give you power' (Egli 1990: 296)

---

16   For Paiwan, Ferrell (1982) writes <m>, Egli (1990) <əm>. Since the latter matches the phonotactics of the language, I assume that it is phonologically more appropriate.

(5) Sediq

    skat-an-i-<u>ku</u>   qəhuni

    cut-LP-AT-PV:1S tree

    'Please cut the tree for me!' (Asai 1953: 46)

The aspect and mood categories "Neutral", "Perfective", "Durative", "Projective", and "Atemporal" used in Table 1 and in the Appendix require some explanation. On both formal and semantic grounds, these five categories are divisible into two higher-order classes, "indicative" and "non-indicative". I return to the division in form below. "Neutral" refers to the finite indicative verb form not marked for tense or aspect. Its functional range depends on the availability of other forms in the language, e.g. on whether there is a special durative form or whether the neutral form functions as both punctual and durative. "Perfective" refers to the finite indicative form used for completed events. "Durative" is used for events viewed as ongoing at some point of time. "Projective" refers to the finite verb form used to express intention, possibility and exhortation, as in these Atayal examples:

(6) Atayal

    a. srx-au-su?   sqani qhoniq

       make.stand-UP.PJ-GEN:2S here   tree

       'You might place the tree here.' (Egerod 1966: 355)[17]

    b. spŋ-au-ta?   lukus   su?

       measure-UP.PJ-GEN:1IP clothes GEN:2S

       'Let's us measure your clothes (sometime)' (Egerod 1965: 277)

    c. p-nbu-au   su?   laqi?

       CS-be.sick-UP.PJ GEN:2S child

       'Don't let your child be sick!'

---

17   Egerod notes this as 'first passive', i.e. LC pv, but this appears incorrect.

d. gal-a-ku?

   choose-AP.PJ-PV:1S

   'Let me choose' (Egerod 1965: 353)

e. sŋŋu-ai-su?   na?   utux

   frighten-LP.PJ-PV:2S GEN spirit

   'You might be frightened by the spirits.' (Egerod 1966: 354)

f. lg-ai-maku?

   accompany-LP.PJ-GEN:1S

   'I will come along' (Egerod 1965: 279)

Atemporal forms have three basic functions in daughter languages (and often have all three functions in the same language). Firstly, they function as plain imperatives. Secondly, they express events in sequence in narrative. Thirdly, they are the forms which occur subordinate to some auxiliaries.

The following examples from Paiwan illustrate the second usage. The first verb has the neutral form, and verbs following it have (apparently optionally) the atemporal form. Another example occurs in (1) above.

(7) Paiwan

a. ribu-in   sa pa-d^{y}ulu-i

   defeat-UP CJ CS-be.simple-UP.AT

   'He defeated and pacified it [i.e. the village]' (Egli 1990: 226)

b. kiqənəc-an sa pa-pa-piriq-i

   look.at-UP CJ RECIP-R-divide-UP.AT

   'He looked at and divided it.' (Egli 1990: 242)

c. vuɬuq-ən sa ka-d^{y}aməq

   spear-UP CJ AT.PASS-hit

   He speared it and it was hit.' (Egli 1990: 226)

The (7c) contains the verb *ka-d^{y}aməq*, the atemporal form of the Paiwan

passive. The passive is unique to Paiwan, and it formally resembles an AC pivot verb of the *ka-* class: its neutral form begins with *ma-*, its atemporal with *ka-*. Atemporal verbs in narrative sequences are also common in the Dusunic languages of Sabah (Kroeger 1991, Miller and Miller 1991).

The third usage of atemporal forms is in subordination to some auxiliaries. As Starosta (1985, 1988, Starosta, Pawley and Reid 1982) has shown, Formosan (and some extra-Formosan) languages make considerable use of sentence-initial auxiliaries, also called "pre-verbs" by some scholars, which carry information on aspect, time, negation, manner, location and so on.

(8) Atayal

    a. ini?-saku?   hŋu?   qsia? lukus

       NEG-PV:1S   AP.AT.soak water clothes

       'I have not soaked the clothes in water' (Egerod 1965: 273)

    b. ini?-sami   kac-i   na? mqu?

       NEG-PV:1EP bite-LP.AT GEN snake

       'We have not been bitten by snakes.' (Egerod 1966:354)

    c. laxi   zŋ-i   snon-an-maku?   isu?

       PROHIB forget-LP.AT message-OBLIQUE-GEN:1S D:2S

       'You must not forget my message.' (Egerod 1966:358)

    d. si-nha?   sr?ag-i   ma   ai.

       ACTUAL-GEN:3P go. along-LP.AT it.is.said IJ

       They were following (the river). (Atayal 032)

The first word in each example above is an auxiliary and the main predicate of the sentence, and in these cases (although in by no means all Atayal sentences beginning with an auxiliary) the subordinate verb is atemporal. In (8a) *hŋu?* is the AC pivot atemporal form (cf *h<m>ŋu?* AC pivot neutral). In the other three examples the subordinate verb is a LC pivot atemporal

marked with -*i*.

PAN auxiliaries have played an important role in the development of the morphosyntax of daughter languages. The sixteen paradigms in the appendix show considerable systemic variety. As Starosta (1985) shows, in Tsou, all aspect and mood functions have been taken over by the auxiliary system with the results (i) that the Tsou paradigm distinguishes only between the four pivots and (ii) that the surviving UG and LC pivot forms are the forms of the PAN atemporals. At some stage in the development of Tsou, all non-AC pivot verbs subordinate to an auxiliary became atemporals, and at a later stage *all* aspect and mood marking was taken over by auxiliaries, so that all verbs were subordinate, and the only non-AC forms to occur were atemporals.

The Tsou case is extreme, but the systemic variety in the appendix is largely attributable to the fact that the role of auxiliaries has increased in some languages and decreased in others. Puyuma, for example, with its paradigms of durative and future forms, makes little use of auxiliaries and uses atemporals only after the past negative auxiliary:

(9) Puyuma

    adi   ku   dirus-i   na enai

    NEG GEN:1S wash-LP.AT PV water

    'I did not wash with that water.' (Cauquelin 1991: 49)

We saw from the examples in (3) that a verb-initial clause in Paiwan is followed by noun phrases marked as pivot, non-pivot or genitive (in some languages genitive and other non-pivot noun phrases form a single category). When these noun phrases take the form of pronouns, these pronouns are often encliticised to the verb (see Starosta 1988), as we see in (5), (6a), (6b), (6d), (6e), and (6f) above. But when the verb is subordinate to an auxiliary, the pronouns are encliticised instead to the auxiliary, as in (8a),

(8b) and (8d). A consequence of this has been that in some languages where auxiliary use has declined, some auxiliaries have disappeared, but the pronoun cliticised to the auxiliary has remained "stranded" and has instead become a proclitic to the verb. Hence we find languages like Atayal with numerous auxiliaries, and with pronouns enclitic to the verb or the auxiliary, and on the other hand Puyuma and Paiwan, with few auxiliaries, but genitive pronouns proclitic to the verb. Compare (10) with (9):

(10)Puyuma

    ku   da-dirus-aw   na guŋ

    GEN:1S DUR-wash-UP.PJ PV ox

    'I was washing the ox.' (Cauquelin 1991: 49)

A rather different change involving auxiliaries has occurred in Rukai. Here a process opposite to that in Tsou has occurred. The auxiliary was apparently downgraded from its status as main verb to that of mere adverb, and its subordinate verb instead became the main verb. As a result, pronouns were no longer cliticised to the auxiliary, and the auxiliary itself, now no more than a tense/aspect marking particle, became cliticised to the verb. Thus Rukai *wa*-'past' is derived from the non-AC form of the PAN auxiliary *ua* (AC pivot *m-ua*) 'past', or as a main verb 'come', reflected in the past auxiliaries Atayal *ua-l*, Sediq *wa-da*, Tsou *o/m-o*. Rukai *ay*- 'future' is derived from the non-AC form of the PAN auxiliary *aSa* (AC pivot *m-aSa*) 'past', reflected in the future auxiliaries Sediq *m-aha* and Amis *asa* and in the future prefixes Thao *ʔa*-and Saaroa *a*-. Thus a Rukai phrase like

(11a)Rukai

    wa-kanə-aku   'I ate'

    PAST-eat-PV:1S

    is derivable from a PAN sentence something like

(11b)*ua-Saku   kā?ən   'I ate'

    PAST-PV:1S eat

where *ua* is the PAN past auxiliary, *Saku* 'I' the enclitic pivot pronoun, and *ka?ən* 'eat' the subordinate verb. Over time, *ua* became a past-tense particle, *ka?ən* became the main verb and attracted the enclitic *Saku*.

2.3. Discussion.

There have been several previous reconstructions of PAN verbal morphology, and the one in Table 1 differs in various ways from all of these. They are tabulated here using the same conventions as in Table 1. The earliest was Wolff's (1973), which I return to below. Table 2 sets out the pivot morphemes proposed by Dahl (1976: 119), which correspond to the "Neutral" and "Atemporal" sets in Table 1.

Table 2: PAN pivot morphemes according to Dahl (1976)

	Actor	Undergoer	Location	Accessory
Neutral	<um>√	√-ən	√-an	Si-√
Imperative	√-a	...	√-i	...

The next reconstruction is by Pawley and Reid (1980), who set themselves the task of reconciling the very different verbal morphologies of the Formosan/Philippine languages and Oceanic languages. The system they arrive at, shown in Table 3, resembles that of Toba Batak. Implicit in their reconstruction, however, are very different subgrouping assumptions from those made here. Although they do not directly say so, their methodology presupposes that Formosan/Philippine and Oceanic are first-order subgroups of Austronesian. This explains their reconstruction of *-akən, a reflex of which is reconstructible in Proto Oceanic, and which is also widely reflected across Indonesia – but has no reflexes in Formosan or Philippine languages

and is assumed here to be a later local innovation. Pawley and Reid refer to the infixes < *um* > and < *in* > as AC pivot and perfective respectively, but it is not clear to me where they fit into the paradigm.

Table 3: PAN pivot morphemes according to Pawley and Reid (1980)

	Undergoer	Location	Accessory
Active	√	√-i	√-akən
Passive	√-ən	√-an	i-√

The reconstruction in Table 3 was quickly superseded by that in Table 4, much closer in form to mine in Table 1.

Table 4: PAN pivot and aspect morphemes according to Starosta, Pawley and Reid (1982)

	Actor	Undergoer	Location	Accessory
Neutral	<um>√	√-ən	√-an	iSi-√
Perfective	<um><in>√	<in>√	<in>√-an	<in>iSi-√
Imperative/subordinate	<um>√?	√-a	√-i	(√-akən)

The suffixes *-a* and *-i* were recognised as serving two of the three functions described above for atemporal verb forms. However, whereas Table 4 has *-a* in the UG slot, Table 1 has it in the AC slot. The suffix *-akən* was regarded as doubtful because it did not appear in the Formosan and Philippine languages, whose significance for reconstruction has increased. The IN prefix *iSi-* was awarded an initial *i- to account for Philippine forms which lacked expected initial *h-. However, as Starosta, Pawley and Reid point out themselves, there are Philippine reflexes with *h when the IN prefix is preceded by another prefix. Hence loss of initial *h can be attributed to that cross-linguistically common phenomenon, *h*-dropping.[18]

---

18   Dahl (1986) also presents evidence for *Si- rather than *iSi-.

Starosta's most recent revision of this scheme is shown in Table 5. The addition of *-*i* as an alternant in the UG slot, is motivated by internal reconstruction from within the reconstructed PAN system itself rather than by standard comparative reconstruction. However, inspection of the appendix shows *-*i* reflected here in Thao, Puyuma and Paiwan. I will return to Starosta's reconstruction of two Accessory (IN) forms below.

Table 5: PAN pivot morphemes according to Starosta (1992)

	Actor	Undergoer	Location	Accessory
Neutral	\<um>√	√-ən	√-an	Si-√, Sa-√
Imperative/subordinate	\<um>√	√-a, √-i	√-i	_

Interestingly, the reconstruction which matches mine most closely is the oldest, Wolff's. The resemblance does not have much to do with the span of years, however, but arises from the fact that we have both employed the classical comparative method. Wolff used materials from Atayal and Tsou (both Formosan), from Samar-Leyte Bisayan (Philippine) and from Javanese. The differences between my reconstruction and his are directly attributable to the use of different languages, which in turn reflects different subgrouping hypotheses.

Table 6: PAN pivot, mood and aspect morphemes according to Wolff (1973)

	Actor	Undergoer	Location	Instrument
Non-past	\<um>√	√-ən	√-an	i-√
Past	\<inum>√	\<in>√	\<in>√-an	\<in>i-√?
Future	...	R-√-ən	R-√-an	...
Dependent	√	√-a	√-i	√-an
Subjunctive	√-a	...	√-ay	...

Comparison of Table 1 with the tables above indicates several aspects of

my reconstruction which need justification. The areas of difference include:

(a) aspect/mood categories: did PAN have durative forms? future forms?

(b) the number of "voices": did PAN have IN pivot verb forms?

(c) the scope (and antiquity) AC pivot *<*um*>: did it occur in the non-indicative paradigm?

(d) non-indicative forms marking UG pivot: are * √ -*aw* 'projective' and * √ -*u* 'atemporal' reconstructible? and what is the place of * √ -*a* in the paradigm?

2.3.1. Aspect/Mood Categories.

The reconstruction of aspect/mood categories in Table 1 has been carried out by conventional means on the basis of the comparative data in the appendix. In the light of the reconstructions in Tables 2, 4, 5 and 6, the reconstruction of the neutral forms (other than the absence of the IN pivot form, discussed in section 2.3.2) is uncontroversial. Scholars who have reconstructed the perfective (Wolff calls it "past") have also agreed on its function and forms. Starosta's imperative/subordinate (Tables 4 and 5) and Wolff's "dependent" categories correspond functionally with my atemporal category, although there is some disagreement about the forms, to which I return in section 2.3.3. Wolff's "subjunctive" matches my "projective", and the forms here are also discussed in section 2.3.3.

Thus the one aspect/mood category reconstructed in Table 1 which occurs in no previous reconstruction is the durative, and the one category in others' reconstructions which is absent from Table 1 is Wolff's "future" (Table 6).

Among the fifteen languages in the appendix, durative forms occur in Thao, Kanakanavu, Saaroa, Rukai, Puyuma and Paiwan. In each case the dura-

tive is formed by reduplication, and except in Rukai its form is C_1a-C_1V... This is somewhat unusual (extra-Formosan single-syllable reduplication is usually C_1V_1-C_1V_1... or $C_1V_1C_2$-$C_1V_1C_2$...) and suggests that the form occurred in PAN. Even more unusual is the AC pivot durative pattern which occurs in Kanakanavu, Saaroa, and Puyuma with some verbs formed from a verbalising (derivational) prefix and a root. Here reduplication is replaced by the insertion of the vowel -a- between the verbaliser and the root. Hence we find forms like those from Puyuma (from Cauquelin 1991: 52-55):

(12) AC pivot neutral   AC pivot durative

ki-beray	ask for	*ki-a-beray*
ki-ləŋaw	listen	*ki-a-ləŋaw*
kua-luŋ	be sick	*kua?-a-luŋ*
mi-kipiŋ	get dressed	*mi-a-kipiŋ*

The strangeness of the form indicates that it is a shared inheritance - whether from PAN or from a lower-order proto language within Taiwan is not clear.

Of the languages in the appendix, Kavalan, Sediq, Thao, Saaroa, Rukai and Puyuma have a future category.[19] The Thao, Saaroa and Rukai forms are derived from an auxiliary, as mentioned above. The future in both Sediq and Puyuma is formed by reduplication. The Sediq pattern is $C_1ə$-C_1V ... (Asai 1953: 61), the Puyuma C_1a-C_1V ..., i.e. the same reduplicative pattern as in the durative.[20] I have taken this to mean that the future is not separately

---

19   Wolff (1973) gives examples implying that Atayal, like its close relative Sediq, has a future formed by reduplication. Egerod (1965) also records a 'future', but it is not Wolff's. It is in fact a particular use of the atemporal form after the auxiliary which expresses prohibition.

20   According to Sung (1969), Kanakanavu also has a future with (on my interpretation of her data) the reduplicative pattern C. She does not mention a durative

reconstructible in PAN, but rather that the durative pattern has been functionally extended to cover the future (as in English *I'm going tomorrow*).

2.3.2. The Rise of PAN Indicative Forms.

Questions (b) and (c) above are both bound up with the rise of the PAN indicative/non-indicative division, and it is to this that I now turn.

It has long been observed that in Formosan and Philippine languages the morphemes which occur in indicative verb forms are also used to form nominalisations. Thus in Paiwan we find the following derivations from the verb root *kan* 'eat' (Ferrell 1982: 17, 106):

(13)	verb form	nominalisation
k<əm>an	AC pivot neutral	'eater', 'someone who eats'
kan-ən	UG pivot neutral	'food', 'something to be eaten'
k<in>an	UG pivot perfective	'consumed food', 'something eaten'
kan-an	LC pivot neutral	'place where one eats'
si-kan	IN pivot neutral	'eating utensil', 'something to eat with'

Pawley and Reid (1980) suggested that the nominalisations were derived from their "passive" verb forms (Table 3), but Starosta, Pawley and Reid (1982) argue – convincingly, I think – for the opposite derivation. This entails a series of diachronic steps somewhat as follows (this is my interpretation, and Starosta et al. should not be held responsible for it):

(a) The neutral verb forms were originally those of the atemporal series.

(b) As in modern Formosan languages, attributes were connected to their nouns by a "ligature".[21] The attribute preceded or followed the head noun in accordance with discourse constraints. Hence in Paiwan

---

usage, Mei (1982) does not mention it at all. It is not included in the Appendix, as Sung's presentation does not show how it co-occurs with other morphemes.

21   The ligatures of most Formosan languates are formally similar to construction markers in the same languages, and presumably derived from them.

(14) a aɬak a vavaian

PV child LIG woman

either means 'the female child' (aɬak is the head) or something like 'the young female' (*vavaian* is the head) (Ferrell 1982: 13). Again as in the modern languages, a relative clause was formed with a noun or nominalisation as its predicate, and the relative clause was connected to its head noun with a ligature like any other attribute. Again from Paiwan:

(15) ku-kama a uqaɬay

GEN:1S-father LIG man

'the man who is my father' (Ferrell 1982: 32)

(16) inu, qali-an, a su-d^y <in>umak a gaŋ.

where friend-VOCATIVE PV GEN:2S-<PF>find LIG crab

'Where, my friend, are the crabs you found?' (Paiwan 047)

In (16), *su-d^y<in>umak* is a nominalisation 'something found of yours', used as a relative clause attribute of the noun *gaŋ* 'crab'.

(c) As a strategy of diathesis, i.e. of making the pivot a noun phrase refer-ring to a non-agent, a nominalisation was used as predicate. We may illustrate this using the Paiwan sentences in (2) and (3) above. (2a) is repeated below.

(2)a. təkəl-ən a vaua

drink-UP PV wine

'the wine will be drunk' ('s/he/they will drink the wine')

If *təkəl-ən* in (2a) is interpreted as a nominalisation, i.e. 'something to be drunk', then the example can be reglossed as:

(17) təkəl-ən a vaua

*drink-NOM PV wine

*'the wine is something to be drunk'

an interpretation which catches something of how such nominalisations were used in Pre-PAN. We turn now to the expansion of (2a) given as (3a) above:

(3)a. təkəl-ən   nua qała   a   vaua

    drink-UP GEN stranger PV wine

    'the wine will be drunk by a/the stranger' ('a/the stranger will drink the wine')

Again interpreting *təkəl-ən* as a nominalisation, we regloss the example as:

(18) təkəl-ən   nua qała   a   vaua

    *drink-NOM GEN stranger PV wine

    *'the wine is something of a/the stranger's to be drunk'

The fact that *təkəl-ən* in (2a) and (3a) and other non-AC pivot verbs are derived from nominalisations explains why their agent is expressed in the genitive in most daughter languages and evidently in PAN: it is the result of its earlier status as possessor of a nominalisation.

(d) This once highly marked strategy became decreasingly marked until the nominalisations were reinterpreted as verb forms and ousted the original neutral verb forms from main and relative clauses, leaving them as atemporals in imperatives, in narrative sequences where nominalisation was discoursally inappropriate, and in other subordinate clauses. However, the same morphemes continued (and continue) to be used to form nominalisations, with the result that sentences like (2a) and (3a), at least when taken out of context, can be ambiguous in modern Formosan and Philippine languages.

The verbs in (2a) and (3a) are UG pivot forms, but similar considerations apply at least to LC pivot forms. Example (2c), *təkəl-an a kakəsan* 'the kitchen will be drunk in' ('s/he/they will drink it/them in the kitchen'), is

derived from 'the kitchen is the place of drinking'.

If all five of the morphemes which offer themselves as candidates for reconstruction both as nominalisers and as verbal pivot morphemes (*<um>, *<in>, *-ən, *-an, *Si-) were used in both functions in PAN, then we should expect to find them evenly distributed in both functions in the modern languages. But Table 7 shows that this expectation is not supported. The table sets out the functions performed by reflexes of these five PAN morphemes and of a sixth, *Sa-, in eleven of the fifteen Formosan languages considered in this paper. Pazeh, Thao, Bunun and Siraya are omitted because the data available to me do not provide sufficient information on derivational morphology.

Table 7 shows these forms to have a rather patchy distribution among Formosan languages, and when it is compared with such neat tabulations as Ferrell's (1979: 201) of Formosan verbal pivot morphemes, one is left wondering whether the sources on which Table 7 is based are simply littered with gaps, or whether Ferrell's work was based on the assumption that all Formosan languages would fit into the Philippine four-pivot mould. The latter seems possible when it is observed that the Rukai morpheme sequence *ta-√-an* which Ferrell cites as LC pivot is analysed only as a nominalisation, not as a verb form, by Li (1977a: 202).

The patchiness is reduced a little if it is recognised that the two uses of these morphemes (as nominalisers and as verbal pivot markers) are mutually reinforcing. For reasons discussed above, Tsou and Rukai have lost the pivot marking function of PAN *<in>, *-ən, and *-an, and Rukai has also lost *<um>. It is not surprising, therefore, that they have also lost *<in> and *-ən in their nominalisation function.[22]

---

22   Starosta (1985) asks why the nominalising function of the morphemes has been

Taking this into account, Table 7 shows a fairly consistent distribution of PAN *<*in*>, *-*ən*, and *-*an* in both their nominalisation and pivot marking functions, and both functions may reasonably be reconstructed for PAN:[23]

(19)		verb form	nominalisation
*<*in*> √		UG pivot perfective	'something √-ed'
*√-*ən*		UG pivot neutral	'something to be √-ed'
*√-*an*		LC pivot neutral	'place where one √-s'

However, when we try to reconstruct a PAN IN pivot morpheme and its corresponding instrument-forming nominaliser, we are confronted by a problem. *Si-*, the usual candidate for the IN pivot marker, is unambiguously reflected in this function in Table 7 only in Saisiyat, in Paiwan, and in the Mayrinax dialect of Atayal (it also occurs in Thao and Bunun). It is reflected in its nominalisation function in Kavalan, Kanakanavu, Puyuma, Paiwan and in the Mayrinax dialect of Atayal. But as Dahl (1978, 1986) recognised, reflexes both in and outside Taiwan also compel us to reconstruct a form PAN *Sa-*. Its reflexes in Amis and Rukai form instrument nominalisations, and it has also acquired a verbal pivot marking function in Saaroa (where *saa-* may mark IN pivot – the data are insuffcient to tell) and in Malagasy (where it marks *inter alia* IN pivot). We cannot tell directly whether Sediq

---

lost in Tsou, as well as their verbal pivot marking function. This paragraph provides a possible answer.

23   Table 1, however, also shows the infix *<*in*> as the verbal perfective marker co-occurring with *<*um*> and* -*an* in the perfectives respectively of the AC and LC pivot verb forms. There is evidence that at least the LC perfective form may also have been derived from a nominalisation, since Puyuma shows a contrast between perfective and non-perfective nominalisations in *ka-kərut -an* 'place to be dug' and *k<*in*> ərut -an* 'place dug'. However, evidence from one language is certainly not sufficient for reconstruction.

Table 7: Morphemes used to form nominalisations and verbal pivot forms

PAN	*<um>√	*<in>√	*√-an	*<in>√-an	*√-an	*Si-√	*Si-√	*Sa-√
Kavalan	AC	PF	...	...	non-AC	...	instrument	...
Saisiyat	AC	PF	UG	s. t. to be√-ed: ka-√-an	place of √-ing; <in>√-an	instrument	IN	...
Atayal	AC	PF	UG	s. t. to be√-ed	place of √-ing	IN	instrument	...
Sediq	AC	PF	UG	s. t. to be√-ed	place of √-ing	IN	IN <---> have N: s-N <---> have N: s-N	IN
Amis	AC	s. t. √-ed	UG	...	place of√-ing: R-√-an	IN	have N: si-N	IN / instrument
Tsou	AC	...	...	...	place of√-ing	...	...	...
Kanakanavu	AC	PF	UG	...	place of√-ing: ta-√-an; s. t.√-ed:√-an	instrument	...	...
Saaroa	AC	PF	...	...	place of√-ing	...	'Special pivot'	...
Rukai	...	...	...	...	place of√-ing: [ta-]√-an; s. t.√-ed: <in>√-an	have N: si-N	...	instrument
Puyuma	AC / √-er	...	...	...	place√-ing will occur: R-√-an; place√occurred: <in>√-an	instrument	...	...
Paiwan	AC / √-er	PF	UG	s. t. to be√-ed	place of√-ing	IN	instrument, reason...	have quality of N

Note: <---> indicates that morphemes cannot unambiguously be assigned to *Si- or *Sa- because of vowel deletion in Sediq. Similar vowel deletion occurs in most Atayal dialects, but the Mayrinax Atayal dialect form si- shows that Atayal forms are descended from *Si-.

s-, which marks the IN pivot verb form, is derived from *Si- or from *Sa-, but may infer that it reflects *Si-, since the Sediq dialect chain is closely related to the Atayal chain, where the Mayrinax evidence points to *Si-.

What are we to make of this seeming confusion? It is clear from available descriptions that Formosan languages have a plethora of derivational prefixes (but far fewer infixes and suffixes) used to form nouns and verbs, and that PAN was probably similar. It is therefore eminently likely that *Si- and *Sa- were both PAN derivational prefixes, that each may have had several uses, and that these may have overlapped with each other. From reflexes in Sediq, Amis and Rukai noted in Table 7 we know that PAN *Si- was added to a noun N to make a verb meaning "have, possess, wear N". In the light of the diachronic rise of the indicative morphemes sketched above, it is likely that nominalisations formed with affixes other than PAN *<in>, *-ən, and *-an were used as nominal predicates as part of the diathesis strategy outlined in (c), and that these nominalisations had the potential to be reinterpreted as verbs. However, although we assume in the cases of *<in>, *-ən, and *-an that this reinterpretation had already occurred in PAN, there is no reason to infer that such reinterpretations stopped with the break-up of PAN. Daughter languages inherited two functions, nominalisation and verbal pivot marking, for the reflexes of *<in>, *-ən, and *-an, and this would have provided the matrix for the reinterpretation of other nominalisations as verbs in these languages.[24] The seeming confusion in the reflexes of

24 I am assuming in this paper that in PAN nominalisations and homophonous verb forms were morphologically identical but syntactically distinct. However, it is possible that there was also no syntactically relevant difference between them, as Himmelmann (1987: 73-78; 1991) claims for Tagalog. If this were the case, then we would expect that *Si- and *Sa- nominalisations would at least occasionally have been used as predicates (i.e. as event expressions) in PAN. But there is a

Si- and *Sa-* suggests that, whilst they were both nominalising prefixes at the break-up of PAN, they had not yet undergone reinterpretation as verbal pivot markers. This occurred separately in various daughter languages, by analogy with the dual function and potential ambiguity of *<in>*, *-ən*, and *-an*. Amis *sa-*, for example, occurs in some nominalisations which are perhaps on the way to reinterpretation as IN pivot verbs, but it cannot be labelled 'IN pivot', to judge from Chen's (1987) account or the comments and examples in Fey (1986: 372-375).

Future research may well show that *Si-* or *Sa-* (or both) is reconstructable as the PAN IN pivot morpheme, but the data available to me now do not support such reconstruction. This conservatism is reinforced by the fact that only two of the languages in the appendix, Atayal and Paiwan, have a complete set of IN pivot verbal forms based on *Si-* or *Sa-*. This is not especially surprising, since (i) the functional load of the IN pivot in languages which possess it is much lower than the loads of the AC, UG and LC pivots, and (ii) there is little Formosan comparative evidence at all on which to base the reconstruction of non-indicative PAN forms for the IN pivot. Indeed, the non-indicative IN pivot forms listed in the appendix appear to be derived from compounds of two morphemes.[25] This latter fact

---

complication in projecting the Tagalog situation onto PAN: Tagalog has no reflexes of non-AC pivot atemporals or projectives and therefore no forms which are unambiguously verbal rather than nominal, whereas PAN clearly had such forms and therefore had a distinction between nouns and verbs.

25   These compounds appear to be * √ -*an-ay*, reflected in Puyuma √ -*anay* 'IN pivot neutral' and * √ -*an-i*, reflected in Saisiyat √ -*ani* 'IN pivot atemporal' and Tsou √ -*[n]əni* 'IN/Benefactive pivot neutral'. It is possible, but not certain, that Puyuma and Paiwan √ -*an* 'IN pivot atemporal' also reflect * √ -*an-i* rather than plain * √ *an*. Atayal *anay s-* √ 'Circumstantial projective' and *an s-* √ 'Circumstantial atemporal' may reflect * √ -*an-ay* and * √ -*an-i* forms of a (now) zero auxiliary. Given the positions their reflexes occupy in the modern paradigms, * √ -an-ay can

suggests that when the (Pre-PAN) situation in (a) above prevailed, i.e. the neutral verb forms were those of the later atemporal series, there were only the three pivot categories of AC, UG and LC. Thus there was no pressure for the use of instrument nominalisations as predicates.

Table 7 reveals a situation with regard to PAN *<um> which is almost the opposite of that described for *Si- and *Sa-. Here we find ample evidence for the reconstruction of PAN *<um> as an AC pivot marker, but reflexes of its use in agent nominalisations are found only in Puyuma and Paiwan. In other words, the evidence in Table 7 suggests that, unlike *<in>, *-ən, and *-an, PAN *<um> was primarily the AC pivot marker, and *not* a nominalising affix. There are two facts which support this assertion. Firstly (see below), the reconstruction of the PAN AC projective form *<um> √ -a in Table 1 is well supported by the data in the appendix. If *<um> was part of a projective, i.e. a non-indicative form, then its verbal use belongs to stage (a) above: it predates stages (c) and (d), when nominalising morphemes became pivot markers. Secondly, if the strategy of diathesis outlined in (c) has been formulated correctly, i.e. it was a strategy for making a noun phrase referring to a *non-agent* into the pivot, then there was no reason for agent nominalisation to figure in the development described in (c).

We still need to account for the Puyuma and Paiwan reflexes of *<um> in agent nominalisations. Given the frequent use of nominalisations in relative clauses discussed in (b), however, it is not difficult to see that AC

---

be glossed 'IN pivot projective' and * √ -an[-i] 'IN pivot atemporal'. It may yet prove that these forms are reconstructible for PAN. For the moment, however, I will assume that these compounds arose in part of the post-PAN Formosan dialect network to fill a newly arisen paradigmatic gap. I have found no non-Formosan cognates.

pivot verbs in relative clauses were open by analogy to reinterpretation as agent nominalisations, i.e. the opposite process to that hypothesised for *<*in*>, *-*ən*, and *-*an*.

The discussion here also accounts for the formal properties of the indicative and non-indicative subsystems reconstructed in Table 1. The non-indicative subsystem consists of a fairly orderly paradigm of suffixes, as we might expect of the original verbal system. The indicative subsystem, on the other hand, is a rather untidy collection of infixes and suffixes (and prefixes if we include *Si-* and *Sa-*), an untidiness which is satisfactorily explained by the fact that these morphemes were drawn from different places and did not originally form a paradigm.

2.3.3. The Non-Indicative Forms.

Table 8 summarises the reconstructions of non-indicative forms from Tables 2, 4, 5 and 6 above and compares them with the reconstructions given in Table 1.

Table 8: Reconstructions of PAN non-indicative forms

	Actor	Undergoer	Location	Instrument
Dahl (1976)	√-a	...	√-i	...
Starosta et al. (1982)	<um>√?	√-a	√-i	(√-akən)
Starosta (1992)	<um>√	√-a, √-i	√-i	...
Wolff (1973)				
Dependent	√	√-a	√-i	√-an
Subjunctive	√-a	...	√-ay	...
Table 1 of the present paper				
Atemporal	√	√-u (√-a)	√-i	...
Projective	<um>√-a	√-aw	√-ay	...

I shall not pursue the reconstruction of IN pivot morphemes further here.

Supporting evidence for my atemporal reconstructions is set out in Table 9, for my projectives in Table 10.

It is convenient to examine the atemporal morphemes first. Evidence for Formosan languages is drawn from the appendix. Because the reconstructed *forms* of the non-indicative morphemes are potentially more controversial than those of the indicatives, I have added evidence from a number of extra-Formosan languages which reflect appropriate forms. However, I have not pursued extra-Formosan evidence systematically, and the languages are simply those for which evidence was immediately at hand.[26] The convention I have used in Tables 9 and 10 is as follows: if the language concerned reflects the form at the top of the column, then the pivot(s) marked by its reflex are indicated (AC, UG, LC). If a change in the aspect/mood has occurred (e.g. the Tsou reflex of $*\sqrt{}$ -*i* marks the neutral instead of the atemporal), then this is indicated. If no aspect/mood is indicated, then the reflex marks the atemporal. If a change in form has occurred, i.e. the reflex of *<um>* has been added to $*\sqrt{}$ -*i* in Bunun, then this is also shown. Otherwise the reflex has the expected form.

The majority of the evidence in Table 9 supports the reconstruction of PAN $*\sqrt{}$ -*i* 'LC atemporal' and Table 8 shows that scholars are agreed on this. The evidence is equally strong for reconstructing the plain stem PAN $*\sqrt{}$ 'AC atemporal', on which Wolff and I agree.

---

26   Sources of extra-Formosan material in Tables 9 and 10 are: Proto Bisayan, Zorc (1977: 118, 138, 247); Western Bukidnon Manobo, Elkins (1970: 51, 53, 56); Banggi, Boutin (1988: 42-43); Timugon Murut, Prentice (1971 219-224); Kimaragang, Kroeger (1988: 235-237, 1991:95); Tombonuo, King (1988: 169-170); Malagasy, Dahl (1976: 119-120, 1986); Proto Malayic, Adelaar (1992: 163-164).

Evidence for the reconstruction of PAN $*\sqrt{}$ -u 'UG atemporal' is weak, the more so as there is only one possible reflex known outside Taiwan (see below). However I reconstruct it here for two reasons. The first has to do with internal reconstruction. The non-indicative morphemes reconstructed in Table 1 form a pattern of two elements, shown in (20), the first of which distinguishes projective -a from atemporal zero, and the second marks the pivot. It is most unlikely that this pattern is the result of accident, and its existence gives strong support to the reconstruction of PAN $*\sqrt{}$ -u 'UG atemporal'.

(20)		Actor	Undergoer	Locative
	Atemporal	-ø -ø	-ø -u	-ø -i
	Projective	-a -ø	-a -u	-a -i

The second reason for reconstructing $*\sqrt{}$ -u is that it seems unlikely that the alternate candidate for UG atemporal, $*\sqrt{}$ -a, was both UG atemporal *and* AC projective, as Wolff reconstructs. However, other scholars have also reconstructed $*\sqrt{}$ -a in this slot, and, as Table 10 indicates, there is quite a lot of support for this. Indeed the support is strong enough to suggest that there was a form $*\sqrt{}$ -a with some kind of UG pivot atemporal function by the break-up of PAN. If there was a difference in function betwee $*\sqrt{}$ -u and $*\sqrt{}$ -a, then the evidence in Table 9 suggests that $*\sqrt{}$ -u was an imperative, whilst $*\sqrt{}$ -a had the other atemporal functions. However, in the one language outside Taiwan with a possible reflex of $*$-u, namely Lun Daye (Apo Duat group, Borneo), -u? and -a? mark an UG pivot imperative for actions respectively near to and at a distance from the speaker; it is just possible that this is an inherited contrast. I have no other explanation for this breech in the paradigm, except to observe that Tables 9 and 10 show functional drift from both AC and LC to UG, presumably resulting from the

fact that the UG was and generally still is the unmarked pivot choice.

Table 9: Supporting evidence for atemporal reconstructions

PAN	AC:*√	UG:*√-u	LC:*√-i
Kavalan	AC Future, Imperative	_	non-AC Imperative
Saisiyat	AC	_	UG
Atayal	AC	_	LC/UG
Sediq	AC	_	LC
Pazeh	_	_	LC/UG
Thao	AC	AC Request	UG
Bunun	_	_	AC Imperative:<um> √-i
Amis	AC	_	LC
Tsou	_	_	LC Neutral
Kanakanavu	UG, LC Neutral	_	LC Projective
Saaroa	AC	UG Imperative	LC Imperative
Rukai	AC Neutral	_	_
Siraya	_	_	_
Puyuma	AC Imperative	UG Imperative	LC/UG
Paiwan	AC	UG Imperative	UG
Proto Bisayan	AC	_	LC
W. B. Manobo	AC	_	LC
Banggi	AC	_	_
Timugon Murut	AC	_	LC
Kimaragang	AC	_	_
Tombonuo	_	_	LC
Malagasy	_	_	UG/LC
Proto Malayic	AC	_	LC Neut

In Table 9 Paiwan √-u is shown as UG atemporal, in spite of the fact

that Ferrell labels it AC (1982: 36). Available unambiguous examples show that it usually marks UG pivot, e.g. *guguc-u a ku-uquł* 'Scratch my back!' (Ferrell 1982: 97); *inəpətʸ-u a lavək* 'Sea be calm' (Egli 1990: 181); *kan-u a kataua* 'Eat the pawpaw' (Egli 1990: 181). However, AC pivot examples do occur (Egli 1990: 181). It appears that the effect of the imperative is to promote the unmentioned second person participant to pivot, resulting in an AC pivot interpretation for this form. The same change has perhaps affected the Thao reflex of *√-u*.

Of the projective forms in Table 8, my *-ay* 'LC pivot projective' agrees with Wolff's. The supporting evidence in Table 10 is quite strong.

Where other scholars have reconstructed *√ -a* 'AC pivot projective', I reconstruct *<um> √ -a* with the support of the Atayal, Bunun, Kanakanavu, Saaroa, Siraya and Puyuma reflexes. This reconstruction also makes the presence of the alternant UG atemporal form* √ a more plausible, as it avoids attributing the same form to the AC projective and UG atemporal slots.

The reconstruction of *√ -aw* 'UG projective' is to my knowledge new. However, the support for it in Table 10 is quite strong, and it fits the pattern in (20).

Table 10: Supporting evidence for projective reconstructions

PAN	AC: *<um> √ -a*	UG: *√ -aw*	LC: *√ -ay*
Kavalan	non-AC Future: √ -a	_	_
Saisiyat	AC: √ -a	UG	_
Atayal	AC	UG	LC
Sediq	UG Atemp: √ -a	UG	LC
Pazeh	_	UG	AC/UG/LC Future
Thao	_	_	_
Bunun	AC Imperative	AC, UG Imperative	UG

Amis	_	UG ?	LC ?
Tsou	UG Neutral: √ -a	_	_
Kanakanavu	AC	UG	UG/LC Atemp
Saaroa	AC	_	_
	UG Neutral: √a		
Rukai	_	_	_
Siraya	AC	UG	LC
Puyuma	AC	UG Neutral	LC Neutral
Paiwan	_	UG	LC
Proto Bisayan	UG Atemp: √ -a	_	_
W. B. Manobo	UG Atemp: √ -a	_	_
Banggi	UG Atemp: √ -a	_	LC Atemp
Timugon Murut	UG Atemp: √ -a	_	_
Kimaragang	UG Atemp: √ -a	_	LC Atemp
Tombonuo	AC Atemp: √ -a	UG Atemp[27]	_
Malagasy	AC	UG/LC Imperative	_
Proto Malayic	AC: √ -a	_	_

3. Alternative Hypotheses

Nikolaus Himmelmann (pers. comm.) has pointed out that there is an apparent inconsistency in the set of hypotheses presented in Section 2. I have adopted the claim that the UG and LC indicative PAN forms presented in Table 1 were derived from nominalisations. This implies that the pre-PAN

---

27 The Tombonuo imperative suffixes are -a 'AC', -o 'UG' and -i 'LC'. I have assumed that -a and -i reflect *-a and *-i, and that -o reflects *-aw. However, much more work is needed on the diachronic development of Dusunic verbal morphology.

system of pivot and mood morphemes included only the non-indicative and the AC indicative forms, perhaps as shown in Table 11. The 'neutral' and 'projective' labels of Table 1 are here probably better replaced by the more conventional 'realis' and 'irrealis', whilst the unaffixed root was probably (as in many languages) an imperative. The perceived inconsistency is associated with the claim (in 2.3.2) that nominalisations were used in predicates as a strategy of diathesis. This claim presupp-oses that Pre-PAN *lacked* a diathetic system distinguishing AC, UG and LC pivots. If it had already had such a system, there would have been no need for the new one based on nominalisations. In any case, it is the nominalisation hypothesis which explains the rise of the three- or four- (or more) way diathetic system of many modern AN languages.

Table 11: A possible reconstruction of Pre-PAN pivot and mood morphemes

	Actor	Undergoer	Location
Neutral	\<um>√ , √	√ -a, √ -u	√ -i
(= realis, imperative)	*k< um> áRaw, káRaw	*kaRáw-u	*kaRáw-i
	*k< um> aRáC, kaRáC	*kaRaC-ú	*kaRaC-i
Projective	\<um>√ -a	√ -aw [< √ -a-u]	√ -ay [< √ -a-i]
(irrealis)	*k< um> aRáw-a	*kaRáw-aw	*kaRáw-ay
	*k< um> aRaC-á	*kaRaC-áw	*kaRaC-áy

The upshot of these arguments is that the labels 'Undergoer' and 'Location' in Table 11 (and their association with the same forms in Table 1) are called into question, and that it would be sensible to consider alternative hypotheses about the Pre-PAN functions of the forms *-u, *-a 'UG neutral' and *-i. (The *-a(-) of the lower row of Table 11 emerges as the Pre-PAN irrealis marker.) The labels 'Undergoer' and 'Location' make a semantic

distinction among non-Actor pivots. Alternative hypotheses must necessarily offer some other distinction. Possibilities which present themselves are distinctions in verbal tense or aspect, a distinction in the definiteness/referentiality of the pivot, or a distinction in pivot deixis. But none of these possibilities offers a plausible account of the fact that that reflexes of *-i and *-ay in Tables 9 and 10 are most often LC, whilst reflexes of *-u , *-aw and *-a almost never are. That is, it is difficult to avoid the conclusion that Pre-PAN *-i and *-ay were indeed LC pivot markers. However, the fact that both also have a number of reflexes which mark an UG pivot suggests that the semantic distinction between LC and UG pivot marking was becoming blurred around the time that PAN broke up, and that the new diathetic strategy based on nominalisations arose to compensate for this blurring.

Starosta, Pawley and Reid (1982) and Starosta (1992) have suggested that the Pre-PAN verbal suffixes originally arose through the capture of noun-phrase-initial morphemes. To account for the suffixes *-a, *-u (and *-aw), and *-i (and *-ay) in Table 11, this suggestion would require the reconstruction of the Pre-PAN noun-phrase-initial morphemes *a, *u and *i with relevant functions. Interestingly, there is some evidence for the reconstruction of just these morphemes: *a and *u were common determiners ('common' as opposed to 'personal') and *i a locative preposition. The distinction between *a and *u was evidently one of deixis, with *a as the unmarked member of the set and *u probably marking the referent of its noun phrase as not visible or as distant.[28] These reconstructions would

---

28 My interpretation differs somewhat from Starosta's (1992), but our interpretations of *a and *i are somewhat similar.
    Detailed justification of my reconstructions and of the system of which they formed a part, and discussion of the literature on such reconstructions, lie beyond the scope of this paper. Suffice to say here that (i) *a is reflected in

support the inference that *-*i* did indeed mark LC pivot, and would account '
for the *-*a*/-*u* distinction as one of deixis, i.e. between a deictically
unmarked pivot referent and one which was invisible or far away. It must be
stressed, however, that the reconstruction at this time-depth of morphemes
consisting of a single vowel is risky (because of the high probability of
homophonous morphemes arising) and that the explanation offered in this
paragraph is therefore speculative.

Table 11 raises two more questions. The first concerns Pre-PAN aspect
morphology. In Table 1, I have reconstructed durative indicative forms, the
durative aspect being marked by reduplication, but no durative non-indica-
tive forms. But the PAN durative indicative forms were derived from
Pre-PAN nominalisations. Does this mean that nominalisations could be
marked for durative aspect in Pre-PAN, but verbs like those in Table 11
could not? Typologically, this seems unlikely, and scattered reflexes of redu-
plicated Pre-PAN verbs (*R- √ in Pazeh, Puyuma and Tagalog, *R- √ -*a* in
Saaroa, *R- √ -*i* in Puyuma, and *R- √ -*ay* in Pazeh) suggest that the forms in
Table 11 had reduplicated durative counterparts.

The second question concerns the status of *<*um*> in Pre-PAN. In
Table 1 PAN *<*um*> is reconstructed not only in the indicative AC forms

---

Kavalan *a* 'common pivot' and *[y]a* 'ligature', Pazeh *a* 'common pivot, ligature',
Thao *[w]a* 'ligature', Amis *a* 'ligature', Puyuma *a* 'ligature', Paiwan *a* 'common
pivot, ligature', Proto Malayo-Polynesian *[y]a* 'common pivot', reflected in
Philippine languages; (ii) *u* is reflected in Amis *u* 'common pivot'., perhaps
Pazeh *u* 'common non-pivot, and PMP *[y]u* 'common pivot', with reflexes in
Philippine languages of the Batanic and North Cordilleran groups; and (iii) *i* is
reflected in Pazeh, Thao, Amis, Puyuma, Paiwan *i* 'locative preposition'. The
inference that *u* marked a referent as either distant of absent is made from
distinctions recorded in the Batanic language Itbayat and the North Cordilleran
language Casiguran Dumagat.

but also as part of the projective AC form *$<um>$-$\sqrt{}$-a. It is thus the only affix to straddle the divide between PAN indicative and non-indicative forms and, as a result, the only non-suffix to be reconstructed among the Pre-PAN verb forms of Table 11. In view of this distributional oddity, one must ask: did *$<um>$, like the other infix *$<in>$, occur in Pre-PAN only as a nominalising affix? and was *$<um>$-$\sqrt{}$-a perhaps the result of later analogical formation? If this is so, then perhaps the forms with *$<um>$ should be eliminated from Table 11 and *$<um>$-$\sqrt{}$-a emended to *$\sqrt{}$-a . Against this, however, speak the shortage of reflexes of *$<um>$ as a nominaliser in Table 7 and the already noted presence of reflexes of non-indicative *$<um>$-$\sqrt{}$-a in the Appendix. At the moment I have no explanation for this seeming conflict between data and distribution.

4. Where to Next?

My main reason for undertaking this reassessment of the verbal morphology of PAN is to provide a secure basis for identifying morphological innov-ations shared by groups of daughter languages, as these innovations may provide evidence for subgrouping hypotheses and thereby cast more light on early Austronesian prehistory. Some of these innovations themselves may also contribute to our understanding of morphosyntactic change. For example, in this paper the radical restructurings of the verb systems of Tsou and Rukai have been touched on. The verb system of Puyuma, where earlier projective morphemes have ousted their neutral counterparts, also needs investigating.

This search for innovations has yet to be pursued systematically, and I would like here only to suggest a few research directions. These can be

divided into intra-Formosan and extra-Formosan.

Within Taiwan, especially as more detailed descriptions of languages become available, it may be possible to identify the spread of innovations through the early dialect network which resulted from the break-up of PAN and thereby to improve our understanding of the earliest stages of Austronesian linguistic prehistory. It seems likely that Proto Malayo-Polynesian, the language ancestral to all extra-Formosan Austronesian languages, may subgroup with a small number of Formosan languages, probably in the south of Taiwan, and research is needed to identify innovations which may be shared by south Formosan languages and Proto Malayo-Polynesian. Possible points of investigation are the drift of forms in *-an from marking LC pivot to marking UG pivot, the rise of *-*an* as the IN pivot atemporal form, and the loss of *<*um*> from *<*um*> √-*a* 'AC projective'.

Outside Taiwan, it seems likely that Proto Malayo-Polynesian itself is characterised by innovations in verbal morphology. The rise of *maN-* as an AC pivot morpheme seems to be a Proto Malayo-Polynesian innovation, with reflexes as far away as Oceania (Ross 1988: 40-43). This probable innovation was recognised by Dahl (1976: 127-128), but to my knowledge it has never been satisfactorily investigated. Whilst the PAN verbal derivational prefix *ma-* is certainly reflected in Formosan languages, I have found no reflex attributable to *maN-*,[29] the origin of which may well be associated with the

---

29   The *-*N-* of PMP *maN-* indicates that the stem-initial consonant undergoes nasal substitution. Shigeru Tsuchida and John Wolff (personal communications) point out that there is a small scattering of cases in Formosan languages where nasal substitution occurs, sometimes following an *m-* initial prefix. There are also adjectival or stative verbal forms in Atayal and Bunun which have a prefix *man-*(without nasal substitution). These all need further investigation. However, they do not form large-scale paradigms like those formed by verbs with and without *maN-* in a number of Malayo-Ploynesian languages.

(as yet unexplained) origin of the so-called "optional nasal" in Malayo-Polynesian languages (Ross, 1994). Patterns of reduplication also need investigation, since, as mentioned above, Formosan and extra-Formosan patterns appear to differ.

I also hope that the reconstruction of PAN verbal morphology may provide a basis for identifying subgroups among western Austronesian languages, where our understanding of linguistic prehistory is poor and alternate hypotheses abound (see Ross, 1994, for a recent discussion). For this purpose, however, it will first be necessary to understand how Proto Malayo-Polynesian verbal morphology differs from PAN. It would be useful, for example, to define the extent and functions of reflexes of PAN *maR-. Certainly this morpheme is reflected in Formosan languages, but apparently only marking reciprocal verbs. In many Philippine languages, it has become an AC pivot marker: this is an innovation in need of pinning down. More controversially, the reconstruction of *-akən in Table 3 needs examination. It is evidently not reconstructible for PAN, nor apparently for Proto Malayo-Polynesian, but its reflexes span much of Indo-Malaysia and Oceania. If they reflect a shared innovation, then it is an innovation which confounds most present subgrouping hypotheses by separating many Indo-Malaysian languages from their Philippine and other neighbours and placing them in a group with the Oceanic languages.

The reference to *maR- above touches on another research direction.Within Taiwan, it is clear that there are complex systems of verbal derivational morphology[30] on which no comparative work has yet been done. Their investigation would also extend the basis for the enterprise outlined in the preceding paragraphs.

---

30 See, for example, Ferrell (1982:15-27) on Paiwan, and Tsuchida (1990) on Tsou.

APPENDIX

Pivot, mood and aspect morphemes in Formosan languages

Key:

√   verb root

¶   derivational prefix

<X> X is infixed, normally after the root-initial consonant

-X   X is suffixed to the root

R   reduplication: Ca-√, where C is a consonant identical to the root-initial consonant

[X]   X is sometimes but not always present

KAVALAN (Sources: Tsuchida, personal communication; Li 1978b, 1982)

	Actor	Non-actor	Location	Instrument
Neutral	<m>√	√-an (see note 1)	_	ti-√ (see note 2)
Perfective	<m>√	<n>√	<n>√-an (see note 2)	_
Future	√	√-a	_	_
Imperative	√	√-i	_	_

1. The form √-an 'non-AC neutral' may reflect a conflation of UG and LC forms. PAN *ə in final syllables is sometimes reflected as Kavalan /ə/, sometimes as /a/, in Li's (1982) examples. The conditioning is not clear. Hence Kavalan -an may reflect both PAN *√-ən and PAN *√-an.

2. Kavalan properly has only AC and non-AC verbal pivot forms, as the forms <n>√-an and ti-√ occur only in nominalised clauses, never in independent clauses. It is not clear whether ti- reflects PAN *Si-, as the expected Kavalan reflex of the latter is **si-.

SAISIYAT (Sources: Tsuchida 1964, Starosta 1974, Li 1978a, 1978b, Ye 1991)

	Actor	Undergoer	Location	Instrument
Neutral	\<om\>√	√-ən	√-an (see note)	si-√
Perfective	\<om\>\<in\>√	\<in\>√	...	...
Projective	√-a	√-aw	...	...
Atemporal	√	√-i	...	√-ani

Ye (1991) notes that -an is found only in nominalisations.

ATAYAL (Sources: Egerod 1965, 1966, 1969, Li 1980, 1981, Huang 1991, Rau 1992)

	Actor	Undergoer	Location	Instrumental
Neutral	\<m\>√	√r-un	√r-an	s-√ (Mayrinax si-√)
Perfective	\<m\>\<in\>√	\<in\>√	\<in\>√r-an	n-s-√
Projective	\<m\>√r-a	√r-aw	√r-ay	an-ay s-√ (see note 3)
Atemporal	√	√r-i (see note 4)	√r-i	[an] s-√ (see note 3)

1. Huang (1991) shows that the "Instrumental" pivot forms entailing *s*-√ have a semantic range which includes Instrument pivot and Beneficiary pivot, but is far wider than this. Hence she labels it "Circumstantial". Rau retains the term "Instrumental". In the conservative Mayrinax dialect of Atayal the form is *si*-√.

2. √*r* Egerod's (1965: 255) "reduced stem"

3. /an/, occurring in the IN pivot projective and atemporal forms, is apparently an auxiliary. Square brackets indicate that it is not always present. Rau (1992: 56) did not find an IN pivot projective form.

4. Use of √*r-i* as UG pivot (as well as LC pivot) is noted by Rao (1992: 53). Egerod has √here

SEDIQ (Sources: Asai 1953, Starosta 1974, Pecoraro 1979, Li 1981)

	Actor	Undergoer	Location	Instrument
Neutral	<um>√	√-un	√-an	s-√
Perfective	<mun>√	<in>√	<in>√-an	...
Future	<um>R-√	R-√-un	...	...
Projective	...	√-aw	√-ay	...
Atemporal	√	√-a	√-i	...

PAZEH (Sources: Ferrell 1968, Li 1978b)

	Actor	Undergoer	Location	Instrument
Neutral	mə-√	√-ən	√-an	...
Perfective	<in>mə-√	...	...	...
Durative	R-√	R-√-ən	R-√-an	...
Future	mə-√-ay	R-√ay	R-√-ay (?)	...
Projective	...	√-aw	...	...
Atemporal	...	√-i	√-i (?)	si/sa-√-i (?)

THAO (Sources: Li, Chen and Tang 1956, Li 1976, 1977b, 1978b)

	Actor	Undergoer	Location	Instrument
Neutral	<m>√	√-in	√-an	ʃi-√
Perfective	<m><in>√	<in>√	<in>√-an	...
Durative	<m>R-√	R-√-in		
Future	ʔa-<m>√	ʔa-√-in	...	...
Imperative	√	√-in	...	...
Request	√-u	√-i	...	...

BUNUN (Takitudu dialect) (Source: Wu 1969)

	Actor	Undergoer	Location (see note 1)	Instrument
Neutral	√	√-un	√-an	is-√ [-un, -an]
Past	<in>√	<in>√-un	<in>√an	...

Imperative   ...   √-i (see note 2)   ...

1.According to Wu (1969: 26), the distinction between √-*un* and √-*an* is phonologically conditioned, with √-*un* occurring after a nasal or /a/ and √-*an* elsewhere, but an anonymous reviewer has drawn my attention to data showing that this is incorrect.

2.Wu does not make a clear distinction between AC pivot and UG/LC pivot imperatives, but his examples (1969: 29, 33) are almost all clearly UG/LC pivot.

BUNUN (Takbanuað dialect) (Sources: Jeng 1977, Li 1987, 1988)[31]

	Actor	Undergoer	Location	Instrument
Neutral	mV-√	√-un	√-an	is-√
Perfective	m<in>V-√	<in>√	...	...
Projective	√-aw	√-ei	...	...
Imperative	mV-√-i, ma-√-a √-aw		...	...

AMIS (Sources: Fey 1986, Chen 1987)[32]

	Actor	Undergoer	Location	Instrument
Neutral	<əm>	√-ən	√-an (see note 1)	sa-√ (see note 1)
Projective	...	√-aw (see note 2)	√-ay (see note 2) ...	
Imperative	√	...	√-i	...

1.There is no consistent means of forming LC pivot or IN pivot verbs, but a variety of derived forms alongside these diachronically "regular" reflexes.

2.These forms are recorded by Ferrell (1972), who evidently took them

---

31 Jeng's concern with case grammar means that almost all his examples are declarative; non-indicative verb forms are very rare in his data.

32 Ferrell (1972) also gives some Amis data, but they are significantly different (and diachronically more 'regular') than Fey's and Chen's data.

from Ogawa and Asai. I have not found them in Fey (1986) or Chen (1987).

TSOU (Sources: Tung 1964, Tsuchida 1976, 1990, Zeitoun 1992)

	Actor	Undergoer	Location	Instrument/Beneficiary
Neutral	\<m>√, mV-√	√-a	√-i	√-[n]əni

KANAKANAVU (Sources: Sung 1969, Tsuchida 1976, Mei 1982) [33]

	Actor	Undergoer	Location
Neutral	\<um>√	√	√
Perfective	\<inəm>√	\<in>√	\<in>√-an
Durative	\<um>R-√	√-ən	√-ən
	\<um>¶-a-√		
Projective	\<um>√-a (see note 1)	√-aw (see note 1)	√-i
Atemporal	...	√-ay (see note 2)	√-ay (see note 2)

1.Sung (1969) records -aw as -o. She recognises -a and -o as imperatives, but assumes them to be phonological alternants whose conditioning she cannot determine.

2.This is Tsuchida's (1976: 49-51) "special" pivot. It is functionally UG/ LC pivot, and its distribution – in narrative sequences and subordinate to a negative auxiliary – fit the atemporal category.

SAAROA (Source: Tsuchida 1976)

	Actor	Undergoer	Location	"Special"
Neutral	\<um>√	√-a	...	saa-√[-a] (see note)
Perfective	ɬi-\<um>√	ɬi-√[-a]	ɬi-√-a[na]	...
Durative	\<um>R-√	R-√-a	R-√a[na]	...
	\<um>¶-a-√			
Future	_	a-√[-a]	a-√-a[na]	

33 Kanakanavu forms in /-ən/ and /-an/are underlying forms. Surface forms either delete final /-n/ or add schwa.

Imperative	<um>√-a	√-u	√-i, √-ani	...
Atemporal	√	...	...	...
~~A~~temp. Dur.	a-√	...	...	...
	¶-a-√			

The function of Tsuchida's (1976: 75-76) "special" pivot is unclear. Its form, and some of his examples, suggest that it was originally IN pivot.

RUKAI (Source: Li 1973)

	Actor			Non-Actor		
	Present	Past	Future	Present	Past	Future
Neutral	√	wa-√	[a]y-√	ki-√	ki-a-√	ay-ki-√
Durative	R-√	wa-R-√	[a]y-R-√	ki-R-√	ki-a-R-√	ay-ki-R-√
Completive	√-ŋa	wa-√-ŋa	[a]y-√-ŋa ...		ki-a-√-ŋa	ay-ki-√-ŋa
Imperative	[<u>]√-a	...	...	...	...	...
Atemporal	[<u>]√	...	...	...	...	...

SIRAYA (Source: K.A. Adelaar, pers. comm.)

	Actor	Undergoer	Location	Instrument
Neutral	<m>√	√-ən	√-an (see note)	(i-√) (see note)
Perfective	...	ni-√-ən	ni-√-an	...
Projective	<m>√-a	√-aw	√-ei, √-anei	...

Adelaar considers that the forms marked here as LC and IN pivot do not perform these functions, at least not in the language of the translators of Matthew's Gospel. I have seen only one possible case of *i-*√ and it is not clear that it functions as IN pivot.

PUYUMA (Sources: Ting 1978, Tsuchida 1980, 1983, Cauquelin 1991)

	Actor	Undergoer	Location	Instrument
Neutral	<m>√	√-aw	√-ay	√anay
Durative	<m>R-√	R-√-aw	√-ay	R-√anay
	¶-a-√			
Future	R-√	R-√-i	R-√-i	R-√-an
Projective	<m>√-a	...	...	...
Imperative	√	√-u	√-i	√-an[ay]
Atemporal	<m>√	√-i	√-i	√-an

PAIWAN (Sources: Ho 1977, 1978, Ferrell 1982, Egli 1990)

	Actor	Undergoer	Location	Instrument
Neutral	<əm>√	√-ən	√-an	si-√
Perfective	na + <əm>√	<in>√	<in>√-an	<in>si-√
Durative	<əm>R-√	R-√-ən	R-√-an	si-R-√
Imperative	√-u (see note)	√-u, √-i	√-an	√-an
Projective	...	√-aw	√-ay	...
Atemporal	√	√-i	√-an	√-an

Ferrell and Egli both label √-*u* as AC pivot, but both also give examples where the UG is marked as the pivot (Ferrell 1982: 97, Egli 1990: 181).

REFERENCES

Adelaar, K. Alexander

    1992   Proto Malayic: the Reconstruction of its Phonology and Parts of its Lexicon and Morphology. Pacific Linguistics C-119. Canberra: Australian National University.

Asai, Erin

    1953   The Sedik Language of Formosa. Kanazawa: Cercle Linguistique de Kanazawa.

Bellwood, Peter

    1985   Prehistory of the Indo-Malaysian Archipelago. Sydney: Academic Press.

Blust, Robert A

    1977   The Proto-Austronesian Pronouns and Austronesian Subgrouping: a Preliminary Report. WPLUH 9.2:1-15.

    1985   The Austronesian Homeland: a Linguistic Perspective. Asian Perspectives 26:45-67.

Boutin, Michael E

    1988   Banggi Clause Structure. In C. Peck, ed., 1988, 1-52.

Brewis, Richard and Stephen H. Levinsohn

    1991   Topic and Emphasis in Timugon Murut. In S. H. Levinsohn, ed., 1991, 29-43.

Cartier, Alice

    1976   Une Langue à Double Construction Objective et Ergative: L'indonésien. La Linguistique 12:99-130.

Cauquelin, Josiane,

    1991   The Puyuma Language. BKI 147:17-60.

Chen, Teresa M

    1987   Verbal Constructions and Verbal Classification in Nataoran-Amis. Pacific Linguistics C-85. Canberra: Australian National University.

Cooreman, Anne M., Barbara Fox and Talmy Givón

    1988   The Discourse Definition of Ergativity: a Study of Chamorro and Tagalog Texts. In R. McGinn, ed., 1988, 387-425.

Cumming, Susanna

    1986   Word Order Change in Malay. In P. Geraghty, L. Carrington and S. A. Wurm, eds, 1986, 97-111.

    1988   Syntactic Function and Constituent Order Change in Malay. Ph. D. dissertation, University of California, Los Angeles.

Cumming, Susanna and Fay Wouk

    1987   Is there 'Discourse Ergativity' in Austronesian Languages? Lingua 71:271-296.

Dahl, Otto Christian

    1976   Proto-Austronesian. London: Curzon Press.

    1978   The Fourth Focus. In S. A. Wurm and Lois Carrington, eds Second International Conference on Austronesian Linguistics: Proceedings. Pacific Linguistics C-61, 383-393. Canberra: Australian National University.

    1986   Focus in Malagasy and Proto-Austronesian. In P. Geraghty, L. Carrington and S. A. Wurm, eds, 1986, 21-42.

De Guzman, Videa P

    1988   Ergative Analysis for Philippine Languages: an Analysis. In R. McGinn, ed., 1988, 323-345.

Dixon, R. M. W

1979   Ergativity. Language 55:59-138.

Egerod, Søren

1965   Verb Inflexion in Atayal. Lingua 15:251-282.

1966   Word Order and Word Classes in Atayal. Language 42:346-369.

1969   The Origin of Headhunting: an Atayal Text with Vocabulary. Bulletin of the Institute of History and Philology, Academia Sinica 39:291-325.

Egli, Hans

1990   Paiwangrammatik. Wiesbaden: Otto Harrassowitz.

Elkins, Richard E

1970   Major Grammatical Patterns of Western Bukidnon Manobo. Norman, Oklahoma: Summer Institute of Linguistics.

Ferrell, Raleigh J

1968   The Pazeh-Kahabu Language. Bulletin of the Department of Archaeology and Anthropology, National Taiwan University 31-32, 73-97.

1972   Verb Systems in Formosan Languages. In Jacqueline M. C. Thomas and Lucien Bernot, eds Langues et techniques I: Approche linguistique, 121–128. Paris: Klincksieck.

1979   Construction Markers and Subgrouping of Formosan Languages. In Nguyen Dang Liem, ed. South-east Asian Linguistic Studies. vol. 2, 199-211. Pacific Linguistics C-45. Canberra: Australian National University.

1982   Paiwan Dictionary. Pacific Linguistics C-73. Canberra: Australian National University.

Fey, Virginia

    1986   Amis Dictionary. Taipei: The Bible Society in the Republic of China.

Foley, William A. and Robert D. van Valin jr

    1984   Functional Syntax and Universal Grammar. Cambridge: Cambridge University Press.

Geraghty, Paul, Lois Carrington and S. A. Wurm, eds

    1986   FOCAL II: Papers from the Fourth International Conference on Austronesian Linguistics. Pacific Linguistics C-94. Canberra: Australian National University.

Gerdts, Donna B

    1988   Antipassives and Causatives in Ilokano: Evidence for an Ergative Analysis. In R. McGinn, ed., 1988, 295-321.

Gibson, Jeanne D. and Stanley Starosta

    1990   Ergativity East and West. In Philip Baldi (ed.), Linguistic Change And Reconstruction Methodology, 195-210. Berlin: Mouton de Gruyter.

Halim, Amran, Lois Carrington and S. A. Wurm, eds

    1982   Papers from the Third International Conference on Austronesian Linguistics. vol.2 Tracking the Travellers . Pacific Linguistics C-75. Canberra: Australian National University.

Himmelmann, Nikolaus P.

    1987   Morphosyntax und Morphologie – Die Ausrichtungsaffixe im Tagalog. München: Wilhelm Fink.

    1991   The Philippine Challenge to Universal Grammar. Institut für Sprach Wissenschaft, Universität zu Köln, Arbeitspapier Nr 15 (Neue Folge).

Ho Dah-an

    1977   The Phonology of the Butanglu Dialect of Paiwan [in Chinese]. Bulletin of the Institute of History and Philology, Academia Sinica 48, 595-618.

    1978   A Preliminary Comparative Study of Five Paiwan Dialects [in Chinese]. Bulletin of the Institute of History and Philology, Academia Sinica 49:565-681.

Hopper, Paul J

    1977   Observations on the Typology of Focus and Aspect in Narrative Language. NUSA 3:14-24.

    1979a   Aspect and Foregrounding in Discourse. In Talmy Givón, Syntax And Semantics. 12 Discourse And Syntax. New York: Academic Press.

    1979b   Some Discourse Sources of Ergativity. WPLUH 11:137-153.

    1983   Ergative, Passive and Active in Malay Narrative. In Flora Klein-Andreu, ed. Discourse Perspectives on Syntax, 67-88. New York: Academic Press.

    1988   How Ergative is Malay? In R. McGinn, ed., 1988, 441-454.

Huang, Lillian M

    1991   The Semantics of s- in Atayal. Studies in English Literature and Linguistics 17: 37-50.

Jeng, Heng-hsiung

    1977   Topic and Focus in Bunun. Academia Sinica Special Publication No. 72. Taipei: Institute of History and Philology.

King, John Wayne

    1988   Tombonuo Clauses. In C. Peck, ed., 1988, 149-174.

Kroeger, Paul R

    1988   Verbal Focus in Kimaragang. In Hein Steinhauer, ed. Papers in Western Austronesian Linguistics No. 3, 217-240. Pacific Linguistics A-78. Canberra: Australian National University.

    1991   The Event Line in Kimaragang Narrative. In S. H. Levinsohn, ed., 1991, 93-104.

Laycock, Donald C.and Werner Winter, eds

    1987   A World of Language: Papers Presented to Professor S.A. Wurm on his 65th Birthday. PL, C-100. Canberra: Australian National University.

Levinsohn, Stephen H

    1991   Constituent Order in *di* Clauses in Malay Narratives. In H. Steinhauer, ed. Papers in Austronesian Linguistics 1:125-135. Pacific Linguistics A-82. Canberra: Australian National Univer-sity.

Levinsohn, Stephen H., ed

    1991   Thematic Continuity and Development in Languages of Sabah. Pacific Linguistics C-118. Canberra: Australian National University.

Li Fang-Kuei, Chen Chi-lu and Tang Mei-chün

    1956   Notes on the Thao Language [in Chinese]. Bulletin of the Department of Archaeology and Anthropology, National Taiwan University 7:23-51.

Li, Paul Jen-kuei

    1973   Rukai Structure. Academia Sinica Special Publication No. 64. Taipei: Institute of History and Philology.

    1976   Thao Phonology. Bulletin of the Institute of History and Philology, Academia Sinica 47:219-242.

1977a   The Internal Relationships of Rukai. Bulletin of the Institute of History and Philology, Academia Sinica 48:1-92.

1977b   Morphophonemic Alternations in Formosan Languages. Bulletin of the Institute of History and Philology, Academia Sinica 48: 375-413.

1978a   A Comparative Vocabulary of Saisiyat Dialects. BIHP 49:133-199.

1978b   The Case-Marking Systems of the Four Less-Known Formosan Languages,. In: S. A. Wurm and Lois Carrington, eds Second International Conference on Austronesian Linguistics: Proceedings, 569-615. PL C-61. Canberra: Australian National University.

1980   The Phonological Rules of Atayal Dialects. Bulletin of the Institute of History and Philology, Academia Sinica 51:349-405.

1981   Reconstruction of Proto-Atayalic Phonology. Bulletin of the Institute of History and Philology, Academia Sinica 52:235-301.

1982   Kavalan Phonology: Synchronic and Diachronic. In Raine Carle, Martina Keinschke, Peter W. Pink, Christel Rost and Karen Stadtlander (eds), Gava': Studies in Austronesian Languages and Cultures, 479-495. Berlin: Dietrich Reimer.

1987   The Preglottalised Stops in Bunun. In D. C. Laycock and W. Winter, eds, 1987, 381-387.

1988   A Comparative Study of Bunun Dialects. Bulletin of the Institute of History and Philology, Academia Sinica 59:479-508.

McCune, Keith

1979   Passive Function and the Indonesian Passive. Oceanic Linguistics 18:119-169.

McGinn, Richard, ed

1988   Studies in Austronesian Linguistics. Athens, Ohio: Center for

Southeast Asian Studies, Ohio University Center for International Studies.

Mei, Kuang

1982   Pronouns and Verb Inflection in Kanakanavu. Tsing Hua Journal of Chinese Studies, New Series 14:207-231.

Naylor, Paz Buenaventura

1975   Topic, Focus and Emphasis in the Tagalog Verbal Clause. Oceanic Linguistics 14:12-79.

Naylor, Paz Buenaventura, ed

1980   Austronesian Studies: Papers from the Second Eastern Conference on Austronesian Languages. Ann Arbor: University of Michigan, Center for South and Southeast Asian Studies.

Pawley, Andrew K. and Lawrence A. Reid

1980   The Evolution of Transitive Constructions in Austronesian. In P. B. Naylor, ed., 1980, 103-130.

Peck, Charles, ed

1988   Borneo Language Studies 1: Sabah Syntax Papers. Dallas: Summer Institute of Linguistics.

Pecoraro, Ferdinando

1979   Eléments de Grammaire Taroko, Précédés del la Présentation de la vie et de la Culture des taroko. Paris: Association Archipel.

Prentice, D. J.

1971   The Murut Languages of Sabah. Pacific Linguistics C-18.

Rau, Der-Hwa Victoria

1992   A Grammar of Atayal. Ph.D. Dissertation. Cornell University. Taipei: Crane Publishing Co.

Ross, Malcolm D

    1987   A Contact-Induced Morphosyntactic Change in the Bel Languages of Papua New Guinea. In D. C. Laycock and W. Winter, eds, 1987, 583-601.

    1988   Proto Oceanic and the Austronesian Languages of Western Melanesia. Pacific Linguistics C-98. Canberra: Australian National University.

    1992   The Sound of Proto-Austronesian: an Outsider's View of the Formosan Evidence. Oceanic Linguistics 31:23-64.

    1994   Some Current Issues in Austronesian Linguistics. In D. T. Tryon, ed. Comparative Austronesian Dictionary, 45-120. Berlin: Mouton de Gruyter.

Shibatani, Masayoshi

    1988   Voice in Philippine Languages. In Masayoshi Shibatani, ed., Passive and Voice, 85-142. Amsterdam: Benjamins.

Starosta, Stanley

    1974   Causative Verbs in Formosan Languages. Oceanic Linguistics 13: 279-370.

    1985   Verbal Inflection Versus Deverbal Nominalisation in PAN: the Evidence from Tsou. In Andrew K. Pawley and Lois Carrington, eds Austronesian Linguistics at the 15th Pacific Science Congress, 281-312. Pacific Linguistics C-88. Canberra: Australian National University.

    1988   A Grammatical Typology of Formosan Languages. Bulletin of the Institute of History and Philology, Academia Sinica 59:541-576.

    1992   The Case-Marking System of Proto-Formosan. Paper presented at the Third International Symposium on Language and Linguistics:

Pan-Asiatic Linguistics, Bangkok.

Starosta, Stanley, Andrew K. Pawley and Lawrence A. Reid

    1981   The Evolution of Focus in Austronesian. Paper presented to the Third International Conference on Austronesian Linguistics, Bali.

    1982   The Evolution of Focus in Austronesian. In Amran Halim, Lois Carrington and S. A. Wurm, eds Papers from the Third International Conference on Austronesian Linguistics. vol. 2 Tracking the Travellers, 145-170. PL C-75. Canberra: Australian National University.

Sung, Margaret Mian Yan

    1969   Word Structure of the Kankanavu Language. M.A. thesis. Cornell University.

Ting Pang-hsin

    1978   Reconstruction of Proto-Puyuma Phonology [in Chinese]. Bulletin of the Institute of History and Philology, Academia Sinica 49:321-392.

Tsuchida, Shigeru

    1964   Preliminary Reports on Saisiyat: Phonology. Gengo Kenkyu 46:42-52.

    1976   Reconstruction of Proto-Tsouic Phonology. Studies of Languages and Cultures of Asia and Africa, Monograph Series 5. Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

    1980   Puyuma-go (Tamarakao hōgen) goi: fu gohōgaisetsu oyobi tekisuto [Puyuma (Tamalakaw Dialect) Vocabulary: with Grammatical Notes and Texts]. In: Kuroshio Bunka no Kai [Black Current Cultures Committee], eds Kuroshio no minzoku, bunka,

gengo [Ethnology, cultures and languages along the Black Current], 183-307. Tokyo: Kadokawa Shoten.

1983   Puyuma-English Index. Working Papers in Linguistics, University of Tokyo 83:10-63.

1990   Classificatory Prefixes of Tsou Verbs. Tokyo University Linguistics Papers '89, 17-52.

Tung, T'ung-ho

1964   A Descriptive Study of the Tsou Language, Formosa. Academia Sinica Special Publication No. 48. Taipei: Institute of History and Philology.

Verhaar, John W. M

1978   Some Notes on the Verbal Passive in Indonesian. NUSA 6:11-19.

1988   Syntactic Ergativity in Contemporary Indonesian. In R. McGinn, ed., 1988, 347-384.

Wolff, John U

1973   Verbal Inflection in Proto-Austronesian. In Andrew Gonzalez, ed. Parangal kay Cecilio Lopez, 71-91. Quezon City: Linguistic Society of the Philippines.

1980   Verbal Morphology and Verbal Sentences in Proto-Austronesian. In P. B. Naylor, ed., 1980, 153-167.

Wouk, Fay

1986   Transitivity in Batak and Tagalog. Studies in Language 10:391-424.

Wu, Peter A

1969   A Descriptive Analysis of Bunun Language. M.S. term paper, Georgetown University, Washington D.C.

Yeh, Mei-li

  1991   Saisiyat Structure. M.A. thesis, Institute of Linguistics, National
          Tsing Hua University, Hsinchu, Taiwan.

Zeitoun, Elizabeth

  1992   A Syntactic and Semantic Study of Tsou Focus System. M.A.
          thesis, National Tsing Hua University, Hsinchu, Taiwan, R.O.C.

Zorc, R. David

  1977   The Bisayan Dialects of the Philippines: Subgrouping and Recon-
          struction. PL, C-44. Canberra: Australian National University.

  1978   Proto-Philippine Word Accent: Innovation or Proto-Hesper-
          onesian Retention? In S. A. Wurm and L. Carrington, eds, 1978,
          67-119.

Alienable and Inalienable Distinction in Puyuma?

Tsuchida Shigeru

In Puyuma, an Austronesian language spoken on the east coast of Taiwan, there is a distinction between alienable nouns and inalienable nouns, which reminds us of a very similar distinction observed in the Oceanic languages. The inalienable construction in Puyuma, however, occurs only in nouns of kinship terms and of body parts used in a figurative sense. Besides, the first person genitive pronoun -li appearing in this particular construction is totally different from -ku appearing with other nouns. Unfortunately, the origin of this distinction in Puyuma is not clear.

1.   It is well known that there is a kind of noun class distinction between "alienables" and "inalienables" in the Polynesian languages.   Thus in Samoan:

(1)   (a) *'o lo-'u mata* "my eye", and

(b) *'o la-'u ufi* "my yam",

where *'o* is a type of article "the", -*'u* "my", and *lo-/la-*, often simply called "particles", may best be called noun-class markers. Based on the difference of their vowel quality, this distinction is also often called O-class and A-class. It is usually described in terms of nouns designating "body parts, kinship terms, mind/feelings, lands/villages, houses/boats, etc." belonging to O-class, whereas those designating "animals, language, food, conduct/manners, etc." belonging to A-class. Because body parts, kinship terms, etc. cannot be separated from ourselves by our own will, the former are also called

"uncontrollable", or "inalienable", the latter nouns "controllable", or "alienable".

2.   In Melanesian areas the number of noun classes differs from language to language, but the basic distinction between "inalienable" and "alienable" still exists. In Fijian, alienable nouns are further divided into "edible", "drinkable", and "neutral". A genitive pronoun immediately follows an inalienable noun, whereas other nouns require a noun-class marker *ke-*, *me-*, or *no-*. For the convenience of comparison, I will give a table below with examples from Fijian, Samoan, Tagalog, and Indonesian.

Table 1: Alienable and inalienable distinctions in Oceanic languages				
English	Fijian	Samoan	Tagalog	Indonesian
my yam	na ke-qu uvi[1]	'o la-'u ufi	ubi ko	ubiku
my water	na me-qu wai	'o la-'u vai	tubig ko	airku
my house	na no-qu vale	'o lo-'u fale	bahay ko	rumahku
my eye	na mata-qu	'o lo-'u mata	mata ko	mataku
my father	na tama-qu	'o lo-'u tamā	tatay ko	bapaku

As seen in the above table, the Hesperonesian languages such as Tagalog and Indonesian do not have any distinction between the two.[2]

---

1   This expression is rather dialectal, as in Ra. In Bauan, which is considered the standard Fijian, the first sg. genitive pronoun is exceptionally *qau*, thus: *na qau uvi* "my yam". See Milner (1956:65). It has also been observed that the noun classes in Oceanic languages are different from "gender" in European languages in that the same base may be neutral or edible, or else neutral or drinkable, according to the needs of the situation (ibid.): *na no-qu yaqona* "my kava (which I grow or sell)", *na me-qu yaqona* "my kava (which I drink)".

2   In Indonesian and other Malayo-Javanic languages, however, there are other kinds of noun classes, which appear in counting, as in Ind. *budak dua orang* "two

3. In the Formosan languages, generally speaking, just like in all other Hesperonesian languages, there is no distinction between alienable and inalienable. Only in Puyuma, however, is there some kind of difference in nouns designating "kinship terms", and "body parts" when used with a figurative sense. The difference is observed in the appearance of different genitive pronouns and different genitive constructions.

3.1 Puyuma personal pronouns are as follows:[3]

Table 2: Puyuma personal pronouns						
		Nominative		Genitive		Agentive	Oblique
		Indep.	Enclitic	I	II		
1		ingku	-ku	-ngku	-li	ku-	kaningku ~ kanangku
2		inu	-u	-ngnu	-u	nu-	kangnu
3	-remote	iziw	iziw ~ ø	-ntaw	-taw	taw-	kantaw
	+remote						kanTiw
1+		inian	-mi	-niam	-mi	mi-	kaniam
1+2+		inta	-ta	-nta	-ta	ta-	kanta
2+		inmu	-mu	-nmu	-mu	mu-	kanmu
3+	-remote	naziw	naziw ~ ø	-ntaw	-taw	taw-	kanantaw
	+remote						kanaziw

3.2 Nominative pronouns usually occur following a predicative verb or

---

children (lit. two person of children)", *tikus dua ekor* "two rats (lit. two tail of rats)", *kereta dua buah* "two cars (lit. two fruit of cars)", etc.

3   Puyuma (Tamalakaw dialect) phonemes are: p, t, T [t̪], k, ʔ [ʔ], v [β], d [ð], s, z [ʐ] , h, H [ɦ], m, n, ng [ŋ], l, L [ɭ], r [r] (trill), w, y, i, u, a, e [ə].

noun, but when they are emphasized, independent nominative forms appear in the sentence initial position:[4]

> (2) *ingku   mu,   H-em-au   -ku la   za   zanum.*
>   1sg:indep TOP   drink:AF-PunReal.1sg already OBL-unsp water
>   "As for me, I already drank water."

3.3 Agentive pronouns occur preceding a verb in non-actor focus, and designate the agent of that action:

> (3) *ku-   keruT-aw   na   vurasi?.*
>   1sg:agt dig:OF-PunReal NOM-sp   sweet-potatoes
>   "I dug up the sweet potatoes."

Compare (4):

> (4) *k-em-eruT   -ku   za   vurasi?.*
>   dig:AF-PunReal 1sg:nom OBL-unsp sweet-potatoes
>   "I dug up sweet potatoes."

3.4 Oblique pronouns designate objects of a verb:

> (5) *i?man a   veray   za   vurasi*
>   who   NOM-unsp   give:AF-PunReal OBL-unsp sweet-potato
>   *kang-nu.*
>   2sg:obl
>   "Who (is the one who) gave you sweet potatoes?"

3.5 Genitive pronouns I always occur following a relation marker:

---

4   Abbreviations are: AF (Actor Focus); agt (agentive); BF (Benefactive Focus); gen (genitive); IF (Instrumental Focus); indep (independent); LF (Locative Focus); LIG (LIGATURE); nom (nominative (pronoun)); NOM (Nominative); obl (oblique (pronoun)); OBL (Oblique (Relation Marker)); OF (Object Focus); PF (Partitive Focus); pl (plural); pn (personal name); Pun (Punctual); Real (Realis); sg (singular); sp (specific, known); TOP (TOPIC); unsp (unspecific, generic, unknown).

(6) *izu* *i* *Hapurar mu,* *tatlaw a-ntaw* *Hutil zi,*

    that iLIG pn TOP long NOM-unsp-3sg:gen penis and,

    taw *avak-an-ay* *ka-ntaw* *adi* *na-ntaw* *Hutil.*

    3sg:agt put:BF-PunReal OBL-3sg:gen net-bag NOM-3sg:gen penis

    "As for that Hapurar, his penis was long, and so his penis was put

    in his net-bag (when he walked)."

3.6 Genitive pronouns II, however, have been observed to occur imme- diately following a kinship term, as in (7), or a body part word used in a figurative sense, as in (8) and (9):

ama	"father/uncle";	*ama li*	"my _"
ina	"mother/aunt";	*ina ni*[5]	"my _"
imu	"grandchild/grandparent/ ancestor";	*imu li*	"my _"
iva	"older sibling/cousin";	*iva li*	"my _"
wadi	"younger sibling/cousin";	*wadi li*	"my _"
musavak	"son-in-law, adopted son";	*musavak li*	"my _"
kuravak	"husband of one's sister";	*kuravak li*	"my _"

In case of *Hazin* "spouse", it is reported that *Hazin li* "my spouse" sounds a little archaic, and today *ni-ngku Hazin* is more commonly used.

HaLiHa "a man's friend of the same sex" is not a so-called kinship term, but is treated as such in Puyuma: *HaLiHa li* "my friend (of male)".

In case of *alak* "child", which is not a so-called kinship term either, there are two forms with different meanings:

    (n)a-ngku alak "my child";

    alak li "my son-in-law, my adopted son",

the latter being a synonymous expression of *musavak li* above.

---

5 *-ni* instead of *-li* due to assimilation to the preceding *n* in *ina*.

In addition, these words take a personal case marker as in (7):

(7) *ayawan*　　*i*　　　　*ama*　　*li.*

　　.chieftain　　NOM-PER　　father　　my

　　"My father is chieftain."

Examples of words designating body parts in a figurative sense are as follows:

(8) *na-ngnu*　　*k-in-a-sahar mu,*　*Hawlay*　*ziya i*　*Lima li.*

　　NOM-sp:2sg　thing-loved　TOP,　here-exist　still LOC hand　my

　　"What you want is still in my possession."

(9) *na-ngnu*　　*zazek*　*a*　　*lalak*　*u*　　*ziya*　*mu,*

　　NOM-sp:2sg body　LIG　young　2:nom　still　TOP

　　Hawlay　*ziya*　*i*　　*maTa*　*li.*

　　here-exist　still　LOC　eye　　my

　　"Your figure in your youth still lies in my eyes, i.e. I still remember you when you were young."

If the body part noun is used, not in a figurative sense, but in its original meaning, then an enclitic pronoun never appears:

(10) *nana*　　*a-ngku*　　　*maTa.*

　　painful　NOM-unsp-my　eye

　　"My eyes are aching."

4.　　What is the origin of this particular distinction observed in nouns of Puyuma, which very much resembles that between alienable and inalienable in the Oceanic languages? This is unfortunately not at all clear.

4.1　　Linguistically Puyuma is most closely related to Paiwan, which is located to the south of Puyuma. Personal pronouns in Paiwan are as follows (cf. also Ferrell (1982:14)):

Table 3: Paiwan Personal Pronouns					
	Nominative		Genitive		Oblique	Locative
	Indep.	Enclitic	Indep.	Proclitic		
1	ti-aken	-aken	ni-aken	ku-	tyanu-aken	i+Oblique
2	ti-sun	-sun	ni-sun	su-	tyanu-sun	
3	ti-madyu	ø	ni-madyu	ø	tyay-madyu	
1+	ti-amen	-amen	ni-amen	nia-	tyanu-amen	
1+2+	ti-tyen	-ityen	ni-tyen	tya-	tyanu-ityen	
2+	ti-mun	-mun	ni-mun	nu-	tyanu-mun	
3+	tia-madyu	ø	nia-madyu	ø	tyaya-madyu	

There are two genitive forms in Paiwan: one is independent "possessive" form ("mine", "yours", etc.), while the other is proclitic form. Either of these two precedes its head, but an independent form requires a ligature *a* in between:

(11)   ni-aken a kava   "my clothes",   and

         ku kava   "my clothes",

and there is no distinction of the sort found in Puyuma:

(12)   *ni-aken a kama,   ku kama*   "my father"

         ni-aken a matsa,   ku matsa   "my eyes"

4.2   First person singular genitive form -*li* in Puyuma looks rather peculiar.[6] As far as I know, the same form, or even a similar form for that matter, occurs only in Rukai, a language spoken to the west and the south

---

6   It is strange that Cauquelin (1991) did not notice the existence of -*li*, which was already reported by Ogawa and Asai (1935:303-304) and Tsuchida (1980:196-200).

of Puyuma.[7]

Personal pronouns in Rukai (Budai dialect) are as follows (cf. also Li (1973:83) which is of Tanan dialect):

Table 4: Rukai Personal Pronouns				
	Nominative		Genitive	Oblique
	Independent	Enclitic	Enclitic	
1	kunaku	-(a)ku/ -naku/naw-	-li	nakua/ili
2	kusu/asu	-su	-su	musua
3 -remote	kuani	ø	-ini	-inia
3 +remote	kuaDa	ø	-iDa	-iDa(a)
1+	kunai	-nai	-nai	naia
1+2+	kuta	-ta	-(i)ta	mita(a)
2+	kunumi	-numi	-numi	numia
3+ -remote	kulini	ø	-lini	-linia
3+ +remote	kuliDa	ø	-liDa	-liDa(a)

Genitive pronouns occur only after the head, and apparently there are no noun classes in Rukai:

(13) *laymay li*   "my clothes",

   tama li   "my father",

   maca li   "my eye".

---

7   Robert Blust kindly suggested to me that Buli in South Halmahera, Selaru on Tanimbar Islands, and Larike on Ambon Island might have *ni* for the first singular genitive pronoun. I could check only Buli with Maan (1940, 1951), but could not find a similar form in those Buli data.

4.3   Judging from the fact that this particular genitive pronoun *-li* appears in all Rukai dialects, including those of Lower Three Villages, and the fact that there is no genitive form **ku* observed in any Rukai dialect, we can probably assume that *-li* is an innovation which occurred in proto-Rukai, and that the same form *-li* in Puyuma is most likely a borrowing from some dialect of Rukai. If such is the case, however, we should expect many lexical loan words from Rukai in Puyuma. Yet strangely enough we cannot find exclusively shared cognates between these two languages. In Swadesh's basic vocabulary, there is seemingly only one such item (Rukai forms are taken from Li (1977) unless otherwise noted):

(14)   English   "squeeze"

Rukai   (Budai)   *wa-pə:*

(Maga)   *u-pə:*

(Tona)   *wa-pəʔə*

(Mantauran)   *o-pəʔə*

Puyuma   *-pereH*

These forms are restricted to Rukai and Puyuma in Taiwan, but must be cognate to Indonesian *perah* "squeeze", and thus Proto-Hesperonesian **peReq* can be reconstructed.

Unfortunately no good Rukai dictionaries have been published, and therefore no detailed lexical comparison is possible. I investigated the Kochapongan dialect of Rukai in 1985, but due to my tight schedule I have as yet had no opportunity to put the data in order. Therefore I compared only the body part vocabulary (about 300 lexical items) of the two languages, but could find only the following three items exclusively restricted to Rukai (Kochapongan dialect) and Puyuma (Tamalakaw dialect):

(15) English	Rukai	Puyuma
"have one's hair cut"	*ki-a-goLogoLo*	*ki-hu-a-LhuL*
		< *-huLhuL*
"close one's mouth"	*pa:-ka-kameDe*	*mar-kemez*
"toothless (due to age)"	*ma:-ngopaLe*	*ma-ngupaL*

It seems to me that three matches out of three hundred compared sets are too few if Puyuma really had some contact with Rukai in the past and was linguistically influenced by the latter. One possible explanation might be that because the Rukai language shows much dialectal difference, especially between Rukai Proper and Lower Three Villages, if the source dialect of Rukai had become extinct, then it would be rather difficult to prove that Puyuma took over many Rukai lexical elements.

5.   Even if it were proved that Puyuma took over many Rukai words, including the first person singular genitive pronoun -*li*, there still remains another problem: how did the alienable- and inalienable-like distinction develop in Puyuma? For the moment I have no answer to this question. What we need is perhaps more detailed investigation and thorough study of the Rukai language.

REFERENCES

Cauquelin, Josiane

1991   The Puyuma Language. Bijdragen tot de Taal-, Land- en Volken-
        kunde 147.1:17-60. Leiden: Koninklijk Instituut voor Taal-, Land-
        en Volkenkunde.

1991   Dictionnaire Puyuma-Français.   Paris,   Jakarta: Ecole Française
        d'Extrême-Orient.

Ferrell, Raleigh

1982   Paiwan Dictionary. Pacific Linguistics Series C, No. 73. Canberra:
        Dept. of Linguistics, RSPacS, ANU.

Li, Paul Jen-kuei

1973   Rukai Structure.   Institute of   History   and Philology,   Academia
        Sinica, Special Publications No. 64. Taipei.

1977   The Internal Relationships of Rukai. The Bulletin of the Institute
        of History and Philology, Academia Sinica 48.1:1-92.

Maan, G.

1940   Boelisch-Nederlandsche Woordenlijst   met   Nederlandsch-Boelisch
        Register. Verhandelingen van het Koninklijk Bataviaasch Genoot-
        schap van Kunsten en Wetenschappen, Deel LXXIV, Derde Stuk.
        Bandoeng: A.C. Nix & Co.

1951   Proeve van een Bulische Spraakkunst. Verhandelingen van het
        Koninklijk Instituut voor Taal-, Land- end Volkenkunde, Deel X.
        's-Gravenhage: Martinus Nijhoff.

Milner, G. B.

1956   Fijian Grammar. Suva (Fiji): Government Press.

Ogawa, Naoyoshi, and Erin Asai

  1935   Gengo-ni yoru Takasagozoku Densetsushu (The Myths and Tradi-
          tions of the Formosan Native Tribes (Texts and Notes)). Taipei:
          Taihoku Imperial University.

Tsuchida, Shigeru

  1980   Puyuma (Tamalakaw Dialect) Vocabulary — with Grammatical
          Notes and Texts — In Kuroshio no Minzoku, Bunka, Gengo,
          pp.183-307. Tokyo: Kadokawa Shoten.