THE PEPET IN SAMA-BAJAW

A. KEMP PALLESEN

1. Topic of discussion
2. Sama-Bajaw as a distinct sub-grouping
3. The data
4. Normal Proto Sama-Bajaw reflexes of PAN *æ
5. Irregular reflexes
6. Conditioning factors for the PSB reflexes of PAN *æ
7. The development of gemination in PSB
8. Phonemic status of the PSB reflexes *æ and *e
9. Processes by which æ or e is innovated in Sama-Bajaw
10. Reflexes of PSB *æ and *e in daughter languages
11. Conclusion

1. TOPIC OF DISCUSSION

PAN *æ, the so-called 'pepet' vowel of Proto-Austronesian, described by Conant (1912) as the "original indifferent vowel... colorless and indefinite in pronunciation", is notable for the variety of its realisations in the daughter languages. This variety is described for Sama-Bajaw¹, a sub-grouping of Austronesian, where it finds especially rich expression. Some explanatory hypotheses are proposed for the splitting of PAN *æ into three Proto Sama-Bajaw phonemes.

2. SAMA-BAJAW AS A DISTINCT SUB-GROUPING

Sama-Bajaw languages occur in greatest number and diversity along the Sulu Archipelago (Philippines) and up into Sibuguey Gulf on the western end of Mindanao. Other languages of the group are Abaknon (AB)² in the
Central Philippines; Sabah Bajau (NB) on the West Coast of Sabah, Malaysia; and a number of closely related but geographically dispersed groups (IB) in Indonesia - in Sulawesi, the Moluccas, and the Lesser Sunda Islands.

The position of the Sama-Bajaw subgrouping within Austronesian has not yet been satisfactorily determined. It is a Hesperonesian sub-group, but the data do not clearly support the generally accepted assignment to the Philippine Hesion. Central Sulu (CS) languages share 31% of cognates with both Malay and Cebuano on a 372 word list, and 38% with Siocon Subanon, a Meso-Philippine language of western Mindanao. Sabah Bajau (NB) shares 37% with Malay and 32% with Siocon Subanon. (The 54% which Central Sulu languages share with Tausug, a displaced Southern Bisayan language spoken in Central Sulu, is due to considerable linguistic convergence.) The sub-group is atypical, though perhaps not unique (if Sangir is a member of the Philippine Hesion), in having *h as the regular reflex of PAN *R.

Within the sub-grouping Sama-Bajaw is divided into six sections on the basis of cognacy, syntactic typology, and reflexes of PSB phonemes. The dividing lines of these sections correspond to discontinuities in chains of mutual intelligibility. Since it is not important for the topic of this paper, this internal subdivision is not discussed in further detail, beyond noting that apart from relationships with (AB), which shows a considerable degree of convergence with Central Philippine languages, the lowest internal Sama-Bajaw cognate relationship is between (CS) and Indonesian Bajau (IB), with 54%.

3. THE DATA

Accurately recorded data based on the Elkins word list, and considerably augmented in several cases, are available for all sections of Sama-Bajaw and form the basis of this study. The notation is essentially phonemic unless otherwise indicated. The Proto Sama-Bajaw forms cited are from a corpus of about 2500 reconstructions currently in preparation.

4. NORMAL PROTO SAMABAJAW REFLEXES OF PAN *g

Sama-Bajaw languages exemplify Conant's 'law of gemination'. Gemination in SB is the result of two processes: (1) the gemination of a single medial consonant, and (2) the assimilation of the first consonant of a heterorganic cluster formed by the reduplication of a monosyllable. Non-geminate homorganic clusters are formed by the assimilation of an infixed *h to the point of articulation of the following consonant.
A following homorganic cluster, geminate or otherwise, is the essential environment for the reflection of PAN *ə as *ə⁷. Elsewhere, i.e. where this environmental condition is not met, the reflex of PAN *ə is *ə⁸ or *a. PSB is thus like Malagasy in having diverse reflexes of PAN *ə in penult and ultima, but it does not conform to Conant's⁵ observation that "the Philippines languages, wherever they show this double vocalism, have a in the penult". The phonemic status of *ə and *u in PSB is discussed in Section 8.

4.1 The three PSB reflexes of PAN *ə are now presented and exemplified for the penultimate and ultimate syllables, i.e. for the disyllabic words which comprise the great proportion of the Proto-Austronesian corpus. The proposal of this paper is that the differentiation of PAN *ə into the PSB reflexes *ə, *u and *a is most satisfactorily explained by positing contrastive stress in the dialect of PAN⁹ from which PSB most immediately derives, at least in those words which contain PAN *ə. This stress hypothesis is discussed in Section 6.

*ə is the reflex of unstressed PAN *ə in the penult; *a is the reflex of unstressed PAN *ə in the ultima, when the vowel of the penult is other than PAN *ə; *u is the reflex of PAN *ə elsewhere, i.e. of stressed PAN *ə in penult and ultima, and of unstressed PAN *ə in the ultima when the vowel of the penult is also PAN *ə:

PAN *rebah 'to throw down, destroy' > *hebbaq 'to topple'; PAN *səpsəp 'to suck' > *səsəsəp; PAN *lipət 'to fold' > *lipət 'to fold in two'; PAN *lōnajah 'kind of plant (sesame)' > (*lōnajah⁸ 'Sesamum orientale'; PAN *kəməkem 'to keep close' > (*kəməkem 'to hold in closed fist'; PAN(B) *sərə(CT) 'constriction, to tighten' > (*hugut⁸ 'to be secure'; PAN *həjəq 'charcoal' > *ərəq; PAN *lə(ŋ)bəŋ 'grave, ditch' > *ləbbəq 'to bury'.

4.2 Lack of data makes it impossible to give an adequate description of the PSB reflexes of PAN *ə in the antepenultimate syllable. The one PAN example which has a PSB reflex is unsatisfactory because of the ambiguity of the Ja. and Ml. evidence for PAN *ə⁹:

PAN *serampəŋ 'sprout, time' (the citations in VLA suggest 'fish spear' as a more satisfactory gloss) > *saḥapaŋ 'multi-tined fish spear'. There are numerous comparisons, however, between Malay and Sama-Bajaw which show that SB regularly has *a where Ml. has ə in the antepenult (the possible loan status of some of the following forms does not affect the point of the relationship between the vocalisms):

PSB *jambatan 'bridge, pier' : Ml. jombatan, jambatan 'bridge';
*kəranjaŋ 'open-weave basket' < PAN *kəranzaŋ : Ml. kəranjaŋ 'basket';
(\*)salassay \< (salassay) 'to settle a dispute' : Ml. selesay 'to solve, settle, complete'. It seems probable that the correspondence of PSB \*a and Ml. \*e in the antepenultimate reflects PAN \*a.

Although the PAN > PSB evidence is sparse, the internal evidence of Sama-Bajaw supports the reconstruction of five vowels in the antepenultimate syllable: \*a, \*i, \*u, \*e, \*w. Three of these are of interest here\(^{12}\); \*a is included because of the partial overlap of its reflexes with those of \*e and \*u.

4.21 \*a is posited for the correspondence: e (occasionally a in (YK), \*e (occasionally a) in (NB)\(^{13}\), a in other languages:

(#) bagunbun 'dust'. (ZB, NS.2-3, CS.5, SS.4) bagumbun, (YK) begumbun, (NS.4-9, WS, CS.2, SS.1-4, Tsg.) bagunbun.

*sarudun 'roofed area'. (YK) sebudun 'extension of roof overhang', (CS.2) saudun 'roofed, unmwalled part of house', (NB.1) sarudun 'ooking area'.

4.22 \*e is posited for the correspondence: a in (AB), i (with a few exceptions) in (ZB) and (NS-SS), e or i in (YK)\(^{14}\), e in (NB), and a or i in (IB)\(^{15}\):

*dabuhiq 'last night'. (AB) dabuqi, (YK) debuhiq, (NS-SS) dibuhiq, (NB) debuqi, (IB) dibuqi.


4.23 \*u is posited for the correspondence: a in (AB), u in (ZB, YK, NS-SS and JM), u or \*e in (NB), a or i in (IB). Only the (AB) and (IB) reflexes distinguish this set from \*u.

*kupapoq 'groupers, rock cod'. (YK, NS.3, WS.1, CS.2) kupapoq, (SS.4) kupapuq, (IB.1) kuapuq.

*tauolaq 'bone'. (AB) taqulan, (ZB.3, CS.2-3, SS) toqolaq (*tauolaq).

The internal PSB evidence thus suggests that PAN \*e in the antepenult split into the reflexes \*e and \*u, a split analogous to that described for the penult. The primary conditioning factor in the antepenult, however, does not appear to be contrastive stress, as is posited for the penult (in Section 6). The small number of examples permits only the following tentative conclusion at this stage:

In the antepenultimate syllable \*u is the reflex of PAN \*e before the consonants \*h, \*w and \*l; \*e is the reflex of PAN \*e otherwise.

\*h and \*w are consonants which do not geminate, and this limitation inhibits the reflex of PAN \*e as \*o in the penult (discussed in
Section 6.1). It is reasonable to suppose that the same inhibiting function is exercised in the antepenult also. *l follows *e in two of the six examples where *e is reconstructible in the antepenult, but there is no equivalent restriction on the gemination of *l. The inclusion of *l in the above distribution statement remains unexplained at present.

No gemination takes place following *e in the antepenult, contrary to what might be expected from the constraints on the realisation of PAN *e as *e in the penult. There is apparently an overriding constraint which may be informally stated as "no word may have two sets of geminates". Such a constraint is current in (CS.2), a language which reflects penultimate *e as e: buhat 'to serve refreshments' becomes (with stative prefix mag- and initial CV reduplication) magbubuhat 'one who serves refreshments'. tettaq 'to chop firewood' becomes, by the same affixation process, magtatetqaq 'one who chops firewood'. Where a homorganic cluster (nasal plus stop) follows *e in the antepenult the reflex e does occur in (NS-SS), but not otherwise:

*entello 'egg'. (AB) antallu, (NS.1-2, CS.2-3,6, SS.1,4-5) entallo, (IB.1-2,4) antillo.

*daqilaw 'yesterday'. (AB) daqilaw, (ZB.1,3, NS, WS.2, CS, SS.1,3-4) dqiilaw, (IB.1,5) dilaw.

5. IRREGULAR REFLEXES

Three hypotheses are considered as possible explanations for reflexes of proto-phonemes which are aberrant insofar as they do not conform to the rules stated in Section 415: (1) that the words in question are loans, (2) that the reconstructed proto-forms are incorrect, or incomplete insofar as they fail to indicate phonological variants of the proto-form (doublets), and (3) that there are further developments within PSB, either regular processes of sound changes for which the conditioning factors must be discovered and described, or innovations (apparently unconditioned variants of protoforms). Each of these hypotheses has its legitimate function16, but of the three borrowing is the least rigorous and the most open to misuse. Dyen's cautionary word (Dyen 1953) in this respect is worth keeping in mind:

"If phonetically similar phonemes occur in the words of related languages in contact, and their correspondence does not agree with any formula that has been otherwise established, it is not safe to conclude that the words are related only by borrowing. ...it may be discovered that this (borrowing) explanation is insufficient"
5.1 Almost all the exceptions to the rules given in Section 4 are those where PAN *a > *a, a situation similar to that noted by Llamzon (1966) for Tagalog. In this section the general problem of *a reflexes of PAN *a is considered, for ultima as well as penult, since there are plainly possible alternatives to the interpretation that *a in the ultima is the reflex of unstressed PAN *a.

5.11 Malay is an obvious possible source for SB loans; it has exerted considerable influence on Sulu languages for many centuries, and it reflects PAN *a as a in the penult. There is, furthermore, evidence that other Philippine languages replace the penultimate or antepenultimate o of Ml. loans with a (a replacement noted by Dyen (1953) for Ngadju Dayak loans from the same source)\(^{17}\). In the following examples from non-Sama-Bajaw languages the supposition that we are dealing with loan words is supported by the occurrence of non-inherited features in addition to the replacement of PAN *a with a, i.e. by the presence of a Malay prefix or of a second irregular prefix of a PAN phoneme:

Mar. parampuan 'woman' : Ml. perampuan; Tsg. jorat 'noose' : Ml. jorat (the expected Tsg. reflex of PAN *gorat is (dj)umat). SB languages show the same pattern:

CS.3 halbaña 'of noble birth' : Ml. borbaña.

The foregoing evidence justifies consideration of the possibility of borrowing from Malay as an explanation for PAN *a > *a irregularities. It also suggests an alternative source for Tg. katām 'carpenter's plane' (if it is in fact a borrowing), for which Llamzon\(^{18}\) proposed Kapampangan as a plausible source\(^{19}\). A Malay source would also explain the presence of katām in Sulu languages.

Such a hypothesis would be strengthened if a Philippine language were to exhibit the expected regular reflex of PAN *a in PAN *katām (e.g. Tg. kitām, CB kutum or FSB (*katūm/kottum)). This is the case with FSB (*g)anap and *gannap 'to become complete' (Tg. ganap and Ml. ganap) for which VLAW gives only PAN *ganap as a possible source. Alongside the FSB doublet, which may be a borrowing, we have the CS.2 form kēnnap 'to add to' < *(kēnnap). Apart from the devoicing of the initial consonant this is a regular reflex of PAN *ganāp.

There is, however, no proof of the borrowing hypothesis for such problem cases. The apparently irregular form may be a regularly reflected doublet of the proto-language or of some meso-language. As more data show the same 'irregular' reflex occurring in languages not in contact, and as such forms prove to be unambiguously reconstructible for the proto-languages of distinct sub-groupings, a doublet hypothesis gains strength. A borrowing hypothesis, while not ruled out, requires more substantive evidence - or greater ingenuity.
5.12 The problem of PAN *a > *s is one for which alternatives (2) and (3) must be thoroughly investigated before a borrowing explanation is entertained. Alternative (2) — doublets in a proto-language — is supported by a significant number of PAN doublets which show an alternation between PAN *a and PAN *a in both penult and ultima:

PAN *basuR/*basuR 'satisfied'; PAN *basih/*basih 'iron'; PAN *dampul/*dampul 'adhesive'; PAN *kaRat 'to bite' / *kaRat 'to cut off'; PAN *laŋkap 'to be complete' / *raŋkap 'to be combined'; PAN *tanaq/*tanaq 'ground, earth'; PAN *zazag/*zazag 'to gain a firm footing'.

Of the corpus of PSB words which reflect a PAN form with an infixed *ŋ a remarkably high proportion — over half — show the reflex *a for PAN *a in the penult. Among them are:

PAN *gandaŋ 'kettledrum' > *gandaŋ 'drum; to beat time'; PAN *qentih 'to stop, finish' > (*h)aŋtiq 'to break a journey' (Ml. menanti 'to wait for', hanti 'to stop'); PAN *qentih 'a girdle' > *kan(d)jiti 'scarf; saash'; PAN(B) *lañtiŋk 'graceful curve' > (*l)antiŋk (Tg., Mar. lantik); PAN *temat 'domicile' > (*t)ampat 'site where psychic power is concentrated' (Mar. tampat 'tomb'); PAN *to(n)tu 'certainly' > *tantu (Mar. tanto, CB tantu 'to an extreme degree'; (cf. also PKr. tantu 'certainly' and mutu tantu 'too, overly', and the suggested Port. origin for this disputed reconstruction).

It can hardly be the case that there is any greater inherent probability that such forms will be borrowed from Malay; rather it is evidence of the facility with which *a and *a alternate in this environment, at some stage in the development of PSB.

The *a/*a alternation before an infixed *ŋ was evidently an established pattern at the PAN level. Examples given below are those which have a Sama-Bajaw reflex:

PAN *lamba(r)/*lamba(r) 'material, "piece" as a numerical classifier' > *lamba 'strand, piece'; PAN *ma(n)tah/*mentah 'unripe, raw' > *mataq; PAN *sambah 'respect' / *sambah 'reverence' > (CS.2) samba 'to revere'.

5.13 One further set of data remains to be considered, where PAN *a > *a in the final syllable:

PAN *li(q)ker 'to circle, wind' > (*)(ka)liŋkar 'to surround'; PAN *lipat 'to fold' > *lipat 'to fold in two'; PAN *rapat 'joined together' > (*)(r)apat 'to be joined without any gap' (Mar. lapat 'airtight', Tg. lapat 'fitted in'); PAN *sanat 'excessive' > *sanat 'intensity; severity'; PAN *tabaq 'to be fat' > (*t)abaq 'fat which adheres to flesh' (CB tabaq 'pork fat', Mwa tabaq, Tg. tabaq 'to be fat').

It is proposed in Section 6 that PSB *a, when it corresponds to PAN *a in the ultima, is the regular reflex of PAN [a], the unstressed allophone
of PAN *e in the ultima. Such a hypothesis must contend with explanations in terms of borrowing and doublets (the alternatives (1) and (2) above). For the borrowing hypothesis it may be argued that the SB forms in the above examples are exactly what would be expected if they were in fact loans from Malay. One reflex at least, however, cannot be explained as a borrowing from ML (nor from a Philippine language which reflects PAN *R as g):

PAN *báraq 'ulcer, abscess' > *bahaq 'swelling, a boil'.

As has been shown in 5.11 there is ample precedent from PAN for positing the existence of doublets which exhibit the a/o alternation in the ultima. Examples like PAN *rapot above, for which diverse Philippine languages reflect a in the ultima, also suggest the possibility of doublets, at least at the level of a Proto-Philippine meso-language.

The conclusions indicated at this stage are: (a) that the presence of undoubted Malay loans in SB and in other Philippine languages, and the pattern of e to a replacement in such words, support the possibility of borrowing as an explanation; (b) that there is evidence from PAN in support of the hypothesis that some PAN *e > *a forms are the regular reflex of previously unidentified PAN doublets; (c) that in view of the high frequency of forms showing the a/o alternation before a homorganic nasal-stop cluster some at least of these must be inherited from a language ancestral to PSB, and (d) that the data suggest that some of the a forms are either PSB innovations, or the result of regular phonological processes in the ultima. The arguments for the latter proposal are presented in Section 6.3 and are not repeated here.

The choice between the competing hypotheses is not resolved. Each case must be considered separately; but any borrowing hypothesis can be sustained only by good evidence.

5.2 Irregularities other than PAN *e > *a make up a small residue, all of which is now considered. In the following material the expected reflex, indicated by [...], follows the actual PSB or daughter language form:

(*) bakkus *[bak(aw)s], (CS.2) bekos [bak(ao)s] 'to be marked by transverse bands': PAN(B) *bakes 'belt'. This is probably a PSB innovation, with *e replacing the vowel of the penult. The (CS.2) form is a secondary development, perhaps via *(bekws).

(C) (YK) dellet *[d(e)llet] 'noose': PAN *dirat 'rope' (cf. PAN *zorat 'loop'). Probably a PSB innovation with vowel replacement and stress shift, from *(derret).

*hadun *[haddan/hudan] 'ladder': PAN *hözan. A PSB innovation by vowel metathesis.
*
i\text{ntaq} *\text{entaq}"to eat raw" : PAN *qa\text{n}taq "to be raw". A PSB innovation which is unusual with regard to the replacement of PAN *o by *i.

*kehet *\text{kuhut}"to cut up; to slice" : PAN *ka\text{ret} "to cut off". Probably a PSB innovation, without parallel other than the next example.

*keket *\text{kukut/kakkut}"to bite" : PAN *ka\text{kat} "to hold fast". Ho. hehotra (VLAW) 'grip, bite' confirms the semantic connection. This and the preceding example display the regular (YK) reflex of the posited *(\text{kuhut}) and *(\text{kukut}) , but it is unlikely that the (YK) form was the loan source for other PSB languages, in view of the wide spread of unambiguous reflexes of the reconstructions given.

*kijut *\text{kudut/keddut}"small muscular movement; twitch" : PAN *ka\text{dut} "to smart" (VLAW gives Ja. kudut "muscular jerking"). It is not clear that the PAN and PSB forms are directly related, though a doublet may be involved. Cf. PSB *kudut/keddut "startled movement" , and PAN(B) *ka\text{zzu} (CT) "pinch" (with Singhi kujet noted as metath.).

(NS, CS) lenog *[\text{donog/denog} "noisy, disturbed" : PAN *\text{denog} "to hear". Probably not inherited from this PAN form. The CB data, ligug-\text{ligug} "to ignore by pretending not to hear" and du\text{ng}ug "to hear" , suggest a doublet, PAN *(\text{li\text{ng}g}).

(CS.2) lipaq *[\text{lopaq/lopaq} "to apply cosmetic paste" : PAN *lo\text{pah} "paste". Probably a reflex of a doublet with PAN *(\text{li\text{paq}) which VLAW suggests somewhat uncertainly (cf. MI. lepa "plaster, mortar").

(YK) lu\text{gu}q *[\text{leguq/leguq} "thunder" : PAN(B) *\text{leguq} "booming sound". Probably the reflex of a doublet *(\text{lu\text{gu}q} . Cf. Mar. ruguq and WBM ruguq "thunder", and the same vocalism in CB hgu\text{q} "produce a steady humming sound" and in Tg. hgu\text{g}q "noise like thunder".

*passil *[\text{pusul/pussul}]"to apply pressure with the fingers; to massage" : PAN(B) *pe\text{ce}l "to squeeze in the hands". Possibly a PSB innovation, but cf. the final vowel of CB pis\text{f}l, pisil "to squeeze s.t. with the tips of the fingers wide apart".

(*\text{san}de(\text{lr}) *[\text{san\text{aw})}"to derive support from" : PAN *san\text{De}R "lean against". Probably the reflex of a doublet. Cf. Batak san\text{dir} "to lean against a wall" , and CB sandig "to lean against".

(*\text{sambaliq *[\text{sambaliq} "to slaughter by draining blood" : PAN *sambalih "to slaughter according to Moslem rites". Probably a Philippines innovation. Cf. Mar. sambaliq, WBM sambaliq.

(CS.2) tanog *[\text{tonog/tanog} "resonant" : PAN *t\text{nea}R "voice". Probably a PSB or CS innovation. Cf. CB tunug "widely known, resonant" ; Tg. tinig "voice" ; which are regular reflexes of the PAN form.

*timba\text{k} *[\text{tambak}] "to shoot a firearm" : PAN *tambak. Either a borrowing from Tg. (timbak) (unrecorded), or the reflex of a Philippines innovation. Cf. Mar. timbak.
*tīner *[tunger/taner] 'tanbark mangrove (Ceriops sp.)' : PAN *tēner. Probably a PSB innovation.

(*tumān *[tumān/tamān] 'certain; to fulfill' : PAN *tēman 'to be accustomed'. Possibly a borrowing from CB tumān 'extreme', or from Mar. tumān 'true; to fulfill' (in which the expected Mar. reflex would be (taman)). Possibly also the reflex of a doublet *(tumān). Cf. the Ja. doublet: temān 'to be faithful'; tumān 'accustomed'.

6. CONDITIONING FACTORS FOR THE PSB REFLEXES OF PAN *a

Dempwolff\textsuperscript{22} noted the occurrence of stress in the three primary languages of his study, but concluded that it was not possible to compare stress in a useful way. Zorc (1972), in a paper entitled 'Current and Proto Tagalic Stress', raises the question:

"If... one posits stress as predictable and hence non-phonemic for Proto Malay-Polynesian, how is it that it does become phonemic in so many daughter languages in the Philippines?"

In a concluding paragraph, having noted the need to posit both penultimate and ultimate stress for Proto Tagalic, and the existence of at least one minimal pair, he raises the further question:

"Did Proto Tagalic inherit or innovate its ultimate stresses. A careful investigation of the lexicons of Tagalic and other Philippine languages is called for. As time progresses and areas of research increase we should be able to discover just how far back in time we can postulate phonemic stress."

It is here proposed that differential stress in PAN\textsuperscript{9} was the primary conditioning factor for the splitting of PAN *a into the PSB reflexes *a, *u and *ə. According to this hypothesis, allophonic variations of PAN *a existed as a result of contrastive stress, such that unstressed PAN *a in the penult was manifested by the normative allophone [ə]; unstressed PAN *a in the ultima was manifested by the lowered allophone [ʌ] when the vowel of the penult was other than PAN *a, and by the raised allophone [i]\textsuperscript{22} when the vowel of the penult was also PAN *a. Stressed PAN *a in both penult and ultima was manifested by the raised allophone [i]. The distribution of the posited allophones corresponds to the distribution of the PSB reflexes as stated in Section 4.1, so that PAN [ə] > *ə; PAN [ʌ] > *a, and PAN [i] > *u, as in the following:

<table>
<thead>
<tr>
<th>PAN</th>
<th>PSB</th>
</tr>
</thead>
<tbody>
<tr>
<td>*lōŋa [lōŋa] 'sesame'</td>
<td>&gt; *lūnga 'sesame'</td>
</tr>
<tr>
<td>*gōtāh [gōtah] 'sap'</td>
<td>&gt; *gōttaq 'sticky sap'</td>
</tr>
<tr>
<td>*tārəb [tārib] 'quantity, mass'</td>
<td>&gt; (*təqub\textsuperscript{24} 'heavily loaded'</td>
</tr>
<tr>
<td>*lāŋān [lāŋin] 'forearm'</td>
<td>&gt; *lāŋūn 'upper arm'</td>
</tr>
<tr>
<td>*pūtər [putər] 'to turn around'</td>
<td>&gt; *putar 'to revolve; to unwind'</td>
</tr>
</tbody>
</table>
6.1 Stress is indeterminate for words where other factors supercede stress as the primary conditioning factor. Two such cases are suggested by the data.

6.11 Stress is indeterminate when PAN *e in the penult is followed by a consonant which does not geminate in PSB, e.g. PSB *h. The reflex of PAN *e as *w is inhibited, since a following homorganic cluster is an essential condition for this reflex in the penult:

\[
\begin{align*}
PAN & \quad *b\text{e}r\text{as} \quad 'husked rice' > *b\text{w}h\text{as} \\
PAN & \quad *b\text{e}r\text{s}\text{ay}^{25} \quad 'canoe paddle' > *(b\text{w}h\text{s}\text{ay}) > *b\text{u}s\text{ay}
\end{align*}
\]

There are thus two processes by which *w could have been derived from PAN *a in the above examples. If stress occurred on the penult the posited PAN [+] allophone of PAN *e would have been reflected as *w; had stress occurred on the ultima the unstressed PAN [e] allophone would likewise have been reflected as *w, due to the limitation just described.

6.12 Stress is also indeterminate for forms with an infixed PAN *g following penultimate PAN *e. The present data provides no evidence of contrast between *e and *w in this environment. This appears to be more than a data limitation, and the possible phonetic explanation is suggested in Section 7.2. No stress has been assigned to examples such as those given in 6.11 above, or to the following:

\[
\begin{align*}
PAN & \quad *k\text{o}(m)\text{b}ar \quad 'twin' > *k\text{om}b\text{ar}, (*k\text{am}b\text{ar}; PAN *g\text{enti}n \quad 'thin' > *g\text{enti}n \quad 'slim-waisted'.
\end{align*}
\]

6.2 The differentiation of stress carries a low semantic load:

\[
\begin{align*}
PAN & \quad *t\text{e}g\text{ah} \quad 'middle' \quad > *t\text{e}g\text{a}q \quad 'middle (range)' \\
& \quad *t\text{e}g\text{ah} \quad 'n' \quad > *t\text{e}g\text{a}q \quad 'middle (point)' \\
& \quad *(\text{i}f)\text{g}k\text{or} \quad 'to circle, wind' \quad > (\#)(\text{k}a)\text{l}\text{ink}k\text{al} \quad 'to surround' \\
& \quad *(\text{i}f)\text{g}k\text{or} \quad 'n' \quad > *(\text{l}\text{inkw}) \quad (\text{CS.}2) \text{l}\text{nk}k\text{l} \quad 'halo about sun' \\
& \quad *b\text{a}(n)\text{t}\text{e}g \quad 'stomach' \quad > *b\text{at}\text{u}g \quad 'stomach, bulge, pregnant' \\
& \quad *b\text{a}(n)\text{t}\text{e}g \quad 'n' \quad > (*b\text{at}w)^{26} \quad 'mature, green coconut' \\
& \quad *l\text{e}p\text{as} \quad 'loose, free' \quad > *l\text{app}\text{s} \quad 'emerge; pierce through' \\
& \quad *l\text{e}p\text{as} \quad 'n' \quad > (*l\text{up}\text{s}) \quad 'to be lost' \\
& \quad *(\text{t}e\text{g}\text{\text{e}l}) \quad ('persistent') \quad > *t\text{e}g\text{g}\text{w}\text{l} \quad 'long period of time' \\
& \quad *(\text{t}e\text{g}\text{\text{e}l}) \quad 'n' \quad > *(t\text{e}g\text{w}) \quad (\text{CS.}2) \text{t}\text{og}k\text{l} \quad 'persist'
\end{align*}
\]

In some cases the reflexes of stress-differentiated PAN pairs show no current semantic contrast in SB languages:
6.3 As has been mentioned in Section 5.13 there are competing hypotheses for the proposal that PAN *a in the ultima had stressed and unstressed allophones [e] and [ʌ] which were reflected in PSB as *e and *a respectively. The following arguments are not conclusive, but they warrant serious consideration of the proposal.

6.31 The parallelism between the split in the penult and the posited split in the ultima is phonetically plausible. In both syllables the phonetically higher reflex is that of the posited stressed allophone of PAN *e. The unstressed allophones and their PSB reflexes differ in phonetic height, with the lower occurring in the ultimate syllable.

6.32 If two languages, not closely related and not in contact, share a distinctive phonological phenomenon there is reason to seek an explanation in similar phonological processes. Tagalog, for which the regular reflex of PAN *e is i, and Sama-Bajaw both show significant numbers of words which exhibit ultimate a instead of the expected reflex of PAN *e. Llamzon²⁷ cites some doublets which parallel the SB data:

batás, batás 'to ford a creek' < PAN *batas 'boundary'; hilás, hilás 'cut'; lapit²⁸ 'close, draw near', lapet 'fitted in' < PAN *rapet 'joined together'; sapit 'arrived', sapet 'attached'.

One would expect, if doublets are the correct explanation for these a reflexes in both languages, to find evidence of the same doublets in other languages of the Central and Southern Philippines. This evidence does not appear to be abundant in Cebuano Bisayan, but more data are needed, both from CB and other languages such as Bikol, Maranao, Subanon and Sangir.

If *(baRaq) 'ulcer, abscess' is reconstructed as a doublet (alongside PAN *baRaq) on the evidence of Tg. bagaq 'abscess' and PSB *bahaq 'swelling; boil', we must suppose the persistence of both members of the doublet down to the Proto Tagalic stage at least. Cf. CB bagaq 'disease characterised by... swelling in the region of the stomach' (and also CDu bagaq 'to be swollen').

The persistence of such doublets is not unprecedented, to be sure, but it is significant that the two languages which justify the posited doublet are languages which exhibit similar alternations between a and the regular reflex of PAN *e. It thus seems reasonable to consider the
possibility that we are not dealing here with a doublet but with a distinctive, shared phonological development.

6.33 The hypothesis of contrastive stress in the ultima is consistent with the alternations displayed within PSB, not all of which have an established PAN source. In the first example below the shift of stress to the ultima is accompanied by reduction of original PAN *i in the penult to *ə29:

PAN *lipet 'to fold' > *lipet > 'to fold in two'
*lepêt)30 > *lepêt > 'to fold up'
*táhəp) 'to winnow' > *tahap
*tahép) > *(tahup) > (AB) tahup

7. THE DEVELOPMENT OF GEMINATION IN PSB

Universal penultimate stress is reconstructible for PSB. It is therefore proposed that the posited contrastive stress of PAN became non-contrastive in PSB. In all the PSB daughter languages which reflect it ə is a short vowel. It does not occur with the vowel lengthening which is the usual concomitant of stress in the daughter languages (and presumably reconstructible as a feature of stress in PSB also). It has been proposed in Section 6 that PSB ə only occurred as the reflex of unstressed PAN *ə in the penult. It is further proposed that with the shift of stress from the ultima to the penult the short vowel was retained and that compensatory length was carried by the following consonant, thus causing gemination. Thus PAN *bekás 'inheritance' becomes, with the regularisation of stress, PSB *békːəs 'trace of former occupancy'.

7.1 In (NB), the only languages which do not have universal penultimate stress31, stress falls on the ultima only when the vowel of the penult is ə. In (NB), furthermore, the length of the medial geminate is substantially reduced, so that it no longer carries the length feature of the stressed syllable. Stress thus reverts to the ultima. In the following examples the raised period indicates a very slight lengthening of the consonant; rather more when the ə is initial:

PSB *emmaq 'father' > (NB) omaq32 [m·aq] and > (CS) emmaq [əmaq]; PAN *beb(n)tis 'skin-bone' > *bettis 'leg' > (NB) bētis [bēt·fs] and > (CS) bêtis [bēttis].

(CS) languages display an extreme degree of transference of syllabicity from stressed ə to the following consonant. The phonetics below are simplified with respect to the allophones of the other segments:
FSB *ænum 'six' > (CS.2) ënom [ánom]; PSB *ëmpit 'to hold a child (on the lap)' > (CS) ëmpit [ápit]; PSB (*)ggok 'to make a gulping motion' > (CS.2) gguk ['gguk] (the initial g is slightly imploded); PSB (*)&ttus 'disease which pocks the skin' > (CS) *ttus [ítus] (tongue tip is held in the stop position for the normal duration of a stressed vowel).

7.2 Where PSB reflects a PAN word in which PAN *æ in the penult occurred before an infixed *ŋ, only *ə occurred as the reflex of PAN *æ. This is not surprising; the nasal carried the compensatory length when stress occurred on the penult, and the normative vocalism of PAN *æ was maintained. This environment, it is presumed, conditioned the allophones of PAN *æ so that [æ] (rather than [i]) was the allophone of stressed PAN *æ in the penult before a nasal-stop cluster. Since both stressed and unstressed PAN *æ in this environment have the reflex *ə it is not possible to determine stress.

8. PHONEMIC STATUS OF THE PSB REFLEXES *ə AND *u

We now consider the question of the phonemic status in PSB of the PSB reflexes *ə and *u of PAN *æ. It has been proposed in the previous section that contrastive PAN stress became non-contrastive in PSB. The contrastive function of stress in PAN was thus taken over by phonemically contrastive vowels, as witnessed by the phonemic status of the reflexes of *æ and *u in all daughter languages except (YK)33. PSB thus represents the completion of a transitional period during which contrastive stress plus sub-phonemic vocalic contrast gave place to non-contrastive stress and phonemic vowel contrast. The gemination of the medial consonants was a by-product of the change in the pattern of stress placement, and played only a secondary role in marking phonemic contrast.

9. PROCESSES BY WHICH æ OR u IS INNOVATED IN SB

There are four processes by which æ or u is developed from vowels other than PAN *æ. In some cases the development is reconstructible for PSB; in others it appears to have taken place subsequent to the subdivision of PSB.

9.1 The first of these innovations is based on the analogy of stress-conditioned reflexes of PAN *æ in a reduplicated monosyllable. The effect of contrastive stress on the PSB reflexes is displayed in the
following examples; with ultimate stress the heterorganic cluster of reduplicated monosyllables is assimilated to a geminate and the vowel of the penult is reflected as *ə:

PAN *kémkem 'to keep close' > *(k)umkum 'to hold in closed fist';
PAN *søpsøp 'to suck' > *søssøp; PAN *køpøp 'to grasp' > *(k)øpøp
but PAN *køpøp > *(k)økøp.

A similar process based on this pattern may operate on a reduplicated monosyllable with a vowel other than *ə. The medial cluster becomes a geminate, and the vowel of the penult is sometimes innovated as o:

PSB *kakkab 'to fan' > *(kakkab) > (YK) kekkab; *(k)utkut 'to scratch' > *(kakkut) > (ZB.3) kakkut, and (CS.2) kakkut (where the vowel of the penult in both cases is a secondary development); *søssøk 'narrow gap' and (CS.2) sōsok < *(swksøk) suggest a back formation on the same analogy, although the (CS.2) form may be original, lost in other languages to the innovating modification.

Assimilation of a heterorganic cluster does not, however, obligatorily require this vowel replacement:

*(b)ukbug 'staroh food cooked till mushy' > (YK) būbug 'rice gruel'.

The alternation also occurs within a single daughter language, marking minor semantic modification:

(CS.2) loklok 'to withdraw into a hole in order to avoid detection';
(CS.2) lollipop 'to be just below surface, as a nail driven into wood'.
(The CS.2 vowel o is a merger of two PSB proto-vowels, *u and *o, so that the above forms may be reflections of PSB *(lukluk).)

9.2 The second innovation is based on the alternation between *ə and some other vowel (not a) in the penult, with the regular gemination of the medial consonant after the innovated ə. There are precedents for this pattern in PAN:

PAN *kumur, *kamur 'to move to and fro in the mouth'.

In some cases the innovation is of PSB provenance:

(*b)attad 'to stretch out; to set s.t. out' and (NS, CS) bitad 'to pull taut'. Cf. PAN(B) *bi(ŋ)Otad 'spread out'.

(*b)ulud 'mountain' (in NS.7-10, WS, CS, SS.1-3) and *bollud 'mountain' (in ZB, NS.2-3, SS.4-5, NB, IB).

The alternation is also exploited to mark semantic modifications within a single daughter language, apparently as a fairly recent innovation. A few examples are presented from CS.2, where the alternation is abundant although no longer an active process:

babas 'to abate, as wind'
babbas 'to abate, of swelling'
bakat 'out in flesh'
bakkat 'to break a line'
kilat 'lightning; to flash'
kellat 'to show light, of sun'
kullit 'outer layer, skin'  kellit 'leathery'
tigan 'dried out'  toggan 'barren womb'
tubbud 'to well up, as water'  tebbud 'to billow up, as smoke'

9.3 PAN *a in the ultima before final PAN *h is regularly reflected as PSB *u:

PAN *kimah 'giant-shell' > *kima 'giant clam'; PAN *duwah 'two' > *duwu; PAN *talighah 'ear' > *taligw.

9.4 PAN *a in the penult sometimes becomes *u when the vowel of the ultima is the *u reflex of PAN *a, and the medial consonant is *l or *r:

PAN *balós 'to retaliate' > *balus or *bulus 'to reciprocate an action'; PAN *tazam 'sharp' > *tarum or *taram.

10. REFLEXES OF PSB *a AND *u IN DAUGHTER LANGUAGES

PSB had seven phonemic vowels after the regularisation of stress placement on the penultimate syllable:

```
i  u
 e  o
 a
```

Two daughter languages, (ZB.1, WS.1) have retained this seven-vowel system. (ZB.2-3, NS, WS.2, CS, SS, JM, NB, IB) have all reduced the PSB proto-system to a six-vowel system by merging *u with another phonemic vowel (with either *a, *o or *u). Some (NS) and (IB) languages are in a stage of transition between six and five vowels, as e, a phonetically unstable segment, merges diversely with other phonemic vowels. (YK) has a five-vowel system resulting from the merger of both *a and *u with *e, and (AB) has a three-vowel system resulting from the merger of *o and *u with *u, of *a with *a, and of *e with *i.

10.1 The reflexes of PSB *a and *u are now presented for each of the daughter languages (V₁ refers to the vowel of the penult; V₂ to the vowel of the ultima):

10.11 In the antepenult *a > a in (AB); > i in (ZB, NS-SS, JM), usually; > e or i in (YK); > u or o in (NB); > a or i in (IB):
*təmboroq 'finger'. (AB) tambaruq 'hand', (ZB) tomboq < *(tomboloq) < *(tomboloq), (YK) timbooq 'index finger', (NB) təmboro 'index finger'.

10.12 In the penult *ə > a in (AB); > e in (ZB, NS-SS, JM, NB, IB), with secondary modifications as indicated for (ZB, NS, IB); > e in (YK):

*battuq 'stomach; pregnant'. (AB) battuq, (ZB.3, NS.2, CS, SS.2, JM) bətətq, (WS.1) bətətq, (WS.2) bətətq, (NB.1) bətətq, (IB.8) bətəh.

*aṃmaq 'father'. (ZB.1, NS-SS, JM) aṃmaq, (NB) aṃaq, (IB.1) aṃmaq 'mother'.

10.121 In (ZB.2, NS, IB) e is an unstable segment, and secondary processes are still continuing, by which e is replaced or merged with an existing phoneme. In (ZB.2) and (NS) two processes operate and compete with each other and with the inherited reflex e. e usually persists when it is initial in a word; elsewhere e is sometimes (apparently unpredictably) replaced by e (especially in NS.6), or e assimilates to V₂ (especially in ZB.2, NS.1,9-10):

*laŋgən 'upper arm'. (ZB.1, WS.1) laŋgən, (ZB.2, NS.2-3,7-10, CS.4, SS.4) laŋgon, (ZB.3, CS.1-3,6, SS.1,3) laŋgon, (YK) leŋgon, (WS.2) leŋgon.

*bəlla 'to cook'. (ZB.1, NS.3,6-7, WS.2, CS.2-6, SS.1,4, JM) bəlla, (ZB.2-3, NS.10) bəlla, (YK, NS.3,5,8) bəlla.

10.122 In (IB) e usually persists when it is initial in a word; elsewhere it is usually replaced by i when V₂ is *e or *a, and sometimes replaced by i or u in other environments:

*bəlla 'to cook'. (IB.1) billa.

*bəlli 'to buy'. (IB.1) bəlli, (IB.2) bəlli.

*bənnəd 'muscular cramp'. (IB.1) binaq.

*bərrət 'heavy'. (IB.1) birraq, (IB.4) bərraq.

*dəppə 'fathom'. (IB.1) duppa, [IB.9] dəpe.

*ałłum 'to be alive'. (IB.1) ałlamaq, ałloq.

*ałloq 'neck'. (IB.1) killoq, (IB.2) kulloq, [IB.9] kəlo.

10.21 In the antepenult *u > a in (AB); > u in (ZB, YK, NS-SS, JM); > u or a in (NB); > a or i in (IB):

*dəwaqə 'to descend'. (AB) dəwaqi, duwaqi, (ZB, YK, NS-SS, JM, NB) duaqi.

*kəhəpoq 'grouper; roc kood'. (YK, NS.3, CS.2) kəhəpoq, (IB.1) kiəpuq.
10.22 In the penult *u > u in (AB, WS.2); > u in (ZB.1, WS.1); > o in
(ZB.2-3, NS, CS, SS, JM, NB); > e in (YK); > o in (IB) when both V₁ and
V₂ are *u (there are no data for *u in penult and a different vowel in
the ultima):
*butaq 'to place; to install'. (AB) butaq, (WS.1) butaq, (CS.2-3)
butaq.
*dweq 'to desist from doing s.t.'. (YK) deheq, (NS.1, CS) doheq,
(IB.1) doqoh 'to abate, of a storm'.
*bulus 'to reciprocate an action'. (AB) bulus 'to take revenge',
(NS.1, NB.1) boos 'to borrow', (WS.1) boos 'to borrow'. (NB.4) bohos <
*boos) 'to take revenge'.

10.23 In the ultima, in an open syllable, *u > a in (AB, ZB, NS-SS,
JM, IB); > e in (YK); > o in (NB):
*haya 'big'. (AB, ZB.1-2) haya, (YK) haje, (CS, SS) hea, (NB) oyo.
*sau 'snake'. (AB) sawa, (ZB, NS-SS, IB.1) soa, (YK) sawe, (JM,
NB) soo.

In the ultima, in a closed syllable, *u > u in (AB, WS.2); > u in
(ZB.1, WS.1); > o in (ZB.2-3, NS-SS, JM); > e in (YK); > a in (IB):
*anum 'to weave a mat'. (ZB.1) anum, (ZB.3, NS.3, CS.2-3, SS.2-4)
anum, (ZB.2, WS.2) anum, (NB.1-2) ṇanum, (IB.1,4-5) anaq.
*dakap 'to seize'. (CS.3) dakap 'to abduct a woman', (NB.1-2) ḍakap,
dokap, (IB.9) ḍakap.
*anum 'six'. (AB) anum, (ZB.1, WS.1) anum, (ZB.2, NS.3,5,10)
onum, (ZB.3, NS.2,7-9, WS.2, CS, SS, JM) onum, (NB.1,3-4) onum,
(IB.1,5) onaq.

10.24 In (NB) a secondary replacement of the o reflex of *u is taking
place, such that o > a. There is considerable alternation between a
and o in (NB), in both directions, possibly under the influence of
cognates with Malay which show a regular correspondence between a and o
in the ultima.

*kettep 'to slice, as meat'. (NB.1) getap 'to chop small'.
*anum 'six'. (NB.1,3-4) onum, (NB.1-2) onum, ML. onam.
*(h)mek 'to tie'. (NB.2) āmek, (NB.1) āmek, āmek, ML. ikat.
*sarret 'tightly wedged'. (NB.1) sorot, (NB.2) sarat, ML. sarat
'jammed'.
*sessor 'to compel s.t. to move'. (NB.1) sasar, sesor 'to drive
away', ML. sesar 'to push s.t. aside'. Cf. also ML. pertama 'first',
borrowed as (NB.1) pərtom. 
11. CONCLUSION

In order to facilitate the description of the Proto Sama-Bajaw reflexes of PAN *e - an interesting and rather complex set of data - we have proposed that contrastive stress was a feature of a language ancestral to PSB, though not necessarily of Proto-Austronesian. Most of the SB data are explicable in terms of this hypothesis; a borrowing or doublet hypothesis is indicated for only a small proportion of the corpus.

However, satisfactory though the hypothesis appears to be for the PSB data, it needs to be tested against the data of other PAN sub-groupings. Of particular value, it seems, are the data of those languages in which PAN *e splits into two or more phonemes, and those languages which exhibit secondary developments such as gemination of the medial consonant after penultimate PAN *e. Another useful set of data may be provided by languages which, like Sama-Bajaw, have phonemic contrast between u and o, and between i and e. Such splits may be due in part to contrastive reflexes of stress-conditioned allophonic variations of PAN *u and PAN *i.
NOTES

1. Sama is a widespread autonym (from Sulu, Sabah and Indonesia), and is reconstructible as the proto-form by which speakers of the language referred to themselves. Bajaw is the Malay designation used throughout Malaysia and Indonesia, and for one group in the Southern Philippines. The compound name of the title has been chosen to facilitate identification, and at the same time to retain the autonym.

Other names which occur currently and in the literature are Orang Laut, Samal (the usual external designation for Sama in the Philippines), Luwaqan (nomadic Philippines Sama) and "Sea Gypsies".

2. The following conventions have been followed: Proto-Austronesian reconstructions have been written in Dyen’s orthography with the exception that the pepet is written as *e, and the symbol e retained to indicate the mid-front vowel [ɛ] which is phonemic in Sama-Bajaw. e indicates a high back-central unrounded vocoid; i indicates a high central unrounded vocoid. I have followed Blust in writing the PAN infixed nasal as *n in his reconstructions, but have not modified the VLAW material. Malay citations have been modified by writing the non-syllabic high vocoids as -w and -y, the digraph ng as η, and e as œ where it represents the mid central vowel.

Since we are dealing with two levels of reconstruction the convention is followed of prefixing all Proto-Austronesian forms with the abbreviation PAN, and only prefixing Proto Sama-Bajaw forms with PSB when there is a possibility of confusion. PSB forms well-attested in Sulu languages, but unattested outside them, are indicated by (*), and hypothetical forms by *(...); thus PAN *bélah 'to split' > *(bułaq) > (CS.2) bułaq 'bamboo lath'. Primary stress is marked by a grave accent over the vowel of the stressed syllable. [...] enclosing a language code indicates that there is some uncertainty about the phonetic interpretation of some part of the following citation. Language citations appear
bold type and are identified by a preceding letter code in the case of Sama-Bajaw languages, or by an abbreviation.

Sources and abbreviations are as follows:

PAN: Proto-Austronesian (Dempwolff); PAN(B): Proto-Austronesian (Blust); PSB: Proto Sama-Bajaw; Batak (Rodda); CB: Cebuano Bisayan (Wolff); CDu: Casiguran Dumagat (Headland & Headland); Ho.: Nova (VLAW); Ja.: Javanese; Mar.: Maranao (McKaynah and Macaraya); Ml.: Malay, Indonesian (Echols and Shadily); Mwa.: Mamanwa (Miller and Miller); Minahasa and Togian Bajaw (Adrian); PKr.: Papia Kristang (Hancock); Proto East Mindanao - a subgrouping of Southern Bisayan which includes Mamanwa and Tausug (author's fieldnotes); SB: Sama-Bajaw daughter languages, collectively considered; Skr.: Sanskrit (Gonda); Tg.: Tagalog (English); Tsg.: Tausug (Tagusuug) (Ashley); VLAW: Vergleichende Lautlehre des Austronesischen Wortschatzes; WBM: Western Bukidnon Manobo (Elkins).

Sama-Bajaw daughter languages have been assigned letter codes, with following letters indicating dialect or sub-dialect variations within each coded section. The expansion of the code is indicated as a mnemonic aid: (AB-Abaknon): Abaknon or Capuleno (Capul Island, Central Philippines, between Masbate and Northern Samar); (ZB-Zamboanga): Bataan, Lutangan and Sibukuq Sama (Sibuguey Gulf and Zamboanga Peninsula, Mindanao); (YK-Yakan): Yakan (Basilan Island); (NS-Northern Sulu): Baanging and related dialects (northeastern Sulu, with migrant groups in other parts of the Philippines); (WS-Western Sulu): Pangutaran and Soang Bunaq (islands about 35 miles north-west of Jolo Island in Sulu); (CS-Central Sulu): Sama Dilaut (nomadic and semi-nomadic, from eastern Sabah to Zamboanga) and several related dialects spoken by sedentary groups (central Sulu); (SS-Southern Sulu): Tanduq Baaq, Simunul and Sibutuq (southwestern Sulu and eastern Sabah); (JM-Jama Mapun): Mapun (Cagayan Sulu Island, north of Sandakan, Sabah); (NB-North Borneo): West Coast Sabah Bajaw (Labuan Island and the coastal fringe from Papar to Banggi); (IB-Indonesian Bajaw): Indonesian Bajaw (Indonesian Timor, Roti Island, Kajoa in the western Halmaheras, and various places around the Sulawesi coast). (CS.2) indicates Sama Dilaut of the Siasa Lagoon area, Siasa, Central Sulu, the language with which the writer is best acquainted. (NS-SS) includes (NS, WS, CS, SS), a single section characterised by a chain of mutual intelligibility. Sulu languages include (ZB, YK, NS-SS, JM).

The following acknowledgements are made of unpublished sources: James Fox (Roti data (IB.3)), Charles Frake and Carol Molony (Kajoa, Halmahera data (IB.2)), Gerard Rioux (Sulawesi data (IB.4-6)), and various members of the Summer Institute of Linguistics, Philippines, especially Dietlinde Behrens and Janet Pack (YK), Jeanne and Helen Miller
(Mamanwa), Charles Walton and William Hall (various Sama word lists) and Siocon Subanon (Hall).

3. Dyen (1965) assigns one language of the sub-group, Yakan, to the Philippine Hesion as a first order sub-grouping, and indicates the problematic nature of the assignment by a double asterisk.

4. The word list is a list adapted by Richard E. Elkins from the Swadesh list for the use of the field workers of the Summer Institute of Linguistics in the Philippines. A comparison of cognate counts between the Swadesh and the Elkins list indicates a difference of ± 4%. The Elkins list forms the basis of Philippine Minor Languages: Word Lists and Phonologies (Reid, ed.).

5. Conant op. cit.

6. Blust's (1970) assumption has been followed that the data are best explained by positing a single segment *q which was sometimes infixed preconsonantally in PAN, but Dempwolf's reconstructions have not been rewritten in accordance with this assumption.

7. A minor exception to this statement occurs in the antepenultimate syllable, where PSB *a occurs without a following homorganic cluster. This case is discussed in Section 4.2.

8. Since the daughter languages show a wide diversity of reflexes of PSB *u (e, a, o, u) there was no clear phonetic identity for this proto-form. *u was chosen for pragmatic reasons: (1) it is a unique symbol for the correspondence set, and (2) in the only two languages which maintain the seven-vowel phonemic system of PSB the current phonetic realisation of this proto-phoneme is [u], a high back-central unrounded vocoid. In view of the other reflexes of PSB *u no assumption is made about the rounding of the proto-phoneme.

9. Since Dempwolf's and Blust's reconstructions are based essentially on languages of Dyen's Hesperonesian Linkage, the proto-language of the forms indicated as PAN in this paper are more correctly of Proto-Hesperonesian provenance. The ancestor of PSB for which contrastive stress is posited is thus Proto-Hesperonesian or some lower order proto-language. More work is needed before such mid-level groupings can be satisfactorily identified; the reconstruction of reliable forms for low level proto-languages is a step in this direction. The label PAN in this paper, when it refers to the stress hypothesis, should be read as shorthand for "PAN or some lower order proto-language of Proto-Hesperonesian".
10. These examples, because of their relatively limited range of occurrence, are susceptible to interpretation as loans from a language which has e or a reflexes of PAN *e in both penult and ultima, i.e. a language such as Proto East Mindanao and its daughter language Tausug. Upland Tausug maintains the contrast between e and u, lost in Lowland Tausug. There are, however, other data showing this vocalism which have good PSB credentials though without an identified PAN source: *dahag 'to desist from doing something'.

11. Ml. and Ja. a in the antepenult may derive from either PAN *e or *a, an ambiguity which Blust indicates where appropriate, in his Adenda. Malay regularly, though not invariably, replaces original a with e in the antepenult: merdeka 'free' (Skr. maradeka); kepala 'head' (Skr. kepāla 'skull'); fatwa (Ar.) 'instructions' and petua 'religious advice', where the Arabic word has been modified in the second form to conform with Ml. phonology.

12. *u and *i do not constitute a problem since their reflexes are invariable in all PSB languages (except (NB), which has extensive vowel reduction in the antepenult).

13. (YK) and (NB) are similar to Malay in having a rule by which phonemic a is modified in the antepenult: (YK) saqi 'mother' and seqinen 'his mother'; (NB.1) papag 'to beat' and pinapagan 'beaten'. Cf. (CS.2) saquina 'his mother' and binabagan 'provided with a cross-beam' (from babag 'cross-beam'). The above illustrations reflect only part of a more complex process in both languages. There are other constraints for (YK) on the reduction of a to e, and for (NB) other vowels and other environments are implicated.

14. (YK) reflexes are about evenly divided between e and i, and (IB) reflexes between a and i. There is, however, no correlation between these sets of variations, and insufficient data in either case to suggest any conditioning hypothesis. It should be noted that in (IB) *e has multiple and largely unpredictable reflex potential in the penult as well. In some cases there are currently competing forms: (IB.1) limbaq, lembaq, lumbaq 'to repeat'.

15. I have not included analogic replacement of phonemes since, except for those cases where innovations can be shown to have spread by some stitable process (Zorc's (1972) morphological regularisation of stress to indicate morphemic contrast between segmentally identical forms is a good example), this explanation is frequently a less satisfactory treatment of the data than simply noting that the problem is unexplained.
16. The correct solution may, of course, require a combination of these hypotheses. Doublets may not be of proto-language provenance, but may be the result of early inter-dialect or inter-language borrowing. Once borrowed they become inherited forms of the daughter languages, subject to native laws of sound change. Again, since sound change processes lose their impetus or change phonological direction, the phoneme X of a loan word may undergo different (though regular and describable) processes from those which operated earlier on a phonemically identical phoneme in an inherited word.

17. An alternative explanation is that the Malay source was a dialect of M1. which realised standard Malay e as a, although this is not the case for current contacts between Sulu languages and Malay.

18. Llamzon, op. cit.

19. It is not necessary to suppose that a source language form was borrowed in the form in which it currently appears. Lowland Tausug, which reflects PAN *e as u in all positions, and once assimilated the e of borrowed Sama forms in the same way (Tsg. kuddum 'to frown' probably from PSB *kaddum 'to close the eyes'), now assimilates borrowings from Sama languages in a different way, generally by harmonising Sama e with the final vowel of the word:

Sama danda 'woman' becomes danda; Sama pessi 'fishhook' becomes pessi; Sama emboq 'ancestor' becomes umbuq when spoken by Tausug who do not control the phonology of Sama.

20. PAN *tempet 'domicile' is poorly attested in VLaw.

21. Innovated e is discussed in Section 9. This is a fairly typical example of the process described in 9.2.

22. Dempwolff (1934–8), op. cit.

23. The harmonising of the vowels in this context may have been a PSB process; the evidence is ambiguous.

24. This word is possibly a loan from a Philippine language in view of the *g reflex instead of the expected *h reflex of PAN *R. Medial PAN *R is reflected as PSB *g more frequently than initial and final PAN *R, and there may be conditioning factors. Cf. also such forms as *peggaq < PAN *paRaq 'to squeeze out juice', and the doublet *taguq/ta(h)uq 'to keep', both well attested for PSB. The doublet probably reflects, respectively, PAN(B) *(Ct)aguq 'hidden' and PAN *taRuq 'to set down'.
25. Dempwolff (1925) cites this form, but only *besay is given in VLAQ. The *R inflix is supported by such Philippine reflexes as Proto East Mindanao *bogay *canoe paddle*.

26. Both PSB forms appear to derive from PAN *bọtọ. The form (*bọtọ 'mature, green coconut' though an innovation with this meaning, may be an example of the exploitation of stress contrast. The semantic connection is at least plausible, with a shared component of swelling and fruition.

27. Llamzon, op. cit.

28. Blust (1970) proposes an alternative etymology for Tg. lapiṭ: PAN(B) *ra(ŋ)pi(ct) 'approach'.

29. See Section 9.2 for a discussion of this innovation pattern.

30. The change in stress here may be an innovation of early PSB provenance, and the form in parentheses here is included only to facilitate comparison with the established PAN form. No claim for PAN status is intended.

31. I have no direct data for (AB) stress, though the material was recorded by an experienced field worker who would, I expect, have indicated contrastive stress. In any case there is no evidence in the corpus of some six hundred (AB) words of homophonous pairs for which one might expect stress to be the contrastive feature, and it is assumed with reasonable confidence that the above statement is equally applicable to (AB).

32. The phonemic status of the single versus geminate consonant contrast in (NB) is uncertain. It carries a low contrastive load in any case, even when the vowel of the penult is other than e.

33. (YK) reflects PSB *e and *ε as e in all environments (merging these both with *e), and *a as e under predictable, stress-governed conditions.

34. The PSB reconstruction of this phoneme is based on the correspondence set: -e in (YK), -o in (NB), -a elsewhere. The (YK) form is crucial for the identification of this set, since the (NB) evidence is somewhat ambiguous. In (NB) there is a secondary development whereby a > o in an open final syllable, and penultimate a > o when the final vowel is o. Thus:

*matə 'eye' > (NB) moto; *hadən 'ladder' > (NB) odon.
A similar centralising process takes place with PAN *u and PAN *v before final PAN *h or PAN *q, with these variants reflected in PSB as *o and *e.
REFERENCES

ADRIANI, N.  
1900 Tijdschrift voor Indische Taal-, Land- en Volkenkunde.  

ASHLEY, S.  

BLUST, Robert A.  

CONANT, Carlos Everett  
1912 'The pepet law in Philippine languages'. Anthropos 5:920-47.

DEMPWOLFF, Otto  

1934- Vergleichende Lautlehre des Austronesischen Wortschatzes.
1938 Band I-III. Berlin: Dietrich Reimer.

DYEN, Isidore A.  

1965 A Lexicostatistical Classification of the Austronesian Lan-

ECHOLS, John M., and Hassan SHADILY  
1961 An Indonesian-English Dictionary. New York: Cornell Uni-
versity Press.
ELKINS, Richard E.

ENGLISH, Leo James

GONDA, J.
1952 Sanskrit in Indonesia. Nagpur: The International Academy of Indian Culture.

HANCOCK, Ian F.
1973 'Malacca Creole Portuguese: a brief transformational outline'. Ta Rêo 16:23-44.

HEADLAND, Thomas N., and Janet D. HEADLAND

LLAMZON, Teodoro A.

McKAUGHAN, Howard P., and Batua A. MACARAYA

MILLER, Jeanne, and Helen MILLER

REID, Lawrence A., ed.

RODDA, Rosemary, and Audrey MAYER

ZORC, David R.