TANIMBAR-KEI: AN EASTERN INDONESIAN SUBGROUP

ROGER F. MILLS

1. INTRODUCTION[1]

Three languages of eastern Indonesia appear to form a distinct subgroup: Fordata and Yamdena, spoken in the Tanimbar Islands, and Kei, in the Kei Islands. We owe our knowledge of these languages to the Roman Catholic missionaries Drabbe and Geurtjens, who worked in the early decades of this century. They seem to have had an implicit understanding of the relationship between the languages, but they were not comparativists, and their work was not oriented in that specific direction. Fortunately, the dictionaries and grammars they left us are of very high quality (Geurtjens's peculiar spelling system notwithstanding), and provide excellent foundations for comparative work.

The first modern discussion of these languages is found in Dyen (1965), where on the basis of 'adequate' lists Kei is ungrouped within the Moluccan linkage, its highest percentage (20.9) with Leti. Unfortunately, the two Tanimbar languages were represented by 'subadequate' lists, but showed higher percentages: Fordata-Kei 37.6%, Fordata-Yamdena 31.5%. Dyen (p.45) goes on:

> *Fordat* forms the Fordatic subgroup with Kei...Whether Fordatic is a closed group depends on Fordat's percentage with Jamden. If this percentage is reliable... Jamden is a member of the same subgroup.

Yamdena, he says, "may be no more closely related to Fordat than to Tettum (27.3%)". (Presumably Yamdena-Kei scored even lower than that.) But the resemblances between Fordata and Yamdena, and their divergences with Tettum, are, as we like to say, obvious upon inspection. And frankly, the percentages are not reliable, at least in part because of very conservative judgments as to cognacy,[2] even though in a relative way a correct picture of the relationship happens to emerge.

---

[1]To my regret, I have not been privileged o know Professor Grace except through his many scholarly con ribu ions. It is nonetheless a great honour to present my work to him on his occasion.
Considerations of space have forced upon me an unaccustomed brevity. Thus, some of the presentation may seem qui e summary – but it is certainly not without ample supporting evidence, almost all of which is readily available in well-known publications. If some of my claims seem to require leaps of faith, be assured I do not mean to be dogmatic. The much longer draft from which this paper is condensed is available to anyone in eres ed.
I wish to express my appreciation to the University of Michigan's Center for South and Sou heast Asian Studies for i s assis ance and logistical support over several years, wi hout which my research would have been much impeded.

[2]Given the availability of good dictionaries, whose several thousand en ries afford ample data for detailed comparison, I have never seen the vir ue of sor ing out just 200 words from each. Happily, two recent Indonesian studies (Tetelepta et al. 1985; Pieter et al. 1986) contain word lists recorded from informan s, from which a 'subadequa e' Yamdena-Kei list could be compiled. It contained just 125 out of 197 items but, of these, 61 (or 49.2%) were clearly cognate. Given that Yamdena and Kei are the most distan ly related of the hree, Fordata-Kei must surely score higher than Dyen's 37.6%.

Robert Blust, ed. *Currents in Pacific linguistics: papers on Austronesian languages and ethnolinguistics in honour of George W. Grace*, 241-263.
Pacific Linguistics, C-117, 1991.

It is also probably true that Tanimbar-Kei and Tettum ultimately subgroup together, but at what level is not clear. I hope in this paper to offer a tentative answer to that question, but limitations of space preclude detailed treatment.

A more recent discussion, based on fieldwork and a "cursory examination of Geurtjens and Drabbe", is found in Collins (1982). He considers Kei and Fordata to be members of a "far-flung dialect chain". Since the published sources describe only the two extremes of this putative chain, it is quite possible that intervening areas provide the missing links, but in my opinion Fordata and Kei are now clearly distinct languages (see section 4.2).

2. NEIGHBOURS

Drabbe (1932a) reports on a third Tanimbar language, Selaru. Its vocabulary seems to show strong resemblances to Yamdena, but the word list is short and one has the impression that Drabbe's choices may have been influenced by his greater command of Yamdena. As will be shown, on phonological grounds Selaru is less closely related to Fordata, Kei or Yamdena than they are to each other.

Yet a fourth Tanimbar language, Selwasa, is brought to light in Collins (1982); it is mentioned by Coward (1989:2) as being closely related to Selaru ("at 56% lexical similarity").

A glance at a map shows that Tanimbar and Kei are both quite large and compact island groups; Kei lies some ±200km north-north-east of Tanimbar. To the east (±200km) their only neighbour is the Aru group, whose languages are evidently Austronesian but very divergent (see van Eijbergen 1864; Collins 1982 tentatively groups Aru with Tanimbar-Kei). West of Tanimbar (±150km) the nearest large island is Babar and environs, whose known languages belong to the Lettic group.[3] Between Kei and the eastern tip of Seram lie several small islands. Tetelepta et al. (1985) mention Kur and Teor as languages spoken in that area but not subgrouping with Kei; the brief discussion in Collins (1982:122-123) confirms this view. On Kei itself, a language called Eli-Elat (after the two villages where it is spoken) is said to be the survivor of the original language of the Banda Islands, whose inhabitants were dispersed in the seventeenth century. (See Stresemann 1927:182-190; he held that Eli-Elat did not subgroup with his Ur-Ambon, nor does it appear to subgroup directly with Tanimbar-Kei. Collins, again, confirms this view.)

3. BRIEF PHONOLOGICAL HISTORY OF TANIMBAR-KEI

Comparison of Fordata, Kei and Yamdena allows the fairly straightforward reconstruction of Proto Tanimbar-Kei (PTK) with the following sound system:[4]

Stops: *-p(?), b, mb, t, d, nd, k, ŋg*
Nasals: *m, n, ŋ*
Resonants: *w, r, l, y*

---

[3]That is, Wetan, so well reported in Josselin de Jong (1987, researched in the early 1930s), and 37 words of a language called "Tanimber (Baba)" recorded in Earl (1848), which shows resemblances to both Wetan and Kisar.

[4]Abbreviations used are as follows: AMB – Ambonese, AN – Austronesian, CMP – Central Malayo-Polynesian, EIN – Eastern Indonesian, EMP – Eastern Malayo-Polynesian, F – Fordata, GAL – Galoli, K – Kei, KIS – Kisar, LET – Leti, LS – Luang-Sermata, MN – Melanesian, OC – Oceanic, PBD – Proto Barat Daya, PBDS – Proto Barat Daya-Selaru, PCMP – Proto Central Malayo-Polynesian, PLet – Proto Lettic, PLS – Proto Lettic-Selaru, PSEM – Proto SE (Southern) Maluku, PTK – Proto Tanimbar-Kei, ROT – Roti, SHWNG – South Halmahera West New Guinea, SLAR – Selaru, TET – Tettum, TIM – Timorese, WET – Wetan, Y – Yamdena.

Voiceless continuants: *f, s*
Vowels: *i, e, a, u, o*

The PTK sound system reflects that of AN and CMP (at least as I conceive it) as set forth in Table 1. Crucial to the definition of PTK as a distinct entity are the merger of AN **d,D* and **z,Z* > PTK **d*, the merger of AN **r, j, R* > PTK **r*, the loss of AN **-w* and **-y*, and the treatment of the vowel sequences that arose after loss of the AN/CMP laryngeals. Some of the reflexes in the Table require further comment:

(a) PTK **-p* and **b* are needed to account for Yamdena *-p* and *b*.[5] Their presence here implies their presence in CMP, contrary to Stresemann's position but in line with more recent views (e.g. Collins 1980; Blust 1981). Though rare, stop reflexes of AN **p* and **b* are found in eastern Indonesian languages; their occurrence in Yamdena is indicative of its overall conservatism.

TABLE 1: AN-CMP-PTK REFLEXES

AN	*p*	*b*	*mp,mb*	*m*	*t,T,C*	*s,c*	*n*	*ñ*	
CMP	*p*	*b*	*mb*	*m*	*t*	*s*	*n*	*ñ* (or *n?*)	
PTK	*f-f-p*	*b*	*mb*	*m*	*t*	*s*	*n*		
AN	*d,D*	*z,Z*	*nt,nT,nd,nD,nz,nZ,nc,ns,nj*			*k*	*g*	*ŋk, ŋg*	*ŋ*
CMP	*d*	*Z*	*nd* (+ possible others)			*k*	*g*	*ŋg*	*ŋ*
PTK	*d*		*nd*			*k*	*ŋg*		*ŋ*
AN	*r*	*j*	*R*	*l*	*w*	*y*	*q,S,?,Ø* etc.		
CMP	*r*	*j*	*R*	*l*	*w*	*y*	*q,S,Ø*		
PTK	*r*			*l*	*w-w-Ø*	*y-y-Ø*	*Ø*		
AN	*i*	*u*	*ə*$_1$	*a*	*ə*$_2$	*-ay,-əy*	*-aw*	*-uy*	*-iw*
CMP	*i*	*u*	*ə,e,o*	*a*	*ə,e,o*	*-ay*	*-aw*	*-uy*	*-i?*
PTK	*i*	*u*	*a*		*e*	*-a*		*-u*	*-i*
AN	*a(qSØ)i*		*a(qSØ)u*						
CMP	*a(qSØ)i*		*a(qSØ)u*						
PTK	*e*		*o*						

(b) PTK **nd* (like Stresemann's Ambonese **nd*) in fact clearly reflects only AN **nt, *nd/D* and perhaps **ns;* examples for the other cases are either doubtful or altogether lacking. Collins (1980) gives evidence that at least **ns* shows distinct reflexes in some Ambonese languages.

(c) The merger of AN **r, j, R* > ***r* is also seen in Lettic; Selaru has *r* < **r,R*, but contradictory evidence for **j*.[6]

(d) PTK **y* is traceable to AN **y* only in medial position. PTK **y-* has multiple origins, among them (a) reduction of two homophonous prefixes ***i-* '1. personal article, 2. locative marker' and (b) 'y-accretion' preceding initial **(qSØ)a* (more frequent in Fordata and Kei than in Yamdena). This last is found in several EIN languages as well as in Fijian and perhaps other MN languages. But

[5]Yamdena also has a [p] in other positions. In certain cases, it is a conditioned variant of *mb;* in others, a free variant thereof (see Drabbe 1926b:4). This latter apparently reflects influence from the 'northern' dialect of Yamdena, where *p* < **mb*.

[6]Collins (1982:127) claims SLAR *s* < **j;* no example is cited, but I suspect he had in mind SLAR *waras* 'liana' < **waRej*. At least in Drabbe, there are no unambiguous witnesses for **j:* we see *-Ø* twice, *-Ø-* twice, *-l-* once, and all can be explained. SLAR *waras* is probably < Yamdena *waras*, itself a likely loan (like *fase* 'rice') from Seran Laut (Geser) where **j* > *s* is regular.

most instances of PTK **y-* are of uncertain origin. In no case does **y-* occur preceding **i* or **u*. These statements apply equally well to Lettic **y*.

(e) PTK **o* (except when < AN **a(qSØ)u*) is of obscure origin, as it is, I feel, in most EIN languages, aside from those few, like Tettum, where it reflects AN **ə*. It is likely due to fairly heavy influence from some outside source – languages like Gorontalo or Mongondow of North Sulawesi, or Banggai, Mori or Toraja of East/Central Sulawesi are likely candidates, but possible very early influence from a Proto Oceanic-type language would be a tempting hypothesis.

(f) Where AN **a()i* and **a()u* > PTK **e,o*, the resulting monosyllable is restored to bisyllabicity in one of two ways: in Fordata and Kei, by 'lengthening' with a following *-a-;* in Yamdena by addition of -e (=a) to the base (e.g. AN **baSu* > PTK **bo-n* > F *voa/n*[7] 'smell', K *woan*, Y *bone* 'to kiss'). The treatment is the same for those rare monosyllables containing AN **ə – *ənəm* 'six' > PTK **nem* > F *nean*, K *neyen*, Y *neme*. Thus nothing in these languages compels the reconstruction of a separate reflex of the AN sequences; that is not the case, however, for Lettic and other presumably related neighbours.

(g) Final consonants. PTK shows final **p t k m n ŋ r l s*. Yamdena generally retains these. The situation in Fordata and Kei is less clear, since all finals are liable to loss, depending on various factors: (a) inalienable nouns lose finals, replacing them with possessive markers; (b) many verbs lose their final, either outright or by replacement with *-t, -k, -n* or *-ŋ* (Yamdena also has occasional replacement of verbal finals). These 'suffixes' apparently had some as yet undetermined morphological value: at present only *-t* and *-k* can be traced back to higher levels; (c) **-p* and **-m*, if not otherwise lost, shift > *-t* and *-n*, and likewise **-ŋ* > *-n;* (d) **-s* > *Ø*, probably via F-K ***h*, with rare *-s* likely borrowed; (e) also **-k*, via F-K ***?*, > *Ø*. Thus in Fordata and Kei, only *-r, -l* unambiguously reflect original finals.

4. DEVELOPMENTS FROM PTK

4.1 YAMDENA

The phonological conservatism of this language has already been noted; it retains the PTK sound system with just two changes, that is, merger of **ŋg* and **ŋ* > *ŋ*, and surface realisation of **-a#* as *e* (with *-a-* before suffixes). It also retains traces of what were apparently allophones of PTK **e* (see section 4.1.1a). Other developments that distinguish Yamdena from Fordata and Kei are: infrequency of *y*-accretion; sporadic shift of ultima **u* > *i* (see section 4.1.1b); use of ***-a* to lengthen monosyllables; and a sizeable body of words with 'o of obscure origin' without F-K counterparts.

4.1.1 TWO PECULIARITIES OF YAMDENA PHONOLOGY

(a) The allophones of PTK **e* were evidently a higher [ɐ]/__C{i,u} versus a lower [ɛ]/__C{a} (i.e. elsewhere). Because of its frequency, [ɛ] was probably the norm. Drabbe's 'dictionary' dialect maintains the distinction with written *'é'* versus *'e'*, as does Geurtjens's Kei, writing *'ï'* corresponding to Yamdena *'é'* versus *'e'* otherwise. Compare K *rïŋ*, Y *réŋi* 'carved figures'; K *tïl*, Y *téli* 'three'; K *ref*, Y *defe* 'fathom'. (There are discrepancies, some surely due to omission of the

[7]In this paper a / is used in examples to separate an irrelevant ending or portion of the word cited.

diacritics.) In Drabbe's Yamdena, the distinction has become phonemic, probably through sporadic lowering of **i >é/e* together with borrowing from languages whose */e/* norm was closer to [e] or [ɛ] in 'wrong' environments. Neither of the recent Indonesian studies, however, notes the distinction in Yamdena or Kei. It is certainly possible for it to have been lost in the past sixty-plus years, but in view of the overall quality of these works, one cannot rule out simple oversight.

(b) Shift of ultima **u > i* is observed in some 30 forms, several of them, like *téli* 'three' < **təlu*, 'basic' vocabulary. There is no apparent phonetic conditioning – note *féni* 'full' < PTK **fenu* versus *fenu* 'wear something around the neck' < AMB **ənu* 'neck', *manik* 'bird' < **manuk*. Several languages in the area show the same shift: Leti/Moa sporadically, regularly in Buli, Bonfia, various Ambonese languages and (Lettic) Wetan of Babar which, being geographically closest to Tanimbar, is a likely source.

In contrast, Fordata shows **-u > i* in just a handful of forms, only a few of which match Yamdena. The implication is that the two languages received their *i*-forms from different sources or at different times.

4.2 FORDATA AND KEI

These two languages clearly underwent a period of common development during which the following changes took place:

(a) loss/replacement of final C, as discussed above;

(b) **s* > F-K ***h;*

(c) **k* > F-K ***ʔ* (with later ***h, ʔ* > 0/__#);

(d) (ordered) 1. **b,d* > F-K ***v,r;* 2. **mb,nd* > F-K ***b,d;*

(e) (ordered) 1. **ŋ* > F-K ***n;* 2. **ŋg* > F-K ***ŋ;*

(f) retention of PTK **e* allophones;

(g) 'lengthening' of monosyllables in **e,o* by addition of *-a-;*

(h) Diphthongisation of final high vowels in the environment *** {e,a} {h,ʔ}*__(C)#,

such that, for example, ***-ahu > -ahaw* as in K *yahâw*, F *yáha* 'dog' < **asu*. There are questionable examples that suggest the environment may have included other vowels. This rule evidently began to operate about the time the two languages were separating, as it is seen sporadically and often irregularly in Fordata.

4.2.1 KEI

From the stage just outlined, Kei exhibits further changes:

(a) loss of ***ʔ-ʔ-*. Although Geurtjens's spellings are unclear, and Tetelepta offers contradictory forms, *loss* seems the better assumption.

(b) loss of **l/{a,u}*__# and intervocalically between any *a-u* sequence. I posit that *l* developed a velarised *w*-like allophone in these environments, which subsequently dropped/blended with

neighbouring *u*, for example, *n-a* < ***ala* 'give, take'; *u* ~ *uu-n* 'head' < ***ulu(-n); wowt = wawt* 'image' < ***walut; wuan* 'moon' < ***bulan.*

(c) loss of final vowels. This affected original final V, as well as those exposed by loss of final C during the F-K stage. Final V loss evidently took place while 'l-loss' was on-going, but before **[ɨ] was lost – vowels exposed by 'l-loss' are retained.

(d) reduction of VV sequences > diphthongs in the case of V+*{i,u}*, long V in the case of like vowels. This is one of the knottier areas of Kei phonology, since Geurtjens's orthography not only over-differentiates, but is also rather inconsistent. Tetelepta's 'phonemicization' seems more consistent, but is, I suspect, under-differentiated.

(e) Development of final stress – in fact a corollary of the predominantly monosyllabic structure of most bases.

4.2.2 FORDATA

From the F-K stage, F shows only slight further change:

(a) loss of ***e* allophones;

(b) retention (development?) of penultimate stress;

(c) The 'dictionary' dialect shows sporadic loss of medial ***h,ʔ*, general loss of ***h-* (versus retention in the Molu dialect). Where ***s* > *h/e*__V, the outcome in every case is *-ahi-*, as in *ahil* 'regret' < PTK ***sesal; fahi* 'fart' < PTK ***fesu; vahi* 'paddle', K *wehe*, AN ***bə(R)say*. In all cases I believe diphthongisation was responsible: ***ehV* > *ehe$\check{V}$* > *éheØ* > *éhi* (constraint on final *e*) > *-áhi-* (dissimilation).

We might hypothesise that Fordata failed to continue developing toward a Kei-like stage by contact with conservative Yamdena after moving into the Tanimbar area. Interestingly, however, there is clearer evidence for Fordata influence on Yamdena than vice-versa, for example, occasional Y *f* for ***b*, likely < F v; Y *Ø-Ø-* for ***s* surely < F *h;* Y *r* for ***d*, < F *r.*

5. SHARED GRAMMATICAL FEATURES

5.1 PRONOUNS

With appropriate sound changes, the three languages have almost identical sets of personal pronouns, as seen in Table 2.

TABLE 2: FORDATA-KEI-YAMDENA PERSONAL PRONOUNS

	1SG	2SG	3SG	1PL.INC	1PL.EXC	2PL	3PL
PTK	***yaku***	***ko***	***ia***	***kita***	***kami***	***kimi***	***sira***
F	*ya?a*	*oa*	*ia*	*ita*	*ami*	*mia*†	*(h)ira*
K	*yāw, ya?aw*	*o*	*i*	*it*	*am*	*im*	*hir*
Y	*yaku*	*kou* (< *?kowe* =/ko+a/)	*iye* (=/ia/)	*kite*	*kami*	*kmi*†	*sire*

†(Apocope in F, syncope in Y, unexplained. Perhaps F *mi+a* < ***?mi* by influence of Y *kmi.*)

Table 3 displays the possessive forms. Yamdena shows some analogical extension of the **ni-* formative; Fordata and Kei have unexplained restructurings in their plurals.

TABLE 3: FORDATA-KEI-YAMDENA POSSESSIVES

(a) *Preposed, alienable forms.*

	1SG	2SG	3SG	1PL.INC	1PL.EXC	2PL	3PL
PTK	**niŋgu*	*?*mu*	*?*ni*	**(ni)nda*	**mami*	*?*	*?*
F	*niŋ(u)*	*mu*	*ni*	*did(a)*	*mam(i)*	*bir*	*rir*
K	*niŋ*	*mu*	*ni*	*did*	*mam*	*bir*	*rir*
Y	*niŋu*	*nime*	*nia*	*ninde*	*mami*	*mir*	*nir*

(b) *Suffixed, inalienable forms.*

F	*ŋ*	*m*	*n*	*d*	*m,mami*	*b*	*r*
K	*ŋ*	*m*	*n*	*d*	*b*	*b*	*r*
Y	*ŋ*	*m*	*n*	*nind(ar)*	*mamyar*	*mir*	*nir*

5.2 VERBAL PERSON-MARKING PREFIXES

In Fordata and Yamdena, the verbal personal prefixes have two forms: *long*, as cited in Table 4, and *short*, in which final *i,u* undergo 'binding' (see section 8), while *-a* drops out. The distribution of short versus long forms is not entirely clear; Drabbe cites cases of syntactic conditioning (long forms are emphatic), lexical conditioning (some verbs require the long form), as well as phonological conditioning (long forms precede #CC-).

TABLE 4: FORDATA-KEI-YAMDENA VERBAL PERSONAL PREFIXES

	1SG	2SG	3SG	1PL.INC	1PL.EXC	2PL	3PL
F	*u*	*mu*	*na*	*ta*	*ma*	*mi*	*ra*
K	*u*	*um*	*en~na*	*it*	*am(u)*	*im*	*er~hir*
Y	*ku*	*mu*	*na*	*ta*	*ma*	*mi*	*ra*

5.3 PLURALISATION

Fordata and Yamdena have a productive plural suffix, F *-r*, Y *-r ~ -ar*. It has been lost in Kei, but is seen fossilised in the possessives *bir* and *rir*.

5.4 DERIVATIONAL AFFIXES

Drabbe's analyses of Fordata and Yamdena derivational affixes (all prefixes) show that with minor exceptions the two systems are identical. Adjectival (stative, intransitive) prefixes include *ma-*, *ŋa-*, *ka-* in both and F *ba-*, Y *mba-*. Both show the causative prefix *fa-*. Further, F, Y *ra-*, *sa-*, *si-*, though described in rather vague terms, seem to have similar functions. The vowel of these prefixes usually

deletes when the (long) person markers are added, so (for example in third person singular) *nam-*, *naŋ-*, *naf-*, *nat-* etc.

For Kei, Geurtjens gives an equivalent list: stative *nam-*, *naŋ-*, *nak-*, *nab-*; causative *naf-*, *nat-* ~ *nad-* and *nas-*. There are variants written with *'ĕ'* and *'e'*, possibly 'euphonic' or else lexically determined. Geurtjens's discussion of all these is disappointing and rather confused; it seems unlikely that morphemes of such frequent occurrence can truly be as 'meaningless' as he often says. On the other hand, Tetelepta's analysis is no more illuminating; perhaps Kei has indeed lost some of the grammatical or lexical motivation for its morphology.

5.5 BINDING

Fordata and Yamdena show the morphophonemic process I term 'binding'; Kei has only a remnant of it (see section 8).

6. A UNIQUE (?) PHONOLOGICAL INNOVATION

Based on the following:

F *roal* '(go) seaward (away from the speaker)'
K *roa* 'the sea as a thoroughfare'
Y *dole* 'seaward; easterly; foreign, not Yamdena'

we can construct PTK **dol* 'seaward', presumably by consonantal metathesis < **lod* < AN **laSud*. Fordata and Kei further show reflexes of ***ndol* – F *doal* 'seawards (toward speaker)', K *en/do* 'come from the south or west; come or do something in the direction of speaker'. It cannot be determined whether ***ndol* is a F-K innovation (by analogy with other *r* ~ *nd* forms?) or a reflex of PTK ***ndol*, lost in Yamdena.

Only one other case of a metathesised reflex of **laSud* is known to me: Aru (Ujir dialect) *adohl* 'east' (van Eijbergen 1864), in all likelihood a borrowing from the Tanimbar-Kei area.[8]

At least one other language has a **laSud* reflex with similar directional senses: Buli *lalau* 'seaward, away from speaker', *malau* 'seaward, toward speaker' It is possible that *la-* and *ma-* in these words reflect forms meaning 'go' and 'come'.

If this consonantal metathesis is indeed unique to PTK, it is an innovation that serves to define the subgroup. More interesting, perhaps, is what the form implies: in order for PTK to have **dol* < ***lod* < **laSud*, the final **d* must have survived into the immediate ancestor of PTK, and so on back to CMP. A similar argument for CMP **-b* can be made on the basis of Leti *tutwu* 'to close, cover' reflecting CMP *?*tutub*. The question whether voiced final stops were retained in CMP has not been researched, largely, I suspect, because most of us have assumed that devoicing of final stops would be a more expected development. The question is made more difficult, too, by the relative rarity of AN **-b* and **-d*, and the loss of finals in so many EIN languages. A handful of examples for **-b* in Lettic, and PTK **dol*, are the only evidence so far discovered.

---

[8]The metathesised form may have arisen in the first place by deliberate change in a special 'seafarers' language': one of those ritualised speech forms full of taboo-substitutes and circumlocutions often reported from various Indonesian areas (though not specifically from the T-K area). Such a 'seafarers' language' could well have taken this word to Aru.

7. THE POSITION OF SELARU

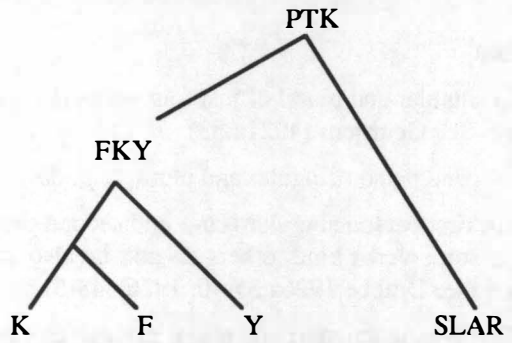
Selaru is spoken on the island of the same name at the southern end of the Tanimbar group. According to Coward (1989:2) there are also Selaru-speaking areas on Yamdena and some offshore islands. Coward estimates the number of speakers at 7000-9000, considerably more than Drabbe's estimate (1932a:1) of about 3600.

In view of its location, and the obvious cognacy of so many forms in Drabbe's word list, one might hastily conclude that Selaru is a close relative of Yamdena. I suspect, however, that Drabbe tended to concentrate on vocabulary that clearly resembled that of the (to him) more familiar Yamdena.

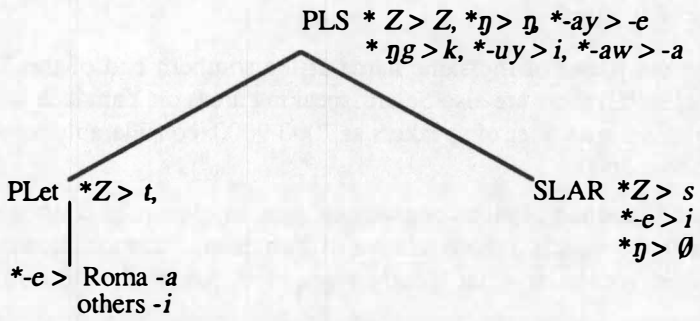
Like Yamdena, Selaru has the *é:e* distinction under the same conditions, but, as in Yamdena, with many 'wrong' occurrences; clearly some of the *é-* or *e*-forms are borrowed from Yandena. Coward notes the distinction, but inexplicably does not consider it phonemic. Clearly borrowed is Selaru *mb*, in a handful of forms almost all of which recur in Yamdena. Selaru has 'binding' in all possible environments, including two that Yamdena lacks – with the singular suffix *-ke*, and in third person singular of verbs, where the prefix is *i-* (unlike Y *n-*). Several phonetic changes, however, show that Selaru has had a quite different history than Fordata, Kei or Yamdena: AN **b* > SLAR *h*, AN **p* > SLAR *Ø* (with rare *f*, *h* probably borrowed), CMP **ŋg* > SLAR *k*, AN **ŋ* > SLAR *n*, *Ø*/V__V, SLAR *s* < AN ** l/_{i,u}*. Further, the base forms of CVCV nouns usually show loss of the final V, as do most verbs; with nouns, the final V reappears when suffixes are added (thus it is present underlyingly). Coward (1989) symbolises these finals with 'glide' phonemes – it remains unclear whether Drabbe's *Ø*-final, or Coward's glide-final forms can occur freely.

Two developments in particular show that any conclusion about 'close relationship' with Yamdena is indeed not only hasty but incorrect: Selaru has (1) *s* < AN **Z* versus PTK **d*, and (2) *-i* < AN **-ay/ey, -uy* versus PTK **-a* and **-u*. (Selaru like PTK has *-a* < **-aw*, but in just one example.)

One alternative is to amend PTK to include **Z*, **ay* and **uy* reflexes; Fordata, Kei and Yamdena still remain a discrete subgroup, but coordinate with Selaru, thus:



The other alternative is to align Selaru with some other group – but which? Its phonology compares well with that of Galoli (eastern Timor, Wetar I.), but yet another reflex is required for *R (GAL *Ø*, SLAR *r*). Further, since Galoli and Tettum are closely related, we should also need **ə* (GAL, SLAR *e*, TET *o*). But this possibility is vitiated by lack of binding and productive pluralisation in Tettum and Galoli. The other local group where 'binding' and 'plural' are found is Lettic, which also has distinctive reflexes of **Z* and **-ay/uy*, with **-a* < **-aw*. I therefore posit Proto Lettic-Selaru as follows:



As we will show later, Proto Lettic-Selaru and Proto Tanimbar-Kei descend in turn from a common node – thus, all these south-eastern Indonesian languages which exhibit binding are related (as we might suspect in any case!).

8. BINDING

Binding is a morphophonemic process whereby at certain junctures (1) a final *{i,u}* of the first element and the initial C of the second element metathesise; (2) the shifted V reduces to *{y,w};* (3) the original V is deleted. (Before a V-initial base, the final *{i,u}* simply shifts to *{y,w}*.) Thus, the two elements are in a sense interlaced, bound together into a single entity. Very schematically:

#...(C) *{i,u}* + (C)V...===> #...(C)Ø+(C)*{y,w}*V...

In an earlier paper (Mills and Grima 1980) this was termed 'pseudometathesis' in view of its phonetically explicable nature since it proceeds via two rules, (1) progressive assimilation of the vowel's palatal or round quality, (2) V-deletion in pretonic, interconsonantal position.

Binding occurs actively to one degree or another in Fordata and Yamdena (only fossilised in Kei), Selaru and all Lettic languages except Kisar. The relevant environments are as follows.

8.1 PERSONAL PREFIX + VERB

(a) Kei: in second person singular and plural of just four verbs *(ba* 'go', *mat* 'die', *ma* 'come', *an* 'eat'; as in *um+ba > umbwa).* See Geurtjens (1921a:33).

(b) Fordata: regularly in second person singular and plural (e.g. *dava* 'seek', *mdwava, mdyava*).

(c) Yamdena: regularly in first person singular *(ku-),* and second person forms. In both Fordata and Yamdena, Drabbe notes, some verbs bind, others do not; he also cites verbs which apparently take both types of conjugation (see Drabbe 1926a:35-40; 1926b:48-51).

(d) Selaru: regularly in first person singular and plural exclusive *(mi-),* second person forms and third person singular *(i-).* Some verbs bind, others do not (Drabbe 1932a:7-10).

(e) Lettic: Examples in Jonker (1932, passim) show binding in first and second person forms in Leti and Moa; there are scattered citations from Roma and Luang-Sermata. It is regular in Wetan, where all cases involve *i* only, due to regular shift of **-u > i.* It is not found in Kisar (Rinnooy 1886, passim). Variation in all the Lettic data suggests that binding may be (is becoming?) optional; Wetan also shows apparent extension of the rule by analogy, that is, *i* inserted in 'wrong' places, as in 'correct' *ira/dena* 'they stay' beside *ira/diena* 'they stay'.

8.2 COMPOUNDING

(a) Not found in Fordata or Kei. In Fordata, most compounds simply drop the final V of the first member; Kei shows simple conjoining. Loss of binding in all the non-verb environments could be a F-K innovation, or independent in each language. Final-V loss in Kei, in any case, makes binding impossible.

(b) Yamdena: present but of uncertain productivity. A few cases are clear – *bali* 'side' + *duwe* 'two' > *bal'dyu* 'both sides'. Others, such as the many tree names, may be frozen forms (e.g. *kabyarat, kafyalu*); as expected, no binding before #CC- (e.g. *kai-plasan*). A free form *kai* 'tree' is not listed. Other apparent compounds also contain undefined elements – *balnyote* 'echo' presumably < *bali* 'answer' + ?. Still other forms, as in Fordata, show loss of the final V.

(c) Selaru: present and apparently productive. In *r/sus twahar/ke* 'they arrange thatching' we see it within the verb phrase *(susu + tahar).*

(d) Lettic: present and productive, but variable, as Lettic *pipdiuma ~ pipi-duma* 'sheep' (< Lettic *pipi* 'goat' + Malay *domba* 'sheep'). Very regular in Wetan; not found in Kisar.

8.3 SUFFIXING

(a) Absent in Fordata and Kei.

(b) Present in Yamdena, involving the plural marker: *lutur* 'wall', pl. *lutrwar; manik* 'bird', pl. *mankyar.* Inalienably possessed nouns behave somewhat differently.

(c) Present in Selaru, involving the singular marker *-ke,* as *hah* (Coward *hahy*) 'pig', *hahkye* (pl. *hahire*); in inalienables, cf. *lurik-kwe* 'my bone' *(luri+ku+ke).* The affected suffixes are first and second person possessives. There is a sizeable minority of forms which add *-ke* without binding.

(d) Lettic: The changes seen in Leti/Moa base forms and possessives are not examples of binding, but suggest a possible origin. Suppose that in Yamdena, final C had (or even now has, subphonemically) an echo-vowel, like Leti/Moa – thus, Yamdena *lútur* = [lúturu], *lútur+ar* > *lúturŭar* > *lútØruar* = *'lutrwar'.* Such an explanation, however, is not so easily applied to Selaru.

Wetan possessives and plurals show V-deletion but no binding. Binding does occur in some phrases (perhaps better subsumed under 'Compounding', section 8.2), as in *nmati loe ~ nmatlyoe* 'he died there', or *ima ma* (*sic* for *imi ma?*) *~ immi* 'you all and...'. Josselin de Jong (1987:159) found this *'sandhi'* to be dependent on the speed of speech.

8.4 FOSSILISED FORMS (?)

A number of Fordata, Yamdena and Selaru forms are cited with bound *y/w.* Drabbe suggests these are due to 'alternate' forms of some prefixes (e.g. *fa- ~ fau-*); perhaps too, some are due to loss of the first member of an original compound. A few examples: F *kyavu* 'dust'; Y *nay/bwar* 'glow (of fire, iron)' cf. *au bware ~ au bare* 'ember'; SLAR *hwalu* 'a dove' likely < ***manu+halu.* In Lettic all such examples involve *-i-*, reflecting prefixes, some known, as in Lettic *tialla* 'road' (locative **i-* + **talan* < **Zalan*) and *siona* 'digging stick' (instrumental **i-* + **son* < **suqan*), others uncertain as *kdioa* 'twin' (?**ki-* + **do* < **nDuSa*) and *kdieli* 'ring' (?**ki-* + ***deli*).

8.5 EFFECTS ON VERBAL CONJUGATION

Both Stresemann and Collins have reconstructed an 'Ambonese' (perhaps = CMP) verbal conjugation of Prefix + Verb. Using **tanis* 'weep' as a model, we have:

SG	1. **ku-tanis*	2. *mu-tanis*	3. *na-tanis*
PL	1. *ta-tanis, mi-tanis*	2. *mi-tanis*	3. *ra-tanis*

Note that this structure is essentially the underlying conjugation (allowing for sound changes) of Fordata-Kei-Yamdena, Selaru and Lettic; it is also close to the surface forms of, for example, Tettum and Galoli (not to mention many other AN languages). Ambonese languages deleted the vowel of the prefix; nasals then assimilated to stem-initial C and went on to develop the same as original *NC (see Stresemann 1927:119ff.); thus:

SG	1. **ktani*	2. *ndani*	3. *ndani*
PL	1. *ttani, ndani*	2. *ndani*	3. *rtani*

With later sound changes (e.g. **kt- > **t, *nd- > **r)*, the result was consonantal alternations in second and third person singular, and in first and second person plural, entailing changes in underlying forms and probable loss of some information.

Binding is at once innovative and conservative: innovative as to the sequence of rules that produce it (assimilation, deletion – which most likely began as fast-speech phenomena), yet conservative in that underlying forms are not changed, and little or no information is lost.

Both processes introduce asymmetries into the verbal paradigm. The Ambonese set (second and third person singular, first and second person plural) must have seemed particularly unmotivated, and it is no surprise that many of those languages have eliminated the irregularities by generalising one or another of the alternating consonants. The Yamdena-Selaru-Lettic procedure (reanalysable as a simple insert *-u-* in first and second singular, and *-i-* in first and second plural) may in some sense be more logical, which perhaps accounts for its better preservation.

Since all these languages show verbal binding (even if only a remnant, as in Kei), it is assumed to have been present in their proto-language; very likely compound-binding was also present. It is hard to imagine, moreover, that morphophonemics of this complexity could be borrowed.

Selaru's *-ke* 'singular marker' is a notable innovation. As to the other environments, it would be tempting to view Yamdena, so conservative in other respects, as a model for the original situation, with individual losses and changes elsewhere, but we cannot be certain.

8.6 WIDER DISTRIBUTION OF BINDING

I recall encountering binding in Biak many years ago, before my interest had turned to the Lettic and Tanimbar languages, but the data were no longer to hand. My vague memories did not, I felt, warrant further discussion at this time of possible wider distribution, but Prof. Blust has brought to my attention the material in Dyen (1978, appendix), and the question must now be addressed.

Dyen (1978:248-249) cites evidence from Numfor (closely related to Biak), Patani (South Halmahera) and Kabhubhaka (West Sumba), which clearly shows 'verbal binding' in these three languages, quite comparable to that seen in Lettic and Tanimbar. In addition there is a West Sumba example involving the personal prefix **i* (*byuulu* < **i+buulu* 'pers. name') and two clear cases in the

Numfor numeral system. It is impossible to judge how extensive or productive the process may be in Patani or Kabhubhaka, whose examples could well be fossils (consider Kei's four verbs); in Numfor/Biak it appears at least as productive as in Fordata. Incidentally, Steinhauer (1985:471ff.) gives the necessary documentation for binding in Biak.

I might also cite two suggestive Misool examples from Blust (1978:202-203): *syop* = Buli *sisop* 'bathe' and *falyan* = Buli *failan* 'mast'; Misool is said to belong to the South Halmahera West New Guinea (SHWNG) group. Binding is definitely absent from Buli, though syncopation of pretonic vowels is frequent there. It appears highly unlikely, I fear, that we have anything like a complete inventory of languages with binding – a major problem, of course, in the discussion that follows.

As Dyen (1978:248) says, binding "constitute[s] a reasonably strong argument for subgrouping together the languages exhibiting it", and I make just such use of it in positing the 'SE Maluku' subgroup. The need to explain the presence of binding in two additional widely separated areas would appear to weaken my argument, but does not necessarily do so.

If we proceed for now under the assumption that all languages with binding subgroup together, two hypotheses can be ignored, though not rejected out of hand:

(a) *Borrowing* from whatever direction, including non-AN source(s). This not only begs the question, but also violates one of historical linguistics' near axioms, to wit, that morphophonemic processes are hardly ever borrowed.

(b) *Parallel development* – but if, as I have suggested, binding originated in fast-speech rules, it may well be that details yet to be discovered could favour this possibility.

It is not utterly impossible that the SHWNG group, or some other group containing at least Numfor/Biak and Patani, could be an early offshoot of the 'Post-CMP' stage outlined in Table 9. None of the sound changes detailed in Blust (1978:192-193) is inconsistent with such a view, despite some local oddities (e.g. AN *tTC > Buli *c*, Numfor *k*). We could then posit binding as an innovation in 'Post-CMP' which was totally lost in Proto Timor (if it is correctly subgrouped here), retained and perhaps extended in Proto SE Maluku, but only selectively retained in the Halmahera-New Guinea area. Perhaps there, as in Kei, final-V loss led to its restriction to word-initial environments.

The West Sumba case is more difficult, for these languages are very poorly documented: except for Wielenga's 1917 word list, where 'Kabhubhaka' is not mentioned (Dyen's *reŋe* 'hear' possibly = Wielenga's Laura dialect *rèngè* under the word *hooren*, no. 547). Wielenga speculates, from cultural evidence, that the languages of West Sumba may have a different origin from those of Eastern Sumba. The data in his word list, in my opinion, do not bear this out, as all Sumbanese languages appear to share a common ancestor. And even Wielenga points out that West Sumba has been isolated from the major trading centres and horse culture to the east. Interestingly, the languages of Sumba bear more resemblance to Savu (and so Roti?) than to languages of Flores (for instance, Sumba has merged the voiced and voiceless NC).

The weakest hypothesis here is:

(a) In West Sumba there is at least one area settled, somewhere in the past, by migrants who spoke a SE Maluku language with binding.

Slightly less weak would be:

(b) At least one language of West Sumba is the remnant of a formerly more widespread SE Maluku-type language with binding. The languages of East Sumba are interlopers, as also would be the main languages of Timor, Roti and Savu. I am uncomfortable with this because, aside from Dyen's two words, there is absolutely no other evidence for binding languages west of the Leti Archipelago.

The most attractive hypothesis would be:

(c) All languages of Sumba, along with Proto Lettic, Proto Tanimbar-Kei, Numfor-Patani (and perhaps Proto Timor and Savu) ultimately descend from the same 'Post-CMP' stage; binding has been lost in most of Sumba and SHWNG (and all of Timor/Savu, if included).

Thus we can see that it is not too difficult to create a reasonably coherent CMP subgroup, with binding, that embraces all the known descendants. The result, however, is to negate Blust's (1978) argument that SHWNG and Oceanic are coordinate under a node labelled 'Eastern Malayo-Polynesian'. Perhaps the situation could be salvaged by reformulating the SHWNG group without Numfor/Biak and Patani, though I suspect not. Blust's arguments for this subgroup seem very strong, however one may feel about its alignment with Oceanic.

More likely we should have to abandon our 'strong assumption' of a single subgroup with binding, and allow after all for parallel development within a relatively restricted geographic area in north-eastern Indonesia.

But if we reject parallel development, insisting on a single subgroup, and yet wish to retain Blust's EMP hypothesis, then we are forced into one of these uncomfortable positions:

(a) PTK/PLet/WSumba as well as SHWNG, are *all* coordinate with OC, and binding is an innovation within EMP; or

(b) PTK/etc. are indeed CMP languages and SHWNG is indeed EMP. Therefore, binding in both is *retained* from a higher level (Blust's East-Central MP at least) and has been lost in a very large number of daughter languages (most CMP and apparently all OC).

I do not know the answer.

9. WIDER RELATIONSHIPS

9.1. PREVIOUS ATTEMPTS

Only three attempts at classifying the languages of south-eastern Indonesia are known to me, all flawed in one way or another.

(a) S. J. Esser, on his well-known language map in the *Atlas van Tropisch Nederland*, posited a Timor-Ambon group, but without any further subdivision and without any record of his evidence. We can be sure, of course, that he was fully conversant with all published materials of the time, and probably had access to manuscript material and personal contacts with coworkers in the area. His published works on other languages demonstrate linguistic talent and insight of the highest order, so that even if his 'Timor-Ambon' group was impressionistically or intuitively based, it represents a very well-educated guess.

(b) Dyen (1965), in my view, gives short shrift to the Moluccan area. The manifold problems of his work are understandable: old and haphazard data, often amateurish and difficult to interpret; reliance on dictionaries; 'subadequate' lists; and extremely conservative decisions as to cognacy. The omission of Roti, one of the best described languages in all Indonesia, is inexplicable and a major flaw. The resultant 'Moluccan Linkage' is no more illuminating than Esser's grouping. And of course, the validity of lexicostatistics itself remains debatable, so that even if all the percentages are reliable (and they are not), the method offers little more than guidelines for further research and becomes irrelevant once enough data are available for proper comparison.

(c) It is wrong to call Collins (1982) 'flawed'; it is merely too brief vis-a-vis my area of interest. A few details aside, in fact, Collins reaches the same conclusion regarding Fordata-Kei-Yamdena as I do, and correctly notes the distinct position of Selaru. He has unearthed a new Tanimbar language, Selwasa, not mentioned by Drabbe. His comments on Teun, Nila and Serua are another first; these are apparently Lettic languages. He concludes:

> ...[T]he geographic boundary which separates Serua and Banda parallels a linguistic boundary. North of that line lie the numerous languages of Central Maluku; south of it are those languages most closely related to those studied by Jonker... (1982:124)

This boundary almost certainly extends eastwards to include at least Tanimbar-Kei; whether it includes Aru is not clear on the basis of available data.

9.2 SOUTHEAST MALUKU

I shall concern myself for now with justifying the alignment of Lettic, Selaru and PTK, mentioned already in section 7 and section 8 above. Space will permit only a brief glance in the direction of Roti, Timorese, Tettum and Galoli.

9.2.1 LETTIC

Based on Jonker (1932), Rinnooy (1886) and Josselin de Jong (1987), I have prepared a reconstruction of Proto-Lettic (Mills:n.d.). All of these languages (Leti, Moa, Roma, Luang-Sermata, Wetan and Kisar) share not only much vocabulary but also these phonological features:

(a) changes in the vowel sequences produced by early loss of AN/CMP laryngeals. Ordered prior to loss of **p;*

(b) loss of AN **p;* shift of AN **b > w (*v);*

(c) merger of AN **d/D, r, j, R > r;*

(d) merger of AN **Z, t > t;*

(e) shift of CMP **mb, ŋg > p, k; *nd > d;*

(f) AN **-uy > i,* but **-aw > a;* distinct reflex of **-ay;*

(g) merger of AN **n,ñ > n;*

(h) binding: see section 8 above (presumed lost in Kisar);

(i) retention of all possible final C by addition of an echo-vowel; along with frequent syncope of the original ultima V, this produces the 'pseudometathesis' discussed in Mills and Grima (1980), for example, LET *ulti ~ ulit* 'skin' < **kulit,* or KIS *hoplo ~ hopolo* 'to sail' < ***sombal.* Occurrence of

these forms is most likely governed by syntactic factors, though speech tempo may be a factor in some cases (Rinnooy's data are especially variable).

Features (a)-(h) are reconstructible for Proto Lettic; the question whether Proto Lettic had echo-vowels need not concern us here. The Proto Lettic sound system is set forth in Table 5.

TABLE 5: PROTO LETTIC SOUND SYSTEM

Consonants:	*p < **mb*	*t < *t,Z*	*k < **ŋg, g*	*ʔ < *k*
	*v < *b*	*d < **nd, ns, nZ?*		
	m	*n < *n, ñ*	*ŋ < *ŋ*	
	*w < *w-w-*	*r < *r,j,R,d/D*	*l*	*y < *?-y-*
		s		
	*(Ø < *p, -w;* for **-y* see below)			
Vowels:	*i < *i, -uy, -iw*		*u < *u*	
	e < ə2, i()a, -ay, (i)		*o < *u()a, (u,??)*	
		*a < *a,ə1,-aw*		
Vowel sequences:		*ei < *a()i*	*ou < *a()u*	

In the modern languages, it is worth noting only these developments:

(a) PLet **ʔ > Ø* everywhere except Moa *-ʔ*, and questionably Roma *h* preceding a consonant.

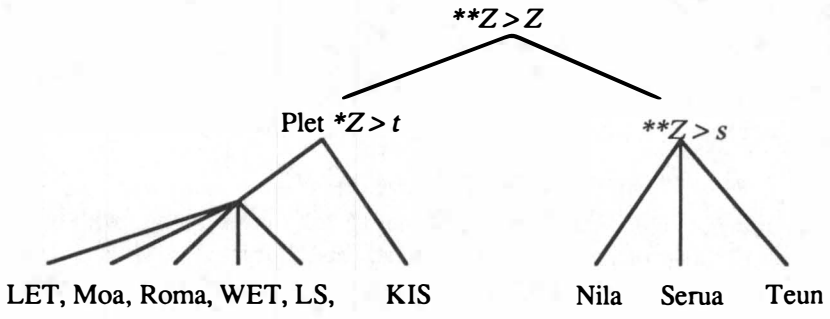
(b) PLet **ŋ* is reflected only in the correspondence KIS *-m-:* LET *-kr- < *ŋ-r#* in a handful of examples; otherwise **ŋ > n* everywhere.

(c) AN **-ay* > PLet **-e* > Roma *-a*, LET, Moa, KIS, WET *-i;* no LS examples.

(d) **ei* and **ou* are retained as sequences or > *e,o* depending on environment (language specific). Most striking is KIS *oi < *ei,* as in *koi* 'excrement' < **taqi, in/hoi* 'who' < **sa()i.*

This last feature, plus others (PLet **k > Ø, *t > k, *s > h,* lack/loss of binding) sets Kisar apart from the other five, which are not differentiated much, and probably represent the remains of a former dialect chain.

Both Jonker and Brandes (see his introductory remarks to Rinnooy) held that the languages of the 'Damar Islands' were also related to Lettic; Collins' two pages (1982:123-125) on Teun, Nila and Serua are the first published data from the area. They evidently have *r < *r,j,R,d/D.* He refers to "consistent metathesis of the final consonants" which I take to indicate forms like LET *ulti ~ ulit* 'skin'. Further, "certain verbs...indicate inflectional change by internal vowel alternation," from which I infer binding. Serua, however, is said to reflect **Z* as *s;* if that is also true of Nila and Teun, then these languages can only be coordinate with Proto Lettic, related at a higher level, thus:



Since this area is known as Kepulauan Barat Daya ('Southwest Islands'), this higher node can be labelled 'Proto Barat Daya' and its two branches Proto Lettic and Proto Damar.

9.2.2 SELARU

In section 7 we have already discussed this language and its divergences from its Tanimbar neighbours, to wit. *Ø < *p,k, h ~ f < *b* (note Nila *Ø ~ f*), *-i < *-ay,uy,* and *s < *Z.* If we assume that **p,* both here and in Lettic, was lost via ***f ~ ɸ ~ h > Ø,* and that the reflexes *t* and *s < *Z* more likely stem from a voiceless than a voiced affricate, then Proto Lettic-Selaru would have had the following sound system:

TABLE 6: PROTO BARAT DAYA-SELARU (PBDS)

Consonants:	*f or ɸ < *p*	*t*	*c < *Z*	*ʔ < *k*
	*v < *b*			
	*b < **mb*	*d < **nd*		*g < **ŋg, g*
	m	*n*		*ŋ*
	w	*r < *r,j,R,d/D*	*l*	*y*
		s		
Vowels:	as in PLet above, except **ua, *ia* remain as sequences.			

9.2.3 PROTO SE MALUKU

At this point we can compare PTK, and note that only three mergers and a few phonetic shifts distinguish PTK and PBDS. The mergers involve **d/D* and **Z,* the vowel sequences **ei* and **ou,* and the diphthongs **-ay* and **-uy;* the phonetic shifts involve **p, *b* and the **NC.* We can therefore unite PTK and PBDS under 'Proto SE Maluku' with a sound system as set forth in Table 7.

We have seen that PTK merged **d,Z > *d,* while PBDS merged **d,r > *r* and presumably devoiced **Z* to ***c;* PTK lost **-y* while PBDS changed vowel quality; PTK simplified **ei, ou* to **e,o,* PBDS retained them.

TABLE 7: PROTO SE MALUKU

Consonants:	*f-f-p*	*t*			*k*
	b	*d < *dD*		*Z*	
	mb	*nd*			*ŋg < *ŋg,g*
	m	*n*			*ŋ*
	w	*r < *r,j,R*	*l*	*y*	
Vowels:	*i < *i, -iw*				*u < *u*
	*e < *ə2,(i)*				*o < (u,* loans?)
		*a < *a,ə1, -aw*			
Diphthongs:	*-ay*	*-uy*			
Sequences:	*-ei- < *a()i*	*-ou- < *a()u*			

As shown in section 8, binding affects all the languages of this subgroup, and is viewed as a distinctive innovation.

9.3 DESCENT FROM PROTO CENTRAL MALAYO-POLYNESIAN

Let us now compare Proto SE Maluku with the probable Central Malayo-Polynesian proto-language, assuming its sound system as in Table 8.

TABLE 8: PROTO CENTRAL MALAYO-POLYNESIAN (?)

Consonants:	*p*	*t*		*k*	*q*	Vowels: *i ə a u*
	b	*d*	*Z*	*g*		
	mb	*nd*		*ŋg*		Diphthongs: *-ay, -aw, -uy*
	m	*n*		*ŋ*		
	w	*r,l*	*y*	*R*	*S*	
		s	*j*			

NOTE: This assumes, crucially, the merger of AN **d* and **D* > CMP **d*, and merger of the voiced and voiceless nasal-grade consonants (though likely not all the dental/palatal NC had merged). Reduction of the number of laryngeals, merger of **t,T,C* > **t*, **s,c* > **s*, **n,ñ* > **n* and **-iw* > **i* are also assumed to have taken place in pre-CMP times, but even if they had not, they are of slight diagnostic value in eastern Indonesian languages.

For now, note that PSEM shows a) loss of **-w, b)* merger of **r,j,R, c)* loss of **q,S* and creation of the new sequences ***-ei-* and ***-ou--, d) *ə* > ***e-a-*.

9.4 LANGUAGES OF TIMOR

If we are on anything like firm ground in positing common ancestry for PTK, PLet and Selaru, the footing becomes slippery indeed as we approach Timor. The island's principal languages – Tettum, Galoli, Timorese (Dawan) – are adequately described, even exhaustively so in the case of Roti, which is not, of course, spoken on Timor but appears to subgroup better with its languages than with any others. Capell's (1944) survey offers bits of data for other Timorese languages, including some non-AN ones, but his 'Vaikenu' is clearly a dialect of Timorese, 'Belu' probably a dialect of Tettum. The others cannot be classified at this point.

9.4.1 CHANGES IN COMMON

Aside from Tettum and Galoli, there is little obvious evidence of a close relationship between these four languages. There are enough shared developments, however, that a case for a Timorese subgroup can be made.

(a) All four show *ø* < AN **R*.

(b) AN diphthongs: **-ay* > *e*, **-aw* > *o*, **uy* > *i* in all.

(c) AN **q,S* > *ø; V(qSø)V* sequences > **a()i* > *e*, **a()u* > *o* in all.

(d) Following from (b) and (c), all four languages *probably* have 7-vowel systems *i e ɛ a u o ɔ* (*ɛ* and *ɔ* hereafter 'è' and 'ò'). The evidence for contrastive mid/low vowels is ambiguous: all the writers describe such contrasts and indicate them orthographically, but also cite 'rules' for determining where the vowels occur.

(e) Laxing/devoicing of AN **p* and **b*, likely via ***ɸ* and ***v*: ***ɸ* > ROT *h-* and *-ø-* or *-ʔ-*, TIM *h-* and *-h-* or *-ø-*, TET *h-ø-s*, GAL *ø-ø-s; **v* > ROT, TIM *f-f-*, TET *f-h-*, GAL *h-h-*.

(f) AN **n, *ñ, *ŋ > n* in all.

(g) None of the languages has binding; conjugation is by prefix + verb.

(h) None of the languages shows *y-*.

9.4.2 DISTINCTIONS

To my view, these are at least as weighty as the shared sound changes, involving as they do the crucial elements **j, *Z,* and **d/D,* along with **ə* and the **NC.*

(a) Roti merges **j,Z > d,* and **dD,r > r ~ l* depending on dialect. The **NC* appear as /NC/, voiced /C/ or /N/ depending on position and dialect. AN **ə > e* in both penult and ultima – the latter in contrast to all the languages discussed in this paper.

(b) Timorese merges **j,dD,r,l,n > n,* probably ordered 1) **j>d,* 2) **d>r,* 3) **r>l,* 4) **l>n.* AN **Z* (via ***d* ?) *> r ~ l* depending on dialect (there is also a final *-l* of unclear origin). **ə > e-a-*; *mb, ŋg *> p,k; *nd > k,* perhaps dial. *r,* as Middelkoop cites *k ~ r* variants.

(c) Tettum shifts **ə > o-a-* and **Z > d; Galoli* has **ə > e-a-* and **Z > s.* Otherwise they share **dD,r > r, *nd > d; *mb, ŋg >* TET GAL *b, k,* though GAL also has *p, g* reflexes. All TET/GAL examples for **j* present problems, but *r* is the most likely reflex (the alternative is *l*).

It is theoretically possible, of course, to create a proto-language for any group of ultimately related languages (e.g. 'Proto Swedish-Hindi'!), but such a creation will have little value except as an intriguing hypothesis, or, to put it more bluntly, linguistic game-playing. In that spirit, I offer the following tentative thoughts on 'Proto Timor' and its possible connection with Proto SE Maluku (PTK-PBD-SLAR).

The developments of **j, *dD, *Z* and **ə* are not as confused as they appear. **Z* and **ə* obviously survive well down into the family tree; **dD* evidently *> **r* very early on, after which it can be posited that **j > **d;* many of the shared changes outlined in section 9.4.1 also reflect early change. Given four languages, multiple alignments are possible: the three *e < *ə* languages versus interloper Tettum with *o;* or, the three 'mainland' languages versus Roti, among others. But there is sufficient other evidence (from vocabulary and grammar) that Tettum and Galoli do group together; if that assumption is made, the question then is, what of Roti and Timorese?

Roti and Timorese share a development in common that may be relevant: at some point in their history, both languages lost all final consonants, with subsequent replacement. In Roti, the replacements included the possessive markers *(-k, -m, -n* etc.*)* on inalienable nouns, a 'verbal' *-k,* 'nominal' *-k* and *-s.* Presumably there were once transparent alternations between CVCV ~ CVCV+C forms, but nowadays, all but the possessives seem fossilised, and the +C forms often show semantic changes.

Timorese shows similar developments, but has additional finals, some of which are quite productive: *-f* creates 'generics' out of inalienables, for example, *atèk* 'my liver', *atèf* 'liver in general'; or *-b* 'causative/transitive'; or *-s, -t, -k* for verbal or nominal derivatives. Many plant names have *-l.* Not only are the origins of these very obscure; they contribute to the overall impression that Timorese has undergone influences unseen anywhere else in the area – in particular, the metathesis of final -CV, evidently syntactically conditioned (cf. *teun* 'three', *katenu* 'third'), which is so reminiscent of that seen in Rotuman.

The vowel system underlying all four languages must have had **i, *u, *a* and **ə* (retained < CMP), with ***i < *-uy; **-e < *-ay* and ***-o < *-aw;* and likely ***-e- < ?*ei* < **aØi* and ***-o- < ?*ou < *aØu*. Since 'diphthongal' *-e* and *-o* only occurred unstressed and finally, perhaps they were [-tense] *è ò*, whereas 'sequential' *-e-* and *-o-* were stressed, in #C__(C)# position, and were possibly contrastive *e:è, o:ò* depending on open versus closed syllables. Perhaps too, *e o* were present in Ambonese and outside loans. These intriguing ideas, I must confess, have yet to be researched fully, but they might account for the seven-vowel system *i e è a u o ò* which, after shift of **ə > e/è* or *o/ò* (similarly conditioned?), these languages appear to share.

From all the foregoing, I posit 'Proto Timor', with two branches: West Timor (Roti, Timorese) and East Timor (Tettum, Galoli). But I must re-emphasise the tentative nature of the alignment, with its totally phonological basis, unlike 'Proto SE Maluku' (PTK-PBD-SLAR) where binding, a unique and complex phenomenon, offers an additional and convincing basis for subgrouping.

9.5 PROTO SOUTHERN MALUKU

It is now possible to conceive of a 'Post-CMP' stage from which both Proto Timor and Proto SE Maluku can derive. Minimally, this stage must show CMP **q,S > Ø* with **a()i > **-ei-* and **a()u > **-ou-*, but CMP **j, *R, *r*, the three diphthongs and **ə* remain distinct. Perhaps ***e* and ***o* are also present due to borrowing plus sporadic lowering of **i* and **u*. Only minor contortions are needed to derive PSEM; the number of changes required for Proto Timor suggests rather strongly that additional intervening stages are probably needed here. If not, however, then this 'Post-CMP' stage can be labelled 'Proto Southern Maluku', as is depicted in Table 9.

Viewed against other languages of the area, 'Proto Southern Maluku' is at least distinguished by loss of the laryngeals and resultant vowel changes. A laryngeal reflex and the vowel sequences remain in Bima and Manggarai, for instance; Savu loses the laryngeals but retains the vowel qualities. Bima and Manggarai, further, even show *apparent* reflexes of both voiced and voiceless NC, and on that basis cannot derive from CMP as it is presently conceived. Yet without the NC merger, CMP almost ceases to be a unique entity. Clearly, much interesting work is yet to be done in the Lesser Sunda/Moluccan area.

TABLE 9: FROM 'POST-CMP' TO THE TIMORESE AREA

CMP (See Table 8)

**p > f /_ V; *g,ŋg > *ŋg; *q,S > Ø: *a()i > **ei, *a()u > **ou*

'Proto Post-CMP'

Consonants:					Vowels:		
	f-f-p	*t*		*k*	*i*		*u*
	b	*d*	*Z*		*e?*	*ə*	*o?*
	mb	*nd*		*ŋg*		*a*	
	m	*n*		*ŋ*	*-ei-*		*-ou-*
	w	*r l*	*y*	*R*			
		s	*j*		*-ay*	*-aw*	*-uy*

PROTO TIMOR

**R > Ø; *-p > f: all *f > ɸ(?);*
**b > v; *d > r: *j > d; *ŋ > n;*
**y- > Ø; *ay > è, *aw > ò, *uy > i;*
**k* begins? > *ʔ; *ei > e/ê?, *ou > o/ô?*

f or *ɸ*	*t*		*k, ʔ*	*i*		*u*
v	*d*	*Z*		*e?*	*ə*	*o?*
mb	*nd*		*ŋg*	*-è*	*a*	*-ò*
m	*n*			*ei* or *e/ê*		
w	*r l*	*-y-*		*ou* or *o/ô*		
	s					

PROTO SOUTHERN MALUKU

**j,R,r > r; *-w > Ø;*
**ə > e-a-*

PROTO BARAT DAYA-SELARU (see section 9.2 above)

PROTO TANIMBAR-KEI

PROTO WEST TIMOR

-C > Ø: replaced; **ə$_2$ > e/ê*
**f/ɸ > h; *v > f;* 8 vowels

ROTI

**h > h-ʔ-*
**r > r ~ l*
**Z,d > d*
**ə$_1$ > e*
*NC > NC ~ C ~ N

TIMORESE

**h > Ø/V_V*
**d>r>l>n*
**Z >(d?)> r ~ l*
**ə$_1$ > a*
**mb,nd,ŋg>p,k,k*

PROTO EAST TIMOR

**-f/ɸ > s; *ə$_1$ > a; *d > r(l?)*
**f/ɸ > h; *v > f;* 8 vowels
**mb, nd, ŋg > b, d, g*

TETTUM

**h > Ø/V_V*
**f > h /V_V*
**Z > d*
**ə$_2$ > o*
**g > k*

GALOLI

**h > Ø*
**f > h*
**Z >(c?)> s*
**ə$_2$ > e*
**b,g > b ~ p, g ~ k*

BIBLIOGRAPHY

ABBREVIATIONS

BKI *Bijdragen tot de Taal-, Land-, en Volkenkunde*
TBG *Tijdschrift voor Indische Taal-, Land- en Volkenkunde*
VBG *Verhandelingen van het Batavi-aasch Genootschap van Kunsten en Wetenschappen*
VKI *Verhandelingen van het Koninklijk Instituut voor Taal-, Land- en Volkenkunde*

BLUST, R. A., 1978, Eastern Malayo-Polynesian: a subgrouping argument. In Wurm and Carrington, eds 1978:181-234.

1981, The Soboyo reflexes of Proto-Austronesian *s. In R.A. Blust, ed. *Historical linguistics in Indonesia*, part I, 21-30. *NUSA* 10. Jakarta: Universitas Atma Jaya.

CALON, L.F., 1890, Woordenlijstje van het dialekt van Sikka. *TBG* 33:501-530; 34:283-363.

CAPELL, A., 1944, Peoples and languages of Timor. *Oceania* 14/3:191-219, 14/4:311-337, 15/1:19-48.

1968, *A new Fijian dictionary*. 3rd edn; 1st, Sydney, 1941. Suva, Government Printer/London: Eyre & Spottiswoode.

COLLINS, J. T., 1980, The historical relationships of the languages of Central Maluku, Indonesia. PhD dissertation, University of Chicago (since published as *PL*, D-47, 1983).

1981, Preliminary notes on Proto-West Central Maluku: Buru, Sula, Taliabo, and Ambelau. *NUSA* 10:31-45.

1982, Linguistic research in Maluku: a report of recent field work. *Oceanic Linguistics* 21/2:73-146.

COWARD, N.E., 1989, A phonological sketch of the Selaru language. In: Wyn Laidig, ed. *Workpapers in Indonesian languages and cultures*, vol.7:1-32. *Maluku*. Ambon, Pattimura University/SIL.

DEMPWOLFF, O., 1934-37-38, *Vergleichende Lautlehre des austronesisches Wortschatzes. Zeitschrift für Eingeborenen Sprachen*, Beihefte 15, 17, 19.

DORES, R. das, 1907, *Diccionario Teto-Português*. Lisboa: Imprensa Nacional.

DRABBE, P., 1926a, Spraakkunst der Fordaatsche taal. *VBG* 67/1.

1926b, Spraakkunst der Jamdeensche taal. *VBG* 67/2.

1932a, Beknopte spraakkunst en korte woordenlijst der Slaroeësche taal. *VBG* 71/4.

1932b, Woordenboek der Fordaatsche taal. *VBG* 71/2.

1932c, Woordenboek der Jamdeensche taal. *VBG* 71/3.

DYEN, I., 1965, *A lexicostatistical classification of the Austronesian languages*. *IJAL* Memoir 19. Baltimore: Waverly Press.

1978, The position of the languages of eastern Indonesia. In Wurm and Carrington, eds, 1978:235-254.

EARL, G. W., 1848, Specimens of the dialects of Timor and the chain of islands extending thence to New Guinea. *Journal of the Indian Archipelago and Eastern Asia* 2:695-704.

EIJBERGEN, H. C. van, 1864, Korte woordenlijst van de taal der Aroe- en Kei-eilanden. *TBG* 14:557-568.

GEURTJENS, H., 1921a, Spraakleer der Keieesche taal. *VBG* 63/2.

1921b, Woordenlijst der Keieesche taal. *VBG* 63/3.

JONKER, J. C. G., 1908, *Rottineesch-Hollandsch woordenboek*. *VBG* 48/1-2. Leiden, E. J. Brill.

1932, Lettineesche taalstudiën. *VBG* 69/1.

JOSSELIN DE JONG, J. P. B. de, 1987, *Wetan fieldnotes: some eastern Indonesia tests with linguistic notes and a vocabulary*. Dordrecht: Foris. *VKI* 130.

MATHIJSEN, A., 1906, Tettum-Hollandsch woordenlijst. *VBG* 56/2.

MIDDELKOOP, P., 1939, *Amarasisch-Timoreesche teksten*. *VBG* 74/2.

1950, Proeve van een Timorese grammatica. *BKI* 106:375-517.

MILLS, R. F., n.d., Reconstruction of Proto-Lettic. MS.

MILLS, R.F. and J. GRIMA, 1980, Historical developments in Lettinese. In: Paz Buenaventura Naylor, ed. Austronesian studies: papers from the Second Eastern Conference on Austronesian Languages, 273-284. Ann Arbor: University of Michigan, Center for South and Southeast Asian Studies.

PIETER, C. et al., 1986, *Struktur Bahasa Tanimbar.* Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Dep. Pendidikan dan Kebudayaan.

RINNOOY, N., 1886, Maleisch-Kissersche woordenlijst. *TBG* 31:149-213.

SILVA, M. M. A. da, 1905, *Diccionario Portugues-Galoli.* Macau.

STEINHAUER, H., 1985, Number in Biak: counterevidence to two alleged language universals. *BKI* 141/4:462-485.

STOKHOF, W.A.L., ed. ... with Lia SALEH-BRONKHORST and Alma E. ALMANAR, 1981, *Holle lists: vocabularies in languages of Indonesia,* vol. 3/1: *Southern Moluccas; Central Moluccas: Seram (1). (MLIn 4). PL,* D-35.

STRESEMANN, E., 1927, *Die Lauterscheinungen in den ambonischen Sprachen. Zeitschrift für Eingeborenen Sprachen,* Beiheft 10.

TETELEPTA, J. et al., 1985, *Struktur Bahasa Kei.* Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Dep. Pendidikan dan Kebudayaan.

WARD, J. H., 1969, Some background for a linguistic survey of the Indonesian Lesser Sunda Islands. *Working Papers in Linguistics, University of Hawaii* 1/4:169-212.

WIELENGA, D.K., 1917, *Vergelijkende woordenlijst der verschillende dialecten op het eiland Soemba. VBG* 61/5.

WURM, S.A. and Lois CARRINGTON, eds, 1978, *Second International Conference on Austronesian Linguistics: proceedings. PL,* C-61.