B. Nothofer

The subgrouping of the languages of the Javo-Sumatra Hesion; A reconsideration

In: Bijdragen tot de Taal-, Land- en Volkenkunde 141 (1985), no: 2/3, Leiden, 288-302

This PDF-file was downloaded from <http://www.kitlv-journals.nl>

BERND NOTHOFER

THE SUBGROUPING OF THE LANGUAGES OF THE JAVO-SUMATRA HESION: A RECONSIDERATION¹

Dyen, in his 'A lexicostatistical classification of the Austronesian languages', has suggested a subgroup which he has called the 'Javo-Sumatra Hesion'. The tree-configuration of the languages of this subgroup is as follows:

In my dissertation I attempted the reconstruction of the proto-language of the Javo-Sumatra Hesion and called the ancestor 'Proto-Malayo-Javanic' (Nothofer 1975). For the reconstruction I selected Malay, Sundanese, Javanese and Madurese as criterion-languages. I tested

BERND NOTHOFER, who obtained his Ph.D. from Yale University, is at present professor of Southeast Asian Studies at Frankfurt. Specialized in comparative Austronesian linguistics, he is the author of *The Reconstruction of Proto-Malayo-Javanic* and *Dialektgeographische Untersuchungen in West-Java und im westlichen Zentral Java*. Prof. Nothofer may be contacted at Johann Wolfgang Goethe-Universität, Südostasienwissenschaften, Postfach 11 19 32, 6000 Frankfurt am Main.

Dyen's subgrouping by recalculating the cognate percentages among these four languages on the basis of a thorough knowledge of the reflexes of the proto-phonemes in them. In my lexicostatistical calculations the cognate percentage between Javanese and Sundanese is much lower than in Dyen's (Nothofer 1975:3f.). This difference is due to the fact that Dyen considered both Snd. *a* and *eu* as reflexes of PAN *e*, whereas I treat Snd. *eu* as the regular reflex of PAN *e*, while considering Sundanese words with *a* as loans from Javanese and/or, much more rarely, Malay. The other percentages diverge only insignificantly.

The tree-diagram based on my figures still agrees with Dyen's classification:

This subgrouping theory, which is wholly based on lexicostatistics, constituted my basic working assumption.

Blust (1981) suggests a rather different subgrouping of the languages of western Indonesia and mainland Southeast Asia by including some languages not appearing in Dyen's and my subgroup and by excluding others. The new member languages are: Middle-Malay (Besemah), Selako, Iban, Maloh, Rejang, Cham and Jarai. Excluded are: Javanese, Madurese and Lampung. Blust's new subgroup, which he calls 'Malayic', has the following membership: 1) Malay, Minangkabau, Kerinci, Middle-Malay; 2) Selako, Iban; 3) Sundanese, Maloh, Rejang (+ Land Dayak languages); 4) Achinese, Cham, Jarai.

Blust presents the following lexicostatistical and qualitative data in support of his subgrouping.

Let us first consider his lexicostatistical arguments. Blust argues that all cognate percentages except those between Sundanese-Malay and Sundanese-Madurese are lower for the following reasons: there has been a considerable amount of borrowing between Malay and Javanese; Madurese has borrowed from both Malay and Javanese; and Sundanese has been influenced by Javanese. My percentages are too high, because the languages dealt with include a large number of lexical items which are ambiguous as to their history: because of their identical or very similar shapes, they can be considered as being either inherited or borrowed material. When making decisions about cognates in the lexicostatistical lists, I classified ambiguous cases as cognates. Therefore the percentages between Sundanese-Javanese, Malay-Madurese and Javanese-Madurese are distorted to an undetermined degree. Blust con-

cludes that the lowering of the percentages in the above-mentioned pairs supports a grouping that includes Malay and Sundanese but not Javanese.

I agree with Blust's hypothesis regarding the mutual influences of these languages except for his statement that Sundanese "appears to have been effectively isolated from Malay . . . influences until the nineteenth century" (Blust 1981:461). Bosch (1941) presents convincing evidence that Sunda was culturally and politically dominated by Srivijaya from the late 7th until the beginning of the 11th century and in the 13th century. It is because of these fairly intensive historical relations between Srivijaya and Sunda that in my dissertation no proto-forms were reconstructed in the case of words existing only in Malay and Sundanese whose shapes were so similar as to lend themselves to a borrowing hypothesis. It seems reasonable to assume that Sundanese was influenced by Malay, but that Malay borrowed very little from Sundanese. Sundanese nevertheless borrowed a relatively small proportion from Malay compared with the material it borrowed from Javanese. It thus appears that the percentage between Malay-Sundanese is much less distorted than that between the above-mentioned pairs. The only virtually undistorted percentage is that between Sundanese and Madurese.

Considering these facts, one might argue that Sundanese, Malay and Madurese are closely related to each other, since, after the correction of the cognate percentages between Malay-Madurese and Malay-Sundanese, these three languages have very similar percentages, at around 35%. The Malay-Madurese percentage may be somewhat higher at about 38%. Javanese seems less closely related to any of these three languages.

The figures of Table 1 confirm this hypothesis. Table 1 shows the cognate percentages among the four languages in question after elimination of all words from the Swadesh-lists which are so similar as to lend themselves to a borrowing hypothesis. This procedure was followed for *all* possible language pairs.

TABLE 1			
	JAV	MAL	MAD
SND	11.6	14.5	14
	JAV	13.1	11.6
		MAL	20

The table shows that Madurese only scores high with Sundanese and Malay, and that Javanese scores relatively low with all the other languages under consideration. The lexicostatistical evidence seems to support the hypothesis that the closest relatives of Malay among these languages are Sundanese and Madurese. Moreover, Madurese seems to

be more closely related to Malay. Whether Javanese can be included under the next node in a family-tree diagram can only be determined by further lexicostatistical calculations and/or by qualitative evidence.

The figures of Table 1 cannot be taken at face value, however, since the inequality of the retention rates from Proto-Malayo-Polynesian between these four languages has not been taken into account. Blust (personal communication) provides the following data showing that Malay and Madurese have the highest retention percentages among these four languages, viz. Mal. 57.5, Mad. 44.5, Snd. 34.5, Jav. 29.5. This means that Javanese and Sundanese would be expected to share only about $29.5 \times 34.5 = 10.36\%$ basic vocabulary as a result of common retentions from PMP, whereas Malay and Madurese would share some 25.59% (57.5×44.5). These data suggest that the relationship between Malay and Madurese is not significantly closer than that of either language to Sundanese.

Blust's qualitative evidence for his new subgrouping can be summarized as follows. The replacements of the PAN numerals 7-9 (*pitu* '7', *walu* '8', *siwa* '9') in some western Indonesian and mainland Southeast Asian languages suggest that Sundanese, a number of languages in south-central Sumatra, Iban, Maloh and other languages of southwest Borneo, Achinese and the Chamic languages are more closely related to Malay than Javanese, Madurese or Lampung. Blust presents the following table:

TABLE 2

PAN	pitu '7'	walu '8'	siwa '9'
1) Malay	tujoh	delapan	sembilan
Minangkabau	tujueh	salapan	sambilan
Kerinci	tujeuh	salapan	sambiler
Middle-Malay	tujoh	delapan	sembilan
2) Selako	tujuh	lapan	sembilan
Iban	tujoh	lapan	semilan
3) Sundanese	tujuh	dalapan	salapan
Maloh	tuju	lapan	samilan
Rejang	tujua	delapan	semilan
4) Achinese	tujōh	lapan	(sikureuën)
Cham	tijuh	dalapan, salapan	samilan
Jarai	təjuh	səpan	dua rəpan

Blust argues that these numerals are innovations of his 'Malayic'. Javanese, Madurese and Lampung are excluded from this grouping, since they do not show similar innovations, Javanese having *pitu*, *wòlu*, *saṅa*, Madurese having *petto(h)*, *ballu(h)*, *saṅa*, and Lampung having *pitu*, *walu*, *siwa*. Blust reconstructs the following 'Proto-Malayic' numerals:

**tujuh* '7', **dua lapan/dalapan* '8' and **salapan*, **sambilan/samilan* '9'.

According to Blust, the cited forms for 7-9 are best regarded as shared innovations, since:

- a) if they were loans it would be surprising that Javanese and Madurese, which both have been heavily influenced by Malay, did not borrow the Malay forms;
- b) neither Malay *satu* 'one' nor *tiga* 'three' appear in Sundanese, Maloh, Rejang, Achinese or the Chamic languages;
- c) Malay has a nasal cluster in '9' (*səmbilan*), whereas the Iban, Maloh, Rejang and Cham forms have a simple medial nasal.

Let us critically review this qualitative evidence. Several questions arise in this connection.

1) *Are these replacements 'Malayic' innovations?*

Makassarese has *tuju* '7', *sagantuju* '8' (according to Cense (1979) "een meer dan zeven"), *salapan* '9'. Blust's statement that "no other non-borrowed lexical innovation [but Makassarese *salapan* '9' – B.N.] is known to be shared exclusively by Makassarese or other languages of South Sulawesi with the languages of Table 1 [my Table 2 – B.N.]" is not right, since Makassarese has *tuju* both in '7' and in '8'. Two hypotheses are possible: 1) Makassarese borrowed the Malay numerals and *sagantuju* '8' is a secondary development; 2) Makassarese shares the replacements of the PAN numerals 7-9 with Malay and some other languages. The form *sagantuju* '8' is a secondary replacement of the numeral of the proto-language of Malay and Makassarese (**walu* > **dua lapan/dalapan* > *sagantuju*). – It is interesting to note that the Kolo-dialect of Bimanes also has *lape* for 'nine', but has *pitu* for 'seven' and *álu* for 'eight'.

2) *Do Blust's three above-mentioned arguments regarding the decision to consider the numerals 7-9 in his 'Malayic' languages as shared innovations hold?*

- a) To Blust it would be surprising that neither Javanese nor Madurese borrowed the Malay numerals, if the numerals in, say, groups 3) and 4) in Table 2 are the result of borrowing. I do not understand why the borrowing of certain lexical items by certain languages implies the borrowing of these same items by other languages. To me, it would be just as surprising that Malay did not borrow the Javanese numerical system with the lexical innovation *sana* for 'nine'. Madurese, which borrowed more heavily from Javanese than from Malay, did borrow this Javanese numeral.
- b) The fact that the Malay innovations *satu* 'one' and *tiga* 'three' do not appear also in some languages such as Maloh can be explained by the hypothesis that lower numerals tend to be less readily borrowed than higher numerals, since they are more basic². If we look at the dialects

of Sakai and Semang whose numerals are listed in Skeat and Blagden (1966²), we observe that in most of them the numerals for 'one' and 'two' are of non-Malay origin, while in all of them the forms from 4-9 are borrowings from Malay.

- c) The argument that Malay has *-mb-* in the word for 'nine' (*səmbilan*) whereas the Iban, Maloh, Rejang and Cham forms have a simple medial nasal is not strong, since 1) Blust reconstructs **samilan* as one of the forms for 'nine' in 'Proto-Malayic', 2) Mal. *səmbilan* could have been borrowed before **-mb-* became **-m-*, and 3) a restructuring of the form according to phonological patterns of the borrowing language could have taken place.

3) *What are the criteria for including a language in the 'Malayic' subgroup?*

- a) All languages whose forms for 'seven', 'eight' and 'nine' can be considered as cognates of **tujuh*, **dua lapan/dalapan*, **səmbilan/samilan* are treated as belonging to the 'Malayic' subgroup (see Table 2).
- b) Blust argues that the proto-language of the languages in groups 1)-3) (i.e. 'Proto 1)-3)') of Table 2 innovated a metathesized form **hidup* (< PAN *ma-quDip* 'living, alive'). Since this metathesis is found in some Land Dayak reflexes, these languages are regarded as descendants of 'Proto 1)-3)' (Blust 1981:463 and fn. 8). However, the Land Dayak languages only have a form for 'seven' that can be regarded as being cognate with Malay *tujoh*. In most Land Dayak languages the numerals 'eight' and 'nine' are represented by forms that have a geographically limited distribution, e.g. Sadong *mahi*, *prie*. Blust assumes that they are historically secondary replacements of **dua lapan/dalapan*, **salapan*.

This second argument, however, does not hold, since the metathesized form for 'living, alive' is older than 'Proto-Malayic': Balinese has *idup*³ and Sasak has *idup* and *irup*. Consequently, *other* evidence has to be used for the subgrouping of languages which do not have all the reflexes of the so-called "'Malayic' innovations" for 7-9 (if we assume that the presence of these forms is sufficient evidence for classifying a language as 'Malayic'). If we restrict our evidence to that presented by Blust, Land Dayak languages such as Sadong cannot be considered to be closer to the 'Malayic' languages than Makassarese.

If we only consider the numerals, we might argue that languages should be regarded as being less close to Malay if they have forms similar to Mal. *tujoh* for 'seven', but a form derived from **walu* for 'eight' or **siwa*, **siam* for 'nine'. It can be inferred that in these languages the forms for 'seven' are borrowings. This interpretation seems to be supported by the fact that in the languages of Borneo listed in Ray (1913) the numerals for 'eight' and 'nine' are always developments from **walu*

(and *siwa or *siam?) if the numeral for 'seven' is a continuation of *pitu.

4) *Are there any arguments against a subgrouping of languages on the basis of the numerals 7-9?*

There are several arguments against Blust's hypothesis, as follows.

a) *Shared lexical innovation vs. product of borrowing*

The forms for 'seven', 'eight' and 'nine' cited in Table 2 can all be interpreted as either loans from Malay or shared innovations. The history of all these forms is ambiguous. It is not possible, therefore, to determine the degree of relationship on the basis of the presence or absence of these numerals in a given language. The fact that *salapan*, which originally meant 'nine', appears as the word for 'eight' in Minangkabau, Kerinci and Cham and that forms originally meaning 'eight' and 'nine' have been interchanged in Jarai (*dua sapan* '9', *sapan* '8') supports a hypothesis of borrowing of these forms. It was probably lack of "knowledge" of the original morphology of the forms for 'eight' and 'nine' which led to this usage.

b) *Level of reconstruction of the replacement of the PAN numerals 7-9*

Besides the problems in connection with the existence of Mak. *tuju* and *salapan*, there are Javanese forms which suggest that at least *salap(an) and perhaps *dua lapan/dalapan are older than Blust's 'Proto-Malayic'. Pigeaud (1938) lists a bookish Javanese form (*satus*) *salap* '(100)-1:99'. Zoetmulder (1982) has the following entry under *lapan* I (*alapan*?): *salapan* '(with one subtracted) nine?' (cf. Mal. *sambilan*). Under this same entry we also find *dwalapan* 'eight'. Zoetmulder (1982) has one entry which can hardly be treated as a borrowing from Malay, namely OJv. *ḍolapan* 'eight'⁴. OJv. *tuju* in *maləm tuju likur* is marked by Van der Tuuk (1897-1912) as a Malay form.

To conclude the discussion on the numerals 7-9, I would like to briefly describe the system for these numerals as they appear in western Indonesia and Southeast Asia. There seem to exist three major numerical systems. System 1 is that described by Blust, which may be called the 'Malayic' system: *tujuh '7', *dua lapan/dalapan (with occasional secondary replacement) '8', *salapan, *sambilan/samila (with occasional secondary replacement) '9'. System 2, which may be called the 'Javanese' system, has the PAN numerals for '7' and '8', namely *pitu and *walu. The numeral for '9' is the lexical innovation *sanja* < *sanja. System 3 comprises reflexes of the original PAN numerals for 7-9. System 2 is found in, among others, Javanese, Madurese, Bukit, Bajau of Kota Baru, and in Banjar. System 3 occurs in, among others, Lampung, Karo-Batak and Simalungun.

With regard to System 2, I would argue that *sanja* is a Javanese

innovation which was borrowed by Madurese, Bukit, Bajau, and Banjar, and that its occurrence in these languages cannot be taken as evidence for a closer relationship among these languages, particularly since Banjar is genetically very close to standard Malay.

Comments on Blust's internal subgrouping of his 'Malayic' group

Blust's highest-order internal subgrouping of his 'Malayic' group is based on the reflexes of PAN *ma-quDip* 'living, alive'. As explained above, Blust argues that the metathesized form **hidup* is a 'Proto 1)-3)' innovation. I have shown that reflexes of this metathesized form also occur in non-'Malayic' languages. Hence it is pre-'Proto-Malayic'.

With regard to the subgrouping of Iban and Selako on the one hand and Sundanese on the other hand, Blust's evidence is contradictory. He posits that on the one hand Selako and Iban *satu* 'one' and *tiga* 'three' are cognates of Malay and that on the other hand Snd. *anjin* 'dog' is a cognate of Malay. Since the relevant Selako and Iban forms are low numerals, one might tend to treat these as shared innovations rather than as loans. The Sundanese form, however, can be considered as either a borrowing or a shared innovation, so that this material cannot be used as a basis for subgrouping. Furthermore, there seems to exist a Sundanese word which directly continues PAN Wasu 'dog'. Coolsma (1913) lists Snd. *basu*: 'a legendary monster in the form of a wild boar with the head of a dog with scabies'. It should also be noted that Old Javanese has a form *anjin* 'dog'.

Qualitative evidence in support of a grouping together of Malay, Madurese, Sundanese and Javanese

There exists qualitative evidence suggesting that Malay, Madurese, Sundanese and Javanese form a single subgroup which also includes the languages of Blust's 'Malayic' subgroup. On the following pages I will only attempt to demonstrate the membership of Madurese and Javanese of a group to which both Malay and Sundanese belong. Therefore I will only consider material from these four languages, with occasional references to Lampung data to show that this language probably also is a member of this putative subgroup. In accordance with the usage in my dissertation, I shall call this subgroup 'Malayo-Javanic', although a more appropriate name might be suggested.

On the basis of the new lexicostatistical classification as set out in Table 1, I will treat as Proto-Malayo-Javanic innovations those sets of likely cognates which are found in Javanese and in at least one of the following languages: Malay, Madurese, Sundanese, or Lampung, and which have no extra-Malayo-Javanic cognates. I will present only those sets in which the relationship between the Javanese form and the form in one or more of the other languages cannot be treated as products of a borrowing relationship. These lists will illustrate that one is indeed

justified in positing a subgroup which has Malay, Madurese, Sundanese, Javanese and Lampung among its members.

TABLE 3

Innovations shared by Malay, Sundanese, Madurese and Javanese

PMJ	Malay	Sundanese	Madurese	Javanese
*keRiŋ ⁵ 'dry'	kəreŋ	keŋ	kərreŋ	a-kéŋ
*kuciŋ 'cat'	kuceŋ	uciŋ	kocəŋ	kucəŋ
*helaŋ 'eagle'	(hə)laŋ	heulaŋ	laŋ	OJv. (hə)laŋ

There is an abundance of comparable cases of which there exist forms in all four languages which are ambiguous as to whether they are shared innovations or borrowings. An example is the comparison Mal. *jantŋ*, Snd. *jantun*, *jajantun*, Mad. *jhantŋ*, Jav. *jantŋn* 'heart'. I would refer to my dissertation for further examples. It seems reasonable to assume that at least in some of these cases the Javanese form and the form in another Malayo-Javanic language are cognate.

TABLE 4

Innovations shared by Malay, Sundanese and Javanese

PMJ	Malay	Sundanese	Javanese
*kihuŋ 'land-snail'	kioŋ	keoŋ	kéyoŋ
*(mb)entigi 'Pemphis acidula'	məntigi	cantigi	təmtigi

TABLE 5

Innovations shared by Sundanese, Javanese and Madurese

PMJ	Sundanese	Javanese	Madurese
*duqum 'distribute'	duum	dóm	dhuqum
*keqerŋ 'loneliness'	keueŋ	kón	koqorŋ
*DaŋDeR 'cotton tree'	daŋdeur	raŋdu	naŋghər
*laha 'fish-trap'	laha	laha	laqa
*lahaŋ 'palm wine'	lahaŋ	lahaŋ	laqaŋ
*taŋhi 'get up'	taŋhi	taŋi	taŋge
*jaraw 'bamboo with holes'	jaro	jaró	jharuy
*galaw 'mixed'	galo	galó	ghaluy

TABLE 6

Innovations shared by Malay, Javanese and Madurese

PMJ	Malay	Javanese	Madurese
*Zeliŋ 'see'	jəleŋ	dələŋ OJv. dəliŋ	jhəlliŋ
*peliR 'scrotum'	pəler	pəli 'penis'	pəller
*ta(m)bi(ŋ)kaR 'shard, potsherd'	təmbikar	wiŋka	tabiŋkar

TABLE 7

Innovations shared by Sundanese, Javanese and Lampung

PMJ	Sundanese	Javanese	Lampung
*pa-waRi 'dry in the sun'	poe	pé	paway
*piteuk ⁶ 'gadfly'	piteuk	pitaq	pitoq

TABLE 8

Innovations shared by Sundanese and Javanese

PMJ	Sundanese	Javanese
*huntu 'tooth'	huntu	untu
*hiji	hiji ⁷ 'one'	iji, ijèn 'alone'
*pV-haney 'arrange warp on the loom'	pihane	pani
*e(n)juk ⁸ 'palm fibre'	injuk, eunjuk (dial.)	dóq
*sideha 'push'	sideuha	siduwa
*waya	aya 'be'	OJv. wwaya 'be' bóya, mbóya 'not be'

TABLE 9

Innovations shared by Malay and Javanese

PMJ	Malay	Javanese
*Zeles 'true; plain'	jəlas	dələs
*pear 'pedestal tray'	pahar	pòr

TABLE 10

Innovations shared by Javanese and Madurese

PMJ	Javanese	Madurese
*ari-an 'sugar palm'	arèn	areyan
*kiZuŋ 'song'	kidónj	kejhuŋ
*unZaŋ 'invite'	undaŋ	onjhaŋ
*kuncaR 'loose end of garment'	kunca	koncar
*sehaR 'fibre'	suwa	saqar
*laqas 'unhusked rice grain'	las	laqas

The internal subgrouping of the Malayo-Javanic languages

It appears that Madurese is closer to Malay than Sundanese. Iban, which is genetically closest to Malay apart from Minangkabau, Kerinci, and Middle-Malay (see Blust 1981:464), and Madurese both share the following innovations: Iban *səmak*, Mad. *səmmaq* 'near, adjacent'; Iban *angup*, Mad. *aŋhuq* 'tweezers'; Iban *tasi*, Mad. *tase(h)* 'cord'; Iban *mua*, Mad. *muwa* 'face'; Iban *igi* 'numerical classifier for fruit, eggs, seeds, teeth, and the like', Mad. *bighi* in *sabighi* 'a piece'. Both Malay and Madurese and also Iban (and Achinese) share reflexes of the innovation *katupat 'packet of glutinous rice': Mal., Iban *kətuPAT*, Mad. *katopaq*

'id.'. The appearance of Malay, Sundanese and Madurese *tara* 'flat, even' seems to support the hypothesis that these three languages have an immediate ancestor, since this form appears to be a metathesized form of an early Javanese form *rata* 'id.' < Pan DataR 'id.'.⁹

The exact position of Sundanese is difficult to determine. It seems closely related to Lampung, with which it shares the metathesis of the consonants in the reflexes of PAN *lapaR* 'hungry, exhausted': Snd. *palay* 'desire, tired', Lmp. *palay* 'hurt of tired feet'. In both languages (one type of) *R becomes *y*. Sundanese shares a considerable number of innovations with Malay, Iban and Madurese. I therefore agree with Blust's hypothesis that the development of some languages in western Indonesia "has not always followed the abrupt cleavages implied by the family-tree model" (Blust 1981:464). Further in agreement with Blust, I would like to posit the existence of a dialect chain in which Malay intervened between certain languages. According to Blust it is likely that "Malay intervened between Sundanese and the Bornean communities with which it was in closer contact" (Blust 1981:464). I would argue that Malay intervened perhaps between Sundanese, Lampung, Iban (and perhaps other Bornean languages) and Madurese, being in closer contact with the latter two languages. In disagreement with Blust, I propose a proto-language common to Lampung, Malay, Iban, Madurese, Sundanese and Javanese whereby Lampung, Malay, Iban and Madurese and perhaps Sundanese are closer relatives. As evidence in support of the hypothesis of a closer relationship among Sundanese, Malay and Madurese one should consider the following lists of innovations which appear to be shared exclusively between Madurese-Malay, Madurese-Sundanese and Malay-Sundanese(-Lampung).

TABLE 11

Innovations shared by Malay and Madurese

PMJ	Malay	Madurese
*lemukut 'broken rice tips'	ləmukot, məlukot	malokoq
*juay 'brother-(sister-)in-law'	ipar duay	luway
*pala(jt) 'penis'	palat	palaq
*u(rR)i 'afterbirth'	uri	ore
*ge-Rahem 'molar tooth'	gəraham, gərham	ghərrəm

TABLE 12

Innovations shared by Sundanese and Madurese

PMJ	Sundanese	Madurese
*l-a(n)cuŋ 'k. of plant'	acuŋ	laccoŋ
*kanTit 'put together'	kantet	kanteq
*apiŋ 'close to'	apiŋ	apeŋ
*k-uriŋ	kuriŋ 'I'	oreŋ 'person'
*heay 'yawn'	heuay	oway

*harjet kuku 'luke-warm'	haneut kuku	ṅa-aṅaq koko
*yuryun 'lovely'	ka-yuryun	ka-junjun
*kaliki 'Palma christi'	kaliki	kaleke
*kalin/kilaṅ 'k. of red ant'	kilaṅ	kaleṅ

TABLE 13

Innovations shared by Malay and Sundanese (and Lampung)

PMJ	Malay	Sundanese	Lampung
*dian 'dry, toast'	məndian	sidean	badian
*ruay 'variety of the <i>lalab</i> '	ruay <i>in</i>	roay <i>in</i>	
*birem 'red'	biram	beureum	
*(zZ)uney 'k. of plant'	juni	jone	
*huRem 'shadow'	m-uram	hieum	

Besides the metathesis of the Javanese form *rata* which is shared by Malay, Sundanese and Madurese (*tara*), the following comparisons also appear to represent innovations shared exclusively by these three languages.

TABLE 14

Innovations shared by Malay, Sundanese and Madurese

PMJ	Malay	Sundanese	Madurese
*liat 'tough'	liat	liat	leyaq
*kuliāt 'stretch'	kəliat, gəliat	kuliāt	kaliyaq
*saDap 'tapping knife'	sadap	sadap	saḍaq

I am well aware that some of the comparisons cited in the preceding tables will turn out to have cognates in languages outside the putative subgroup. But there may remain enough cognate sets which make it probable that we are dealing with a subcollection of cognates that 1) connect Javanese with Malay, Sundanese, Madurese and Lampung, and 2) more closely connect Malay, Sundanese, Madurese and Lampung.

Thus, there now exist three hypotheses regarding the classification of the western Indonesian and mainland Southeast Asian languages, namely Dyen (1965) and Nothofer (1975), Blust (1981), and the one just put forward. Why is there no agreement regarding the subgrouping of these languages? The answer is that we are faced with serious and possibly insurmountable problems which have their origin in two closely connected historical developments. 1) These languages and other languages of western Indonesia such as Sasak and Balinese have identical reflexes for many of the most frequently occurring PAN phonemes (namely the nasals, the liquids (except *R, if it was a liquid), *p, *t, *k in non-final position, *s, *c, and the vowels except *e). The result of this is the existence of an abundant number of comparable cases which are

ambiguous as to whether they are cognates or borrowings. 2) Both Malay and Javanese have heavily influenced most other languages of western Indonesia, including languages such as Balinese, Sasak, Makasarese, Buginese or Batak and Ngaju-Dayak. These historical developments not only make it difficult to trace the history of comparisons within a putative subgroup, but also give rise to uncertainty as to the level of reconstruction of comparisons. For example, in a comparison such as Malay *li(h)at*, Lampung *liaq*, Madurese *leyaq*, Balinese *liat* 'to see', it cannot be determined whether the Balinese form is borrowed or whether it is inherited. Thus it is not determinable whether *lihat is Proto-Malayo-Javanic or whether it is an older form. Furthermore, even some of those morphemes in the languages which have been influenced by Malay and/or Javanese which contain inherited reflexes may have to be considered as borrowings. For example, Dyen (1953) reconstructed four tentative sub-*R's and Nothofer (1975) three tentative sub-*R's without there being a significant match among these *R's. This indicates to me that we are mistakenly analysing some comparisons as cognates. Take the following example: Nothofer (1975) reconstructs a sub-*R because of comparisons such as Snd. *beurat*, Mal. *bərat*, Jav. *abət* 'heavy' (< PAN *beRat*) and posits that this type of sub-*R has *r* as regular reflex both in Sundanese and Malay (Javanese having $\emptyset$). The most frequently occurring reflex of Dempwolff's *R, however, is Snd. *y*. Today I am inclined to argue that Snd. *beurat* is not an inherited form, although it contains the *eu* which is considered – as explained above – to be the inherited reflex of PAN *e*, as opposed to Snd. *a* which is interpreted as occurring only in loans. I would argue that Sundanese borrowed Mal. *bərat* at a time before **e* became Snd. *eu*. Let us look at another example. It might be posited that some of the Madurese morphemes with final *q* as the regular reflex of *-p, *-t and *-k are not cognates but borrowings, since these words may have been borrowed from Malay or Javanese at a time before *-p, *-t and *-k became Mad. -*q*.

It also has to be mentioned in this context that the languages dealt with here appear to contain no clear morphological evidence which would allow of our putting forward solid hypotheses regarding their subgrouping.

The study of the history of the languages of greater western Indonesia and of mainland Southeast Asia as well as of their relationships will remain a challenge to all those involved in the historical comparative analysis of western Austronesian languages.

NOTES

- 1 An earlier version of this paper was presented at the XV Pacific Science Congress at Dunedin in 1983. I am grateful to J. C. Anceaux, R. A. Blust and J. Noorduyn for information and critical comments offered during the writing of this article.

- 2 Noorduyn (personal communication) points out that Sundanese, in fact, has *sat*, but only in the expression 'one o'clock', which is always *pukul satu*, and never *pukul hiji*.
- 3 Blust cites Bal. *hidup*, which is taken to be a Malay loan (Blust 1981:467, fn. 9). I was unable to find this form in my sources.
- 4 Van der Tuuk (1897-1912) lists two forms for 'eight': *ḍwalapan* (*ḍo* + *alapan*) and *ḍwambilan* (*ḍo* + *ambilan*).
- 5 Blust (personal communication) points out that Kalamian Tagbanwa has *kagin* 'thoroughly dry' from *kaRin, which can be considered as a doublet of *keRin.
- 6 *pitek is a metathesized form of an older *piket.
- 7 Blust treats Snd. *hiji* as an apparently "irregularly altered borrowing of Javanese *siji*" (Blust 1981:464).
- 8 The PHN form reconstructed by Dempwolff (1934-8) is *izuk.
- 9 Malay *datar* 'smooth, flat' directly continues *DataR.

BIBLIOGRAPHY

- Abdul Djebbar Hapip
 1977 *Kamus Banjar-Indonesia*, Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.
- Abdul Djebbar Hapip, Darmansjah, and Basran Noor
 1979 *Bahasa Bajau*, Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.
- Abdurachman Ismail, Abdul Kadir Ismail, Ahmad Aini, Muhammad Yusni, and Nazamuddin Nasufi
 1979 *Bahasa Bukit*, Jakarta: Pusat Pembinaan dan Pengembangan Bahasa.
- Blust, R. A.
 1981 'The reconstruction of Proto-Malayo-Javanic: an appreciation', *BKI* 137: 456-69.
- Bosch, F. D. K.
 1941 'Een Maleische inscriptie in het Buitenzorgsche', *BKI* 100:48-53.
- Cense, A. A.
 1979 *Makassaars-Nederlands woordenboek*, 's-Gravenhage: Martinus Nijhoff.
- Coolsma, S.
 1913 *Soendaneesch-Hollandsch woordenboek*, 2nd. ed., Leiden: A. W. Sijthoff.
- Dempwolff, O.
 1934-8 *Vergleichende Lautlehre des austronesischen Wortschatzes*, ZfES, Supplements 15, 17, 19, Berlin: Reimer.
- Djajadiningrat, H. R. A.
 1934 *Atjehsch-Nederlandsch woordenboek*, Batavia: Landsdrukkerij.
- Dyen, I.
 1953 'Dempwolff's *R', *Language* 29:359-66.
 1965 'A lexicostatistical classification of the Austronesian languages', *International Journal of American Linguistics*, Memoir 19 (vol. 31, no. 1).
 1971 'The Austronesian languages and Proto-Austronesian', in: T. A. Sebeok (ed.), *Linguistics in Oceania*, 1971:5-54.
- Gonda, J.
 1953 'Varia over Indonesische telwoorden', *BKI* 109:23-31.
- Goris, R.
 1938, 1940 *Beknopt Sasaksch-Nederlandsch woordenboek*, Singaraja.
- Jonker, J. C. G.
 1896 *Bimaneesche spraakkunst*, Batavia: 's Hage.
- Kiliaan, H. N.
 1904 *Madoereesch-Nederlandsch woordenboek*, Leiden: E. J. Brill.
- Nothofer, B.
 1975 *The reconstruction of Proto-Malayo-Javanic*, Verhandelingen KITLV 73, The Hague: Nijhoff.

Panitia Penyusun Kamus Bali-Indonesia

1978 *Kamus Bali-Indonesia*, Dinas Pengajaran Propinsi Daerah Tingkat I Bali.

Pigeaud, Th.

1938 *Javaansch-Nederlandsch handwoordenboek*, Groningen: J. B. Wolters.

Ray, S. H.

1913 'The languages of Borneo', *Sarawak Museum Journal* 1.4:1-196.

Richards, A.

1981 *An Iban-English dictionary*, Oxford: Clarendon Press.

Ronkel, Ph. S. van

1936 'De telwoorden "acht" en "negen" in de Indonesische talen', *Verslag 8e Congres Oost. Gen. Ned.*: 43-45.

Sebeok, T. A. (ed.)

1971 *Linguistics in Oceania*, Current trends in linguistics, vol. 8, The Hague: Mouton.

Skeat, W. W., and C. O. Blagden

1966² *Pagan races of the Malay Peninsula*, London: Frank Cass.

Thaib St. Pamoentjak

1935 *Kamoes Bahasa Minangkabau-Bahasa Melajoe-Riau*, Batavia: Balai Pustaka.

Tuuk, H. N. van der

1897-1912 *Kawi-Balinesch-Nederlandsch woordenboek*, 4 vols., Batavia: Landsdrukkerij.

Walker, D. F.

1976 *A grammar of the Lampung language: the Pesisir dialect of Way Lima*, Jakarta, Badan Penyelenggara Seri NUSA.

Wilkinson, R. J.

1959 *A Malay-English dictionary* (Romanised), London: Macmillan.

Zoetmulder, P. J.

1982 *Old Javanese-English dictionary*, 's-Gravenhage: Martinus Nijhoff.