# MALAY CONSONANT-HARMONY: AN INTERNAL RECONSTRUCTION

K. A. ADELAAR

University of Leiden

Although it is perhaps the best studied Austronesian language, the phonotactics of Malay remains poorly described. Following from some of Fokker's (1895) observations on the details of cooccurrence of certain consonants, all the disyllabic lexemes displaying one of five patterns of homorganic initial and medial consonants are studied. After eliminating cross-dialectal and interlingual borrowing, contractions and frozen derivations as well as acknowledging that *d*, a dental stop, and *t*, an alveolar stop, do not share the same place of articulation, it appears that, with very few exceptions, at an earlier stage of Malay there was a constraint in disyllabic lexemes against the occurrence of a medial stop or nasal, homorganic with, but not with the same voicing as, the word initial stop. It is suggested that this phonological constraint was found at least as early as Proto-Malayo-Polynesian as well.

This article has been made possible by a grant from the Netherlands Foundation for the Advancement of Tropical Research.

## 1. INTRODUCTION

Malay phonotactics has never been described in full. Scholars have distinguished inherited from borrowed phonemes, but the phonotactic rules which apply to the inherited vocabulary are generally not mentioned. Two important exceptions have to be made: in his doctoral thesis, Fokker (1895) gives a description of the Malay sounds and their positions in a lexeme, and Emeis (1955:196-7) draws a pattern for the possible combination of vowels in inherited (disyllabic) lexemes. The phenomenon Emeis calls vowel-harmony could be more appropriately described as 'height-harmony', since it asserts that the mid-vowels *o* and *é* never co-occur with high vowels (*u* and *i*), and co-occur with *a* only in very narrowly defined positions.

In this article I would like to draw attention to the phonotactic rules which apply to consonants. Although Fokker's phonetic description is rather weak even for its time, his phonotactic remarks are fairly accurate and will serve as a basis for part of the conclusions drawn in this article.

The following remarks are particularly interesting:

> p.24: In a Malay disyllabic stem *k*[1] as an initial of one of the syllables is never met along with *g* as an initial of the other.
> p.40: *c* and *j* never occur together in one Malay stem.
> P.61: The only instance of *b* as initial of a stem containing *p* as initial of the root-syllable[2] are: *bapa* 'father' and *bopeŋ* 'pock-pitted' (sic).
> p.62: Not a single instance is quotable of *p* in a Malay stem followed by *b* or *mb* at the beginning of the root-syllable[2].

If Fokker's statements are correct, it follows that inherited Malay lexemes show a tendency to consonant-harmony in that when initial stops are followed by stops of the same articulation place, the latter must agree in voicing. Other statements made by Fokker suggest that such a tendency to articulation-type harmony also applies to inherited lexemes with initial (voiced as well as unvoiced) stops followed by homorganic nasals:

> p.43: *ñ* and *c* or *ñ* and *j* never occur together in an unchanged[2] Malay stem.
> p.63: *p* as initial of a stem is never followed by *m* as single initial of the root-syllable[2].

There are, however, two sets of facts that should be taken into account before concluding that such a tendency exists in the directly inherited vocabulary:

1. the obvious gap in the pattern is that it does not apply to the distribution of *d* and *t*, which are combined in many inherited lexemes like *tidur* 'to sleep'. *tudoh* 'to accuse', *dataŋ* 'to come', *datok* 'chief, head; ancestor, grandfather', etc.;
2. there are, together with *bapa* and *bopeŋ*, a few more exceptions which require an explanation.

In the following pages I will discuss the articulation-type harmony observable in the inherited vocabulary of Malay and try to determine the extent of this phenomenon. The dictionary gives many lexemes that contradict the existence of

articulation-type harmony on the synchronic level. In Sections II and III I point out that most of these lexemes are loanwords, or were shaped through secondary developments. In Section IV I discuss the internal evidence for such a harmony in an earlier stage of Malay and also the agreement with regard to it between the structure of inherited Malay lexemes and the structure of the Proto-Malayo-Polynesian (henceforward PMP) morph[3] (Chrétien 1965).

I will go beyond Fokker's remarks by checking in initial and medial position all possible combinations of homorganic stops, nasals, and semivowels which differ in articulation type[4]. For practical reasons I will confine myself to an inquiry into disyllabic lexemes. In the Malay lexical structure disyllabicity is the norm[5]. Lexemes of more than two syllables are often subject to syllable reduction, and, if not borrowed, usually contain a petrified affix.

As a lexical source I will use Wilkinson's (romanized) *Malay-English Dictionary* (1932)[6], and will modify its spelling in the following way:

| Wilkinson: | modified spelling |
|---|---|
| ch | c |
| ng | ŋ |
| ny | ñ |
| -au | -aw |
| -ai | -ay |
| -ui | -uy |

For the apparent inconsistency given sub 1), I propose the following explanation: initial *t* allows following *d* and vice-versa, because they differ not only in voice, but also in place of articulation. Linguists generally concur that in Malay *d* is alveolar and *t* dental, but since it is phonemically irrelevant, this fact has never seemed worth mentioning. To my knowledge, the only authors who have described these phonemes are N. F. Alieva et al. (1972:34):

> "The consonants *t* and *d* .... voiceless and voiced apical stops. The closure is produced by raising the tongue-tip, which for *t* is pressed against the upper teeth, whereas for *d* it is pressed against the alveolar ridge". (par.21)

Furthermore, Henderson (1965:420-21) mentions this phonetic distinction as an areal feature, common to many Southeast-Asian languages.

As to *n*, this is an alveolar nasal (and thus homorganic with *d*), which becomes a supradental when prenasalized to *t*, cf. Alieva et al. (p.36):

> "The consonant *n* is an apical nasal sonant with oral closure. The tongue-tip is pressed against the alveolar ridge, as with the pronunciation of *d*, ... but before *t* the tongue-tip is pressed, as with the pronunciation of *t*, against the upper teeth, for instance *bantu* 'to help, assist'." (par.27).

## 2. LEXICAL EVIDENCE

The material under discussion will be treated according to the pattern to which each lexeme belongs. Since Malay consonant-clusters and semivowels do not occur initially[7], and given a consonant-chart that looks as follows:

| | labial | supra-dental | alveolar | palatal | velar |
|---|---|---|---|---|---|
| Stops voiceless | p | t | | c | k |
| voiced | b | | d | j | g |
| Nasals | m | | n | ñ | ŋ |
| Semivowels | w | | | y | |
| Others | h, s, l, r | | | | |

I distinguish the following patterns of homorganic consonants in initial and medial position:

I. $S_h\ V\ S_h^+\ V\ (C)$, $S_h^+\ V\ S_h\ V\ (C)$ = initial stop followed by homorganic stop differing in voice

II. $S_h\ V\ N_hS_h^+\ V\ (C)$, $S_h^+\ V\ N_hS_h\ V\ (C)$ = initial stop followed by prenasalized homorganic stop differing in voice

III. $S_h^{(+)}\ V\ N_h\ V\ (V)$ = initial stop followed by homorganic nasal

IV. $N_h\ V\ S_h^{(+)}V\ (C)$, $N^h\ V\ N_hS_h^{(+)}\ V\ (C)$ = nasal followed by simple or prenasalized homorganic stop

V. $S_h^{(+)}V\ w/y\ V\ (C)$, $N_h\ V\ w/y\ V\ (C)$ = stop or nasal followed by homorganic semivowel.

The symbols in these patterns have the following meaning:
S = stop, V = vowel, C = consonant, N = nasal, + = voiced, h = homorganic, ( ) = optional, / = or.

Before continuing with the material, I give as a matter of comparison the

approximate numbers (rounded off downwards) of disyllabic lexemes with an initial stop or nasal:

*b-* : 1200; *d-* : 850; *j-* : 600; *g-* : 880;
*p-* : 1000; *t-* : 1200; *c-* : 900; *k-* : 1600;
*m-* : 900; *n-* : 240; *ñ-* : 60; *ŋ-* : 75;

The numbers of disyllabic lexemes with identical initial and medial stops or nasals are:

| | | |
|---|---|---|
| bVbV(C) : 59; | jVjV(C) : 35; | gVgV(C) : 51; |
| bVmbV(C): 21; | jVnjV(C): 12; | gVŋgV(C): 19; |
| pVpV(C) : 37; | cVcV(C) : 56; | kVkV(C) : 63; |
| pVmpV(C): 15; | cVncV(C): 12; | kVŋkV(C): 18; |
| mVmV(C) : 40; | ñVñV(C) : 16; | ŋVŋV(C) : 16; |

Having provided a rough numerical breakdown of the number of lexemes with initial stop or nasal as well as the number of disyllabic lexemes with identical initial and medial stops or nasals, I proceed with the lexical data relevant to the five patterns described above.

## I. $S_h$ V $S_h^+$ V (C), $S_h^+$ V $S_h$ V (C)

(bVpV(C)) : *bapa, bapaŋ, bapak* 'father'; *bĕpaŋ, bipaŋ, bĕpa,* in *kueh* -- '(k.o.) sweetmeat'; Wilkinson guesses that this is a Chinese loanword; *bopeŋ* 'pock-marked', from Chinese (Leo 1975:8);
(pVbV(C)) : *pabu,* in *main* -- 'somersaults and other tricks by Chinese tumblers', a Chinese loan; *pĕbin* '(k.o) teetotum', from Chinese; *pobien* (disyllabic?) 'wharf', from Chinese;
(jVcV(C)) : *jicuy* 'opium dross doctored a second time for consumption', from Chinese;
(cVjV(C)) : ----
(gVkV(C)) : *ge'kok, go'kek* 'gecko, house-lizard, an onomatope from Javanese;
(kVgV(C)) : *kagak* 'no', from Jakartanese; *kaget* 'startled', from Jakartanese, ultimately from Javanese; *kagum* 'astonished', from Jakartanese, ultimately from Javanese; *kugah* '(k.o.) shrub',

Most of these lexemes are loanwords; the only ones that cannot be explained as such are *bapa* and *bĕpaŋ* (and their variants) and *kugah*.

*Bapa* (*bapaŋ, bapak*) may originally have been a complex form, but I am not certain: in Minangkabau *apa'* and *pa'* 'father' are more usual than *bapa'*, and Iban has *apay* (in which the final *-y* is a secondary development)[8] for 'father, and *bapaq* for 'father-in-law'. There are cognates for *bapa* as well as for *apa'/apay* as far away as Eastern Indonesia[9]. Furthermore, I cannot make out whether *bapa* was originally a complex form (*+b-apa*), or *apa'* and *apay* have lost initial *+b*, or that the forms with *b-* developed independently from those without *b-*.

*Bĕpaŋ (bipaŋ, bĕpa)* is probably from Chinese, like so many other culinary terms, but Wilkinson is not sure.

It is noteworthy that, apart from *bapa*, none of the exceptions belong to the more basic vocabulary of Malay.

## II. $S_h$ V $N_h S_h^+$ V (C), $S_h^+$ V $N_h S_h$ V (C)

(bVmpV(C)) : *bimpaw, bimpo* 'handkerchief, towel', a Chinese loan;
(pVmbV(C)) : *pombak* 'dove, pigeon', a Portuguese loan;
(jVncV(C)) : *jinceŋ* 'grateful', a Chinese loan;
(cVnjV(C)) : ----
(gVŋkV(C)) : ----
(kVŋgV(C)) : ----

The only lexemes belonging to this pattern are loanwords.

## III. $S_h^{(+)}$V $N_h$ V (C)

(bVmV(C)) : *bami* '(k.o.) noodle-dish', from Chinese; *beman,* in *kabur* -- 'tax for the support of royal bandsmen in Old Perak', probably from English 'bandsman'; *Bima* '(proper name)', from Sanskrit; *bomo, bomoh, bomor* 'mage, sorcerer'; *bumi* 'earth', from Sanskrit;
(pVmV(C)) : *pama* 'police-informer', probably from English 'informer'; *pa'ma* (with a medial cluster) 'the plant Rafflesia hasseltii', a Javanese name, the true Malay name being *kerubut*; *pamah* 'low-lying (land)'; *paman* 'uncle', from Javanese; *pamer* 'talking big, boasting', from Javanese; *pamit* 'to beg leave or depart', from Javanese; *pamor, pamur* 'alloyed (iron)', from Javanese; *pomaŋ* 'timber-tree', probably a Jakun loan: its only source is the journal of a traveller in Johore, who found the term in a predominantly Jakun area (Hervey 1881:131);
(jVñV(C)) : *jaña* "to say, think', from Minangkabau *ja+ña (jaño),*

cf. Kendayan-Dayak *jaña* 'his/her words, what she/he says' (Dunselman 1949:69);

(cVñV(C)) : *cuña* '(k.o.) boat', from Jakartanese, ultimately from Chinese;

(gVŋV(C)) : *giŋin* '(k.o.) herb';

(kVŋV(C)) : *kaŋa*, in *damar* -- '(k.o.) tree';
*kaŋar*, *laŋ* -- 'bird of prey';
*kaŋěn* 'to long or pine for', from Jakartanese, ultimately from Javanese;
*koŋek* '(belonging to the) public', from Jakartanese, ultimately from Chinese.

Of these 20 lexemes, 12 are given as loanwords by Wilkinson. Of the remaining lexemes, *beman*, *pama*, and *pomaŋ* must be considered as loanwords as well; five lexemes remain: *bomo (bomoh, bomor)*, *pamah*, *giŋin*, *kaŋa*, and *kaŋar*.

## IV. $N_h$ V $S_h^{(+)}$ V (C), $N_h$ V $N_h S_h^{(+)}$ V (C)

mVbV(C) : *mabaw* 'evil spirit of disease';
*mabay* '(k.o.) tree';
*mabir*, in *tabir-* -- 'all kinds of wall-drapery', derived from *tabir*, cf. *kueh-mueh* 'all k.o. cakes', from *kueh*, *sayur-mayur* 'all k.o. vegetables', from *sayur*, *sendeŋ-mendeŋ* 'tilted this way and that', from *sendeŋ*;
*mabok* 'intoxicated';
*mabub* '(proper name)', from Arabic (*maḥbūb* 'darling');
*mabur* 'to fly', from Javanese;
*mobil* 'car', from Dutch;
*mubal* 'to shoot up', from Javanese;
*mubeŋ* 'to circle, revolve', from Javanese;
*mubyar* (written *mubiar* in Wilkinson) 'to strike the eye, to "scream" (of colour)', from Javanese;

mVmbV(C) : *mambaŋ* I 'spirit of the Indonesian pantheon';
II *ikan* -- 'a snapper', derived from *bambaŋ*;
III title given occasionally to negrito-headmen
*mambu* I = *(semambu)* 'Malacca cane';
II = *mabaw* ('evil spirit of disease');
III 'to smell', from Javanese;
*mambul* 'k.o. climber';
*mimbar* 'pulpit in a mosque', from Arabic;

mVpV(C) : *mapag*, *mapak* 'to go out and meet and then escort to one's house', from Jakartanese, ultimately from Javanese or Sundanese;
*mapar*, *gaŋ* 'flat-ended brazier's chisel or punch', derived from *papar*;
*mapat* '(k.o.) tree';
*mapok* '(k.o.) herb';
*mepas*,=*mempas* 'to fish with the fly';
*meper*, *mipir* 'to edge away under a blow', derived from *peper*;
*mipis* 'thin, tenuous', a variant of *nipis* and *tipis*;
*mopeŋ* 'pock-marked' cf. *(bopeŋ)*, from Chinese (Leo 1975:8);
*mupar*, in *ular tědoŋ* -- 'black cobra', derived from *upar*;
*mupoh*, *mupus* '(k.o.) tree';

mVmpV(C) : *mampat* I 'tightly packed';
II '(k.o.)tree';
*mampir* 'to stop; to touch at', from Javanese, the usual Malay term being *siŋgah*;
*mampoh*, *mampu* 'having the means for', from Jakartanese, ultimately from Sundanese;
*mampoŋ* 'light and spongy in texture', which must be derived from *(h)ampoŋ* 'adrift, afloat; light, lively, cork-like', with a petrified prefix *m-* (see Conclusion);
*mampus* 'to die (vulgar); to be wiped', according to Klinkert derived from *(h)ampus* 'dead, wiped out' (with a petrified *m-*);
*mempas* I 'to fly-fish' = *mepas*;
II -- *jantan* '(k.o.) tree';
*měmpat* '(k.o.) tree';
*mempěr* 'to bear some resemblance to', from Javanese;
*měmpoŋ* '(k.o.) tree';
*mimpi* 'dream', derived from *impi* (with a petrified *m-*);
*mompoŋ* 'used up (*nipah*, *pandan*, etc.), of which the fronds have been taken';
*mumbaŋ* 'coconut in its earliest stage of growth';
*mumpun* 'blunt, stumpy', from Minangkabau;

(ňVjV(C)) : ----
(ňVnjV(C)) : ----
(ňVcV(C)) : ----
(ňVncV(C)) : ----
(ŋVgV(C)) : ----
(ŋVŋgV(C)) : ----

(ŋVkV(C)) : *ŋakak* 'to cackle (fowl)', from *kakak*, a Javanese onomatopoeia;
*ŋěkek* 'to giggle', derived from *kek*;

(ŋVŋkV(C)) : ----

There is a large number (45) of

lexemes which fit into pattern IV. Of these, 14 are loanwords, 9 are derivations, and two are variants of each other (*mepas* and *mempas*) according to Wilkinson. But also *mabok* is originally a derivation, which developed from PAN *ma-buSek (Blust 1976:fn.18), and *mipis* is a variant of the more original *nipis* (Dempwolff 1938). Consequently there remain 18 lexemes which cannot be reduced to loanwords or to lexemes which underwent secondary developments.

V. $S_h^{(+)}$V w/y V (C), $N_h$ V w/y V (C)

(bVwV(C)) : *bawa, bawa', bawak* 'to convey', from PMP *baba[10];
*bawab* 'gatekeeper, porter', from Arabic;
*bawah* 'position under or below', from PMP *babaq[10];
*bawal, ikan* -- '(k.o.) fish, pomfret', which could be from Tamil *voval* according to Wilkinson;
*bawan* 'comrade, playfellow', which must be derived from *bau* 'smell; to smell', cf *sebau* 'of one smell, used to each other, not shy towards each other';
*bawaŋ* 'bulb';
*bawar* I 'customs barrier', a variant of *gawar*;
II ' sword of office', a variant of *baur*;
*bawat* I *payoŋ* -- 'state-umbrella';
II 'drooping, inclining downwards';
III *tali* -- 'braces';
*bewah* 'to feast for the dead', a contraction of *ber-arwah*, and *arwah* is from Arabic *arwāḥ* 'spirits';
*bewak* 'monitor-lizard', a Kedah and Pattani contracted variant of *biawak*;

(pVwV(C)) : *pa'wa* 'eldest uncle' (actually with a consonant-cluster), a contraction of *pa' ua* (which again is an abbreviation for *bapak tua*);
*pawah* I *pawahkan* 'to hire or lend on the metayer system, debtor and creditor sharing the proceeds';
II *rĕmpah-* -- 'all kinds of curry-stuff';
*pawan* I 'title or appellation for Malay-speaking Indians';
II '(k.o.) plant';
*pawaŋ* 'expert in any art believed to need the use of magic; guide; navigation officer, shipmaster; conductor', originally derived from *awaŋ, Minangkabau *puawaŋ, pawaŋ* 'id.'; cf. Timugon Murut *pa-awaŋ* 'chieftain' (D. J. Prentice, p. c.), and Kadazan *pu-avaŋ* 'rich, well-off' (Anthonissen 1958);
*pawas* 'k.o. freshwater fish', from Minangkabau;
*pawat, payon* -- '(k.o.) umbrella', a variant of *bawat* (see above);
*paway* 'insignia borne after a prince; insignia-bearers';
*pawon* 'kitchen', from Javanese;

(mVwV(C)) : *mawa* 'gibbon', a variant of *wa'wa'*;
*mawar* I 'rose-water'; Wilkinson puts a question-mark behind the (grammatically wrong) Arabic etymology *ma' ward*, but I do not see any reason to doubt an Arabic origin *mā' al-ward* (or Persian *mā'-i-ward*);
II *tawar* -- 'harmless, nullified', derived from *tawar*;
*mawas* 'orang hutan', with a variant (in Borneo) *mayas*;
*maway* '(k.o.) shrub';
*mawin*, in *kawin* -- 'marriage-festivities of all sorts', derived from *kawin*;
*méwah* 'plenteous';
*méwék* 'pursing up the mouth, pouting', from Jakartanese according to Iskandar, and ultimately from Javanese according to Kähler;
*méwér* 'to sob', from Javanese;

(jVy(C)) : *jaya* 'triumphant', from Sanskrit;
*jayeŋ* 'victorious over', from Javanese;

(cVyV(C)) : *caya* 'lustre, glow, brightness', short for *cahaya*, from Sanskrit;
*cayah* 'careless, neglectfull', a variant of *cuay*; from Minangkabau;
*cayar* 'diluted, watery (of viscous things)', a variant of *cair*, which developed from PMJ *caiR[13][10];
*cayu* 'sitting-mat', a variant of *siu*;

(ñVyV(C)) : *ñaya* 'injustice', a contraction of *aniaya*, from Sanskrit;
*ñeyag* 'to snap (as a dog)', from Jakartanese, ultimately from Sundanese.

Out of 40 lexemes belonging to pattern V, 13 are loanwords, four are derivations, (*bawaŋ, mawar* II, *mawin*, and *pawaŋ*), three developed from proto-

forms without semivowel (*bawa, bawah, cayar*), two are contracted forms (*bewak, pa'wa*), and two lexemes are variants of each other (*pawat* and *bawat* I). This leaves us with 17 lexemes which cannot be reduced to loanwords or to lexemes which underwent secondary developments.

## 3. COMBINATIONS OF *d, t* AND *n*

For the sake of completeness it would have been warranted to include all combinations of *d, t,* and *n*, but this requires too much space, and is not really necessary. (I count at least 177 lexemes with such combinations, dialectal variants and other variant forms not included.) Since it is relatively easy for the reader to check the data in Wilkinson, I will restrict myself to giving the numbers I found for each pattern. I will also give those lexemes which are thought to be inherited, i.e. lexemes for which a form in PMP or Proto-Malayo-Javanic has been reconstructed. In this way the reader can verify the qualitative difference between lexemes with combinations of *d, t,* and *n*, and lexemes with other combinations of homorganic consonants.

I found the following totals for each pattern:

| | | | |
|---|---|---|---|
| I : 39; | dVtV(C) : 18, | tVdV(C) : 21; |
| II : 52; | dVntV(C): 8, | tVndV(C): 44; |
| III : 61; | dVnV(C) : 40, | tVnV(C) : 21; |
| IV : 25; | nVdV(C) : 9, | nVtV(C) : 5, |
| | nVndV(C): 6, | nVntV(C): 5. |

The total number of lexemes with combinations of *d, t,* and *n*, is 177, which is more than the total sum (112, variant forms subtracted) of lexemes with other combinations of homorganic consonants differing in type of articulation. The numbers given above strongly suggest that there was no constraint on combinations of alveo-dentals in initial and medial position.

The following lexemes are inherited from PMP and/or Proto-Malayo-Javanic. Those reconstructions which are followed by the abbreviation (Bl.) are taken from Blust (1970); *tondoŋ* is taken from Nothofer; and all other reconstructions are taken from Dempwolff[10]:

I : *dataŋ* 'to come' *dateŋ 'id.';
*datar* 'level, flat' *DataR 'id.';
*datok* *(dato', datoŋ, datu)* 'chief, head of the family; grandfather, ancestor' *datu 'chief, head of the family';
*dětek* 'ticking sound' *detik 'throb' (Bl.);
*tědoh* 'abated, calm (wind)' *te(n)duq 'id.';
*tidur* 'to sleep' *tiDuR 'id.';
*tudiŋ* 'aslant, at an angle', and (Iskandar) 'to indicate, point with the finger' *tudiŋ 'to indicate';
*tudoh* 'to accuse' *tuDuq 'to indicate';

II : *tanda* 'sign' *ta(n)da 'id.';
*tandas* 'to state explicitly *(CtT)a(n)(dDj)es 'id.' (Bl.);
*tandiŋ* 'division into equal parts, compare' *(Ct)anDiŋ 'equal; compare' (Bl.);
*tandu* 'hammock-litter' *tanDu 'id.';
*tindas* 'to push' *tiDes 'id.' (Bl.);
*tindeh* 'to lie in heaps' *ti(n)diq 'id.';
*tunda* 'to drag' *tunDa 'id.';
*tundok* 'to bend down, to bow' *tu(n)Duk 'id.';
*tondoŋ* 'to chase away, oust' Proto-Malayo-Javanic *tunduŋ 'id.';

III : *danaw* 'lake, pool' *Danaw 'id.';
*tanah* 'land, earth' *taneq 'id.';
*tanak* 'to cook rice' *tanek 'to cook';
*tanam* 'to plant' *tanem 'id.';
*těnar* 'to be publicly known' *teneR 'voice' (Bl.);
*těnun* 'to weave' *tenun 'id.';
*těnoŋ* 'to gaze fixedly, diagnose illness, to divine' *(Ct)e(nN)uŋ 'to find by divination (Bl.);
*tuna* 'eel' *tuna 'id.';
*tunas* 'shoot, bud' *tunas 'id.';
*tunay* 'cash, ready money' *tu(nN)ay 'id.';
*tunu, tunun* 'to burn up' *tunu 'to burn, fry'.

IV : ----

## 4. DISCUSSION OF THE MATERIAL

Section II shows an unequal distribution of the material among the five patterns. It is striking that there are almost no lexemes belonging to the patterns I, II, or III, which cannot be identified as loanwords or derivations. The only exceptions are *bapa* (<*b-apa?), *běpaŋ* (<Chinese?), and *kugah* belonging to pattern I, none belonging to pattern II, and *bomo, pamah, giŋin, kaŋa,* and *kaŋar* belonging to pattern III. Out of these eight lexemes, only one (*bapa*) is part of everyday speech.

However, the number of such lexemes is much higher for the patterns IV and V (18 and 17 respectively, if one also subtracts *bawa, bawah,* and *cayar* from the former). Another striking feature about the apparently non-borrowed and non-derived lexemes of the five patterns is that they are generally of marginal use:

the botanical and zoological terms, for instance, number 19 out of 46, or 41,3%.

Finally for Section II, it should be noticed that among such lexemes those with initial labials outnumber by far those with non-labial consonants, and, more particularly, those with initial *m* are 22 out of 45, or 48,8%. The occurrence of a high number of initial *m*'s is the more significant if one adds to it the fact that in PMP there was an affix *um, which is still represented in Malay as a petrified prefix *m-* or infix *-ĕm-*, and also a prefix *ma, which still occurs as petrified *m(a)-* as well[11]. Although at present I will not draw any definite conclusion from this, it is possible that some of the lexemes with initial *m* can be explained as forms with a petrified prefix.

From Section III it appears that the total number of lexemes with combinations of *d*, *n* and *t*, is 177, which is more than the total sum lexemes with other combinations of homorganic consonants differing in type of articulation. Out of these, there are 28 lexemes displaying alveo-dental combinations and belonging to the patterns I, II, or III, for which a proto-form has been reconstructed. Of the lexemes displaying all other combinations of homorganic consonants differing in articulation-type (and belonging to one of the patterns I, II, or III), the only one with a corresponding proto-form is *bapa*. It appears that comparative evidence is in agreement with quantitative evidence in that there has not been a constraint on combinations of *d*, *n*, and *t*, ever since PMP.

On the other hand, the only combination of *d* and *n* with a corresponding proto-form is found in *danaw*<*Danaw[12], where as ten lexemes with a proto-form have a combination of *t* and *n*. In other words, comparative evidence seems to favor the assumption that in an earlier stage of Malay, there was a tendency to avoid combinations of initial *d with following *n. Although quantitative evidence from Wilkinson does not support this assumption, it is quite in agreement with the fact that *n*, an alveolar, is homorganic with *d* and not with supradental *t*.

The main picture acquired so far from the above material and discussion is that in early Malay there must have been a constraint on initial stops followed by homorganic nasals or by homorganic stops differing in voice, and that this constraint did not apply to *d and *n on the one hand, and *t on the other. The latter may be due to the different place of articulation of *t. Whether this constraint also applied to combinations of initial nasals followed by homorganic medial stops, or to initial stops or nasals followed by homorganic medial semivowels, is not clear because of the large number of unexplained counterexamples. This picture is in striking agreement with Chrétien's statistical study of combinations of PMP initial and medial consonants (Chrétien 1965:264).

Dempwolff (1934:64) reconstructed the following consonant system for PMP[13]:

| | bilabial | alveolar | retroflex | palatal | velar |
|---|---|---|---|---|---|
| Stops voiceless | *p | *t | *ṭ | *t' *k' | *k |
| voiced | *b | *d | *ḍ | *d' *g' | *g |
| Nasals | *m | *n | | *n' | *ŋ |
| Semivowels[14] | *v | | | *j | |
| Liquids[14] | | *l | *l̤ | | *ɣ |
| (Laryngeals: *c, *h) | | | | | |

Chrétien investigated the phonotactic structure of 1950 disyllabic proto-lexemes taken from Dempwolff (1938) which were not reduplicated monosyllables (p.244-5). For the co-occurrence of homorganic stops, nasals and semivowels in initial and medial position, he gives the following numbers (p.264):

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| I, II : | *bVpV(C) | : 1 | III : | *bVmV(C) | : 0 | IV : | *mVbV(C) | : 1 | V : | *bVvV(C) | : 2 |
| | *pVbV(C) | : 0 | | *pVmV(C) | : 0 | | *mVpV(C) | : 0 | | *pVvV(C) | : 0 |
| | *dVtV(C) | : 2 | | *dVnV(C) | : 1 | | *nVdV(C) | : 0 | | *mVvV(C) | : 1 |
| | *tVdV(C) | : 7 | | *tVnV(C) | : 7 | | *nVtV(C) | : 0[15] | | | |
| | *ḍVṭV(C) | : 0 | | | | | | | | | |
| | *ṭVḍV(C) | : 0 | | | | | | | | | |
| | *ḍ'Vṭ'V(C) | : 0 | | *d'Vn'V(C) | : 0 | | *n'Vd'V(C) | : 0 | | *d'VjV(C) | : 1 |
| | *t'Vd'V(C) | : 1 | | *t'Vn'V(C) | : 0 | | *n'Vt'V(C) | : 1 | | *t'VjV(C) | : 3 |
| | *g'Vk'V(C) | : - | | | | | | | | *n'VjV(C) | : 0[16] |
| | *k'Vg'V(C) | : 0 | | | | | | | | | |
| | *gVkV(C) | : 0 | | *gVnV(C) | : 0 | | *nVgV(C) | : 0 | | | |
| | *kVgV(C) | : 0 | | *kVnV(C) | : 0 | | *nVkV(C) | : 0 | | | |

Combinations of initial semivowels with following homorganic stops or nasals:

| | | | |
|---|---|---|---|
| *vVbV(C) | : 0 | *jVd'V(C) | : 0 |
| *vVpV(C) | : 0 | *jVt'V(C) | : 0 |
| *vVmV(C) | : 0 | *jVn'V(C) | : 0 |

Note that Chrétien ignored prenasalization (p.245), hence the combination of lexemes of patterns I and II; and that *g' is not reconstructed in initial position.

From these numbers it appears that homorganic voiceless stops, voiced stops, nasals, and semivowels almost never combine in the initial and medial positions, unless this combination consists of *t and *d or *n, or *t' and another palatal. This implies that the articulation-type harmony is even more far-reaching than in Malay, since in PMP it also applies to lexemes of pattern IV and V. Exceptions are: *bapa 'father', *bavaŋ 'union', *buvaŋ 'to throw away', *mabuk 'to be mentally ill', *mavas 'orang hutan', and *dunuŋ 'to be clever, smart'.

The PMP consonant system and lexicon has however been subject to much revision since Dempwolff published his *Vergleichende Lautlehre*. But this does not alter the validity of the articulation-type harmony which I propose, and in fact it even favors it:

1. nowadays *v and *j (or *w and *y in Dyen's orthography) are regarded as semivowels which never precede or follow a vowel of the same coloring (i.e. *u and *i respectively)(Dahl 1976:15-18). Consequently, we can eliminate one exception *buvaŋ, which is now reinterpreted as *buaŋ;
2. another exception, *mabuk, can also be eliminated because this is reinterpreted as a polymorphematic PAN *ma-buSek>PMP *ma-buhek (Blust 1976: note 18);
3. The palatals *d', *k', *n' and *j[17] (which became *z, *c, *ñ, and *y respectively in Dyen's orthography (Dyen 1947:fn.2) are now reinterpreted as members of one series (Dahl 1976: 86). This reinterpretation yields the following numbers of combinations of these segments:

| | | | |
|---|---|---|---|
| I, II: | *d'Vk'V(C) : 0 | III: | *d'Vn'V(C) : 0 |
| | *k'Vd'V(C) : 0 | | *k'Vn'V(C) : 0 |
| IV : | *n'Vd'V(C) : 0 | V : | *d'VjV(C) : 0 |
| | *n'Vk'V(C) : 0 | | *k'VjV(C) :(1) |
| | | | *n'VjV(C) : 0 |

Initial *j with following *d', *k', *n' : *jVd'V(C) : 0
*jVk'V(C) : 0
*jVn'V(C) : 0

The only combination of the above palatals is *k'ijum (to smell at; to kiss the Oriental way), but this is now reinterpreted as *k'ium (*cium) (see sub 1)).

The phonetic value of *g' and *t' is still unclear. Dyen (1947: fn.2, and 1949;fn.3) uses the symbols *s and *j respectively, but does so for practical reasons, without implying any particular phonetic characteristics to them. Dahl (1976:84) considers them as affricates. *t' and *g' co-occur in one lexeme: *t'ug'a 'sting, mantrap' (As noted above, *g' never occurs initially).

Dyen (1951) introduced another palatal, *z, but as far as I know it does not co-occur with other palatals.

Chrétien was not aware of an articulation-type harmony. He did not make a correlation between the number of co-occurrences of consonants of the same place of articulation and the number of co-occurrences of consonants of the same articulation-type. Nor did he question the phonetic likeness of *t with *d and *n. He observed that "there is avoidance in varying degree" between consonants which share their place of articulation, and that this avoidance is stronger if the consonants are in close proximity (i.e. the avoidance is stronger if the consonants occur in initial and medial position, or in medial and final position, but less strong if they occur in initial and final position) (p.263). He also found that there was a strong avoidance of identical consonants in a lexeme, but he did not make any other observations, cf. p.269:

> "With respect to sequences of non-identicals, whether of place or of type, no pattern is discernible to me. This is a way of saying that I can make no generalizations, as I did with identicals, I may, however, draw a conclusion: classification by place of articulation is of value since one generalization has been drawn from it; classification by type of articulation, however, does not seem very useful. This may seem disappointing but it should not be surprising: not every possible classification will serve us. Sometimes we may draw a negative conclusion."

## 5. CONCLUSION

So far, the conclusion I would like to draw is that in an earlier stage of Malay there appears to have been a tendency to consonant-constraint in the lexicon, such that in a disyllabic lexeme, an initial stop was not followed by a homorganic nasal, or by a homorganic stop, whether prenasalized or not, which did not agree in voice. The constraint did

not apply to combinations of initial *t* with medial *d* or *n* and vice-versa, which may be due to the fact that *t* is supradental, whereas *d* and *n* are alveolars. The only evidence against such a constraint are nine lexemes which cannot be explained with certainty as loanwords, as derivations, or as contracted forms (viz. *bapa*, *bĕpaŋ*, *kugah*, *bomo*, *pamah*, *giŋin*, *kaŋa*, *kaŋar*, and *danaw*). It is likely that this constraint of articulation-type harmony also applied to other combinations of homorganic consonants, but at present there are too many unexplained counter-examples. This constraint, which synchronically has ceased to be active in Malay, can probably be traced back to Proto-Malayo-Polynesian. As it appears from Chrétien's statistical study of the PMP morph, homorganic voiced stops, voiceless stops, nasals and semivowels, almost never occur together in initial and medial position, with the exception of *t with *d and *n, and of *t' (Dyen's *s) with other palatals.

NOTES

1. The italicizing of phonemes and lexemes is mine.
2. By 'root-syllable' is meant the last syllable of a lexeme. An 'unchanged Malay stem' is a Malay lexeme (or 'root-morpheme').
3. The hypothetical language ancestral to all Austronesian languages is called 'Proto-Austronesian' (PAN). The first languages to branch off from PAN were supposedly the Formosan ones, and the term 'Malayo-Polynesian' is now used by Blust (1982:233) for the remaining Austronesian languages (it corresponds to Dahl's Western and Eastern Austronesian, see Dahl p.128). The Malayo-Polynesian languages are derived from a hypothetical 'Proto-Malayo-Polynesian'. The major works referred to in this article (Dempwolff, Chrétien, Blust 1970) are based on non-Formosan Austronesian languages, hence my use of the term Proto-Malayo-Polynesian (PMP).
4. Other consonants are left out, because no phonotactic rule seems to predict their occurrence in inherited lexemes. Also, the exact phonetic realization of some of them (*h*, *r*, *s*) is difficult to establish, since it differs from dialect to dialect.
5. See Emeis p. 191-2. From a rough estimation of Wilkinson's dictionary, it appears that 67 % (or 2/3) of the lexicon is disyllabic.
6. This is the most comprehensive dictionary of Malay before its standardization as Bahasa Indonesia and Bahasa Malaysia. Fokker's observations are based on field-notes which he collected during his stay in West-Borneo.
7. Except in *yaitu* 'that is, namely', *yaŋ* I (relative pronoun)', *yaŋ* II 'divinity', and *yu* 'shark'. In all these cases the initial semivowels originated through the loss of syllabicity of an earlier *i*: *yaitu* and *yaŋ* I are analyzable as *ia+itu* and *ia+ŋ* respectively (Prentice, unpublished material, and others, cf. Wilk. for *yaitu*), whereas *yaŋ* II and *yu* are variant forms of *hiaŋ* and *hiu* respectively.
8. Diphthongization of ultimate syllables containing *a* is a common phenomenon in Iban, cf. *jalay* 'road, way', *palay* 'return; go, come home', *duay* 'two', etc. (See Scott 1956).
9. Cognates of Malay *bapa* and/or Minangkabau *apa'*/Iban *apay* in other Austronesian language:

   Toba-Batak *apa* '(one's own) father', as against *amá/ŋ* 'father', a term which 'can be used towards young people, and hence by a father to his son' (van der Tuuk);
   Ngaju-Dayak *apa* '(one's own) father', as against *bapa* '(someone else's) father' (Hardeland);
   Soboyo *n-apa* 'grandmother' (Fortgens);
   Kapuas *apa/ŋ* 'father' (Hudson);
   Hanunŏo *bápa/q* 'uncle' (Conklin);
   Maranao *bapa/q* 'uncle, father-in-law' (McKaughan & Macaraya);
   Solor *bapa* 'father' (Arndt)
   Mualang-Dayak (a dialect very close to Iban) *apay* 'father';
   *bapak* is a loan from Malay (Dunselman 1950:3).

   Notice the agreement between Iban *apay*, Toba-Batak and Ngaju-Dayak *apa* in that they directly refer to one's own father, in opposition to Iban *bapaq* ('father-in-law'), Toba-Batak *amá/ŋ*, and Ngaju-Dayak *bapa*.
10. Reconstructions are taken from Dempwolff (1938), unless indicated otherwise. They are represented in Dyen's orthography. (Dyen 1947:fn.2, and 1949:fn.3). PMJ *caiR is taken from Nothofer (1975).
11. See Brandstetter p.169-71; cf. also Malay *m/inum* 'to drink'<PMP *um-*inum, Banjarese Malay *k/inum*, Sundanese *inum* 'id.'; *um is still productive in Hanunóo: *?inúm* 'drink; drinking', *?um-inúm* 'drink' (Conklin 1953), and Iloko: (*inúm*) *um-inúm* 'to drink', *inum-én* 'to drink; beverage, drink'.
    For PMP *ma- which became petrified in Malay, cf. also *m/erah* 'red'<*ma-iRaq; *ma/bok*<PMP *ma-*buhek<PAN *ma-*buSek (Blust 1976:note 18).
12. PMP *D (a retroflex) and *n (an alveolar) are heterorganic, hence their co-occurrence in *Danaw is regular, in contrast to the co-occurrence of *d and *n in *dunun. It is however striking that there is only one case in which *D and *n are found together in initial and medial position.
13. Here I use Dempwolff's orthography, as does Chrétien. This orthography also reflects the mutual relationships between consonants as implied by Dempwolff.
14. I split Dempwolff's fricatives (*Reibelaute*) into semivowels and liquids, as does Chrétien (p.245). This is also in agreement with the interpretation of *v and *j by later Austronesianists (see Dahl 1976:15-18).
15. Chretien's data (p.264) indicate that there is one lexeme of this pattern, but I was not able to find it in Dempwolff (1938).
16. There is no lexeme of this pattern in Dempwolff (1938): Chrétien's numbers for lexemes of the patterns *n'VjV(C) and *n'Vn'V(C) must have been interchanged, because he gives zero for the latter, whereas *n'an'i occurs.
17. Dahl (p.86) categorizes *d', *k' and *n', as members of one series of palatals. He does not mention *j but I include it here.

BIBLIOGRAPHY


Alieva, N.F., Arakin, V.D., Ogloblin, A.K. & Sirk, Ju. X. 1975. *Grammatika Indonezijskogo Jazyka.* Moscow: Akademija Nauk SSSR.

Anthonissen, A. 1958. *Kadazan-English and English-Kadazan Dictionary.* Canberra: Government Printing Office.

Arndt, P.P. 1937. *Grammatik der Solor-Sprache.* Ende, Flores: Arnoldus.

Blust, Robert A. 1970. "Proto-Austronesian Addenda." *Oceanic Linguistics* Vol. IX no.2, 104-162.

_____. 1976. "Dempwolff's Reduplicated Monosyllables". *Oceanic Linguistics* Vol. XV nos. 1 and 2, 107-130.

_____. 1982. "The Linguistic Value of the Wallace Line". *Bijdragen tot de Taal-, Land- en Volkenkunde.* 138, 231-250.

Brandstetter, Renward. 1916. *An Introduction to Indonesian Linguistics.* (Translated by C.O. Blagden). London: the Royal Society, Monograph Vol. XV.

Carro, Andres. 1956. *Iloko-English Dictionary.* (Translated, revised and augmented by Morice Vanoverbergh), Manilla.

Chrétien, C. Douglas. 1965. "The Statistical Structure of the Proto-Austronesian Morph". *Lingua* 14, 243-70.

Conklin, Harold C. 1953. *Hanunóo-English Dictionary.* Berkeley and Los Angeles: Univ. of California Press.

Coolsma, S. 1913. *Soendaneesch-Hollandsch Woordenboek.* (2nd ed.) Leiden: Sijthoff.

Dahl, O.C. 1976. *Proto-Austronesian.* (2nd revised ed.) Scandinavian Institute of Asian Studies Monograph Series No.15, Lund.

Dempwolff, O. 1934. *Vergleichende Lautlehre des austronesischen Wortschatzes. I Band: Induktiver Aufbau einer indonesischen Ursprache.* Fünfzehntes Beiheft zur Zeitschrift für Eingeborenen Sprachen, Berlin.

_____. 1938. *Vergleichende Lautlehre des Austronesischen Wortschatzes. III Band: Austronesisches Wörterverzeichnis.* 19. Beiheft zur ZES, Berlin.

Dunselman, Donatus. 1949. "Bijdrage tot de Kennis van de Taal en Adat der Kendajan Dajaks van West-Borneo". *Bijdragen tot de Taal-, Land- en Volkenkunde.*

_____. 1950. "Over de Huwelijksadat der Moealang-Dajaks van West-Borneo. *Bijdragen tot de Taal-, Land- en Volkenkunde* 106, 3.

Dyen, I. 1947. "The Malayo-Polynesian Word for 'two'". *Language* 23, 50-55.

_____. 1949. "On the History of the Trukese Vowels". *Language* 25, 420-36.

_____. 1951. "Proto-Malayo-Polynesian *Z". *Language* 27, 534-70.

_____. 1953. *The Proto-Malayo-Polynesian Laryngeals.* Baltimore: Linguistic Society of America.

_____. 1965. "Formosan Evidence for Some New Proto-Austronesian Phonemes". *Lingua* 14, 285-305.

Emeis, M. G. 1955. "Vocaal-harmonie in de Maleise Kern van de Indonesische Vocabulaire". *Bijdragen tot de Taal-, Land- en Volkenkunde* 111, 191-201.

Fokker, A.A. 1895. *Malay Phonetics.* Leiden, Brill.

Fortgens, J. 1921. *Bijdragen tot de Kennis van het Sobojo.* The Hague, Royal Institute of Linguistics and Anthropology.

Hapip, Abdul Jebar. 1977. *Kamus Bandjar-Indonesia.* Jakarta: Pusat Pembinaan dan Pengembangan Bahasa (Dep.P & K.).

Hardeland, Aug. 1859. *Dajaksch-Deutsches Wörterbuch.* Amsterdam: Frederik Muller.

Henderson, Eugénie J.A. 1965. "The Topography of Certain Phonetic and Morphological Characteristics of South East Asian Languages". *Lingua* 15, 400-38.

Hervey, D.F.A. 1881. "Itinerary from Singapore to Source of the Sĕmbrong and up the Mâdek". *Journal of the Straits Branch of the Royal Asiatic Society* VIII, 126-132.

Hudson, Alfred B. 1967. *The Barito Isolects of Borneo.* Ithaca, New York.

Iskandar, T. 1970. *Kamus Dewan.* Kuala Lumpur: Dewan Bahasa dan Pustaka.

Känler, H. 1966. *Wörterverzeichnis des Omong Djakarta.* Berlin: Dietrich Reiher.

Klinkert, H.C. 1916. *Nieuw Maleisch-Nederlandsch Woordenboek.* (3d. revised and enlarged ed.). Leiden: E.J. Brill.

Leo, Philip. 1975. *Chinese Loanwords Spoken by the Inhabitants of the City of Jakarta.* Jakarta: Lembaga Research Kebudayaan Nasional (L.I. P.I.) Seri data dasar no.7.

McKaughan, P. & Macaraya, A. Batua. 1967. *A Macarao Dictionary.* Honolulu: University of Hawaii Press.

Nothofer, Berndt. 1975. *The Reconstruction of Proto-Malayo-Javanic.* Verhandelingen van het Koninklijk Instituut voor Taal-, Land- en Volkenkunde 73, The Haque: Nijhoff.

Pamoentjak, M. Thaib. 1935. *Kamoes Bahasa Minangkabau - Bahasa Melajoe Riau.* Batavia: Balai Poestaka.

Scott, N.C. 1956. *A Dictionary of Sea Dayak.* London: School of Oriental and African Studies.

Toorn, J.L. van der. 1891. *Minangkabausch-Maleisch-Nederlandsch Woordenboek.* The Hague: Royal Institute of Linguistics and Anthropology.

Tuuk, H.N. van der. 1861. *Bataksch - Nederduitsch Woordenboek.* Amsterdam: Frederik Muller.

Wilkinson, R.J. 1932. *A Malay-English Dictionary* (romanised). Mytilene: Greece.

* * *