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 acknowleding that a, 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 phonomenon Emeis calls vowel-harmony could be more appropriately described as 'height-harmony', since it asserts that the mid-vowels o and e never co-occur with high vowels (u and /), 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 as an initial of one of the syllables is never met along with g as an initial of the other.

p.40: a and i 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 are: bapa 'father' and bopoq '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.

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: R and / or R and j never occur together in an unchanged2 Malay stem.

p.63: p as initial of a stem is never followed by m as single initial of the root-syllable.

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 a and t, which are combined in many inherited lexemes like tidur 'to sleep', tudoh 'to accuse', dataq 'to come', datok 'chief, head; ancestor, grandfather', etc.; 2. there are, together with bapa and bopoq, 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

57
articulation-type harmony on the synchronous 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-Malay-Polynesian (henceforward PMP) morphs (Chretien 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. For practical reasons I will confine myself to an inquiry into disyllabic lexemes. In the Malay lexical structure disyllabicity is the norm. 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), and will modify its spelling in the following way:

Wilkinson: modified spelling

\[
\begin{array}{c|c|c|c|c|c|c|c|c}
\text{Stop} & \text{voiceless} & \text{voice} & \text{nasal} & \text{semivowel} & \text{Others} \\
\hline
\text{p} & \text{t} & \text{c} & \text{k} \\
\text{b} & \text{d} & \text{j} & \text{g} \\
\text{m} & \text{n} & \text{ŋ} & \text{ŋ} \\
\text{w} & \text{y} \\
\text{h, s, l, r} \\
\end{array}
\]

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

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

II. \(S^+_h V N^+_h S_h V (C)\), = 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. \(N^+_h V S^+_h V (C)\), = nasal followed by simple or prenasalized homorganic stop

V. \(S^+_h V w/y V (C)\), = 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:

-\( b^- = 1200; d^- = 850; j^- = 600; g^- = 880; \)
-\( p^- = 1000; t^- = 1200; s^- = 900; k^- = 1600; \)
-\( m^- = 900; n^- = 240; h^- = 60; q^- = 75; \)

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

-\( bVbV(C) = 59; jVjV(C) = 35; gVgV(C) = 51; \)
-\( bVmV(C) = 21; jVnjV(C) = 12; gVgV(C) = 19; \)
-\( pVpV(C) = 37; cVcV(C) = 56; kVkV(C) = 63; \)
-\( pVmV(C) = 15; cVncV(C) = 12; KVkV(C) = 18; \)
-\( mVmV(C) = 40; hVhV(C) = 16; gVgV(C) = 16. \)

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. \( S^h_s V S^h_s V (C), S^h_s V S^h_s V (C) \)

-\( bVpV(C) = \) bapa, bapaq, bapak 'father';
-\( b\ddot{p}aq, b\ddot{p}aq, b\ddot{p}aq, in kueh --
  (k.o.) sweetmeat'; Wilkinson guesses that this is a Chinese loanword;
-\( bo\ddot{p}eg 'pock-marked', from Chinese (Leo 1975:8); \)
-\( pVbV(C) = \) pabu, in main -- 'somersaults and other tricks by Chinese tumblers', a Chinese loan;
-\( \ddot{p}b\ddot{b}in (k.o) teetotum, from Chinese;
  pobien (disyllabic?) 'wharf', from Chinese;
-\( jVcV(C) = \) jicuy 'opium drop doctor',
  a second time for consumption', from Chinese;
-\( gVvV(C) = \) 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 bapaq (and their variants) and kugah.

Bapa (bapaq, 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 \( b^-\).

B\( b^-\)apa (b\( b^-\)apa) 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. \( S^h_s V N^h_s S^h_s V (C), S^h_s V N^h_s S^h_s V (C) \)

-\( bVmpV(C) = \) bimapau, bimpo 'handkerchief, towel', a Chinese loan;
-\( pVmbV(C) = \) pombak 'dove, pigeon', a Portuguese loan;
-\( jVncV(C) = \) jin\( a\)eq 'grateful', a Chinese loan;
-\( cVnV(C) = \) ---
-\( gVgV(C) = \)
-\( kVgV(C) = \)

The only lexemes belonging to this pattern are loanwords.

III. \( S^h_s V N^h_s V (C) \)

-\( bVmV(C) = \) 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, bomok, bomor 'mage, sorcerer';
-\( bumi 'earth', from Sanskrit;
-\( pVmV(C) = \) pama 'police-informer', probably from English 'informer';
-\( p\ddot{a}ma 'with a medial cluster'
  'the plant Rafflesia hasseltii', a Javanese name, the true Malay name being kerubut;
-\( pamak 'low-lying (land)';
-\( pamak 'uncle', from Javanese;
-\( pamur 'talking big, boasting', from Javanese;
-\( pamur 'to beg leave or depart', from Javanese;
-\( pamor, pamur 'alloyed (iron)', from Javanese;
-\( pomak '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);
-\( jVV(C) = \) ja\( \ddot{a}\)n 'to say, think', from Minangkabau ja\( \ddot{a}\)n (ja\( \ddot{a}\)n),

59
Of these 20 lexemes, 12 are given as loanwords by Wilkinson. Of the remaining lexemes, beman, pama, and pomoq must be considered as loanwords as well; five lexemes remain: bomo (bomoh, bomor), pamah, gisín, kaga, and kagar.

IV. N₇ V₇ S₅⁺ V₇(C), H₇ V₇ N₅⁺ S₅⁺ V(C)

mVb(C) : mabaw 'evil spirit of disease';
      mabag '(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, sendes-mendes 'tilted this way and that', from sendes;
      mabok 'intoxicated';
      mabub '(proper name)', from Arabic (magbāb 'darling');
      mabur 'to fly', from Javanese;
      mobil 'car', from Dutch;
      mubal 'to shoot up', from Javanese;
      mubeq 'to circle, revolve', from Javanese;
      mubyar (written mubiar in Wilkinson) 'to strike the eye, to "scream" (of colour)', from Javanese;

mWmb(C) : mambaq I 'spirit of the Indonesian pantheon';
      II ikan -- 'a snapper', derived from bambaq;
      III title given occasionally to negritoheadmen;

mambu I = (semambu) 'Malacca cane';
      II = mabaw ('evil spirit of disease');
      III 'to smell', from Javanese;

mambul 'k.o. climber';

minbar 'pulpit in a mosque', from Arabic;

mapaq, mapak 'to go out and meet and then escort to one's house', from Javanese, ultimately from Javanese or Sundanese;

mapar, gaq 'flat-ended brazer's chisel or punch', derived from papar;

mapat '(k.o.) tree';

mapok '(k.o.) herb';

mepas = mempas 'to fish with the fly';

mepir, mipir 'to edge away under a blow', derived from peper;

mepis 'thin, tenuous', a variant of nīpis and tīpis;

mepoq 'pock-marked' cf. (bo- peq), from Chinese (Leo 1975:8);

mupar, in ular tādq -- 'black cobra', derived from upar;

mupoh, mupaq '(k.o.) tree';

mupir 'to stop; to touch at', from Javanese, the usual Malay term being saqah;

mampoh, mampu 'having the means for', from Javanese, ultimately from Sundanese;

mampuq 'light and spongy in texture', which must be derived from (h)ampuq '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';

mespat '(k.o.) tree';

mespir 'to bear some resemblance to', from Javanese;

mespoq '(k.o.) tree';

mēppi 'dream', derived from āmpi (with a petrified m-);

mepoq 'used up (nīpah, pandan, etc.), of which the fronds have been taken';

mombag 'coconut in its earliest stage of growth';

mumpun 'blunt, stumpy', from Minangkabau;

mWjV/C) : ----

(mWjC) : ----

(mWcv/C) : ----

(mWvcC) : ----

(qVQv/C) : ----

(qVqV/C) : ----

(qVqVC) : ----

(qVqkC) : ----

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.10), and mpis 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 (t) V w/y V (C), N (t) V w/y V (C) (*awaq, Minangkabau pawaq, pawaq 'id.'); cf. Timugon Murut pa-awaq 'chieftain' (D. J. Prentice, p. c.), and Kadazan pu-awaq 'rich, well-off' (Anthonissen 1938);
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 beside the (grammatically wrong) Arabic etymology ma' ward, but I do not see any reason to doubt an Arabic origin ma' al-ward (or Persian māl-i-ward);
III 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;
mewah 'plenteous';
mewek 'pursing up the mouth, pouthing', from Jakartaese according to Iskandar, and ultimately from Javanese according to Kähler;
méwér 'to sob', from Javanese;
(jVvV(C)) : jaya 'triumpant', from Sanskrit;
jayeq 'victorious over', from Javanese;
(cVvV(C)) : aaya 'lustre, glow, brightness', short for oahaya, from Sanskrit;
aayah 'careless, neglectful', a variant of oay; from Minangkabau;
amaya 'diluted, watery (of viscous things)', a variant of cair, which developed from PMJ *cair;
aayu 'sitting-mat', a variant of sū;
(RVvV(C)) : ņaya 'injustice', a contraction of anlaya, from Sanskrit;
ñeyag 'to snap (as a dog)', from Jakartaese, ultimately from Sundanese.

Out of 40 lexemes belonging to pattern V, 13 are loanwords, four are derivations, (bawaq, mawar II, mawin, and pawag), three developed from proto-
forms without semivowel (bawa, bavah, 
cayar), two are contracted forms (bewak, 
pawala), 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 combina-
tions of homorganic consonants.

I found the following totals for 
each pattern:

I : 39; dVt(V)(C) : 18, tVdV(C) : 21;
II : 52; dVnt(V)(C) : 8, tVndV(C) : 44;
III : 61; dVn(V)(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 conso-
nants differing in type of articulation. 
The numbers given above strongly suggest 
that there was no constraint on combina-
tions 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); toonoq is taken from 
Nithofr; and all other reconstructions 
are taken from Dempwolf.

I : datag 'to come', datag 'id.'; 
datar 'level, flat' *Datarr 'id.'; 
datok (datok, datago, datu) 
'chief, head of the family; 
grandfather, ancestor' 
*datu 'chief, head of the family'; 
dtek 'ticking sound' *dtek 
'throb' (Bl.); 
tdoh 'abated, calm (wind)' 
*te(d)uoq 'id.'; 
tdur 'to sleep' *tDUR 'id.';
tdiq 'aslan, at an angle'; 
and (Iskandar) 'to indicate, point with the 
finger' *tdiq 'to indicate'; 
tudoq 'to accuse' *tuDuq 'to indicate'; 
II : tanda 'sign' *ta(n)da 'id.'; 
tandas 'to stats explicitly' 
*(Ct)ta(n) (d)aj es 'id.' 
(Bl.); 
tandoq 'division into equal 
parts, compare' *(Ct)Daq 
'equal; compare' (Bl.); 
tandu 'hammock-litter' *tanDu 
'id.'; 
tindas 'to push' *tiDes 'id.' 
(Bl.); 
tindeh 'to lie in heaps' *
*tinDiq 'id.'; 
tunsa 'to drag' *tunDa 'id.'; 
tundok 'to bend down, to bow' 
*tu(n)Duk 'id.'; 
tondoq 'to chase away, cut' 
Proto-Malayo-Javanic 
*toonoq 'id.';
III : danaw 'lake, pool' *Danaw 'id.'; 
takah 'land, earth' *taneq 
'id.;
tanak 'to cook rice' *tanek 'to 
cook'; 
tanar 'to plant' *tanem 'id.;
tener 'be publicly known' 
*teneR 'voice' (Bl.); 
tenun 'to weave' *tenun 'id.';
tenog 'to gaze fixedly, diagnose 
ilness, to divine' 
*(Ct)enog 'to find by 
divination' (Bl.); 
tuna 'eel' *tuna 'id.; 
tunas 'shoot, bud' *tunas 'id.';
tunay 'cash, ready money' 
*tu(n)ay 'id.';
tunu, tunun 'to burn up' *tunu 
'to burn, fry';
IV : ----

4. DISCUSSION OF THE MATERIAL

Section II shows an unequal distribu-
tion of the material among the five 
patterns. It is striking that there are 
almost no lexemes belonging to the pat-
terns I, II, or III, which cannot be 
identified as loanwords or derivations. 
The only exceptions are bapua (*b-apa?), 
bupaq (<Chinese?), and kugah belonging 
to pattern I, none belonging to pattern 
II, and boma, pamah, gìn, kaga, and 
kaqar belonging to pattern III. Out of 
these eight lexemes, only one (bapua) 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, bavah, 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 occur-
rence 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 petrifed prefix *m- or infix *em-,
and also a prefix *ma, which still occurs
as petrifed 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 petrifed
prefix.

From Section III it appears that
the total number of lexemes with combi-
ations 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 combi-
ations of homorganic consonants differ-
ing 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 combina-
tion of d and n with a corresponding
proto-form is found in danaw*Danaw12,
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 supra-
dental 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 con-
straint 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 combina-
tions of initial nasals followed by homorganic
medial stops, or to initial stops or na-
sals followed by homorganic medial semi-
vowels, 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 PMP13:

<table>
<thead>
<tr>
<th>bilabial</th>
<th>alveolar</th>
<th>retroflex</th>
<th>palatal</th>
<th>velar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stops voiceless</td>
<td>*p</td>
<td>*t</td>
<td>*t'</td>
<td>*k</td>
</tr>
<tr>
<td>voiced</td>
<td>*b</td>
<td>*d</td>
<td>*d'</td>
<td>*g'</td>
</tr>
<tr>
<td>Nasals</td>
<td>*m</td>
<td>*n</td>
<td>*n'</td>
<td>*q</td>
</tr>
<tr>
<td>Semivowels</td>
<td>*v</td>
<td>*j</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liquids</td>
<td>*l</td>
<td>*l</td>
<td>*j'</td>
<td>*γ</td>
</tr>
</tbody>
</table>

(Charyngeals: *c, *h)

Chrétien investigated the phonotactic
structure of 1950 disyllabic proto-lexemes
taken from Dempwolff (1938) which were
not reduplicated monosyllables (p.244-5).

<table>
<thead>
<tr>
<th>I, II</th>
<th>*bVpV(C)</th>
<th>1</th>
<th>III</th>
<th>*bVmV(C)</th>
<th>0</th>
<th>IV</th>
<th>*mVbV(C)</th>
<th>1</th>
<th>V</th>
<th>*lVwV(C)</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>*pVbV(C)</td>
<td>0</td>
<td>*pVmV(C)</td>
<td>0</td>
<td>*mVpV(C)</td>
<td>0</td>
<td>*pVwV(C)</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*dVvV(C)</td>
<td>2</td>
<td>*dVnV(C)</td>
<td>1</td>
<td>*nVdV(C)</td>
<td>0</td>
<td>*mVwV(C)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*tVdV(C)</td>
<td>7</td>
<td>*tVnV(C)</td>
<td>7</td>
<td>*nVtV(C)</td>
<td>015</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*dVtV(C)</td>
<td>0</td>
<td>*tVdV(C)</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*dVvV(C)</td>
<td>0</td>
<td>*d'VnV(C)</td>
<td>0</td>
<td>*n'VdV(C)</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*t'VdV(C)</td>
<td>1</td>
<td>*t'VnV(C)</td>
<td>0</td>
<td>*n'VtV(C)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*g'VkV(C)</td>
<td>-</td>
<td>*k'VgV(C)</td>
<td>0</td>
<td>*k'VjV(C)</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*gVkV(C)</td>
<td>0</td>
<td>*gVnV(C)</td>
<td>0</td>
<td>*nVgV(C)</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>*kVgV(C)</td>
<td>0</td>
<td>*kVnV(C)</td>
<td>0</td>
<td>*nVkV(C)</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For the co-occurrence of homorganic stops,
nasals and semivowels in initial and
medial position, he gives the following
numbers (p.264):

| I, II | *n'VjV(C) | 016 |

63
Combinations of initial semivowels with following homorganic stops or nasals:

\[ *vVbV(C) : 0 \quad *jVd'V(C) : 0 \]

\[ *vVpV(C) : 0 \quad *jVt'V(C) : 0 \]

\[ *vVmV(C) : 0 \quad *jVn'V(C) : 0 \]

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, and 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', *bavaq 'union', *bavaq 'to throw away', *mabuk 'to be mentally ill', *mavas 'orang hutan', and *dun '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 *bavaq, which is now reconstructed as *buaq;
2. another exception, *mabuk, can also be eliminated because this is reinterpreted as a polymorphic PAN *ma-bu Sekh>PM *ma-buhek (Blust 1976: note 18);
3. The palatals *d', *k', *n' and *j' (which became *z, *c, *h, 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:

\[ I, II:*d'Vv'V(C) : 0 \quad III:*d'Vn'V(C) : 0 \]
\[ *k'Vd'V(C) : 0 \quad *k'Vn'V(C) : 0 \]
\[ IV : *n'Vd'V(C) : 0 \quad V : *d'VjV(C) : 0 \]
\[ *n'Vv'V(C) : 0 \quad *k'VjV(C) : 0 \]
\[ *n'VjV(C) : 0 \]

Initial *j with following *d', *k', *n':

\[ *jVd'V(C) : 0 \quad *jVt'V(C) : 0 \]
\[ *jVkV(C) : 0 \quad *jVn'V(C) : 0 \]

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

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, *s, 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, *d̪əpəq, *kugah, *boma, *pamah, *gəqən, *kaqəd, *kagar, and *danəq). 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 Chretien'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 this hypothetical 'Proto-Malayo-Polynesian'. The major works referred to in this article (Dempwolff, Chretien, 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 Emenem 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, Pokker's observations are based on field-notes which he collected during his stay in West-Borneo.
7. Except in *ga1u 'that is, namely', *gəq I (relative pronoun), *gɔ1 'divinity', and *gər 'shark'. In all these cases the initial semivowels originated through the loss of syllabicity of an earlier i: *gaiu and *gəq I are analyzable as ia1+iu and ia1+ə respectively (Prentice, unpublished material, and others, cf. Wink. for *ga1u), whereas *gəq II and *gəq are variant forms of *həq and *həq respectively.
8. Diphthongization of ultimate syllables containing a is a common phenomenon in Iban, cf. *jalay 'road, way', *palaq 'return; go', *come home', *mu1u 'two', etc. (See Scott 1956).
9. Cognates of Malay *apa and/or Minangkabau *apa'/Iban *apay in other Austronesian languages:

Toba-Batak *apa 'one's own' father', as against amã/q '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 apa 'someone else's' father' (Hardland);
Soboyo n-apa 'grandmother' (Fortgens);
Kapuas apa/q 'father' (Hudson);
Hanunoo *bapa/q 'uncle' (Conklin);
Maranoa *bapa/q 'uncle, father-in-law' (McKaughan & Macaraya);
Solor *bapa 'father' (Ardnt);
Mualang-Dayak (a dialect very close to Iban) *apay 'father'; *bapak is a loan from Malay (Dunsheim 1950:13).

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 *amã/q, and Ngaju-Dayak *bapaq.
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). PMW *ca1R is taken from Nothofer (1975).
11. See Brandstetter p.169-71; cf. also Malay m/inum 'to drink'<PMP *um-inun, Banjarase Malay k/inum, Sundanese inum 'id.', *um is still productive in Hanunoo: inum 'drink; drinking', inum-inum 'drink' (Conklin 1953), and Tolo: (inum) um-inum 'to drink', inum-en 'to drink; beverage, drink'. For PMP *ma- which became petrified in Malay, cf. also meraq 'red' < *ma-iraq; ma-bok<PMP *ma-buhek<PAN *ma-bu Sek (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 *dumun. 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 Chretien. 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 Chretien (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): Chretien's numbers for lexemes of the patterns *nV*jV(C) and *nV*nV(C) must have been interchanged, because he gives zero for the latter, whereas *nV*jV occurs.
17. Dahl (p.86) categorizes *d', *k' and *n', as members of one series of palatals. He does not mention *) but I include it here.


* * *