TABA AND ROMA: CLUSTERS AND GEMINATES IN TWO AUSTRONESIAN LANGUAGES

John Hajek‡ and John Bowden*
‡*School of Languages, University of Melbourne*
**Research School of Pacific and Asian Studies. Australian National University*

ABSTRACT

Some lesser-known Austronesian languages, with relatively simple phonological inventories, nevertheless have very complex phonotactic structures. The presence of a wide range of typologically unusual cluster combinations in word-initial position has important implications for generally accepted notions about Austronesian languages and about segment sequencing and the sonority hierarchy.

1. INTRODUCTION

It is often considered that Austronesian language have straightforward sound systems and structures with relatively little to say for phonological theory, especially in the area of phonotactics. Such an observation may be justified for languages such as Hawaiian, Tahitian and Indonesian/Malay, but closer inspection of Austronesian shows great variation across this very large language family. It is not widely known that typologically unusual word-initial geminates are found in different languages scattered across the Austronesian area, including many Peninsular Malay dialects, such as Pattani, and numerous Micronesian languages such as Trukese. Initial geminates are also frequently found in the little-known Austronesian languages of the Maluku area of Central-Eastern Indonesia, e.g. Taba /'llo/ 'blood' v. /'lo/ 'inside'. In these same Maluku languages such geminates form only a small part of much larger and more complex systems of unusual clusters in word-initial position.

2. MALUKU LANGUAGES, BASIC PHONOLOGY AND INITIAL GEMINATES

Whilst the Maluku languages referred to in this paper are all Austronesian, it needs to be noted that they are not for our purposes necessarily all very closely related, nor geographically in any direct contact. Taba is a member of the major East Malayo-Polynesian branch, and is spoken in the north Maluku area [2]. Roma, and other languages, including Leti also cited her, belong to the Central Malayo-Polynesian branch, and spoken mostly in the southwest Maluku area [5,6]. All the Maluku languages in question have relatively simple and small phonological inventories. If one excludes clearly borrowed phonemes, Taba appears to have only 15 phonemes, ie /p t k b d g s h m n ŋ l r w j/ and Roma only 12, /p t k d s h m n l r w j/ [2,5]. By way of comparison, Leti has even fewer, with 10 phonemes /p t k d s β m n l r/ [6]. However, as the tables below demonstrate, all three languages permit a wide range of homorganic and heterorganic clusters at phonemic level. Statistically speaking, Taba allows the greatest set of combinations, Roma the least, with Leti occupying a middle position. Such a ranking is consistent with a number of other interrelated observations, all of which suggest that overall Roma conforms the most with accepted notions of phonotactic patterning & Taba the least. Taba allows 11 different geminates in word-initial position: /bb dd gg tt kk mm nn ŋŋ ll hh ww/, and Roma only 3: /tt nn rr/. Initial geminates are relatively rare in Roma, far outnumbered by their more frequent use and wider range in word-medial position. Moreover, whilst /nn/ surfaces as [n:] medially, in initial position the first /n/ is always syllabic, with the second /n/ forming the onset of the following syllable, e.g. /nnehu/ [n̩'nehu] 'he jumped', cf. /nunnun hai/ ['nun'nun'ha^{i}] 'what's he doing?'. This particular surface outcome is not unusual to Roma, being widespread amongst many other languages spoken in the area. In Taba, on the other hand, geminates are very frequent word-initially. Most unusual amongst the set of initial geminates in Taba are /hh/ and /ww/, e.g./wwe/ 'leg' & /hhan/ 'you (pl.) go'. It is not known to these authors whether they are reported in other languages of the world.

Much work remains to be done on establishing the historical origins of initial geminates, especially in Taba. It is true that in these languages, many examples of initial gemination result directly from a range of morphophonemic processes associated with prefixing, especially of proclitics, e.g. /l+dod/ -> ['ldod] 'they ask'. However, other items are monomorphemic and are more difficult to explain, although reduction of earlier productive reduplication appears to be important, e.g. Taba ['ddoba] 'earth' < [dad'doba] < [dab'doba] < /'doba/ 'garden'.

3. NON-GEMINATE CLUSTERS, SEQUENCING AND THE SONORITY HIERARCHY

Fundamental to phonotactics are basic notions of permissible segment sequencing and the so-called sonority hierarchy which has been frequently observed as governing the permitted organization of consonants clusters, and of syllables as a whole [1,3,7]. Whilst there may be some dispute about the fine tuning, phonologists are in general agreement that speech sounds can be scaled in terms of relative sonority in the following manner: vowels > glides > liquids > nasals > obstruents. Vowels are the most sonorous, and obstruents least so. It has long been noted that relative sonority rises at the left edge of a syllable, peaks in the nucleus and then falls away at the right-edge. This principle has obvious consequences for the sequencing of consonants in onset and coda positions. It predicts that the sequence CC in onset position, the consonant closest to the vocalic nucleus will be more sonorous than its neighbour to the left. And indeed the internal structure of initial consonant clusters in many languages is entirely consistent with these outcomes. Even in those, much rarer, languages where clusters of so-called reverse sonority may be found to breach this principle, it has been argued that the mirror, and appropriately sequenced cluster will also always be found, ie iff /#lb-/, then always /#bl-/.

The existence of reversed sonority clusters in the world's languages is not in doubt, even if they remain relatively rare, They have been reported in languages such as Polish, Russian,

Zapotec and Pashto, and the phonetic basis of their existence has been confirmed experimentally [1,4].

In a detailed study of the cluster phonotactics of 104 languages, Greenberg [3] made a number of important cross-linguistic generalizations about the permissible structure of initial and final clusters, generally consistent with accepted notions of sonority sequencing. However, examination of facts observed in Maluku languages suggest that the universal applicability of some of the stated generalizations cannot be confirmed:

• Greenberg's Principle 7: the presence in initial position of at least one combination of stop + stop implies the presence of at least one stop + fricative combination in the same context. However, Roma permits /tt, kt, kd/ but no sequence of stop + fricative.

• Greenberg's Principle 15: in initial and final systems, a lateral is never followed by an r-type sound. However, initial /lr/ is not uncommon in Taba, e.g. /'lraŋi/ 'they bind something'.

• Greenberg's Principle 21: except for voiced nasals followed by homorganic unvoiced obstruents, a voiceless consonant or sequence of voiceless consonants in initial systems immediately preceding a vowel is/are not preceded by one or more voiced consonants. Taba, Roma and Leti breach this generalization at least at the phonemic level, although the surface outcomes differ in each. In Roma 9 such clusters exist, /mt, ms, np, nk, lp, rp, rt, rk, rs/, although at the surface the sequences are bisyllabic with the 1st consonant becoming fully syllabic, e.g. [ṇ'kiki] 'bite'. Only 4 such clusters are found in Leti: /mt, mk, rs, ms/, with first two treated as bisyllabic, e.g. [m̩'deri] 'name of psn' but ['rsuri] 'they pour'. Taba has ten such clusters, all of which surface as phonetic clusters: /mt, mk, ms, np, nk, ns, lp, lk, lt, ls/, e.g. ['lsobal] 'they sail'.

Greenberg also makes the significant observation that in terms of assimilation homorganic nasal + obstruent clusters are preferred over heterorganic nasal + obstruent clusters. This is not consistent with patterns found in Taba: 6 homorganic N+Obs clusters, /mb, mp, nd, nt, ns, ŋk/, are outnumbered by at least 8 heterorganic counterparts /md, mg, mt, mk, ms, nb, ng, nk,/.

The historical source of most non-geminate clusters in Maluku languages appears to be associated with a series of different morpological processes, involving principally the reduction of prefixes to a single consonant. In most cases unusual clusters are still morphologically transparent, e.g. Taba /'kbabas/ 'I bite', /'tbabas/ 'we (inc) bite', /'lbabas/ 'they bite', /'hbabas/ 'you (pl) bite'.

It has been briefly noted that in the languages examined some clusters may not surface as onset clusters. Indeed on this point there is significant divergence. The existence of reverse sonority clusters does not in itself mean an individual language is not able to eliminate - at least at the surface - the contradiction between such clusters and the cross-linguistically significant sonority hierarchy. Reverse sonority can be eliminated by the breaking up the cluster across two syllables. It has already been noted that in Roma, the first consonant becomes fully syllabic, e.g. /nkiki/ -> [n'kiki] 'bite'. This process applies practically across the board: all sonorant consonants become fully syllabic before all other consonants, except /w j/, in word-initial position. Hence previously noted /nnehu/ [ṇ'nehu] 'he jumped'. Leti is more restrictive in its attempts to reconcile reverse sonority with the sonority hierarchy: /l m n r/ only become fully syllabic before stops, e.g. [l̩'kirna] 'game of marbles'. They remain onset consonants before fricatives and sonorants, .e.g. ['rsuri] 'they pour'. Taba is the most resistent: syllabic consonants do not exist and all phonemic clusters surface as full onset clusters.

With respect to the claim that all reverse sonority clusters will always their mirror pair (even if restricted to non-homorganic clusters), Maluku languages demonstrate that this in fact does not hold universally. Taba /ld, ls/ are not matched by */dl, sl/. Roma /mt, md, np/ and Leti /md/ are also unpaired.

4. CONCLUSIONS

The limited data presented here is sufficient to show that not all Austronesian languages have simple phonotactic stuctures. Some languages spoken in Maluku have extremely complex patterns of consonant sequencing in onset position. Word-initial geminates are not uncommon, especially in Taba. Moreover many of the sequences reported here do not fall within normal phonotactic patterning consistent with the sonority hierarchy. So-called reversed sonority is widespread, and creates problems for generally accepted notions of sequency, including a number of longstanding observations made by Greenberg [3]. There is however a mismatch between phonemic and phonetic realizations in some cases, especially in Roma, whereby clusters of apparently reverse sonority are eliminated by consonant nucleation, in what appears to be an attempt to satisfy the sonority hierarchy. However, Taba appears fully resistant: all phonemic clusters surface as onsets, regardless of their sonority pattern. These languages are not unique in the Maluku - many other Austronesian languages in the same region share similar phonotactic features.

REFERENCES

[1] Bell, A and Saka, M. M. 1983. Reversed sonority in Pashto initial clusters. *Journal of Phonetics*, 11, 259-275.

[2] Bowden, J. 1998. Taba (Makian Dalam). PhD, University of Melbourne.

[3] Greenberg, J. 1978. Some generalisations concerning initial and final consonant clusters. In Greenberg, J. (ed.), *Universals of Human Language*. Stanford University Press, 243-279.

[4] Jaeger, J. J. and Van Valin, Jr., R. D. Initial consonant clusters in Yateé Zapotec. *International Journal of American Linguistics*, 48: 125-138.

[5] Steven, L. 1991. The phonology of Roma, an Austronesian language of Eastern Indonesia. MA, University of Texas, Arlington.

[6] Van Engelenhoven, A. 1995. A description of the Leti language (as spoken in Tutukei). PhD, University of Amsterdam.

[7] Zec, D. 1995. Sonority constraints on syllable structure. *Phonology*, 12, 85-129.

	p	t	k	b	d	g	s	h	m	n	ŋ	l	r	w	j
p		pt					ps	ph				pl		pw	
t	tp	tt	tk	tb	td	tg	ts	th	tm	tn		tl	tr	tw	tj
k	kp	kt	kk	kb	kd	kg	ks	kh	km	kn		kl	kr	kw	kj
b				bb								bl	br		
d					dd										
g						gg						gl			
s							ss								
h	hp	ht	hk	hb	hd	hg	hs	hh	hm	hn		hl	hr	hw	hj
m	mp	mt	mk	mb	md	mg	ms	mh	mm	mn		ml	mr	mw	mj
n	np	nt	nk	nb	nd	ng	ns	nh	nm	nn		nl	nr	nw	nj
ŋ			ŋk					ŋh			ŋŋ				
l	lp	lt	lk	lb	ld	lg	ls	lh	lm	ln		ll	lr	lw	lj
r															
w												wl		ww	
j															

Table 1. Initial Clusters in Taba

	p	t	k	d	s	ß	m	n	l	r
p	pp	pt	pk	pd	ps			pn	pl	pr
t	tp	tt	tk		ts	tß	tm	tn	tl	tr
k	kp	kt	kk	kd	ks	kß	km	kn	kl	kr
d				dd						(dr)
s	sp				ss		sm	sn	sl	sr
ß								ßn	ßl	ßr
m	mp	mt	mk	md	ms	mß	mm	mn	ml	mr
n		nt	nk	nd	ns	nß	nm	nn		nr
l	lp	lt	lk						ll	
r	rp	rt	rk	rd	rs	rß	rm	rn	rl	rr

Table 2. Initial Clusters in Leti

	p	t	k	d	s	h	m	n	l	r	w	j
p									pl	pr	pw	pj
t		tt							tl	tr	tw	tj
k		kt		kd			km	kn	kl	kr	kw	kj
d							dm				dw	dj
s	sp		sk								sw	sj
h								hn	hl	hr	hw	hj
m		mt		md	ms			mn	ml	mr	mw	mj
n	np	nt	nk	nd	ns	nh	nm	nn	nl	nr	nw	nj
l	lp										lw	lj
r	rp	rt	rk	rd	rs	rh	rm	rn	rl	rr	rw	rj
w									wl	wr		wj
j												

Table 3. Initial Clusters in Roma

Table 4. Full Set of Initial Clusters in Taba

bb	bbú	grandchild / grandparent
bl	bléwiŋ	megapode
br	brágis	to be covered in sores
dd	ddáwa	grass
gg	ggónan	itch
gl	glís	species of tree
pt	ptól	three (default classifier)
ps	psós	bug
pl	pláŋ	fly (n.)
ph	phót	four (default classifier)
pw	pwál	eight (default classifier)
tb	tbábas	we (incl.) bite
td	tdód	we (incl.) ask
tg	tgóras	we (incl.) shave

tp	tpóas	we (incl.) row
tt	ttágil	we (incl.) walk
tk	tkútan	we (incl.) ask
tm	tmóŋ	we (incl.) snore
tn	tníŋis	we (incl.) smile
ts	tsóbal	we (incl.) sail
tl	tlúk	we (incl.) bend our heads over
tr	tráŋi	we (incl.) bind something
th	tháŋ	we (incl.) go
tj	tjóg	we (incl.) jump
tw	twágik	we (incl.) sell something
kb	kbábas	I bite
kd	kdód	I ask
kg	kgóras	I shave
kp	kpóas	I row
kt	ktágil	I walk
kk	kkútan	I ask
km	kmóŋ	I snore
kn	kníŋis	I smile
ks	ksóbal	I sail
kl	klúk	I bend my head over
kr	kráŋi	I bind something
kh	khán	I go
kj	kjóg	I jump
kw	kwágik	I sell something
ss	ssó	name
mb	mbúluk	to be rotten (of fish)
md	mɗio	be authentic
mg	mgóras	you (sg.) shave
mp	mpóas	you (sg.) row
mt	mtó	eye
mk	mkútan	you (sg.) ask
mm	mmóŋ	you (sg.) snore
mn	mníŋis	you (sg.) smile
ms	msóbal	you (sg.) sail
ml	mlóŋan	be long / tall
mr	mráŋi	you (sg.) bind something
mh	mhóŋas	be sick
mj	mjóg	you (sg.) jump
mw	mwágik	you (sg.) sell something
nb	nbábas	(s)he bites
nd	ndód	(s)he asks
ng	ngóras	(s)he shaves
np	npóas	(s)he rows
nt	ntònolólan	lizard
nk	nkútan	(s)he asks
nm	nmú	muddy water after rain
nn	nníŋis	(s)he smiles
ns	nsóbal	(s)he sails
nl	nlúk	(s)he bends her / his head over
nr	nráŋi	(s)he binds something
nh	nhík	bat
nj	njóg	(s)he jumps
nw	nwágik	(s)he sells something
ŋk	ŋkól	to be lame
ŋŋ	ŋŋé	kanari nut
ŋh	ŋhóŋ	food
lb	lbábas	they bite
ld	ldód	they ask
lg	lgóras	they shave
lp	lpóas	they row
lt	ltágil	they walk
lk	lkútan	they ask
lm	lmóŋ	they snore
ln	lníŋis	they smile
ls	lsóbal	they sail
ll	llú	leaf
lr	lráŋi	they bind something
lh	lhán	they go
lj	ljóg	they jump
lw	lwágik	they sell something
hb	hbábas	you (pl.) bite
hd	hdód	you (pl.) ask
hg	hgóras	you (pl.) shave
hp	hpóas	you (pl.) row
ht	htágil	you (pl.) walk
hk	hkútan	you (pl.) ask
hm	hmóŋ	you (pl.) snore
hn	hníŋis	you (pl.) smile
hs	hsóbal	you (pl.) sail
hl	hlúk	you (pl.) bend your heads over
hr	hráŋi	you (pl.) bind something
hh	hhán	you (pl.) go
hj	hjóg	you (pl.) jump
hw	hwágik	you (pl.) sell something
wl	wló	heart
ww	wwé	leg