RECONSTRUCTION OF PROTO-BATAK PHONOLOGY

K.A. Adelaar
University of Leiden

INTRODUCTION

The name 'Batak' is used for a group of related dialects spoken in the Batak region, an area in Northern Sumatra comprising Tapanuli and part of the adjacent district (kabupatèn) of East Sumatra, up to the sub-districts (ke-camatan) Langkat, Deli, Serdang and Asahan (ca. 2°- 4° NL and 96°- 100° EL). The Batak dialects belong to the western branch of Aus-tronesian languages; they have a quite exten-sive morphology and are archaic in character as compared with Malay and many other Western Indonesian languages.

In this paper I reconstruct the Proto-Batak (PB) phoneme system on the basis of ma-terial from six main dialects for which gram-matical and lexical data are available: Toba (To), Karo (Ka), Dairi (Da), Angkola (An), Mandailing (Ma), and Simalungun (Si). A se-venth dialect, Alas, was excluded for lack of data. The Batak dialects are divided by Voorhoeve (1955:9) into three groups, the northern dialects (NB: Ka and Da), the south-ern dialects (SB: To, An and Ma) and Simalu-ngun. This last dialect is said to be inter-mediate between the northern and southern groups. I will maintain this division for practical purposes, but in part II I present phonological evidence for a division into northern and southern Batak, with a later split-off of Simalungun from the southern group.

The divisions of this paper are as fol-lows:

part I: an account of the (morpho-)pho-nological data available for each dialect;

part II: a comparison of the dialects, and reconstruction of the PB phoneme system;

part III: a comparison of the PB and Proto-Austronesian (PAN) pho-neme systems;

I start from written Batak because it reflects an earlier stage of the dialects, and I will use the 'ejaan yang disempurnakan', the offi-cial orthography for all Indonesian languages, except for the voiceless (vl.) palatal stop (S), the velar and palatal nasal (N) and the shwa, which I write c,ŋ ,ñ and ə respectively. The pronunciation of the southern dialects, and to a lesser extent their phonemic struc-ture, differs considerably from the orthogra-phy. Where needed I write the pronunciation between square brackets and the phonemic spel-ling between slashes, e.g. Ma *halak*, /alak/, [alak] 'human being'.

There is not much uniformity in the de-scription of the dialects; this is primarily due to the differences between, and the lack of phonological data for each dialect. I have made lists of minimal pairs for Ka and To and I have treated both dialects at great-er length because they stand as a model for the northern and southern groups respectively, and because they are known best to us. I have not attempted to reconstruct consonant clus-ters in PB. Two asterisks indicate a PB lexi-cal reconstruction; two crosses indicate the reconstruction of loanwords, and of forms that do not have a cognate both in NB and SB and/or Si, as such forms are not considered to be PB reconstructions. I use one asterisk for PB phonemes, as well as for PAN phonemes and lexemes, and one cross for hypothetical forms of phonemes and lexemes from any stage between the present and PB.

List of abbreviations

AN	Austrone-sian	par.	paragraph
An	Angkola	PB	Proto-Batak
Dat	Batak	resp.	respective-ly
BI	Bahasa In-donesia	S	stop
C	Consonant	SB	Southern Batak (see Introduc-tion, par. 9.3.1 and par. II)
Da	Dairi		
e.g.	exempli grata		
Jav	Javanese		
Ka	Karo	Si	Simalungun
Ma	Mandailing	Skrt	Sanskrit
Mal	Malay	To	Toba
N	nasal	V	vowel
NB	Northern Batak	vd.	voiced
PAN	Proto-Aus-tronesian	vl.	voiceless

I AN ACCOUNT OF THE (MORPHO-)PHONOLOGICAL DATA AVAILABLE FOR EACH DIALECT

1. *The Karo phoneme system*

1.0 *Vowels*

high	i,ī		u,ū
mid	e,ē	ɔ,ö	o,ō
low		a,ā	

Consonants	glot-tal	velar	pal-atal	den-tal	la-bial
Stops vd		g	j	d	b
vl		k	c	t	p
Nasals		ŋ		n	m
Liquids vibrant				r	
lateral				l	
Fricative	h			s	
Semivowels			y		w

1.1 Vowels

Vowels (V) have long and short variants: vowel length is distinctive, e.g. *idās* 'on, on top of', against *idas* 'one-ply thread'; *mān* 'to eat' against *man* 'to, for'. A long V originates from contraction of two V's: a) by loss of intervocalic C, e.g. *idās* and *idatas*, which is less in use; *mān* and *maŋan*, which is slightly more formal, b) in the process of affixation, e.g. *mbule/mbulēn*; *jandi/pĕrjandīn*.

The long variant of shwa, /ö/ is not distinct from /ə/ in the script; its sound is like the /ö/ in German *schön*. Its status in the phonemic system is unclear. Neumann gives the following lexemes with /ö/: *andö*, *böbö*, *bök*, *candö*, *cöt*, *dö*, *döm*, *dönö*, *dör*, *dös*, *(m)əntö*, *gök*, *göŋ*, *jinö*, *jöhö*, *köm*, *köŋ*, *köp*, *lös*, *möl*, *nö*, *pəsöl*, *pöt*, *pös*, *röh*, *röha*, *səntö*, *söh*, *sökər*, *taŋgö*, *təkö*, *tök*. In his Grammar he also mentions *ŋalömi* (see par. 7.1). The /ö/ occurs in monosyllabic lexemes and in disyllabic lexemes as a final phoneme ((C)VCö-type); it also occurs as V in both open syllables of a disyllabic lexeme, e.g. *dönö*, *böbö*, *jöhö*. Three items do not fit in the lexeme types given here: *pəsöl*, *röha*, and *sökər*, which has a variant *uskir*; in two cases a minimal pair of /ə/ and /ö/ was found, and on account of this I consider both sounds as phonemically distinct: *gəŋ* 'to have lost and to start again, in a panta-play', against *göŋ* 'k.o. deer-trap'; *pəsəl* 'to push on the belly of a laboring woman, in order to accelerate the birth', against *pəsöl* 'k.o. leaf'. Examples of /ə/ in open ultima and monosyllables: *jəh*, *dəh* in *dəhkən* 'to press down, to compress'; *ərgədə* and *ərkətə* 'to be unable to forget s.th., to sulk'.

1.1.1

In his remark on the pronunciation of -/o/# in *naŋko*, and also in a remark on the accent of Sinembah-Serdong Karo, Neumann suggests that the spoken language distinguishes diphtongs from single vowels; these diphthongs are more or less carefully distinguished, according to the village where the native speaker comes from.

1.2 Semivowels

/y/ and /w/ do not occur in initial or final position nor do they have distinctive value in the sequences [iya], [iyo], and [iye], where they are omitted in Roman script. Examples of semivowels with distinctive value are: *ayun*, *mayan*, *ayap*, *buyak*, *buyur*, *tiwən*, *awak*, *awin*.

1.3 Consonants

/s/ and /c/ are distinctive, e.g. *sadur* 'to mix' against *cadur* in *ndərdər cadur* 'to make one's mouth water'; *səpah* 'be struck, caught unawares (by rain e.g.)' against *cəpah* 'chewed betel-nut'; on the other hand there are many instances of free variation of /c/ with /s/ and of /c/ with /t/.

/h/ and /k/ are also distinctive, e.g. *hayaŋ* 'languid, slow feeling after an attack of fever', against *kayaŋ-kayaŋ* 'k.o. lianas with dark fruits'; *bəcik* 'spring-set spear or arrow-trap', against *bəcih* '1) puffed, of rice; 2) the cocoon of a type of yellow butterfly, used for food'; *dakan* 'to cook rice', against *dahan* 'large branch of a tree'. The -/h/# sounds like German -ch ([x]); /h/- is usually not pronounced and is hardly ever distinctive in this position, e.g. *ayam/hayam*; *aŋat/haŋat*; *antu/hantu*. In the process of prenasalisation clusters of S preceded by homorganic N arise, e.g. *gudam/ŋgudam*; *bahut/mbahut*; *jurmak/njurmak*; *dampiŋ/ndampiŋ*[1]; *təkö/ntəkö*.

1.4 Accent:

(non-distinctive) occurs on the penultima, or, when this contains a shwa, on the ultima; in doubled root-morphemes it falls on both syllables equally.

1.5 Syllable structure

V	*i*, *u*, *o*, *e*, *ú-juŋ*, *í-saŋ*, *á-la*
CV	*da*, *dö*, *lú-kat*, *tá-lu*
VC	*ər-ta*, *əl-təp*, *sú-an*
NCV	*ndái-i*, *á-ṃpu*, *tá-nda*
NCVC	*tá-mbah*, *ti-ŋgəl*, *á-ŋkat*

1.6 Minimal pairs of Karo phonemes (consonants)

a. batuk 'cough'
patuk 'to peck, bite'
datuk 'a title'
buaŋ 'to throw away'
tuaŋ 'to pour'
bəlaŋ 'wide, extensive'
gəlaŋ 'bracelet'
baŋke 'corpse'
haŋke 'reverent'
buka 'open'
bara 'shoulder; k.o. plant'
nari 'still, left (used with numerals); only, mere, sheer'
bante, ər -- 'to join in together'
rante 'chain, bridle'
bapa 'kinship term'
sapa 'amortized (debt), pardoned (crime)'
bari, m -- 'to be cold'
baruŋ 'k.o. corral'
ŋari 'to serve at meals'

karuŋ 'plaited rice-bag'
bur, m -- 'fat, thick'
cur 'very fine, very small

d. dampar 'side, lateral face'
tampar 'to ward off a blow, to parry'
daya, ər -- 'to sell'
gaya '(earth-) worm'
dayaŋ 'k.o. plant; title for a maiden'
hayaŋ 'heavy feeling as after an attack of fever'
dimpu 'to be very afraid, cowardly'
dalaŋ, -- ən 'offering, dish to the spirits; idiomatically; to have a dry throat'
malaŋ 'respectful'
dalih, ər -- 'to fail, of a gun'

nalih, sə -- 'quantity of rice'
dəŋat 'to droop, of a branch'
rəŋat 'to radiate; to leap, shift, of pain'
daməl, mə -- 'half sweet and half sour, of taste'
saməl, mə -- 'pleasantly soft to the touch'
daŋa 'to lessen, to stop'
ŋaŋa 'wide open (mouth), opening, gap'
dalaŋ (see above)
jalaŋ 'to wander about, to rove'
dəru, ər -- 'to walk in a flock, of sheep'
kəru 'to be annoyed'
dian 'a candle of beeswax; rather, better'
cian 'having a grudge against s.o.; to chase away, take s.o.'s place'
g. gaga 'underside, lower side of a mangosteen'
haga, mə -- 'respectful'
gadiŋ 'ivory, elephant's tooth'
ladiŋ 'generic for curved knives and swords'
gulaŋ, mə -- 'to roll down a slope'
mulaŋ 'weaned, of animals'
gundaŋ 'float with bait; reciting ənduŋ-ənduŋən
nundaŋ 'to fail, let down, desert'
gumbuŋ 'big earthen cooking-pot'
rumbuŋ, rumbuŋ-rumbuŋ 'hereditary enemy'
gəmbuŋ 'swollen, of belly'
səmbuŋ 'sheath of knife'
galah 'stick, bamboo-stick to knock off fruits'

ŋalah 'heavy, slow, of limbs'
gambut 'k.o. creeper'
jambut 'to take away in the dark, to sneak, to steal'
girik 'to turn a stick quickly between the hands, like a drill'
kirik 'k.o. grasshopper or beetle'
gabak 'small axe'
cabak 'wooden bowl, finger bowl'
h. haga, mə -- 'respectful'
laga, mə -- 'wild, fierce; proud, arrogant, vehement'
hayaŋ (see above)
mayaŋ 'areca-palm, k.o. tree'
hala, hala-hala 'flat shell'
nala 'flame'
haŋke 'to stand in awe of, to be afraid'
raŋke 'plaitwork of rope; bunch, cluster'
hiri, hiri ate 'to bear a person a grudge
siri, siri-siri 'dragon-fly'
hile, ŋe -- 'exult with high-pitched voice'
jile, -- ne 'the beauty of s. th.'
kile, ərkile-kile 'to say one thing today, another thing tomorrow'
hio 'stereotyped word in the bilaŋ-bilaŋ'
cio 'protected from rain (a place)'
huŋ 'Skrt. ōm
ŋuŋ 'onomatopoetic for deep, resonant sound'
p. patuk, datuk (see above, under b.)
puar 'torn off, broken off'
tuar 'small fish-trap'

pulaŋ 'wiped out (by an epidemic), extinguished'
gulaŋ, mə -- 'to roll down a slope'
pitu 'seven'
hitu, si-- 'the suspect in a law-case'
puas 'satisfied'
luas 'evaporated, of luke-warm water'
pulaŋ, mulaŋ (see above)
pande 'skilled, clever'
nande 'mother'
piah 'kidneys; at last, eventually'
riah 'mirth, excitement'
patuk 'k.o. chopping knife; to peck, to bite'
satuk 'the mixing of indigo-paint with chalk'
pərah, pərah-pərah 'to weep heavily, to cry'
ŋerah 'to be conceited'
pulus, ər-- 'to twist, like a rope'
julus 'attentive, indifferent'
pira 'roe, eggs of a frog'
kira 'to count, to reckon'
parik 'ditch'
carik, n-- 'to tear off, rip off'
t. talar 'free, wide (view)'
galar 'what is given in payment for s.th.'
tala 'to turn o.'s face to'
hala-hala (see above)
taki, ər-- 'to conspire, to act secretly against s.o.'
laki 'husband'
talu 'to have lost (against s.o.), to have failed'
malu 'ashamed'
salu 'with (instrumentalis)'

tala, nala (see above)
ramas 'a set (of s.th.)
tamas 'empty, finished, of a fruit-tree'
tuda, mə-- 'impudent, presumptuous'
ŋuda 'tender, young'
təlis 'outcry of peevishness, to a woman'
jəlis 'clean, ritually pure'
tuan 'master, lord, European'
kuan, ər-- 'to talk; to fondle, of girls and boys'
cuan 'pickaxe to work the ground'
l. lalu 'rice-pounder'
malu 'ashamed'
lanam 'folded cloth, small cushion'
nanam 'the taste of s.th., the use of it'
liaŋ 'hole, gap, opening'
siaŋ 'to break, of the day, to become'
riaŋ, riaŋ-riaŋ 'creeper; scorpion'
liat 'smooth, like paper'
ŋiat 'to expand, to swell'
laba, ər-- 'to be bought with profit'
jaba 'barley'
luh, mə-- 'to be loose, e.g. of the hair
kuh 'complete, full'
liho 'thin, watery'
ciho, mə-- 'clean, pure, transparent'
m. mala 'disappeared (smell), faded (color)'
nala (see above)
mədak 'open (eyes); to liquify, of oil, fat'
sedak 'oppressed in the chest; oppressed of mind, because of sorrow'

3

muah, pǝ-- 'curse'
ruah 'ripped off,
of weeds'
momo 'acknowledge-
ment'
ŋoŋo 'totally
quiet, deso-
late'
malaŋ 'respect-
ful'
jalaŋ (see above)
malu (id.)
kalu 'term used at
a panta-play'
mala (see above)
cala, cala-cayak
'k.o. cymbal'

n. ni 'genitive mark-
er'
si 'personal arti-
cle'
name 'deep spot in
river'
ramo, ŋramoi 'to
sort out,
choose'
nilnil 'with pro-
truding eyes'
ŋilŋil, -- ipena
'to show
o.'s teeth'
nai 'formerly,
recently;
(the) late'
jai, ǝr-- 'to be
married, of
women'
kai 'interrogative
pronoun'
nala, cala (see
above)

s. sauk 'to roast in
a pan (maize
or coffee)'
rauk 'a grasp (of
the hand)'
sǝna 'k.o. tree;
isǝnakǝn to
acknowledge,
spread the
news'
ŋena 'to like,
love s.th.'

sauŋ 'what serves
as an um-
brella'
jauŋ 'maize'
sial 'misfortune,
bad luck'
kial 'bent, be-
cause of the
heat etc.; to
ward off the
'panta' of
the adversary'
sulu, isulu 'to
inquire into'
culu 'to be out,
finished, of
fruit'

r. rasi, ŋǝ-- 'to
consult'
ŋasi, la-- 'impo-
tent'
rambe 'k.o. tree'
jambe 'k.o. gourd'
rambiŋ 'term used
for counting
maize-ears'
kambiŋ 'goat'
rio 'with holes,
of a leaking
roof'
cio 'protected
from the rain
(a place)'

ŋ. ŋari 'to serve at
meals'
cari 'to look for
o.'s living'
kari 'presently,
by and by'
ŋadi 'to cease,
stop'
jadi 'to become,
come into
being;
tǝrjadin to
put s.th.
aside'

j. jari 'finger, toe'
kari (see above)
cari (id.)

c. cari, kari

2. The Dairi phoneme system

2.0 Vowels

	high	i,ī		u,ū
	mid	e,ē	ǝ	o,ō
	low		a,ā	

Consonants

		glot-tal	velar	pal-atal	den-tal	la-bial
Stops	vd		g	j	d	b
	vl		k	c	t	p
Nasals			ŋ		n	m
Liquids	vibrant				r	
	lateral				l	
Fricatives		h			s	
Semivowels				y		w

2.1 Vowels

The Da /ǝ/ occurs in all syllables except in the position __(h)# . All V's but /ǝ/ have a long variant which originates from contraction in the process of affixation (see par. 2.4), or is correlated with accent: V is long in (C)VC and also in (C)V́CV(C); examples: gūŋ būk pāntis tūtur pōstǝp. Height harmony: in VCǝC V can never be /i/ or /u/, but /o/ and /e/ do occur, e.g. kōdǝn, ēŋǝt; /u/ does not occur in VCi(C), but /o/ does, e.g. kōciŋ; V=/e/ and ≠ /i/ in VCu(C), e.g. ēnum.

2.2 Semivowels

These occur only between V's; /w/ follows /a/, /e/, /i/, /o/ and /u/, e.g. kawan, siwah, lawi, ǝncuan,[ǝncuwan], kalcan [kalowan], dewan. Van der Tuuk deals only with /w/ in the sequences /awa/ and /awi/, whereas the Dairi-Pakpak — Dutch dictionary also gives six instances of /ewa/ — sequences and one with a sequence of /iwa/: sewah (=sebah), dewan (=deban), rewan (=reban), bewa 'tax on commodities', bewan 'calladium species', rewas 'k.o. clap-net to catch something' against rebas 'destroyed', siwah (=sibah). It appears that three lexemes with an /ewa/-sequence, and siwah have a variant with /b/, and that in only one of the remaining lexemes, /w/ contrasts with /b/ i.e. rewas against rebas. Given only one known example with distinctive /w/ in this environment it is perhaps best to consider rewas as a borrowing, and to assign a distinctive /w/ only to the sequences /awa/ and /awi/. /y/ follows any V ≠ /ǝ/; its value is not distinctive after /i/ and /e/ and in this position it is not written in Roman script, e.g. rayam, loyaŋ, beaŋǝn [beyaŋǝn], suyup, ia [iya].

2.3 Consonants

/h/ and /k/ are distinctive phonemes; examples of minimal pairs: hia 'glad, happy, pleased' against kia, a form of address to younger siblings; pǝ-kia-kǝn 'to treat s.o. in a presumptuous way'; nahan 'soon, in the future' against nakan 'cooked rice'; kalah 'uvula' against kalak 'human being, person'. In initial position /h/ is weak, and is often omitted, as in hasa: [hasa]=[asa]; but in some lexemes /h/ is never omitted, as in hakum [hɨkum] 'if, in case'. Two different liquids

4

never occur within one lexeme. /s/ and /c/ are distinctive; when the next C = /c/ and after a N only /c/ is found, e.g. *kanca*, *cəcədur*; minimal pairs of the /s/-/c/ - opposition: *kəcin* 'k.o. tree' against *kəsin* 'to turn around'; *campur* 'to mix, blend' against *sampur* 'to go down, descend'. Gemination is avoided except at morpheme boundaries, as in *dahan-na* and *dok-kən*.

2.4 Morphonology

Through suffixation of -/ən/ after a final V this V is lengthened and the /ə/ of the suffix is omitted, e.g. *doli*+ən becomes *dolīn*; *ue*+ən yields *uēn*. The prefixes /pə/- and /kə/- lose /ə/ before /ə/, /e/, /i/, and /o/[2] (v.d. Tuuk only mentions /i/, /e/, and /o/, but one example that he cites shows contraction of /pə/- with /ə/: *pə*+*əmpat*+*kən* becomes *pəmpatkən*); examples: *pə*+*idah*/*idah* yields *pidahidah*; *olih*/*polihkən* *eŋət*/*keŋətən*; with prefixation before /u/ and /a/ an /h/ is inferred: *uap*/*pəhuap*; *alo*/*pəhalo*.

2.4.1 Vowel assimilation:

/ə/ of the suffix -/ən/ usually becomes /u/ after -u(r,h)_, e.g. *napur*/*napurun*; *pusuh*/*pusuhun*. This /ə/ can also assimilate with /i/ of the preceding syllable: *kərsik*/*kərsikin*. There are also some cases of assimilation of /ə/ to /i/ in the suffix -/kən/: *arih*/*arihkin*; *bakin*/*bakinkin*. /ə/ often assimilates with /i/ or /u/ of a following syllable that starts with /r/, e.g. *surune* also *sarune* and *sərune* (the Dairi-Pakpak dictionary only mentions *sarune* and *surune*; *sərune* occurs as an example in v.d. Tuuk's Grammar). /ə/ in prefix /kə/- assimilates with /i/ of the infix -/in/-: *k-in-ə*+*guru*+*ən* yields *kinigurūn*; *k-in-ə*+*juah* yields *kinijuah*.

2.4.2

Prof. J.C. Anceaux supposes that the lengthening of a root-final V before -/ən/ involves the same principle as suffixation of -/ən/ to lexemes with -/h/ or -/r/, e.g. *bagīn* would have originated from *bagi*+ən with preceding vowel assimilation. This would also explain why suffixed words can have the accent on the ultima, as this originated from contraction of the penultima and ultima in the earlier form.

2.4.3

The prefix /mə/- becomes /əN/- before a C which is not /r/, /l/, /h/ or N, e.g. /mə/+/pəŋke/ yields *əmpəŋke*; /mə/+/haŋke/ yields *məhaŋke*; /mə/+/tor/ yields *əntor*; /mə/+/gəluh/ yields *əŋgəluh*; /mə/+/lampis/ yields *məlampis*; /mə/+/rincan/ yields *mərincan*; /mə/+/nahan/ yields *mənahan*. Before V the /ə/ is omitted: /mə/+/ela/ yields *mela*; /mə/+/oto/ yields *moto*.

2.5 Sandhi

Nasals in final position do not undergo the influence of sandhi in Dairi; vowel contraction occurs under the same conditions as in Toba (see par. 3.6.5). Two other cases should be mentioned here:

1) the V of the suffix -/mu/ is deleted when followed by a demonstrative pronoun (van der Tuuk does not provide examples of this principle); also the final V of the lexemes *bae*, *(h)ari* and *bage* is deleted when they are followed by a demonstrative pronoun, e.g. *bage* 'like, (same) as, as well' +*idi* 'there, yonder' yields *bagidi* or *bagī* 'so, thus, in such a way'; *sidah (h)ari* 'when' +*en* 'this' yields *sidaren* 'today, at present'; *bae* 's.o. whose name one does not know or who one does not want to call by name' +*en* yields *baen* 'this friend, this companion'; *bage*+*ena* (=*en*) yields *bagena* 'like her/him';

2) *ŋo*, an emphasizer, and *mo* 'word put next to the predicate, which brings the meaning of the latter outside present reality', sometimes lose their V before a pronoun used as a substantive, e.g. *laus* 'to leave, depart' +*mo*+*ia* 'third person singular' yields *lausmia* '(so) she/he went'; *laus*+*mo*+*aku* 'I' yields *lausmaku* '(so) I went'; *ena*+*ŋo*+*aku* yields *enaŋaku* 'this is me'.

The /ə/ of the prefix /əN/- is omitted before final V of the preceding word, e.g. *kuta*+*əntuara* yields *kutantuara*; *kəncan*+*əntasak* yields *kəncantasak*; *maŋan*+*mo*+*əntuara* yields *maŋanmontuara*.

2.6 Accent

The accent is on the penultima, unless it contains /ə/, in which case the accent is on the ultima, even if here the V is /ə/, e.g. *bəlgàh*, *bəlèn*. A V in the ultima takes the accent, which is the case only in polymorphemic words, e.g. *bagīn kinigurūn*. In prefixing /məN/- to a lexeme with initial labial, /mə/- is sometimes omitted, e.g. *buat*/*məmuat*/*muat*; *bunuh*/*məmunuh*/*munuh*; *pahan*/*məmahan*/*mahan*; but there is no such form as *mukar* along with *məmukar*. The suffix -/sù/, which expresses an excessive, demands the accent: /əntor/+/sù/ yields *əntorsù*.

Forms of address (vocatives) also have the accent on the ultimate syllable. In Da lexemes of more than three syllables are unusual, as with the borrowing *əntuara*: this lexeme derives from the Toba words *antu* and *ara* and has a trisyllabic variant *tuara*.

NB. Except in lexemes with a CəCəC structure we see that the place of the accent within the Da lexeme is determined by vowel length with the following hierarchic arrangement: V̄, V, /ə/; it is not the syllable structure which determines the place of the accent but the relative quantity of the vowels (again, with the exception of lexemes with CəCəC structure).

3 The Toba phoneme system

3.0 Vowels

high	i		u
mid		e	o
low			a

5

Consonants

	glottal	velar	palatal	dental	labial
Stops vd		g	j	d	b
vl		k		t	p
Nasals		ŋ		n	m
Liquids vibrant				r	
lateral				l	
Fricatives	(h)			s	

3.1 *Semivowels*

Toba does not have phonologically distinctive semivowels; they occur in pronunciation as predictable glides.[3]

3.2 *Consonants*

/h/ and /k/ are in complementary distribution: /h/ occurs in initial position, between V's and after liquids, whereas /k/ is the final C of a syllable, and can also be the second component of a consonant cluster (see par. 3.3). /h/ is not pronounced in medial position before /i/, e.g. *mahiandu* [*maiyandu*]. Two different liquids do not occur within one lexeme (see also par. 3.6.6); voiced stops do not occur in final position.

3.3 *Realization of consonant clusters within Toba lexemes*

3.3.1

N+vl. stop or /s/ yields gemination of the latter, e.g. *aŋka* [*akka*]; *ompu* [*oppu*]; *tintin* [*tittin*]; geminated /s/ is realized as [ts]: *hinsu* [*hitsu*].

3.3.2

/n/+/r/, /l/ or /m/ yields gemination of /r/, /l/ or /m/, e.g. *ronron* [*rorron*]; *monmon* [*mommon*]; *lanlan* [*lallan*].

3.3.3

(C ≠ /s/) + /h/ yields [kk] in doubled root-morphemes, e.g. *hophop* [*hokkop*]; *haŋham* [*hakkam*] (from hamham+); /s/+/h/ yields gemination of /s/: *hushus* [*huṭsus*]; /t/+/h/ yields [kk] or [tt], e.g. *huthut* can be realized as [hukkut] as well as [huttut].

3.3.4

N+/g/ yields /ŋg/ in doubled root-morphemes, e.g. *goŋgom* from gomgom+.

3.3.5

NB. /r/+/l/ yields gemination in *porlak*, which is realized as [pollak], probably analogous to the rule of prefixation of /mor/- before /l/ (see par. 3.5.2).

3.4 *Accent*

The accent is on the penultima, on the ultima or, in the case of doubled root-morphemes of onomatopoetic character, on both syllables, e.g. /hùndul/, /gàbe/, /tanòm/, /debatà/, /bùmbàm/, /dòkdàk/; its place is distinctive, e.g. /sogòt/ 'early in the morning', against /sògot/ 'soon, later, in the

future'. It is evident that one has to do here with etymologically related lexemes.

3.4.1

The place of the accent of a verbal lexeme can change depending on its attributive or predicative use: in the first case it is on the ultima, in the second case on the penultima. Substantives also get the accent on the ultima when used as predicate, e.g. /hinsu/ 'key' and /maŋhinsu/ 'to close, shut', against /hinsù/ 'closed, shut'; (with nasalization of initial C:) /ma-nànom/ 'to bury' against /tanòm/ 'buried'; /tàntan/ 'hanging down', against /tantàn/ 'to be lowered down'; /ràhat/ 'saw' against /rahàt/ 'to be rough'; /bègu/ 'supernatural being', against /begù/ 'courageous'; /bàu/ 'odour, smell', against /baù/ 'to stink'. Except for words subject to the rules mentioned in 3.4.2, polymorphemic words have the accent on the penultima, e.g. /timbò/ which equals /ma-timbo/ in meaning 'to be high'; /rarà/ and /ma-ràra/ 'to be red'; /sòhot/ becomes /sohòton/ when suffixed.

3.4.2

Polymorphemic words can have the accent on the ultima only in the next two cases: 1) when suffixed with -/hù/ and -/tù/, which express excessiveness, and -/àn/ which expresses the comparative degree, e.g. /magànjan+hù/ yields /maganjanhù/; /rahàn+àn/ yields /rahanàn/; 2) when a monosyllabic lexeme is affixed the accent is always on the stem, e.g. /dòk/, /mandòk/, /dòhonon/.

3.5 *Morphonology:* a schematic survey

3.5.1 *Nasalisation*

N+/d/, /j/ yields /nd/, /nj/, e.g. *dege/ mandege; joŋkal/manjoŋkal*.

N+/b/ yields /mb/, e.g. *buat/mambuat*; sometimes /b/ is lost in the course of nasalisation, e.g. *bunu/mamunu*; this never occurs when the second syllable begins with /b/ as well, e.g. *balbal/mambalbal; bebe/mambebe*.

N+/g/, /h/, V yields /ŋg/, /ŋh/, /ŋV/, e.g. *goar/maŋgoar; hinsu/maŋhinsu; inum/maŋinum; alap/maŋalap*. Some lexemes used transitively and beginning with /h/ lose this phoneme, e.g. *hibul* 'round, cylindrical, globular' and *maŋibul* 'to be compressed; to be closed like a bud, i.e. to be childless'.

N+/r/, /l/ yields /ŋar/, /ŋal/, e.g. *rior/maŋarior; lapa/maŋalapa*.

N+/s/, /t/ yields /n/, except with exclusively intransitive verbs, where it yields prenasalisation, e.g. *surat/manurat; tura/ manura*; (intransitive) *sohot/mansohot*; (v.d. Tuuk does not provide an instance of prenasalisation of an initial /t/ in an intransitive verb).

N+/n/, /ŋ/, /m/ yields /n/, /ŋ/, /m/, e.g. *nolnol/manolnol; ŋatŋat/maŋatŋat; mata/ mamatahon*.

N+/p/ yields /m/, e.g. *pinsaŋ/maminsaŋ*.

3.5.2 *Prefixation*

/mor/-, /por/-: /r/+/l/ yields /ll/, e.g. *porlanja* is pronounced [pallanja] and *morlaŋə* is pronounced [mallaŋe] (the /o/ of /mor/-, /por/-, and /tor/- are usually pronounced as [a]).

3.5.3

Suffixation -/hu/, -/hon/: final C +/h/ yields gemination of C, e.g. *pathu* [pattu]; *solothon* [solotton]; *taraphu* [tarappu]; *uloshu* [ulotsu]; -/n/+/h/ can also become [kk] in words of more than two syllables, e. g. *napuranhu* is pronounced [napurattu] as well as [napurakku].

Suffixation of -/mu/: -/n/+/m/ yields [mm], e.g. *tinaruanmu* [tinaruammu].

Suffixation of -/sa/: -/s/+/s/ yields /ts/, e.g. *tanarossa* [tanarotsa].

3.6 *Sandhi*

3.6.1

A nasal is denasalised as **a final** C and becomes its homorganic vl.S if **it is** followed by a vl.S or /s/, e.g. *tian saba* [tiatsaba]; *indaŋ podo* [indakpodo]; exceptions to this rule are: a) /n/+/h/ yields [kk], e.g. *paŋan hamu* [panakkamu]; **b)** /n/+/p/ yields [pp], e.g. *dibaen pogosna* [dibaeppogosna]; c) /m/+hVpVC yields kkVpVC, e.g. *onom hupaŋ* [onokkupaŋ]; d) monosyllables sometimes change their final N to one homorganic to the next C, e.g. *dan bulan* [dambulan]; /ŋh/ yields [kk], e.g. *hiŋhian* [hikkiyan].

3.6.2

Final N is assimilated to a following lateral or /m/, e.g. *tian ruma* [tiyarruma]; *tuan laen* [tuallaen]; *manuŋkun ma ibana* [manukkummaibana].

3.6.3

Vl.S + /h/ yields gemination of the S, e.g. *hutopot hami* [hutopottami].

3.6.4

Final /s/ + initial /s/ yields /ts/, e. g. *las soada* [latsowada].

3.6.5

Vowel contraction: two similar V's that are in contact through conjunction of two words are contracted to one V; when this V occurs in the last syllable, it gets the accent, e.g. *diida* [dida]; *hu utahon* [hutahon]; *horbon* (=/horbo/+/on/) [horbòn]; *maridi* (=/maridi/+/i/) [maridi] (both last cases are spelled in contracted form).

The pronominal suffixes -/hu/ and -/mu/, and also some lexemes like *aŋkora*[4] and *baio*[4] lose the final V before a demonstrative pronoun, e.g. *hujurmon* which is the contraction of /hujur/+/mu/+/on/; *lambuŋkon* is the contraction of /lambuŋ/+/ku/+/on/ (here too the contracted forms are reflected in the spelling). Interjections do not lose their final V: *ba aman* [ba'aman]; *ale eda* [ale'eda]. The preposition *di+i* yields *disi* (cf. BI ini/di sini; itu/di situ).

3.6.6

Regressive assimilation takes place when two different liquids occur in a word group, e.g. *malampis bibirna* [marappisbibirna]; *ulu poraŋ* [uruporaŋ].

b. batu 'stone, weight'
patu 'subjected'
datu 'sorcerer'
bolon 'big, voluminous'
tolon 'to swallow s.th'
bale 'rice-measure, ¼ of a solup'
gale 'weak, languid'
bunti 'k.o. offering'
hunti 'to carry on the head'
bias 'nausea, boredom, disgust'
lias 'free from guilt, from suspicion'
malu 'to be ashamed of s.th.; not to feel like'
balu, ma-- 'to have become a widow(er); roŋgur-balu thunder without flash (as an omen)'
bandar 'cabin to stay overnight'
mandar 'wrapping, e.g. sarong'
bunu 'to strike, kill'
nunu 'to whisper, to mutter, of many people'
boa 'to proclaim, announce'
roa 'bad, mean'
bere 'sister's child'
sere 'gold'
abat 'obstacle'
aŋat 'trick, artifice; better'
balo 'satisfied'
jalo, manjalo 'to take, receive; jalo-jalo forthwith, immediately, quick'

p. upa 'salary, reward'
uta 'to vomit'
patu 'subjected'
datu 'sorcerer'
lapo 'small shop'
laho 'to go, walk'
sopo 'Batak rice-barn'

sogo 'angry, irritable, reluctant'
paho 'propriety, fashion, custom'
laho 'to go, walk'
lapo 'small shop'
lamo 'pleasant to the touch'
api 'fire'
ani, maŋanihon 'to wind up yarn on a rack'
puntas, ma-- 'to burst open, of cracked leather bags'
runtas, maŋa-- 'pull down, demolish (a house)'
pargut, ma-- 'to peck, of birds; to bite, of, snakes'
sargut, manarguti 'to be many'
palu, mamalui 'to beat, strike'
ŋalu, mor-- 'to ruminate'
palo 'white spots, on one's eyes'
jalo (see above)

d. dandi 'to disobey, out of insolence'
tandi 'prominent (in good or bad sense)'
dao 'wide, broad, vast'
hao 'to itch, scratch with one's hand'
gao 'to move the hands, in swimming'
dabu, mardabu 'to fall'
labu 'to gather, come together'
duhut 'grass'
muhut 'trifles; junk'
dopaŋ, man-- 'to fight, attack'
ropaŋ 'jagged, notched'
deaŋ 'young girl'
neaŋ 'easy to carry'
duruŋ 'scoop net'
suruŋ 'privilege'

7

dae 'unfit, indecent'
ŋae 'watery, of tubers and fruits'
duruk 'to fall, precipitate forward; man-- to ask for, claim'
juruk 'to project, stick out'

t. ata, man-- 'to eat raw meat'
aha 'interrog. pronoun "what"'
tado 'to sit on the ground (with the feet stretched out in front)'
gado 'stick, staff'
buat 'to take'
bual 'to rise, of smoke'
rata 'green, fresh'
rama 'friendly, kind, familiar; to catch fish'
otan, maŋ-- 'to seize'
onan 'market (-place)'
rata (see above)
rara 'red'
ata (see above)
asa 'so, thus, therefore'
tambaŋ 'athwart, across'
tambat 'tied, bound fast'
taha, manaha 'to cleave wood in small pieces'
jaha, manjaha 'to read'

h. halak 'human being'
galak 'to burn brightly, of light, fire'
hambiŋ 'goat'
lambiŋ 'protruding, of a bow'
hata 'word, speech'
mata 'eye'
hiaŋ 'dry'
niaŋ 'to be meager, lean'
habut 'opaque, non-transparent'
rabut 'rough, hairy (of skin)'

hali 'to dig'
sali, mor-- 'to borrow, lend, of money or goods'
ŋali 'cold'
habut (see above)
jabut 'hair on the chest'

g. giaŋ 'mixed, blended'
liaŋ 'hole, gap'
gutu 'struggle'
mutu 'willing, zealous'
gabuŋ 'to mix, blend'
nabuŋ 'to shave'
gara 'live coal'
rara 'red'
gabur 'easy to do'
sabur 'disseminated, scattered'
giak 'dirty, nauseating, disagreeable'
ŋiar 'burning fiercely'
gabuŋ 'to mix, blend'
jabuŋ 'furious'

l. lalu 'to arrive'
malu 'to be ashamed'
laŋka 'pace, step'
naŋka 'kernel of the pinasa-fruit'
lambiŋ 'pot-bellied'
rambiŋ, maŋa-- 'to hang up, suspend s.th.'
lamo 'pleasant to the touch'
samo 'modest, in behavior or way of clothing'
lae 'brother-in-law; agreeable'
ŋae 'watery, of tubers and fruit'
lamuk 'soft, weak, light'
jamuk 'simple'; morjamuk 'to wash one's hands'

m. maga 'majestic'
naga '1 mythical snake 2 = maga'
mabas 'to come off, of small-pox marks'
rabas, maŋa-- 'to lop wood at the top, to

cut wood'
mabuk 'intoxicated'
sabuk 'waist-cloth'
maol 'difficult, expensive'
ŋaol 'tired (said of hands and feet)'
molo 'when'
jolo 'first, before'

n. nama 'only'
rama 'friendly'
niaŋ, mor-- 'to be lean, meagre'
siaŋ 'to become bright, daylight'
nok 'deep (of sleep)'
ŋok, tar-- 'to loose consciousness, to faint'
nama (see above)
jama 'to seize, take'

r. rian 'women's ailment (after childbirth)'
sian '(preposition) from'
riak 'amusement, fun'
ŋiak, mor-- 'to grunt, of pigs'
rahat, ma, a-- 'to file, polish'
jahat 'bad, mean'

s. sali, morsali (see above)
ŋali 'cold'
sora, manorahon 'to hand over'
jora '(willing) to mend one's ways'

n.j. jomjom 'to get stuck in the mud'
ŋomŋom 'lying in front of'

4. *The Mandailing phoneme system*

4.0 *Vowels*

high	i	u
mid	e	o
low	a	

Consonants

	glottal	velar	palatal	dental	labial
Stops vd		g	j	d	b
vl		k	(c)	t	p
Nasals		ŋ	ñ	n	m
Liquids vibrant				r	
lateral				l	
Fricatives	h			s	
Semivowels			y		

The Mandailing phoneme system differs from the Toba system only in minor details, as follows:

4.1

/c/ is an allophone of /s/ which appears 1) after /n/, e.g. *lancat* /lansat/ [lancat]; 2) as the initial C of a syllable in a word which contains [nc][5], e.g. *bincucur* /binsusur/ [bincucur]. Otherwise /c/ only occurs in loanwords, and in those cases it is not a positional variant of /s/.[6]

4.2

Ma has a palatal N which occurs in initial and medial position. It appears also through nasalisation of /s/, unless the second syllable of the prefixed lexeme also begins with /s/, in which case it is realized as /n/, e.g. surat/mañurat; *sosop/manosop*; *susu/manusu*.

4.3.1

/h/ occurs in only the position /a-a/, e.g. *dahan* 'mushroom'. Sometimes a final /h/ is part of the underlying structure and appears on the surface when followed by a suffix beginning with /a/. Thus *hala* 'scorpion' [ala] has a phonemic structure /kalah/, with final C surfacing in the derived form [parkalahan] (for /k/ see next par.).

4.3.2

There is a /k/ which occurs in initial and final position and as a component of a C cluster. Many lexemes also have initial /k/ as part of the underlying structure; it reappears at the surface level when the lexeme follows another C, either through Sandhi or through prefixation, e.g. *huta* /kuta/ [uta], [morkuta]; *halak* /kalak/ [alak] in the phrase *dibuat halak* [dibuwatkalak].

NB. The initial (underlying) /k/ of a verb of which the next syllable also begins with /k/ is dropped when the verb is used transitively,[7] e.g. *hiskis* [iskis] – (transitive) [maŋiskis] 'to scrape s.th.' – (intransitive) [maŋkiskis]; *hilkil* (ilkil) – (transitive) [maŋilkil] 'to swallow s.th.'; (no intransitive form is given by van der Tuuk or Eggink). This also applies to the verb *hoas* [kowas] (written *hohas* in the Batak script) which has a transitive form [maŋowas]. [kowas] is one of the few inherited lexemes with a surface /k/ in initial position. If we assume a retention of initial /k/ in lexemes with a kVkV(C) prototype, we can explain both the transitive form [maŋowas] and the initial /k/ of its unaffixed form. Similarly we can explain the initial /k/ of other lexemes, e.g. *kehe* [ke] 'to go', from ⁺keke, with contraction of like vowels. This structure is also reflected in the Batak orthography where one reads 'hohas' and 'hehe'.

4.4.

Gemination is not usual in Ma, and [kk] in loanwords is always represented as [ŋk]. /raŋkon/ 'with' is probably a borrowing from To raphon⁺, which is pronounced as [rakkon] in some eastern Toba areas. Two /k/'s following each other through suffixation are usually realized as [ŋk], and sometimes as [k], e.g. the circumfixed form *mandokkon* /maN+dok+kon/, and its pronunciation [mandoŋkon] and [mandokon].

4.5

Final nasals before a vl.S are realized as such, e.g. *teman pargarutan* [temanpargarutan].

4.6

/y/ occurs in intervocalic position, and is distinctive after /a/, /o/, /u/, e.g.

poyop 'curse, malediction'; *hayu* 'tree; wood'; *sayom* 'means to safeguard s.o. from illness or bad influences'; *boyu* 'andung word for *boru*,[8] 'daughter'.

4.7

Morphonology: V contraction occurs, but not in final position: *tu huta* [tuta], but *maridi i* is pronounced [maridii]; nasalisation of /s/ yields /ñ/, except if the next syllable also begins with /s/ (see par. 4.2). See par. 4.4 for the realization of two subsequent /h/'s through suffixation of -/hon/; the V of the suffixes -/mu/ and -/hu/ is always pronounced.

5. *The Angkola phoneme system*

5.0 *Vowels*

high	i	u
mid	e	o
low	a	

Consonants

	glottal	velar	palatal	dental	labial
Stops vd		g	j	d	b
vl		k	(c)	t	p
Nasals		ŋ	ñ	n	m
Liquids vibrant				r	
lateral				l	
Fricatives	(h)			s	
Semivowels			y		

The Angkola phoneme system differs from the Mandailing system only in the following respects:

5.1

/k/ and /h/ are in complementary distribution: /h/ occurs in initial and intervocalic position, and /k/ occurs in final position, as a component of a consonant cluster, and in initial position in lexemes if the next syllable begins with /h/, e.g. *hehe* [Kehe]; *hahaŋ* [kahàŋ].[9]

5.2

Geminated C's occur, (as in Toba), e.g. *pantis* [pattis]; *laŋka* [lakka].

5.3

N is replaced by its homorganic vl.S as first component of a cluster with a vl.S or /s/, e.g. *pantis* and *laŋka* in par. 5.2; /s/ is realized as /c/ as second component in a cluster with a vl.S as first component, e.g. *maŋsi* [makci].[10]

5.4

Sandhi: final N is replaced by its homorganic vl.S before a vl.S, e.g. *teman pargarutan* [tematpargarutan]. After a final vl.S /s/ is realized as /c/, e.g. *halak soŋko ni* [halakcokkoni].

C clusters only occur in medial position and consist of two components (see par. 6.3).

/k/ and /h/ are distinct phonemes, but lexemes with a hVhV(C) pattern hardly occur, and retention of *K in a restricted environment, or regressive dissimilation may explain most lexemes with a kVhV(C) pattern.

For C clusters see par. 6.7.

6. *The Simalungun phoneme system*

6.0 *Vowels*

			diphthongs	
high	i	u		-uy
mid	e	o		-ey
low	a			-ow

Consonants

	glot-tal	velar	pal-atal	den-tal	la-bial
Stops vd		g	j	d	b
vl		k		t	p
Nasals		ŋ		n	m
Liquids vibrant				r	
lateral				l	
Fricatives	h			s	
Semivowels			y		w

6.1 *Diphthongs*

Si has final vowel sequences which can be analyzed as diphthongs, e.g. *horbou* [horbow] 'buffalo', *borei* [borey] 'to give', *babui* [babuy] 'domestic pig'. Tarigan analyzes them as two vowels (which implies an additional syllable to the lexeme), but he writes them as diphthongs. P. Voorhoeve informs me that these endings are diphthongs, and in this paper I will analyze them as such (in Si orthography they are written as vowel sequences).

6.2 *Semivowels*

These do not occur finally and are very rare in initial position. Saragih's dictionary gives only three entries with initial /w/ and none with initial /y/: *wajip* 'duty', a Malay-Arabic loan; *waltoŋ* 'a big wasp', which is also spelled *ualtoŋ*; *wewean* 'implements for making rope'.

6.3 *Consonants*

All C's occur in initial, medial, and final position except /j/, which occurs only initially and intervocalically, e.g. *harad, godog, sabsab, godok, sapsap, liat, hapur, hapal*.

6.4 *Accent*

The accent falls on the penult or on the ultima. When a lexeme is used attributively the accent falls on the ultima; examples: *hàpur* 'chalk' against *hapùr* 'crumbly, brittle, delicate'; *tòbu* 'sugar cane' against *tobù* 'sweet'; *hàpal* 'ship' against *hapàl* 'thick, fat'; *bàyu* 'to braid, to plait', against *bayù* 'new'; the first two minimal pairs are semantically related: chalk is a brittle and delicate substance, and sweetness is the main fea-

ture of sugar.

6.5 *Schematic survey of Simalungun morphonology*

6.5.1 *Nasalisation*

N + /d/, /j/, /l/ yields /nd/, /nj/, /nl/; e.g. *duda/manduda/panduda; jalo/manjalo/panjalo; lapah/manlapah/panlapah;* /nl/ is often realized as [ll], as in *manlapah* and *panlapah* ([mallapah], [pallapah] respectively).

N + /b/ yields /mb/, e.g. *buat/mambuat/pambuat.*

N + /g/, /h/, /V/ yields /ŋg/, /ŋh/, /ŋV/; e.g. *gual/mangual/pangual; hadaŋ/maŋhadaŋ/paŋhadaŋ;* the cluster /ŋg/ is realized as a geminated /k/, which yields the pronunciation [makkadaŋ] and [pakkadaŋ]. *alop/maŋalop/paŋalop; inum/maŋinum/paŋinum.*

N + /s/, /t/ yields /n/; e.g. *suan/manuan/panuan; taŋko/manaŋko/panaŋko.*

N + /p/ yields /m/; e.g. *palu/mamalu/pamalu; poroh/mamoroh/pamoroh.*

N + /n/, /m/, /ŋ/ yields /n/, /m/, /ŋ/; e.g. *nunut/manunut/panunut; muŋkah/mamuŋkah/pamuŋkah; ŋatŋat/maŋatŋat/paŋatŋat.*

N + /k/ yields /ŋ/; e.g. *karat/maŋarat/paŋarat;* Tarigan does not give examples of nasalisation of lexemes with initial /r/.

6.5.2 *Prefixation*

/uN/-; /uN/- + /b/, /m/ yields /umb/, /umm/; e.g. *bahàt/umbàhat; muràh/ummùrah.*

/uN/- + /d/, /j/, /l/ yields /und/, /unj/, /unl/; e.g. *daòh/undaoh; jahàt/unjàhat; lojà/unlòja.*

/uN/- + /g/, /h/, /ŋ/ yields /uŋg/, /uŋh/, /uŋŋ/; *garàŋ/uŋgàraŋ; haŋàr/uŋhàŋar; ŋalùt/uŋŋalùt.*

/uN/- + V yields /umV/ as well as /uŋV/, without difference of meaning; e.g. *idòp/umidop/uŋidop; etèk/umètek/uŋètek; alùs/umàlus/uŋàlus.*

/uN/- + /p/ yields /ump/, e.g. *posò/umpòso.*

/uN/- + /t/ yields /unt/, e.g. *togùh/untòguh.*

/uN/- + /s/ yields /uns/, e.g. *susàh/unsùsah.*

The clusters /mp/, /nt/ and /ns/ are realized as [pp], [tt] and [ss] respectively, which yields the phonetic representations [uppòso], [uttòguh] and [ussùsah] for the last three examples.

Tarigan does not give instances of /uN/- before lexemes with initial /n/ or /r/.

Prefixation of /siN/-; becomes [sim]- when prefixed to a lexeme with initial /b/, e.g. [simbandar]; in all other cases it becomes [sin]-; e.g. [singunuŋmariyah].

6.5.3 *Suffixation*

The suffix -/hon/ becomes -/kon/ after /h/; e.g. *tuduh/tuduhkon; bolah/bolahkon;* in all other cases the suffix remains -/hon/.

-/hu/, the pronominal suffix for the first person singular, becomes -/ku/ after /h/ and -/ŋku/ after V; e.g. *rumah/rumahku; jabu/jabuŋku; aŋgi/aŋgiŋku.* In all other cases the suffix remains -/hu/.

-/ta/, the pronominal suffix for the

first person singular, becomes -/nta/ after a V; e.g. *huda/hudanta*; *toŋgo/toŋgonta*. In all other cases it remains -/ta/.

Suffixation of -/mu/ happens in two ways:
1) the accent falls on the suffix, which is a more polite form, e.g.

bapa-mù	'your father'
lombu-mù	'your buffalo'
inaŋ-mù	'your mother'
tulaŋ-mù	'your uncle'

2) the accent falls on the penultima; the suffix becomes -/m/ after lexemes ending on a V, e.g.

bapàm	'your father'
lombùm	'your buffalo'
inàŋmu	'your mother'
tulàŋmu	'your uncle'

6.6 *Structure of the syllable*

V:		VC:	
ba-ù		*àr-ta*	
à-i		*àn-dar*	
ù-sih		*sù-an*	
		bò-i	

CV:		CVC:	
tò-lu		*hir-tàg*	
hù-da		*tàm-bar*	
bà-li		*tàŋ-kap*	
		pàn-sur	

II COMPARISON OF THE DIALECTS AND RECONSTRUCTION OF THE PROTO-BATAK PHONEME SYSTEM

7. *Vowels*

Vowels in the Northern Batak dialects have long and short variants except for the shwa, which has a long variant /ö/ in Ka, but not in Da.; in Da we find /ō/ as a correspondence to Ka /ö/, as is demonstrated in the following instances: Ka *bön*, Da *bōn*, 'to get dark', Ka *döm*, Da *dōm* 'full, filled'.

7.1

In Neumann's grammar we find the word *ŋalömi* 'to reconcile', and his dictionary gives *aləm*, which corresponds to Da *alum* 'quenched, satisfied, cured', SB *alum* 'satisfied, cooled down'. It is conceivable that Neumann was mistaken in his grammar, and that the correct representation should be *ŋaləmi*, which, like *inəm* 'to drink' has /ə/ instead of /u/ as in all other dialects (Da has *ēnum*[11] and *ēnəm*, without difference of meaning).

It is also conceivable that the shwa in *aləm* became /ö/ because of a shift of the accent through suffixation, but our data do not support such a rule.

Other lexemes like *jarum* have a regular /u/ as a correspondence in Ka, but Balinese and Sasak have *inəm* 'to drink', and Da has both forms; (for /ē/ in the penult. see par. 2.1). Until a better solution is found I reconstruct PB doublets **alem/**alum, and **inem/**inum   (1-2)

The long variants develop either through contraction or through working of the accent (positional alternation: see par. 2.1). Hence vowel length will not play a role in the PB phonemic system as it is not distinctive within an unaffixed NB lexeme.

7.2

Vowel correspondences are not always consistent in the Batak dialects: height harmony, assimilation, metathesis, contraction and the working of analogy sometimes confuse the picture, but nevertheless it is not difficult to reconstruct the following vowels and diphthongs by mutual comparison of the lexemes in the respective dialects: *a, *e, *i, *u, *ow, *uy, *ey.

7.2.1

*a from /a/ in all dialects, e.g. NB, SB, Si *dalan* 'way path; means'; PB **dalan   (3)
SB, Si *agoŋ*, NB *agəŋ* 'charcoal'; PB **agəŋ   (4)
NB *bagas* 'deep; house', Si *bagas* 'deep', AnMa *bagas* 'deep', To *bagàs* 'deep', and *bàgas* 'house'; PB **bagas   (5)

7.2.2

*i from /i/ in all dialects, e.g. all dialects *nipi* 'dream'; PB **nipi   (6)
SB, Si *ribut* 'busy, unquiet, unrest, commotion', Da *rēbut* 'great stir, commotion; in commotion, rebellion', Ka *ribut* 'confused, of thoughts; dissension, quarrel'; PB **ribut   (7)
All dialects *dilat* 'to lick'; PB **dilat   (8)
All dialects *laŋit* 'sky, firmament'; PB **laŋit   (9)
SB, Si, Da *bibir* 'lip'; Ka *bibər* id.; PB **bibir   (10)
Lexemes of (C)iCi(r,l) structure do not occur in Ka except for *iɲjil*, an Arabic loan. Corresponding to this type in other dialects Ka has a (C)iCə(r,l) structure.

7.2.3

*u from /u/ in all dialects, e.g. all dialects *urat* 'root'; PB **urat   (11)
All dialects *tuaŋ* 'to pour'; PB **tuaŋ   (12)
SB, Si *mabuk* 'intoxicated, drowsy, drunken', Ka id., also 'furious', Da 'drunken, intoxicated'; PB **mabuk   (13)

7.2.4

*e from /ə/ in Ka, /ə/ and /o/ in Da, /o/ in Si-SB, e.g. SB-Si *jolma* 'human being', NB *jəlma* id., from Skrt. *janma* (from SB-Si /o/ and NB /ə/ in the penult it is clear that this word came into the Batak dialects through other Indonesian languages, probably Malay or Javanese); PB ++*jelma*   (13+)
SB-Si *onom*, NB *ənəm* 'six'; PB **enem   (14)
SB-Si *tanom*, NB *tanəm* 'to bury'; PB **tanem   (15)
Da does not allow /ə/ before fin. /h/ and in monosyll. lexemes, and has /o/ instead. Only Ka confirms reconstruction of *e in this position, e.g. Ka *tanəh* 'ground, land, earth', Da-Si *tanoh* id., SB *tano* id., PB **taneh (for final /h/ in SB: see par. 9.3)   (16)
Ka *darəh*, Da-Si *daroh*, SB *daro*

'blood'; PB **dareh (17)

Ka bɔiəh, Da bətoh, Si botoh, SB boto 'to know'; PB **beteh (18)

Ka rɔh (/röh/), Da roh 'to come; to cook; rise; coming, cause; onset of an illness', Si roh 'to come; coming', SB ro id.; PB **reh (19)

Ka bən, Da bōn, To bot 'to get dark'; also To bod-ari 'night'; PB **bed (see par. 9.1) (20)

For the correspondence of Si vdS:SB homorganic vl. S:NB homorganic N, see par. 9.0 and for final /h/ in NB-Si see par. 9.3.

7.2.5

*-ow from Si -/ow/: -/o/ in other dialects (see par. 1.1.1), e.g. Si rantou [rantow] 'to go wandering', SB-NB ranto 'to go abroad'; PB **rantow (21)

Si pulou, [pulow] SB-NB pulo 'island'; PB **pulow (22)

Si horbou, [horbow] SB horbo, NB kərbo 'buffalo'; PB **kerbow (see par. 9.3 for the correspondence SB-Si /h/- NB /k/). (23)

7.2.6

*-uy from Si -/uy/, -/i/ or -/e/ in other dialects, e.g. Si apuy, [apuy] NB-SB api 'fire'; PB **apuy (24)

Si babuy, [babuy] 'domestic pig', babi occurs as well, with the general meaning of pig, NB-SB babi 'pig'; PB **babuy (25)

Si laŋuy [laŋuy] other dialects laŋe 'to swim'; PB **laŋuy (26)

I find no explanation for the correspondence Si -/uy/: /e/ in the other dialects.

7.2.7

*-ey from Si -/ey/, -/e/ in other dialects, e.g. Si matey, [matey] NB-SB mate 'to die, to be dead'; PB **matey (27)

Si atey-atey, [ateyatey] NB-SB ate 'liver'; PB **(h)atey (for initial ++(h) see par. 9.3) (28)

Si puney [puney], NB-SB pune 'green wood-pigeon'; PB **puney (29)

The original diphthongs are preserved (orthographically) only in Si. They are written as two vowels but must be interpreted as diphthongs because a) other dialects have one sound for each Si vowel cluster, sharing more features with the first component of the diphthong in the case of -/ow/ and -/ey/ and more with the second component in the case of -/uy/, b) this reconstruction is supported by the diphthongic character of some final /o/'s and /e/'s in Ka (see par. 1.1.1), and c) from comparison with PAN.

7.3

/e/ and /o/ occur in all dialects, but an analysis shows that these V's are not original; in NB they arise:

a) from accented high V's /i/ and /u/, e.g. NB òŋkam 'to open, to break open', Si uŋkab 'to open', To uŋkap 'to open, to get open'; PB **uŋkab (30)
NB òbat, SB-Si ubat 'remedy, medicine'; PB **ubat (see also (31)

12

**inum (2) par. 7.1); in Ka we still find some doublets like marèkat and marikat).

b) through vowel assimilation, e.g. Ka pùsuŋ, Da pòson (here a high back V became /o/ under influence of the accent, and subsequently the following pepet assimilated to /o/ as well), To pusok, Si pusog 'navel'; PB **puseg (32)

c) from the diphthongs *-ow, *-ey and *-uy (see par. 7.2.5-6-7);

d) in Da /o/ developed from *e in certain cases (see par. 7.2.4).

e) through contraction and vowel variation, e.g. the following doublets in Ka: lôn/lawən, dôh/dauh, kêh/kalih, kolan/kolən, kepəŋ/kəpiŋ, jelah-jelah/jilah-jilah, jembar/jambar, əmbut/ombut.

7.3.1

In SB /o/ developed

a) from *e (see par. 7.2.4);

b) from *-ow (see par. 7.2.5);

c) through contraction of *w with *a from the preceding or following syllable (see par. 2.8);

d) from *u in an ultima with final velar or liquid, e.g. SB gaol, NB-Si galuh 'banana, banana trunk' (in SB metathesis of +l and +h took place in some earlier stage, and subsequently +u changed to /o/ and +h was lost (see par. 9.3); PB **galuh (33)
SB hasior 'an evil-smelling plant, with roots used as a medicine', Si hasowor, Da kəcēur, Ka kaciwər id.; PB **kaciur (see par. 8.4 for SB-Si /s/: NB /c/). (34)
In Ka there is a tendency to change /Vu/- and /Vi/- sequences to /Vwə/- and /Vyə/- sequences respectively, e.g. the doublets raut/rawət, saip sayəp, liur/liwər. I can not explain the /w/ in Si.

/e/ developed

a) from the diphthongs *-uy and *-ey (see par. 7.2.6)

b) through vowel harmony: +o and +i often became /e/ under the influence of other V's, e.g. To parumaen /pa/- + /ruma/ + -/on/; To lehon/Ma lehen; To iho=ihe; To baen/Ma baon; To hain/haen. To haen comes from (MI) kain.

c) through vowel assimilation: SB, Si deŋke, NB dəŋke 'fish'. In Si there is also a form doŋkei [doŋkey] which means 'meat'. /e/ in the penult of deŋke developed from +o through regressive assimilation; deŋke in Si is a borrowing from SB; PB **deŋkey (35)

7.3.2

In Si /o/ developed

a) from *e (see par. 7.2.4);

b) through contraction, e.g. in gor-an, To goar, AnMa goar/gorar, Da gərar,

Ka *gelar*, 'name'; PB **gelar (36)
In To and Si the medial liquid was
lost under influence of the final
/r/ (apart from AnMa *gorar* no
lexemes of (C)o(r,l)ar-structure
occur in SB and Si).
Si, AnMa, NB *sora*, To *suara*, *soara*
'sound, voice, tone': from Skrt.
svara; PB [+][+]*suara* ([+]36)

c) through vowel variation, e.g. *deos/dias*,
son/sen;

d) in borrowings, e.g. *sekon* 'second' *sikolah*
'school', *mokah* 'Mekkah';

e) as component of a diphthong, see par.
7.2.5;

/e/ developed

a) through contraction, e.g. the contracted
forms *kahean/kahen* from *kahoyan*, *mela* from
maila (both *maila* and *kahoyan* exist togeth-
er with the contracted forms in Si)

b) through vowel variation, e.g. *derder/*
dardar; *darsat/derset*; *jelen/julin*;

c) in borrowings, e.g. *denke* (see par.
7.3.2) and Si *bere*, 'to thank (lit.
to give thanks)'; and the inherited
form is *borei* 'to give', To *i-be/bere*
'a man's sister's child, who is
to be given in marriage to another
lineage-group', according to Batak
custom; AnMa *bore* in *pa/bore/an* 'a
start, a make-weight, an extra' Da
bere 'to give', Ka *bere* 'to give,
allow'; PB **berey (37)

8 *Semivowels*

These do not occur in To, and /w/ is
found only in NB and Si.

8.1

In To [+]y is contracted with a pre-
ceding or following V, and is lost be-
tween /a/ and /u/. In other dialects
there is mutual correspondence of the
respective /y/'s, so that we may recon-
struct *y, e.g. To *hau*, Si-AnMa *hayu*,
NB *kayu* 'tree; wood'; PB **kayu (38)
(see also par. 9.3)
To *saur* in *saur matua* 'to become
old', Si-AnMa *sayur matua*, Da *sayur-ntua*
'to become old'; PB **sayur (39)
To *saop* 'to hold, detain, to lift
up' Da *soap* 'to put away, to store';
PB ** *(a)yap*; see par. 9.4.2 for /n/ (40)
in Da; van der Tuuk also mentions Si
saop, a form which is not given by
Saragih.
To *bean* 'block'; other dialects
iyan, id.; PB **bayan (41)
To *beom* 'to hold in one's mouth for
a long while, e.g. a chew of betel nut'
AnMa *boyom* NB *bonem* 'to stew s.th. in a
leaf, e.g. fish' Si *boyom* 'to eat up,
finish one's eating'; PB ***be(N)yem.* (42)
To *tiun*, SiAnMa *tuyun*, 'Solanum
Melongena'; PB [+][+]*tuyun* ([+]42)
In SB there is a doublet *torun*
with a cognate *terun* in NB, both with
the meaning 'Solanum Melongena
(aubergine, eggplant)'[12]; PB **terun (43)

To *laen* in *tuan laen* 'favourite wife';
van der Tuuk also gives An *laen* and Ma *layan*,
but Eggink's dictionary only gives *laen*. Ac-
cording to v.d.T. To *aek* 'water' has a cognate
ayok in Si, but this word is not given in Sa-
ragih's dictionary, where we find *bah* 'water'.

8.2

/w/ occurs in intervocalic position
in NB and Si. The correspondence in SB
is usually ∅, and in a few cases To /b/
between /a/'s and AnMa /o/ after /a/
(with contraction of following a[+])[13].
If we assume that To /b/ represents an
original *b in the sequence *aba that
merged with [+]w in NB and Si and con-
tracted to /o/ with following a[+] in AnMa,
(and that *w was lost in SB), we are
able to reconstruct *b from To /b/,
NB-Si /w/ and AnMa contracted /o/, and
*w from Si /w/, NB /w/ and SB ∅, e.g. To
haban, Si, Da *kawan* 'cattle;
slave', and AnMa *haon*, *haban* (probably
a To loan) 'herd of white boars, of ele-
phants'; (see also par. 9.3); PB **kaban (44)
To *tabar*, AnMa *taor*, NB-Si *tawar*
'magic remedy, antidote'; PB **tabar (45)
To *ak* (van der Tuuk gives also
aoak[14]), Da *awak* 'waist'; PB **awak (46)
To *dan* (van der Tuuk also: *daoan*[14]
and *daban*), AnMa *dan*, *dahan* (van der
Tuuk) NB *dawan* 'mushroom'; PB **dawan (47)
To *ait*, NB *awin* 'to draw tight, to
draw closer to o.s.'; (see also par.
9.0); PB **awid (48)
To *lai-lai*, Da *lawi* 'the long tail-
feathers of a bird'; PB **lawi (49)
To *sait* 'eye-tooth of humans and
animals', Da *sawit* 'the tusks of a wild
pig'; PB **sawit (50)

8.2.2

Some — mostly trisyllabic — lex-
emes have initial /u/ followed by a V.
Such lexemes occur in all dialects and
their initial /u/ corresponds to PAN
*w. If a cognate shows an initial /u/
before V, I reconstruct PB *w; e.g. Ka
uari, SB, Si, Da *ari* 'day'; NB still has
a semivowel in the compound form *matawari*
'sun'; PB **wari (51)
NB-Si *ualuh*, SB *ualu* 'eight';
PB **waluh (52)
To *alton* 'a spider', AnMa *alton*, Da
ualten, Si *ualton/walton* 'a big and
fiercely stinging wasp'; PB **walten (53)
There is also NB *uili*, SB *aili* 'wild
pig', but I cannot explain initial /a/ in the
SB reflex.

9 *Consonants*

Except for the palatal stop, vl.
stops occur in all dialects in initial,
medial and final position. Vd. stops
occur in all dialects in initial and me-
dial position, and /b/, /d/ and /g/ oc-
cur also in final position in Si. Final
vd. stops in Si correspond to homorganic
nasals in NB and homorganic vl. stops in
SB, e.g. Si *dolog*, SB *dolok*, NB *dolen*
'mountain'; PB **deleg (54)

Si *abab*, SB *abap*, NB *abam* 'ashes';
PB **abab** (55)

Si *saŋkob*, SB *saŋkop*, 'top, cover',
NB *saŋkəm* 'earthenware top of a vessel
used for extinguishing embers';
PB **saŋkeb** (56)

Some SB reflexes of word-compounds
(petrified forms) preserve original
voiced final stops, e.g. To *bot* 'to get
dark', but *bodari* 'night; at night',
from **bed** plus **wari** (see par. 7.0 and
7.2.4); AnMa-Si *sagadon* 'a large piece
of fire-wood', NB *sagan* 'to light,
ignite'. Oddly enough To has only a form
with a homorganic nasal as final C:
sagan 'to be burning (of fire or light),
and with a suffix: *saganon* 'a big piece
of fire-wood'; PB **sagad, sagad-en** (57)

From To *bodari* we know that initial **w**
must have been in a weak position before the
change of final vd.S to vl.S. For the corre-
spondence of Si vd.S: SB homorganic vl.S: NB
homorganic N (in final position) I reconstruct
a homorganic vd.S in PB, because a) a transi-
tion from a vd.S to the present-day corre-
spondences is the most natural alternative,
b) some SB lexemes still represent a vd.S in
petrified forms, c) these correspondences
agree with a vd.S in PAN.

9.1

/b/, /d/, /g/ and /j/ occur in all
dialects in non-final position, where
they correspond to a similar phoneme in
the other dialects, e.g. SB *bagi*, Si-NB
bagi 'part, share'; PB **bagi** (58)

SB *guliŋ*, 'to lean over, incline,
to descend, go down', Si *guliŋ* 'to roll,
to revolve', Ka *guliŋ* 'to fall, to come
down (of s.th. that stands)'; PB **guliŋ** (59)

All dialects *jadi* 'to become; to
be realized'; PB **jadi** (60)

All dialects *ajar* 'lesson, instruc-
tion'; PB **ajar** (61)

See also (3), (5), (7), (8), (10), (13),
(+13), (17), (18), (20), (23), (25), (31),
(35), (36), (37), (41), (42), (47), (54), (55),
(57). Hence follows the reconstruction of
b, **d**, **g**, **j**.

9.2

The vl. stops /p/ and /t/ occur in
all dialects in all positions, e.g. SB
tolu, Si *tolu*, NB *təlu*, 'three';
PB **telu** (62)

To, Si, Ka *tiga* 'market', AnMa, Da
'commodity, merchandise'; PB **tiga** (63)

SB, Si, Da *urup* 'to help, assist',
Ka *urup* 'to do s.th. together (joint ac-
tion), to assist, to stand by';
PB **urup** (64)

AnMa, NB *pake* 'to wear, use, of
clothes', To *pahe*, Si *pakei, pahei*
([pakey], [pahey]) id.; /k/ in AnMa
pake and Si *pakei* is unexplained (see
par. 9.3); PB **pakey** (65)

SB, NB, Si *lupa* 'to forget';
PB **lupa** (66)

See also (6), (7), (8), (9), (11), (12),
(13), (16), (18), (21), (22), (24), (27), (28),
(29), (30), (31), (32), (35), (40), (43), (45),
(46), (50), (53). I reconstruct *p and *t.

9.3

/k/ occurs in all dialects, in NB and Si
in all positions, but in To only after a C and
as a final C of a syllable, and in AnMa only
in initial position, after a C and as final C
of a syllable. To and Ma [k] is in comple-
mentary distribution with /h/ (Ma [ø]) (see
par. 3.2, 4.3.2 and 5). /h/ occurs as a pho-
neme distinctive from /k/ in NB and Si. In To
and An /h/ is an allophone of /k/. Corre-
sponding to NB /h/ we find ø-, -ø-, -ø in SB,
and ø-, -ø-, -/h/ in Si. As NB /h-/ is in a
week position and is sometimes lost, I recon-
struct *h in PB reconstructions which have NB
correspondences with initial /h-/ (sometimes
alternating with ø-), and *(h)- in PB recon-
structions with a corresponding PAN form that
reflects NB /h-/ (=PAN *q); cf. **(h)atey** (28)
with *qaCey, and **(h)ukur** (94) with *qukur.
There is a regular correspondence of NB /k/
with To, An /h⌣k/, Ma /k⌣ø/, and Si /h⌣k/.
In inherited Si lexemes /k/, /h-/ and -/h/-
which agree with NB /k/ occur in conditioned
alternation: /h/ occurs in initial and in in-
tervocalic position, /k/ occurs as final C in
a syllable and in the synchronic patterns
kVhV(C), kV(w,y)Vh, and (C)VkVh.[15] For this
correspondence I reconstruct *k; examples:

SB, Si *hutu* (Ma [utu]), NB *kutu*
'louse'; PB **kutu** (67)

SB, Si *halak* (Ma [alak]), NB *kalak*
'human being'; PB **kalak** (68)

SB, Si *horas* (Ma [oras]), Da *kəras*
'hard, strong' unharmed, healthy', Ka
kəras 'hard, of cooked rice'; PB **keras** (69)

SB, Si *hapur* (Ma [apur]), NB *kapur*
'chalk'; PB **kapur** (70)

SB, Si *ahu* (Ma [au]), NB *aku*,
'first person singular'; PB **aku** (71)

SB *tabo*, Da, Si *taboh*, Ka *tabəh*
'savory, tasty, good'; PB **tabeh** (72)

SB, Si *duhut* (Ma [dut]), NB
dukut 'grass, weed'; PB **dukut** (73)

SB *poro*, Si *poroh*, Da *pəroh*, Ka
pərəh 'to squeeze out, to wring out';
PB **pereh** (74)

The next example shows retention
of Si -/h/ from PB *-h in an irrefuta-
ble way: SB *si-baso*, Si *si-basoh*, Ka
si-baso 'intermediary at the descending
of a spirit; medium', Da *si-baso* 'fe-
male priest, midwife and medium'. The
NB forms are loans from SB hence it is
very unlikely that Si *si-basoh* is bor-
rowed from the BN dialects; PB
++*si-ba(cs)eh* (+74)

9.3.1

Voorhoeve (1955:9) considers Si as re-
presenting one of the three main branches of
the Batak dialects (see Introduction) having
some features in common with the NB group on
the one hand, and with the SB group on the
other. However, agreement with the NB dia-
lects is due to borrowing, and has no bear-
ing on the original Si idiom. I prefer to
group Si with SB, and to draw, in this way, a
basic dichotomy of Batak dialects into NB on
the one hand, and SB plus Si on the other, on
the following grounds:

a) Si agrees with SB in its change of *k to
   /h/ in initial and medial position and in

its loss of *h, except in final position, where it is retained in Si.

b) Si initial and intervocalic /k/ occurs only in loanwords or in lexemes where the next C is /h/.

c) PB *c became /s/ in both SB and Si (see par. 9.4.1)

d) PB *e became /o/ in both SB and Si (see par. 7.2.4)

e) gemination and the working of sandhi occurs in both Si and SB; this argument however is countered by the fact that here Ma does not follow the other dialects (par. 4.4.5).

f) distinctive accent is a common innovation in SB and Si; this is shown by the fact that the accent falls as well on /o/ corresponding to PB *e as on the other vowels, whereas in NB, as in the other Western Indonesian languages the accent as a rule never falls on a shwa, except when the shwa is the only vowel ocurring in a word (see also par. 1.4, 2.6 and 10).

g) this hypothesis is supported by the lexical material: when we look at the lexemes that are less under suspicion of being borrowings (as for instance those ending in a diphthong or a vd.S) we more often find cognates in SB than in NB. Also, the initial and medial /h/ of their SB cognates regularly correspond with /h/ in Si. (see (13), (16), (17), (18), (23), (28), (30), (33), (34), (38), (44), (46), (52), (56), (65).

9.4

Palatals occur in all dialects, /j/ everywhere, /c/ in NB and AnMa, /ñ/ in AnMa; they occur in all dialects in initial and medial position; for the treatment of vd.S /j/ see par. 9.1;

9.4.1

/c/ in AnMa is secondary: see par. 4.1 and 5.3; many NB lexemes with /c/ have a doublet with /s/ or /t/, and lexemes without such doublets often seem to be borrowed from Malay, e.g. *Ka kəcəmbul/kətəmbul, campah/sampah, cəpah/səpah, cihər/sihər, cinəmbu/sinəmbu*; Malay borrowings: *cakap* 'talk, language'; *cap* 'seal, stamp'; *Da ciŋgələm/tiŋgələm, cələbən/sələbən, cilaka/tilaka, cituk/situk, concaŋ/soŋsaŋ, cidahari/sidahari*; Malay borrowings: *cap* 'stamp, seal'; *cakap* 'language, talk'; probably from Ml: *candu* 'prepared opium'; *cərmin* 'mirror'. The Ka borrowing *minak capi* 'beef suet' (from Ml *minyak sapi*) points to the possibility that /c/ is regarded as a foreign sound by Karo speakers, and that hypercorrection might have taken place in other loanwords with +s as well. Si, SB /t/ corresponds to NB /c/ before /i/, e.g.:

SB, Si *hutiŋ*, Ka *kuciŋ*, Da *kōciŋ* 'cat'; PB **kutiŋ* (75)

SB *tiop*, Ka *cikəp*, Da *cēkəp* 'to hold s.th. (in the hand)'; PB **tikep* (76)

For this correspondence I reconstruct *t because 1) the regular correspondence of NB /c/ is /s/ in SB, Si, and 2) because a development from +t to /c/ is much more common

than one from +c to /t/.

On the other hand borrowings from Sanskrit have retained their palatal vl.S in NB, whereas this sound merged with /s/ in SB and Si. There is good reason to suppose a direct influence of Sanskrit on the Batak dialects, rather than through Malay or Javanese[16]; moreover, there are, along with instances of a correspondence To, Si /s/ ∿ AnMa /ñ/ ∿ NB /c/ no instances of a correspondence To, Si /s/ ∿ AnMa /ñ/ ∿ NB /s/ in lexemes without initial nasal substitution. This suggests the possibility of an original AnMa +c with nasal substitute +ñ (for /ñ/ see also next par.). Within the scope of this paper I cannot find sufficient evidence for the treatment of NB /c/ as an innovation, and will reconstruct *c for the regular correspondence To, Si /s/ ∿ AnMa /s∿c/ ∿ NB /c/. I reconstruct PB *(N)c for the correspondence To, Si /s/ ∿ AnMa /ñ/ ∿ NB /c/, because 1) other dialects show no nasal feature for this correspondence; 2) /ñ/ is secondary in AnMa (see next par.), e.g. NB *picat* 'to squeeze s.th., to massage breasts', To, Si *pisat* 'to pinch, to squeeze, to push', AnMa *piñat* 'pressed, pushed in'; PB **pi(N)cat* (77)

To, Si *lisat* 'crushed, squeezed flat', Da *licat*, AnMa *liñat* 'squeezed flat'; PB ***li(N)cat* (78)

To *laso*, NB *laksa*, AnMa *laño* 'ten thousand' is a Sanskrit loanword. The NB form seems to be taken directly from Skrt, while the SB forms are from Minangkabau, as witnessed by their final -/o/. In addition To has another form *loksa* (same meaning), which probably comes from MI or Jav *ləksa*; PB +*la(N)co*, +*laksa*/+*loksa* (+77 - +78)

To *haso*, Ka *kaco*, AnMa *kaño* 'to disturb, to disarrange, make a mess' (I do not have an explanation for AnMa /k/ here); PB ***ka(N)cow* (79)

To *usop* 'to get wet; to put s.th. into water', Si *unob*, AnMa *uñop* 'to sink, to make water, to swallow water'; PB ***ceb*, ***u(N)ceb*, ***c-el-eb* (80)

Ka *ciŋəm*, AnMa *hiñom* 'tightened, of lips'; one of these forms shows metathesis; PB *ci(n)kem/ki(n)cem* I reconstruct (81) *c for the correspondence NB /c/ ∿ Si, SB /s/, e.g. SB, Si *rasun*, NB *racun* 'poison'; PB *racun* (82)

To *hunsi* 'key', Ka, AnMa *kunci* 'lock; to be locked'. These forms are Hindi loanwords; the AnMa /k/- may be explained as a borrowing from MI; PB +*kunci* (+82)

To *suŋkalit* 'difficult, troublesome', AnMa *suŋkalit/kuñalit* 'knotty (wood)' (no Si or NB cognates); PB +*cuŋkalit* (+83)

SB *sampur* 'to mix up; to unite in marriage', Si *sampur*, Da *campur* 'to mix up', Ka *campur* 'to have an intimate relationship, of girls and boys; to mix up'; PB *campur* (83)

Si, SB *pusuk* 'the inner leaf of a palm tree, the young shoot of a tree or plant', NB *pucuk* 'young leaves of a tree or plant, sprout'; PB *pucuk* (84)

9.4.2

AnMa /ñ/ has /y/ as a variant in many lexemes. It also occurs as a nasal substitute of /s/, and corresponds with /n/, /s/, /ns/ or ∅ (see par. 8.1) in To, with /n/, /s/, /ns/ or /y/ in Si, and with /n/, /c/, /nc/ and /y/ in NB. A fuller examination of cognates was made in To because 1) this dialect is more closely related to An and Ma, 2) this was made easier by a better supply of material. Except for nasal substitution of initial /s/ in verbs, it is not predictable which correspondence for /ñ/ other verbs will have. Van der Tuuk's hypothesis of an older +ñ that was maintained in AnMa but changed elsewhere either to /n/ or to /y/ (∅ in To), does not hold because a) it is not predictable when /n/ or /y/ will occur in another dialect, b) AnMa sometimes has /n/ or /y/ (instead of +/ñ/) where other dialects have /y/ (To ∅), and c) the palatal N in PAN lexemes hardly ever corresponds to a palatal N in AnMa lexemes, and when it does, the AnMa forms are usually suspect loans, e.g. *ñañi* ('to sing'; Ml *nyanyi*) and *ñala* ('flame'; Ml *nyala*). On the other hand AnMa *ponuponu* 'turtle' has no palatal N where PAN *peñu* has; with To *ponu*, id.; PB +*penu*   (+84)

In a personal communication Anceaux has suggested that nasalisation of intervocalic /y/, like that in Ngaju-Dayak lexemes prefixed with /maN/-, could have been active in a former stage of the Batak dialects. This suggestion is supported by the development of AnMa *mañabia/manabia* 'Northwest', which comes from Skrt vāyavya, and is reflected in the other dialects as To, Da *manabia*, Ka *maŋabia* (the velar N is peculiar here); here the initial C's of the first two syllables are both nasalized; PB +*ma(N)yabia*   (+85)

The AnMa forms with palatal N do not give enough evidence for a PB palatal N, and I reconstruct *(N)y for the correspondence To ∅ ∼ AnMa /ñ/ ∼ Si, NB /y/, and *(N)c for the correspondence To, Si /s/ ∼ AnMa /ñ/ ∼ NB /c/ (see par. 9.4.1), e.g.: To *heat*, AnMa *hoñat* 'stag-beetle, Ka *kayat* 'beetle' (with an unexplained /a/ in the penultima); PB **keyat/ **kayat   (86)

To *eo*, AnMa *oño/oyo*, Si *oyoh*, Da *(c-)ioh*, Ka *c-iah* (/a/ unexplained) 'urine'; PB **eyeh   (87)

Examples of doublets with /y/ in AnMa: *hañal/hayal*, *oño/oyo*, *huñom/ huyom*, *oñaŋ/oyaŋ*, *pañogon/payogon*, *huña-huña/huya-huya*, the suffix -/muñu/ / -/muyu/. There is also a doublet *uñam/unjam* 'to found a village' that does not have doublets in the other dialects: To has *unjam* 'to seek shelter with someone', but the difference of meaning is considerable.

There is another doublet with difference of meaning: AnMa *muñi* 'k.o. big rat' and *monsi* 'mouse' with cognates To, Si *monsi*, NB *mənci* 'mouse'; PB **menci   (88)

For examples of *(N)y and *(N)c see (34) and (42).

9.5

The liquids /r/ and /l/ occur in all dialects in initial, medial and final position.

Except for Ka and Si, in no dialect do two different liquids ever occur within the same word, and in the process of affixation and word-composition regressive assimilation takes place; see par. 2.3, 3.2, 3.3.5 and 3.5.2. On the base of correspondences between liquids of different dialects we can reconstruct *r and *l, and in reconstructing PB lexemes with different liquids the Ka and Si cognates are criterial for the establishment of each proto-phoneme; e.g. SB *toru*, Si *toruh*, NB *təruh* 'under(neath)'; PB **teruh   (89)

SB *torop*, NB *tərəm* 'number, a lot, many; multitude', Si *torob* 'average, sufficient, mediocre (Ml *sedang*)'; PB **tereb   (90)

SB *sorot* (van der Tuuk) 'to sting, of insects', Si *sorod*, Ka *sərən* 'sting'; PB **sered   (91)

SB, Si *roŋit*, NB *rəŋit* 'mosquito'; PB **reŋit   (92)

SB, Si *ruŋruŋ* 'to loosen, to disengage; to make everything ready', Ka *ruŋruŋ* 'to pour out s.th.', Da *ruŋruŋ golaŋ-golaŋ* 'morning between about 8 and 10 a.m.'; PB **ruŋruŋ   (93)

To *uhur* 'to consider, deliberate, when in gambling the dice land edge-on against some object', NB *ukur* 'thought, opinion, mind', Si *uhur* 'heart', and *maŋuhuri* 'to deceive' (for (h)- see par. 9.3) **(h)ukur   (94)

AnMa *aŋkar*, Si *m-aŋkar* 'half-done, half-cooked, not yet ripe', Ka *aŋkar* 'to restrain o.s., to be moderate, keep within bounds'; PB **aŋkar   (95)

SB, Si *borŋi-n*, Da *bərŋi-n*, Ka *bərŋi* 'night'; PB **berŋi   (96)

(all dialects) *lima* 'five'; PB **lima   (97)

SB *laŋka*, Si, Da *laŋkah* 'step, pace, stride' Ka *laŋkah* 'behavior, way of acting'; PB **laŋkah   (98)

(all dialects) *lubaŋ* 'pit, hole; full of holes'; PB **lubaŋ   (99)

SB *lilit* 'circumference, volume, range', Si *lilit* 'to wrap around, to wind around', NB *lilit* 'the winding, wrapping up of s.th.'; PB **lilit   (100)

SB *cilat*, Si *silad*, NB *cilan* 'to wipe after defecating'; PB **cilad (101)

To *tola*, NB *təlah* 'to settle a dispute, to part combatants'; PB **telah   (102)

To *arut/alut*, AnMa *arut*, Si *alud* NB *alun* 'to massage'; PB **alud   (103)

Ka, Si *salsal* 'to shine (up)on; shining, glow', To *salsal* 'clearly visible, clear'; PB **salsal   (104)

SB, Si *poŋkal* 'hoops around a

fish-trap (*bubu*) or a barrel'; Ka
pəŋkal 'bridle of a horse; PB **peŋkal*   (105)

   SB, Da *rapar*, Ka, Si *lapar* 'to be
in want, to starve'; PB **lapar*   (106)

   SB, Da *ruar*, Ka, Si *luar* 'to come
out of'; PB **luar*   (107)

   SB *borur*, Da *bərur*, Ka *bəlur* 'to
swallow without chewing', Si *bolur* 'to
swallow'; PB **belur*   (108)

9.6

   Except for /ñ/ (which is treated
with the palatals in par. 9.4) nasals
occur in all dialects in initial, me-
dial and final position. On the basis
of their correspondences I reconstruct
*m, *n and *ŋ, e.g. (all dialects)
mata 'eye'; PB **mata*   (109)

   (all dialects) *manuk* 'hen, fowl';
PB **manuk*   (110)

   SB, *dalom* in '*di dalom laut*'; 'to
be in the womb, to be pregnant';
PB **dalem*   (111)

   SB *maos*, Ka, Si *mawas* 'orang
utan'; PB **mawas*   (112)

   (all dialects) *mabar* 'k.o. large
rubber plant'; PB **mabar*   (113)

   NB, To, Si *juma* 'rice-field (usual-
ly dry); To and Si also have *uma*, and
Ka has *(h)uma*. I consider *juma* as the
outcome of a former prefixed form
+di-uma. It may have been borrowed from
a southern dialect into the northern
dialects because NB *juma* shows no trace
of an initial /h/; PB **huma/**juma*   (114)

   (all dialects) *samun* 'to rob';
PB **samun*   (115)

   SB, Si *golom*, NB *gələm* 'to seize,
take, to grasp with the hand', PB **gelem*   (116)

   (all dialects) *ni*, genitive
marker; PB **ni*   (117)

   SB, Si, Ka *nunut* 'lasting, conti-
nuous'; PB **nunut*   (118)

   To *pano-pano* 'freckles', AnMa, Ka
pano, Si *panou* [panow], Da *pano-pano*
'white spots on the skin'; PB **panow*   (119)

   SB *boni*, Si *bonih*, NB *bənih* 'seed';
PB **benih*   (120)

   SB, Si *dahan*, NB *dakan* 'to cook
rice'; PB **dakan*   (121)

   (all dialects) *papan* 'board';
PB **papan*   (122)

   (all dialects) *ŋatŋat* 'to gnaw,
to chew, to nibble'; PB **ŋatŋat*   (123)

   SB, Ka *ŋitŋit* 'moth'; PB **ŋitŋit*   (124)

   SB, NB *aŋin* 'wind'; PB **aŋin*   (125)

   To, *eaŋ* 'illness that causes heavy
feet' Ka *hayaŋ* 'heavy feeling, as after
an attact of fever'; PB **hayaŋ*   (126)

   (all dialects) *timbaŋ* 'to weigh,
to consider'; PB **timbaŋ*   (127)

   SB, Si *bolgaŋ*, NB *bəlgaŋ* 'to cook';
PB **belgaŋ*   (128)

9.7

   The /s/ occurs in all dialects in
initial, medial and final position, and
on the basis of a regular correspond-
ence of this phoneme between the dia-
lects I reconstruct *s; e.g. SB, Si

sahit, NB *sakit* 'illness'; PB **sakit*   (129)

   SB *sia*, Si *siah*, NB *siwah* 'nine';
PB **siwah*   (130)

   SB *sombu*, NB *səmbuh* 'appeased,
satisfied, quenched (thirst), vented
(hatred)'; PB **sembuh*   (131)

   SB *sulu*, NB, Si *suluh* 'torch';
PB **suluh*   (132)

   SB, Ka, Si *surut* 'to go back, to
retreat' Da *surut* 'less; to go back, to
retreat'; PB **surut*   (133)

   SB *tasak* 'finished, prepared (said
of a *luluk*, a k.o. mushroom)', Si *tasak*
'dry; done, ripe', NB *tasak* 'ripe,
done'; PB **tasak*   (134)

   SB *mosok*, Si *mosog*, NB *məsəŋ* 'to
burn, be burnt'; PB **meseg*   (135)

   (all dialects) *asar* 'bird's nest';
PB **asar*   (136)

   See **bagas (5)· and **mawas (112) for /s/
in final position.

10   *Accent*

   In the southern dialects and in Simalu-
ngun the accent is distinctive and its posi-
tion is syntactically determined (see par. 3.4
and 6.3). Polymorphemic words have the accent
on the penultima, irrespective of their syn-
tactical position (except in cases treated in
par. 3.4.2 and 6.4.3). In the Northern dia-
lects accent is not distinctive and has no
syntactic function: it falls on a shwa only in
lexemes with no other vowels than shwa (see
par. 2.6).

   Although clear insight into the origin of
SB and Si accent requires a more thorough in-
vestigation of SB and Si morphology and lexi-
con than is possible within the scope of this
paper, I draw the provisional conclusion here
that the function of accent in SB and Si is an
innovation. This conclusion is based on the
following considerations: a) As was apparently
the case in PAN, accent is not distinctive in
NB; b) the place of accent in SB and Si does
not correlate with the place of /o/ within a
lexeme, whereas in NB the place of accent does
correlate with the place of /ə/.

   Therefore the distinctive accent of SB
and Si was not part of the PB phoneme system.

11   *Schematic representation of the mutual*
relationship of Batak dialects

   The following scheme shows the mutual re-
lationship of Batak dialects according to the
subgrouping on phonological grounds (see par.
9.3.1). I propose the terms

1) Proto-Northern-Batak (PNB) for the proto-
   language of Karo, (Alas) and Dairi;
2) Proto-Southern-Batak (PSB) for the proto-
   language of Toba, Mandailing, Angkola, and
   Simalungun;
3) Proto-Toba (PT) for the proto-language of
   Toba, Angkola, and Mandailing;
4) Proto-Angkola-Mandailing (PAM) for the
   proto-language of Angkola and Mandailing.

Proto-Batak tree: Proto-Batak → PNB, PSB; PSB → PT; PT → PAM. Leaves: (Alas?) Ka Da Si To An Ma
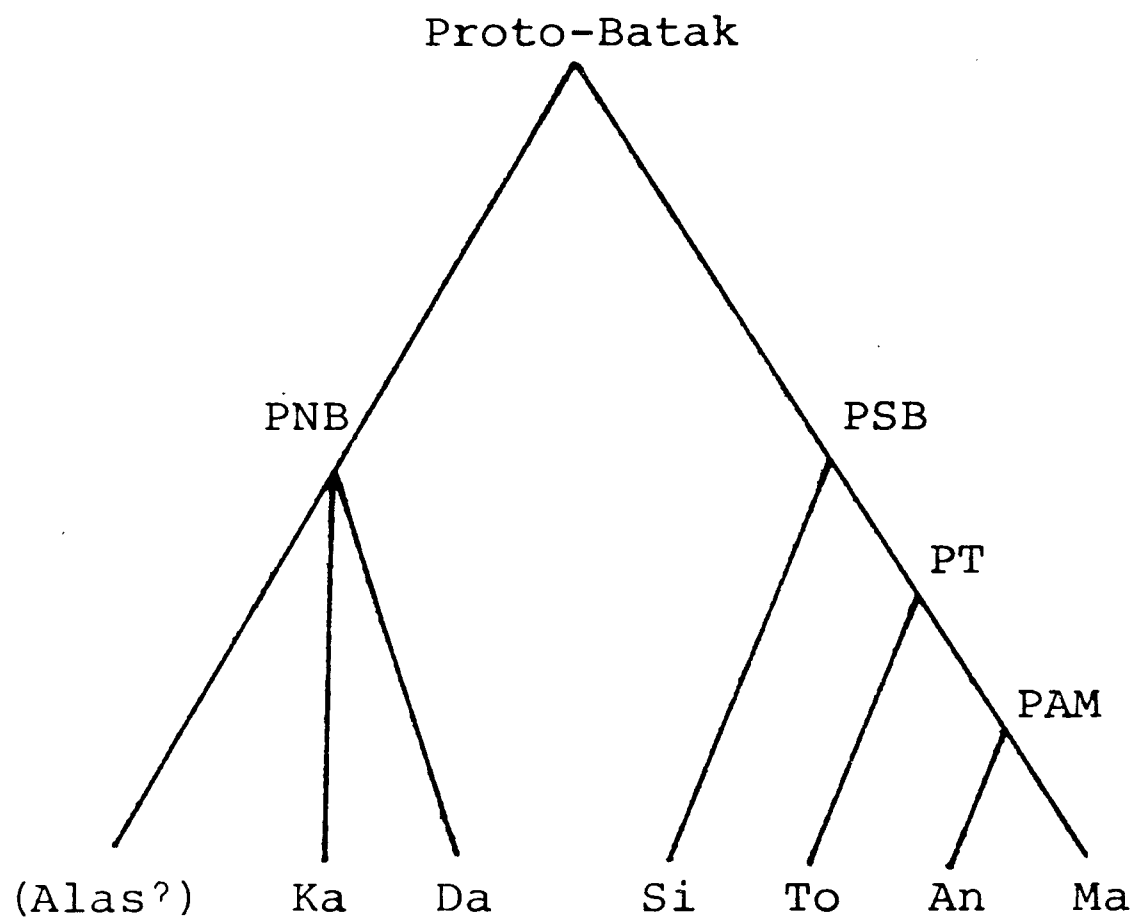
III COMPARISON OF PROTO-BATAK WITH PROTO-AUSTRONESIAN

12 *Proto-Austronesian phoneme system*[17]

Consonants

	uvu-lar	velar	pal-atal	alve-olar	retro-flex	bila-bial
Stops vd		g	j/z/Z	d	D	b
vl		k	s/c	C,t	T	p
Nasals		ŋ	ñ	N,n		m
Liquids						
vibrant	R			r		
lateral				l		
Semivowels			y			w/W
Fricatives			x	S		
laryngeals	q,H,ʔ					

Vowels

high	i	u	diphthongs:	-uy	-iw
mid		e		-ey	-ew
low		a		-ay	-aw

13 *Proto-Batak phoneme system*[18]

Consonants

	glot-tal	velar	pal-atal	dental	la-bial
Stops vd		g(g,j)	j(z)	d(d,D,Z)	b(b)
vl		k(k)	c(C)	t(t,T,C)	p(p)
Nasals		ŋ(ŋ)		n(n,ñ,N)	m(m)
Liquids					
vibrant				r(r,R)	
lateral				l(l)	
Semivowels			y(y)		w(w)
Fricatives	h(q)			s(s)	

Vowels

high	i (i, -iw)	u (u)	diphthongs:	-uy (-uy)
mid		e (e)		-ow (-aw, -ew)
low		a (a)		-ey (-ay, -ey)

14 *Merger of phonemes in Proto-Batak*

The following PAN phonemes merged in PB:

1) *i and *-iw to PB *-i, e.g.
   *zadi   >   **jadi   (60)
   *baRiw   >   **bari,
   (Da b̄āri, AnMa bari 'rotten, musty',
   To bari 'gone bad, spoiled (food)'
   Ka m-bari 'to be cold, of a corpse'.

2) *-aw and *-ew to PB *-ow, e.g.
   *ke(re)baw   >   **kerbow   (23)
   *burew   >   **burow,
   (Ka muro (buro), Da mə-muro-i
   (buro), To, AnMa ma-muro (buro)
   'to chase away birds').
   *panaw   >   **panow   (119)

3) *-ay and *-ey to PB *-ey, e.g.
   *pakay   >   **pakey   (65)
   *punay   >   **puney   (29)
   *qatey   >   **(h)atey   (28)
   *macey   >   **matey   (27)

4) *t, *T and *C to PB *t, e.g.
   *maCa   >   **mata   (109)
   *uRat   >   **urat   (11)
   *telu   >   **telu   (62)
   *laŋiC   >   **laŋit   (9)
   *TukTuk   >   **tuktuk,
   (NB tuktuk 'to knock on s.th.', SB,
   Si tuktuk 'beak; to knock on s.th.').

5) *r and *R to PB *r, e.g.
   *ke(re)baw   >   **kerbow   (23)
   *rantaw   >   **rantow   (21)
   *waRi   >   **wari   (51)
   *luqaR   >   **luar   (107)

6) *d, *D and *Z to PB *d, e.g.
   *Zalan   >   **dalan   (3)
   *DaReQ   >   **dareh   (17)
   *sered   >   **sered   (91)
   *Zilat   >   **dilat   (8)
   *zadi   >   **jadi   (60)

7) *n, *ñ and *N to PB *n, e.g.
   *taNem   >   **tanem   (15)
   *(n)[ui]pi   >   **nipi   (6)
   *peñu   >   (++penu)   (+84)
   *punay   >   **puney   (29)
   *bulaN   >   **bulan,
   ((all dialects) bulan 'moon,
   month').

The following phonemes were lost:
*ʔ, e.g.   *ʔenem   >   **enem   (14)
*x, e.g.   *xapuy   >   **apuy   (24)
*S, e.g.   *ma-buSek   >   **mabuk   (13)
           *iSeq   >   **eyeh   (87)
*H, e.g.   *qumaH   >   **(h)uma/juma   (114)
*W, e.g.   *Wasu   >   **asu,
(**asu is reconstructed on the base of
SB, Si, Da asu 'dog').

PAN *q is lost in initial position in
some cases, but maintained in others;

it is represented by *h in PB, e.g.

        *qukur   >   **(h)ukur   (94)
        *qaCey   >   **hatey
    PAN *z is reflected by *j in PB.

FOOTNOTES

1. /nd/- is realized as [ndr] by Karo woman.

2. Two exceptions should be mentioned to this rule: pə+oto+kən yields pəpotokən, and pə+idah+kən yields pecidahkən.

3. In loanwords original w becomes /o/, and the word gets an additional syllable, e.g. To sara-ó-ar 'trousers' (v.d.Tuuk; Warneck gives saraoal)., from Min. sarawal, (Ml. seluar, serual, sarwal; orig. from Persian); Ml. bawang 'onion' becomes To ba-ó-aŋ, id.

4. Both forms are given by van der Tuuk. Aŋkora is the term of address to a girl; baio means 'human being'. Van der Tuuk does not give compounds of aŋkora or baio with a demonstrative pronoun.

5. The suffix -/sa/ is realized as [ca] after final C, e.g. dapotsa is pronounced [dapotca].

6. Van der Tuuk gives still another example, caciŋ [caciŋ] /sasiŋ/, which he derives from Minangkabau sansiŋ+ which first became sanciŋ+, then canciŋ+ and finally became caciŋ (this lexeme is not given in Eggink's dictionary).

7. This distinction of prefixation between transitive and intransitive verbs is not found in Eggink's dictionary.

8. andung: the language used by women wailing over a corpse.

9. [h] and [k] correspond with /k/ in the deep structure of Mandailing (see par. 4.3).

10. For the replacement of /k/ by a velar N before /s/ see par. 3.3.1 (Toba phoneme system).

11. Written as 'inum' in Saragih's dictionary, see par. 2.

12. In Lampung tiuŋ is found. Van der Tuuk (1872:153—156) gives a couple of words in both Lampung and Batak, which point to mutual influence.

13. See footnote 14.

14. Van der Tuuk sometimes gives doublets with the reflex /o/ (in which case the lexeme has an additional syllable). These doublets are loans, as we can see the same phenomenon in recent loanwords from Ml. See footnote 3 under 3.1: To phoneme system.

15. In Saragih's dictionary there are over 400 lexemes with initial /h/ and 245 with intervocalic /h/. There are 44 lexemes with initial /k/, and 45 with intervocalic /k/; of these 89 lexemes
    a) 18 have a doublet with /h/, and occur also in Bahasa Indonesia (loanwords?)
    b) 25 have /h/ as next C (another two are of the sequence kV(w,y)Vh)
    c) 8 are originally doubled root-morphemes (with a medial C cluster with /k/ as second component)
    d) 5 are onomatopoetic
    e) 10 are loanwords (with medial /k/)
For the remaining forms it is hard to find an explanation within the scope of this paper, but it appears from the relatively few lexemes with initial or medial /k/ that most can be explained as loanwords or as reflexes of a kVkV(C), kV(w,y)Vh, or (C)VkVh prototype. I assume that historically /k/ and initial and medial /h/ are one and the same phoneme and that this phoneme is distinct from final /h/.

16. Gonda (p. 119—131) lists some Indian loanwords in Batak dialects which, according to his information, do not occur in other Indonesian languages. Many of these words however do occur at least in Old or contemporary Javanese (like To humba 'water-jar'; To sarpa 'serpent'; To mintora 'magical formula').
    Tideman gives an account of Indian influence on Batak beliefs and customs. Kern (p. 23—26) wrote about Austronesian numerals in South-Indian trade jargon. Some of these words, and others like pattu pulu 'forty', are also found in the Tamil lexicon.

17. The phoneme inventory presented here is that of Dempwolff as modified by Dyen (esp. 1953, 1965). The phonetic features suggested by the chart are partly chosen for the sake of convenience, and the exact phonetic interpretation of some of the symbols is still open to discussion.

18. Symbols between parentheses refer to PAN sources of PB phonemes.

BIBLIOGRAPHY

Dahl, O.C. 1973. *Proto-Austronesian*. Scandinavian Institute of Asian Studies Monograph Series No. 15, Studentlitteratur. Lund.

Dempwolff, O. 1934-38. *Vergleichende Lautlehre des Austronesischen Wortschatzes* I, II und III. Berlin.

Dyen, I. 1953. *The Proto-Malayo-Polynesian Laryngeals*. Linguistic Society of America. Baltimore.

-------- 1965. Formosan evidence for some new Proto-Austronesian phonemes. *Lingua* 14:285—305.

Eggink, H.J. 1936. *Angkola- en Mandailing-Bataksch- Nederlandsch Woordenboek*. Verhandelingen van het Bataviaasch Genootschap 72/5. Bandoeng.

Gonda, J. 1973. *Sanskrit in Indonesia*. International Academy of Indian Culture, New Delhi.

Joustra, M. 1922. *Nederlandsch-Karosche Woordenlijst*. Uitgaven van het Bataksch Instituut nr. 19. Leiden.

Kern, H. 1915. Over de herkomst van eenige telwoorden in 't Bargoensch der Zuidindische kooplieden. *Verspreide Geschriften* III:23—6.

Liefrink, A.H.J., and Willems, C.P.G. 1947. *Nederlands-Bataksche Woordenboek*. Pro Manuscripto in behalf of the missionaries at the Vicar Apostolic in Medan. Babberich.

Neumann, J.H. 1922. *Schets der Karo-Bataksche Spraakkunst*. VBG 63/4.

------------ 1951. *Karo-Bataks-Nederlands Woordenboek*. Djakarta.

Stap, W.H. 1912. *Nederlandsch-Tobasche Woordenlijst*. Uitgaven van het Bataksch Instituut nr. 8. Leiden.

Tarigan, H.G. 1972. *Pengantar Morfologi Bahasa Simalungun*. Seri bahasa dan budaya Simalungun. Leiden.

Tideman, J. 1936. *Hindoe-invloed in Noordelijk Batakland*. Uitgaven van het Bataksch Instituut nr. 23. Amsterdam.

Tindi Radja Manik n.d. *Dairi-Pakpak-Nederlands Woordenboek*. Manuscript

presented to the Royal Institute of Language and Anthropology. Leiden.

Tuuk, H.N. Neubronner van der. 1971. *A Grammar of Toba Batak*. KITLV translation series 13. The Hague.

------------------------------- 1861. *Bataksch-Nederduitsch Woordenboek*. Amsterdam.

------------------------------- 1872. 't Lampongs en zijne Tongvallen. *Tijdschrift voor Indische Taal-, Land- en Volkenkunde* 18, 118-56.

Voorhoeve, P. 1955. *Critical Survey of Studies on the Languages of Sumatra*. KITLV Bibliographical Series I. The Hague.

Warneck, J. 1906. *Tobabataksch-Deutsches Wörterbuch*. Batavia. Reedited in 1977 at the Royal Institute of Language and Anthropology. Leiden.

Wismar Saragih, J. 1936. *Partingkian ni hata Simaloengoen* (Simalungun-Batak Dictionary). Pamatangraja.

* * *