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 (kabupaten) of East Sumatra, up to the sub-districts (kecamatan) Langkat, Deli, Serdang and Asahan (ca. 2°-4° NL and 96°-100° EL). The Batak dialects belong to the western branch of Austronesian languages; they have a quite extensive 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 material from six main dialects for which grammatical and lexical data are available: Toba (To), Karo (Ka), Dairi (Da), Angkola (An), Mandailing (Ma), and Simalungun (Si). A seventh 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 southern dialects (SB: To, An and Ma) and Simalungun. This last dialect is said to be intermediate 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 follows:

part I: an account of the (morpho-)phonological 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) phoneme systems;

I start from written Batak because it reflects an earlier stage of the dialects, and I will use the 'ejaan yang disempurnakan', the official 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, ŋ and ŋ respectively. The pronunciation of the southern dialects, and to a lesser extent their phonemic structure, differs considerably from the orthography. Where needed I write the pronunciation between square brackets and the phonemic spelling between slashes, e.g. Ma halak, /alak/, [alak] 'human being'.

There is not much uniformity in the description 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 greater 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 clusters in PB. Two asterisks indicate a PB lexical 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 | Austronesian | par. | paragraph |
| An | Angkola | PB | Proto-Batak |
| Bat | Batak | resp. | respective-ly |
| BI | Bahasa Indonesia | S | stop |
| C | Consonant | SB | Southern |
| Da | Dairi | Batak (see Introduc- tion, par. 9.3.1 and par. II) |
| e.g. | exempli grasata | | |
| Jav | Javanese | | |
| Ka | Karo | Skrt | Sanskrit |
| Ma | Mandailing | Si | Simalungun |
| Mal | Malay | To | Toba |
| N | nasal | V | vowel |
| NB | Northern | | |
| PAN | Proto-Austronesian | vl. | voiceless |

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

1. The Karo phoneme system

1.0 Vowels

<p>| high | i, i | u, ŋ |
| mid | e, ĕ | o, ŋ | o, ŋ |
| low | a, ā | |</p>
<table>
<thead>
<tr>
<th>Consonants</th>
<th>glot- velar pal- den- latal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stops v/</td>
<td>g  j  d  b</td>
</tr>
<tr>
<td>v/1</td>
<td>k  c  t  p</td>
</tr>
<tr>
<td>Nasals q</td>
<td>n  m</td>
</tr>
<tr>
<td>Liquids r</td>
<td>l</td>
</tr>
<tr>
<td>Fricative h</td>
<td>y  w</td>
</tr>
<tr>
<td>Semivowels</td>
<td></td>
</tr>
</tbody>
</table>

1.1 Vowels

Vowels (V) have long and short variants: vowel length is distinctive, e.g. īdās 'on, on top of'; against īdās 'one-plied thread'; mān 'to eat' against man 'to, for'. A long V originates from contraction of two Vs: a) by loss of intervocalic C, e.g. īdās and īdātas, which is slightly more formal, b) in the process of affixation, e.g. mbule/mbulēn; jandī/jerjandīn.

The long variant of shwa, /o/ is not distinct from /o/ in the script; its sound is like the /o/ in German øøn. Its status in the phonemic system is unclear. Neumann gives the following lexemes with /o/ and /ø/: ḍaʊ, ṭe, ḍaʊ, ṭe, ḍaʊ, ṭe, ḍaʊ, ṭe, ḍaʊ, ṭe, ḍaʊ, ṭe, ḍaʊ, ṭe, ḍaʊ, ṭe, ḍaʊ, ṭe, ḍaʊ, ṭe.

In his Grammar he also mentions galānt (see par. 7.1). The /o/ occurs in monosyllabic lexemes and in disyllabic lexemes as a final phoneme (VCV0-type); it also occurs as V in both open syllables of a disyllabic lexeme, e.g. ḍaʊ, ṭe, ḍaʊ, ṭe, ḍaʊ, ṭe, ḍaʊ, ṭe, ḍaʊ, ṭe, ḍaʊ, ṭe, ḍaʊ, ṭe, ḍaʊ, ṭe. Three items do not fit in the lexeme types given here: ḍaʊ, ṭe, ḍaʊ, and ṭe, which has a variant uɛkɛr; in two cases a minimal pair of /o/ and /ø/ was found, and on account of this I consider both sounds as phonemically distinct: ʊoɔ 'to have lost and to start again, in a panta-play', against ʊoɔ 'k.o. deer-trap'; ṭe ʊoɔ 'to push on the belly of a laboring woman, in order to accelerate the birth', against ṭe ʊoɔ 'k.o. leaf'. Examples of /o/ in open ultima and monosyllables: ḍaʊ, ṭe in ḍaʊken 'to press down, to compress'; ṭe jolo and ṭe jokọ 'to be unable to forget s.th., to sulk'.

1.1.1 In his remark on the pronunciation of /o/ in naŋkọ, and also in a remark on the accent of Sinemba-Serond Karo, Neumann suggests that the spoken language distinguishes diphthongs 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], [iyọ], and [iye], where they are omitted in Roman script. Examples of semivowels with distinctive value are: ayan, mawan, ayar, bukak, buyak, tio, awak, awir.

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

/β/ and /k/ are also distinctive, e.g. kąkə 'languid, slow feeling after an attack of fever', against kąkə-kąkə 'k.o. lianas with dark fruits'; bənək 'spring-set spear or arrow-trap', against bənək '1l' puffed, of rice; 2) the cocoon of a type of yellow butterfly, used for food'; daka 'to cook rice', against daka '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. ayan/hayam; aqa/aqat; antu/hantu. In the process of prenasalisation clusters of S preceded by homorganic N arise, e.g. guda/m/ugudam; buhut/m/bahut; jumurak/m/jururak; dəmπi/ndəmpiy; təkəni/ntiņkəni.

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, ə, ʊ, ʌ-jo, ʌ-saq, ʌ-la |
| CV | da, də, təkət, tə-lu |
| VC | ə-rtə, ə-ten, sə-an |
| NCV | ndət-i, ˈdəmpu, tənda |
| NCVC | tə-məbəh, tə-ŋəl, tə-ŋaτ |

1.6 Minimal pairs of Karo phonemes (consonants):

| a. | batuk 'cough' | karug  'plaited |
|    | patuk 'to peck, bite' | bur, m  'fat, thick' |
|    | datuk 'a title' | cur  'very fine, very small |
|    | bua  'to throw away' | d. ampaʃ 'side, lateral face' |
|    | tusa  'to pour' | tampar 'to ward off a blow, to parry' |
|    | bala  'wide, extensive' | daya, ər  'to sell' |
|    | gala  'bracelet' | gaya  'earth-worm' |
|    | baka  'corpse' | daya, k.o. plant; title for a maiden |
|    | haʃək  'reverent' | haya  'heavy feeling as after an attack of fever' |
|    | buka  'open' | dimpu 'to be very afraid, cowardly' |
|    | bara  'shoulder; k.o. plant' | dalaq, ən  'offering, dish to the spirits; idiomatically; to have a dry throat' |
|    | nari  'still, left (used with numerals); only, mere, sheer' | malaq  'respectful, dalih, or -- to fail, of a gun' |


nalih, so -- 'quantity of rice'
dagat 'to droop, of a branch'
roqat 'to radiate; to leap, shift, of pain'
damol, ma -- 'half sweet and half sour, of taste'
samol, ma -- 'pleasantly soft to the touch'
daga 'to lessen, to stop'
qaga 'wide open (mouth), opening, gap'
dalalq (see above)
jalalq 'to wander about, to rove'
duru, or -- 'to walk in a flock, of sheep'
kuru '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, ma -- 'respectful'
gadiq 'ivory, elephant's tooth'
ladiq 'generic for curved knives and swords'
gulag, ma -- 'to roll down a slope'
mulaq 'weaned, of animals'
gundaq 'float with bait; reciting' 
endorq-endorqun
unduq 'to fail, let down, desert'
gumbeg 'big earthen cooking-pot'
rumbug, rumbug 'here-ditory enemy'
gombug 'swollen, of belly'
sombug 'sheath of knife'
galagh 'stick, bamboo-stick to knock off fruits'
galagh '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. grass-hopper or beetle'
gabak 'small axe'
cabak 'wooden bowl, finger-bowl'
h. haga, ma -- 'respectful'
laga, ma -- 'wild, fierce; proud, arrogant, vehement'
hayag (see above)
mayaq 'areca-palm, k.o. tree'
hala, hala-hala 'flat shell'
nala 'flame'
haqke 'to stand in awe of, to be afraid'
raqke 'plaitwork of rope; bunch, cluster'
hiri, hiri ate 'to bear a person a grudge'
siri, siri-siri 'dragon-fly'
hile, ge -- 'exult with high-pitched voice'
jile, -- ne 'the beauty of s. th.'
kile, akile-kile 'to say one thing today, another thing tomorrow'
hio 'stereotyped word in the bilaga-bilaga'
crio 'protected from rain (a place)'
huq 'Skrt. 8m
quq 'onomatopoetic for deep, resonant sound'
p. patuk, datuk (see above, under)
p. puar 'torn off, broken off'
tuar 'small fish-trap'
pulaq 'wiped out (by an epidemic), extinguished'
gulaq, ma -- 'to roll down a slope'
pitu 'seven'
hitu, si -- 'the suspect in a law-case'
puaq 'satisfied' 
luas 'evaporated, of lukewarm water'
pulaq, mulaq (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'
parah, parah-parah 'to weep heavily, to cry'
gerah 'to be concealed'
pulus, ar -- '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 (view)'
galar 'what is given in payment for s.th.'
tala 'to turn o.'s face to'
hala-hala (see above)
taka, ar -- 'to conspire, to act secretly against s.o.'
laki 'husband'
talu 'to have lost (against s.o.), to have failed'
mulaq 'ashamed'
salu 'with instrumentalis'
tala, nala (see above)
ramas 'a set (of s.th.)'
tamas 'empty, finished, of a fruit-tree'
tuda, ma -- 'impatient, of pre- sumptuous'
quda 'tender, young'
toliq 'outcry of peevishness, to a woman'
jolis 'clean, virtually pure'
tuan 'master, lord, European'
kuan, ar -- 'to talk; to fondle, of girls and boys'
cuan 'pickaxe to work the ground'
l. lalu 'rice-pounder'
maluq 'ashamed'
lanagan 'folded cloth, small cushion'
nanam 'the taste of s.th., the use of it'
liqag 'hole, gap, opening'
siaq 'to break, of the day, to become'
rigaq, riag-riag 'creepers; scorpion'
liat 'smooth, like paper'
giat 'to expand, to swell'
laba, ar -- 'to be bought with profit'
jaba 'barley'
luh, ma -- 'to be loose, e.g. of the hair'
kuh 'complete, full'
liho 'thin, watery'
chilo, ma -- 'clean, pure, transparent'
m. mala 'disappeared (smell), faded (color)'
nala (see above)
modak 'open (eyes); to liquify, of oil, fat'
sedak 'pressed in the chest; oppressed of mind, because of sorrow'
muan, p -- 'curse'
ruah 'ripped off, of weeds'
momo 'acknowledge-
ment'
ngoi 'totally quiet, deso-
late'
malag 'respect-
ful'
jalaq (see above)
malu (id.)
kalu 'term used at a panta-play'
maia (see above)
cala, cala-cayak 'k.o. cymbal'

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

s. sauk 'to roast in a pan (maize or coffee)'
rak 'a grasp (of the hand)'
sone 'k.o. tree; isomakon to acknowledge, spread the news'
pena 'to like, love s.th.'
saung 'what serves as an umbrella'
jauk 'maize'
sial 'mischance, bad luck'
kial 'bent, because 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, pa 'to consult'
gasi, la 'import-
tant'
rambe 'k.o. tree'
jambe 'k.o. gourd'
rambe 'term used for counting maize-ears'
kambie 'goat'
rio 'with holes, of a leaking roof'
cio 'protected from the rain (a place)'

q. gari 'to serve at meals'
cari 'to look for o.'s living'
kari 'presently, by and by'
nadi 'to cease, stop'
jabi 'to become, come into being; tarjadin 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, i u, ū
mid e, ə o, ɔ
low a, ə

Consonants

Consonants

Consonants

2.1 Vowels
The Da /ə/ occurs in all syllables except in the position (h) #. All V's but /ə/ 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)VVC; examples: gâo bâk pântis tâtur pâstop. Height harmony: in VCVC V can never be /i/ or /u/, but /o/ and /e/ do occur, e.g. kădôn, gâôt; /u/ does not occur in VC(C), but /o/ does, e.g. kâôt; V = e/ and # /i/ in VCu(C), e.g. ñum.

2.2 Semivowels
These occur only between V's; /w/ follows /a/, /ø/, /i/, /o/ and /u/, e.g. kâun, sât, kâun, [kau:n], kâlôkâ [kalomas], ñum.

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 /awa/: sawah (saweh), kâwân (=leban), rewân (rewben), benâ 'tax on commodities', bewan 'calladium species', rewan 'k.o. clap-net to catch something' against rebas 'destroyed', sâwah (=sâbah). It appears that three lexemes with an /ewa/-sequence, and sâwah 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 # /a/; its value is not distinctive after /i/ and /e/ and in this position it is not written in Roman script, e.g. rægæm, beyæg, beyæg [beyæg], akyâp, i [iya].

2.3 Consonants
/h/ and /k/ are distinctive phonemes; examples of minimal pairs: hîa 'glad, happy, pleased' against kîa, a form of address to younger siblings; pa-kîa-kon 'to treat s.o. in a presumptuous way'; nakan 'soon, in the future' against nakân 'cooked rice'; kalâh 'uvalu-
la' against kalak 'human being, person'. In initial position /h/ is weak, and is often omitted, as in hâsa: [hâsa] = [asa]; but in some lexemes /h/ is never omitted, as in hakum [hìkum] 'if, in case'. Two different liquids
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. kanka, caasudur; minimal pairs of the /s/-/c/- opposition: koon 'k.o. tree' against koosn 'to turn around'; ampur 'to mix, blend' against sampur 'to go down, descend'. Gemination is avoided except at morpheme boundaries, as in dahann-na and dok-kon.

2.4 Morphology

Through suffixation of -/ən/ after a final V this V is lengthened and the /ə/ of the suffix is omitted, e.g. doli-ton becomes doli-in; wen-ton yields wen. The prefixes /po/- and /ko/- 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 /po/- with /ə/: pə̐wə̐pat+ken becomes pə̐wə̐patken; examples: po+tiad/ə/d/ əd/ə/d yields pə̐wə̐patken; oit/wə̐gə̐t/ə̐ken/ə̐ken yields: wə̐gə̐tken/ə̐ken with prefixation before /u/ and /ə/ an /h/ is inferred: wə̐pə̐wə̐ap; alo/pə̐halo.

2.4.1 Vowel assimilation:

/ə/ of the suffix -/ən/ usually becomes /u/ after -/ur(h)/, e.g. napur/napurun; puə̐u/h/puə̐u. This /ə/ can also assimilate with /i/ of the preceding syllable: kərə̀k/kərə̀kin. There are also some cases of assimilation of /ə/ to /i/ in the suffix -/ken/: arə̐th/arə̐kin; bakə̐n/bakə̐kin. /ə/ often assimilates with /i/ or /u/ of a following syllable that starts with /r/, e.g. surume also surume and surume (the Dairi-Pakpak dictionary only mentions surume and surume; surume occurs as an example in v.d. Tuuk's Grammar). /ə/ in prefix /ko-/ assimilates with /i/ of the infix /-in/-: k-in-o̍hur-on yields kinihurun; k-in-o̍hjak yields kinihjak.

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. bagin would have originated from bagi+ten 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 /mo/- becomes /oN/- before a C which is not /r/, /l/, /h/ or N, e.g. /mo/+ pɔ̑ke/ yields o̍hpuke; /mo/+ hə̐qə̐kə̐ yields mo̍hə̐qə̐kə̐; /mo/+tor/ yields ə̐wə̐tor; /mo/+go̍lu/h yields o̍gə̐lu/h; /mo/+lampis/ yields mo̍lampis; /mo/+rincan/ yields mo̍rincan; /mo/+ naphanumeric before V the /o/ is omitted: /mo/+ela/ yields melə; /mo/+ 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 baæ, (h)ari and bage is deleted when they are followed by a demonstrative pronoun, e.g. bage 'like, (same) as, well' + idi 'there, yonder' yields bag idi or bag? 'so, thus, in such a way'; sidah (h)ari 'when' + en 'this' yields sidahen 'today, at present'; bae 'so'. whose name one does not know or who one does not want to call by name' + en yields bae+en 'this friend, this companion'; bage+ena 'she' yields bage+ena 'like her/him';

2) so, an emphasis, 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. laka 'to leave, depart' +mo+la 'third person singular' yields lau̍m lau̍m 'lau̍m (so) she/he went'; laka+mo+kə̐u 'I' yields lau̍maku '(so) I went'; ena+mo+kə̐u yields ena+mə̐kə̐u 'this is me'.

The /o/ of the prefix /oN/- is omitted before final V of the preceding word, e.g. kutan+əntua yields kutan+əntua; konə̐ntua yields konə̐ntua; ma̍kə̐ntua yields ma̍kə̐ntua.

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 /e/, e.g. balu̍k, balu. A V in the ultima takes the accent, which is the case only in poly-morphemic words, e.g. bagin kinigurun. In prefixing /moN/- to a lexeme with initial labial, /mo/- is sometimes omitted, e.g. buat/mo̍nuat/muatu; bunu/h/munu/h/munu/muna; pə̐han/mo̍hə̐n/mahə̐n; but there is no such form as mə̐ku̍r along with mə̐mə̐ku̍r. The suffix -/su/, which expresses an excessive, demands the accent: /entor/+su/ yields entorsu.

Forms of address (vocatives) also have the accent on the ultimate syllable. In Dairi lexemes of more than three syllables are unusual, as with the borrowing əntua: this lexeme derives from the Toba words əntu and ara and has a trisyllabic variant tura. NB. Except in lexemes with a CoCoC structure we see that the place of the accent within the Da lexeme is determined by vowel length with the following hierarchical arrangement: V, 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 CoCoC structure).

3 The Toba phoneme system

3.0 Vowels

<table>
<thead>
<tr>
<th>high</th>
<th>i</th>
<th>u</th>
</tr>
</thead>
<tbody>
<tr>
<td>mid</td>
<td>e</td>
<td>o</td>
</tr>
<tr>
<td>low</td>
<td>a</td>
<td></td>
</tr>
</tbody>
</table>

5
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 /mahinsu/ 'to close, shut', against /hinsu/ 'closed, shut'; (with nasalization of initial C): /ma-nānom/ 'to bury' against /tanōm/ 'buried'; /tāntan/ 'hanging down', against /tāntān/ 'to be lowered down'; /rāhāt/ 'saw' against /rāhāt/ 'to be rough'; /bēgu/ 'supernatural being', against /bēgu/ 'courageous'; /bāu/ 'odour, smell', against /bāu/ 'to stink'. Except for words subject to the rules mentioned in 3.4.2, polymorphemic words have the accent on the penultima, e.g. /timbā/ which equals /ma-timbo/ in meaning 'to be high'; /rārā/ and /ma-rārā/ 'to be red'; /sōhōt/ becomes /sōhōton/ when suffixed.

3.4.2
Polymorphemic words can have the accent on the ultima only in the next two cases: 1) when suffixed with -/hū/ and -/tū/, which express excessive suffixation, and -/ān/ which expresses the comparative degree, e.g. /maganjan+hū/ yields /maganjanhū/; /rahān+tān/ yields /rahānān/; 2) when a monosyllabic lexeme is affixed the accent is always on the stem, e.g. /dōk/ /māndōk/ /dōbonon/.

3.3 Realization of consonant clusters within Toba lexemes

3.3.1
N+vl. stop or /s/ yields gemination of the latter, e.g. agka [akka]; ompu [oppu]; tinnin [tittin]; geminated /s/ is realized as [ts]: hinsē [hiteu].

3.3.2
/n+/r/ /l/ or /m/ yields gemination of /r/ /l/ or /m/, e.g. ronron [ronron]; mommon [mommon]; lattān [lallan].

3.3.3
(C ≠ /s/) + /h/ yields [kk] in doubled root-morphemes, e.g. hopkop [hakkop]; harhām [harham] (from hamhām); /s+/h/ yields gemination of /s/: huksu [hutora]; /t+/h/ yields [kk] or [tt], e.g. hukutta can be realized as [hukkut] as well as [huttut].

3.3.4
N+g/yields /ng/ in doubled root-morphemes, e.g. gongom 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 onomatopoeic character, on both syllables, e.g. /hundul/, /gābe/, /tānōm/, /debēta/, /ümmbā/, /dōkkā/: its place is distinctive, e.g. /sogōt/ 'early in the morning', against /sōgōt/ '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 /mahinsu/ 'to close, shut', against /hinsu/ 'closed, shut'; (with nasalization of initial C): /ma-nānom/ 'to bury' against /tanōm/ 'buried'; /tāntan/ 'hanging down', against /tāntān/ 'to be lowered down'; /rāhāt/ 'saw' against /rāhāt/ 'to be rough'; /bēgu/ 'supernatural being', against /bēgu/ 'courageous'; /bāu/ 'odour, smell', against /bāu/ 'to stink'. Except for words subject to the rules mentioned in 3.4.2, polymorphemic words have the accent on the penultima, e.g. /timbā/ which equals /ma-timbo/ in meaning 'to be high'; /rārā/ and /ma-rārā/ 'to be red'; /sōhōt/ becomes /sōhōton/ when suffixed.

3.4.2
Polymorphemic words can have the accent on the ultima only in the next two cases: 1) when suffixed with -/hū/ and -/tū/, which express excessive suffixation, and -/ān/ which expresses the comparative degree, e.g. /maganjan+hū/ yields /maganjanhū/; /rahān+tān/ yields /rahānān/; 2) when a monosyllabic lexeme is affixed the accent is always on the stem, e.g. /dōk/ /māndōk/ /dōbonon/.

3.5 Morphology: a schematic survey

3.5.1 Nasalisation
N+N/, /j/ yields /n/, /nj/, e.g. dege/ mandege; jongkal/jongkal.
N+N/ yields /mb/, e.g. buat/mambut; sometimes /b/ is lost in the course of nasalisation, e.g. bun/mamunu; this never occurs when the second syllable begins with /b/ as well, e.g. bābā/bāmbalā; bebe/mambabe. N+N/, /j/ yields /g/, /gh/, /v/; e.g. goar/magoar; hinsē/mahinsē; inun/magunun; alap/magalap. Some lexemes used transitively and beginning with /h/ lose this phoneme, e.g. hābul 'round, cylindrical, globular' and magul 'to be compressed; to be closed like a bud, i.e. to be childless'.

N+N/, /l/ yields /gār/, /gal/, e.g. rīor/magīor; lapā/magalapa.

N+N/, /t/ yields /n/, except with exclusively intransitive verbs, where it yields prenasalisation, e.g. sūrtum/manurtum; tūru/ tūnanum (intransitive) sōhōt/mahōt; (v.d. Tuuk does not provide an instance of prenasalisation of an initial /t/ in an intransitive verb).

N+N/, /g/, /m yields /n/, /g/, /m/, e.g. nolnol/manolnol; nagger/magagger; mātu/ mamatahun.

N+N/ yields /m/, e.g. pīna/marinaas.

3.5.2 Prefixation
/mor/- /por/- /r+/l/ yields /ll/, e.g. porlāna is pronounced [pallanja] and morlāga is pronounced [mällage] (the /o/ of /mor/-, /por/-, and /tor/- are usually pronounced as [a]).
3.7 Minimal pairs of Toba lexemes

b. batu 'stone, weight'
patu 'subjected'
datu 'sorcerer'
bolon 'big, voluminous'
tolon 'to swallow s.th.'
bale 'rice-measure, ½ 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);
rong-balu 'thunder without flash (as an omen)'
bandar 'cabin to stay over-night'
mendar 'wrapping, e.g. sarong'
bunu 'to strike, kill'
nunu 'to whisper, to murmur, of many people'
boa 'to proclaim, announce'
roa 'bad, mean'
bere 'sister's child'
sere gold'
abat 'obstacle'
agat 'trick, artifice; better'
baio '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'
laaho 'to go, walk'
lapo 'small shop'
lamo 'pleasant to the touch'
api 'fire'
ani, mananihon 'to wind up yarn on a rack'
puntas, ma- 'to burst open, of cracked leather bags'
runtas, mana- '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'
qalu, mor- 'to ruminate'
palo 'white spots, on one's eyes'
jalo (see above)
d. dandi 'to disobey, out of insouciance'
tandi 'prominent (in good or bad sense)'
daio 'wide, broad, vast'
haio 'to itch, scratch with one's hand'
gao 'to move the hands, in swimming'
dabu, marabu 'to fall'
labu 'to gather, come together'
duhut 'grass'
muhut 'trifles; junk'
dopaq, man- 'to fight, attack'
ropaq 'jagged, notched'
deaq 'young girl'
neaq 'easy to carry'
durauq 'scoop net'
suruq 'privilege'
4. The Mandailing phoneme system

4.0 Vowels

<table>
<thead>
<tr>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>u</td>
</tr>
<tr>
<td>e</td>
<td>o</td>
</tr>
<tr>
<td>a</td>
<td></td>
</tr>
</tbody>
</table>

Consonants

<table>
<thead>
<tr>
<th>Glottal</th>
<th>Velar</th>
<th>Palatal</th>
<th>Denal</th>
<th>Alatal</th>
<th>Tal</th>
<th>Bial</th>
</tr>
</thead>
<tbody>
<tr>
<td>g</td>
<td>j</td>
<td>d</td>
<td>b</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>k</td>
<td>(c)</td>
<td>t</td>
<td>p</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Stops

<table>
<thead>
<tr>
<th>s</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>v</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Nsals

<table>
<thead>
<tr>
<th>d</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>m</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Liquids

<table>
<thead>
<tr>
<th>r</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>l</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fricatives

<table>
<thead>
<tr>
<th>s</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>h</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Semivowels

<table>
<thead>
<tr>
<th>y</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>

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

4.1

1. /c/ is an allophone of /s/ which appears
2. as the initial c of a syllable in a word which contains [nc]², e.g. binncur/binsusur/[bincucur]. Otherwise /c/ only occurs in loanwords, and in those cases it is not a positional variant of /s/.⁶
Ma has a palatal N which occurs in initial and medial position. It appears also in syllable nasalisation of /s/, unless the second syllable of the prefixed lexeme also begins with /s/, in which case it is realised as /n/, e.g. surat/mašurat; sosop/manosop; sosu/ mašušu.

4.3.1

/ŋ/ occurs in only the position /a-/e.g. dinaŋ 'mushroom'. Sometimes a final /ŋ/ is part of the underlying structure and appears on the surface when followed by a suffix beginning with /ŋ/. Thus kala 'scorpion' [ala] has a phonemic structure /kalæ/, with final C surfaced 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. kuta/kuta/[utal]; morkutal; halak/kalak/[akal] in the phrase dibuat halak [dibuatkalak].

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, e.g. hikis [iskis] - (transitive) [maškis] 'to scrape s.th.' - (intransitive) [maškis]; hikil [ikil] - (transitive) [maškil]; (no intransitive form is given by van der Tuuk or Egging). This also applies to the verb hoes [kowas] (written hohas in the Batak script) which has a transitive form [mašwos].

[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 [mašxos] and the initial /k/ of its unaffixed form. Similarly we can explain the initial /k/ of other lexemes, e.g. keke [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 [kk]. /raškon/ 'with' is probably a borrowing from To rapthon*, which is pronounced as [rakkon] in some eastern Toba areas. Two /k/’s following each other through suffixation are usually realized as [kk], and sometimes as [k], e.g. the circumfixed form mandokkon /maňdokkon/ and its pronunciation [mandokkon] and [mandokkon].

4.5

Final nasals before a v1.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. pośop 'curse, malediction'; haya 'tree, wood'; sayom 'means to safeguard s.o. from illness or bad influences'; boyu 'andung word for bora, b daughter'.

4.7

Morphology: V contraction occurs, but not in final position: tu huta [tuta], but maridi t is pronounced [maridii]; nasalisation of /s/ yields /R/, 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

<table>
<thead>
<tr>
<th>Glot</th>
<th>Velar</th>
<th>Palat-</th>
<th>Dental</th>
<th>Palatal</th>
<th>Bial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stops</td>
<td>v</td>
<td>d</td>
<td>b</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasals</td>
<td>n</td>
<td>m</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liquids</td>
<td>r</td>
<td>l</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fricatives</td>
<td>h</td>
<td>s</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semivowels</td>
<td>y</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

5.1 /k/ and /ŋ/ are in complementary distribution: /ŋ/ 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 /ŋ/, e.g. keke [Kehe]; háhán [kaháh].

5.2

Geminated C's occur, (as in October), e.g. panti [pattis]; laška [lakka].

5.3

N is replaced by its homorganic v1.s as first component of a cluster with a v1.s or /s/, e.g. panti and láška in par. 5.2; /s/ is realized as /c/ as second component in a cluster with a v1.s as first component, e.g. mašé [makči].

5.4

Sandhi: final N is replaced by its homorganic v1.s before a v1.s, e.g. tešan pargarutan [tematpargarutan]. After a final v1.s /s/ is realized as /c/, e.g. halak aško ni [halakcokkon].
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

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

Consonants

glot velar palatal dental
vocalic t al t al bial

Stops
q j d b
k t p

Nasals
q n m

Liquids
r 1

Fricatives
h s w

Semivowels

6.1 Diphthongs

Si has final vowel sequences which can be analyzed as diphthongs, e.g. horbos 'hurbo' 'buffalo', borot 'bolk' 'to bite', and 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. Saragi's dictionary gives only three entries with initial /w/ and none with initial /j/; wajtp 'duty', a Malay-Arabic loan; watpq 'a big wasp', which is also spelled waltor; wewen 'implements for making rope'.

6.3 Consonants

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

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: hapur 'ch alk against h any 'crumble, brittle, delicate'; tòbu 'sugar cane' against tobu 'sweet'; hapal 'ship' against h apal 'thick, fat'; bayu 'to braid, to plait', against bayù 'new'; the first two minimal pairs are semantically contrasted: chalk is a brittle and delicate substance, and sweetness is the main feature of sugar.

6.5 Schematic survey of Simalungun phonology

6.5.1 Nasalisation

N + /d/, /j/, /l/ yields /nd/, /nj/, /nl/; e.g. duda/manda/danda; jalo/manalo/jalo; lapah/panah/lapah; /nl/ is often realized as [il], as in manalap and panalap ([mallapah], [pallapah] respectively).

N + /b/ yields /mb/, e.g. buat/mambuat.
N + /g/, /h/, /w/ yields /g/, /h/, /w/; e.g. gual/mqgul/pagual; haday/mqhaday/paghaday; the cluster /g/ is realized as a geminated /k/, which yields the pronunciation [mkkada] and [ppkkada]; alop/mqalop/pagalop; inam/mqinam/paginam.
N + /s/, /t/ yields /n/; e.g. suan/manuan; tanjo/panajjok/panajko.
N + /p/ yields /m/; e.g. palu/malalu; poroh/moroh/pamoroh.
N + /n/, /m/, /g/ yields /n/, /m/, /g/; e.g. numut/manumut/pamumut; murgan/mqumgkan/paragqat; gatqat/mqagqat/pagqat.
N + /gi/ yields /g/; e.g. karat/magarat/pagarat; Tarigan does not give examples of nasalisation of lexemes with initial /r/.

6.5.2 Prefixation

/u/n/-; /u/n/- /b/, /m/ yields /umb/, /uum/; e.g. baht/umbhat; muhav/muhavah.
/u/n/- + /d/, /j/, /l/ yields /und/, /unj/, /unl/; e.g. daad/undoad; jahat/undjatjat; lojat/umljat.
/u/n/- + /g/, /h/, /w/ yields /ung/, /ugh/, /uqg/; garah/uggah; haqah/uhqah; gatlut/ugatlut.
/u/n/- + V yields /umV/ as well as /uqV/, without difference of meaning; e.g. idop/umpidop; estek/umpetek; allus/umlu/ugalu.
/u/n/- + /p/ yields /ump/, e.g. posb/umpbso.
/u/n/- + /t/ yields /unt/, e.g. tuguh/untbgh.
/u/n/- + /s/ yields /uns/, e.g. susah/unsasah.

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

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

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

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 /oku/ after /V; e.g. rumah/rumakh; jabu/jabukok; aggi/agigkq. 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; togo/tongonta. 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-mu 'your father
  lombu-mu 'your buffalo'
  inaq-mu 'your mother'
  tulaq-mu 'your uncle'

2) the accent falls on the penultimate; the suf-
fix becomes -/m/ after lexemes ending on a 
V, e.g.
  bapumu 'your father'
  lombumu 'your buffalo'
  inaqmu 'your mother'
  tulqmu 'your uncle'

6.6 Structure of the syllable

V: bɔ- l  k VC: ka-t a
  a-t  bɔ-dar
  a-sa l  bɔ-i

CV: tɔ-lu VC: hir-tɔg
  hɔ-da tɔm-bar
  bɔ-li tɔq-kap
  pɔn-sur

II COMPARISON OF THE DIALECTS AND RECONSTRUC-
TION 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 /ɔ/ in Ka, but 
not in Da.; in Da we find /ɔ/ as a correspon-
dence to Ka /ɔ/, as is demonstrated in the fol-
lowing instances: Ka ɔn, Da ɔn, 'to get 
dark', Ka dɔm, Da dɔm 'full, filled'.

7.1

In Neumann's grammar we find the word 
galɔmi 'to reconcile', and his dictionary 
gives alɔm, which corresponds to Da alum 
'quenched, satisfied, cured', SB alum 'sa-
tisfied, cooled down'. It is conceivable that 
Neumann was mistaken in his grammar, and that 
the correct representation should be galɔmi, 
which, like inam 'to drink' has /a/ instead of 
/ɔ/ as in all other dialects (Da has /a/ and 
isham, without difference of meaning). It is also conceivable that the shwa in 
alɔm became /ɔ/ because of a shift of the ac-
cent through suffixation, but our data do not 
support such a rule.

Other lexemes like jaram have a regular 
/u/ as a correspondence in Ka, but Balinese 
and Sasak have inam 'to drink', and Da has 
both forms; (for /e/ in the penult. see par. 
2.1). Until a better solution is found I re-
construct PB doublets **alum/**alum, and 
**inam/**inam

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 with-
in an unaffixed NB lexeme.

7.2

Vowel correspondences are not always con-
sistent in the Batak dialects: height harmony, 
assimilation, metathesis, contraction and the 
working of analogy sometimes confuse the pic-
ture but nevertheless it is not difficult to 
reconstruct the following vowels and diphthongs 
by mutual comparison of the lexemes in the re-
spective dialects: *a, *e, *i, *u, *ow, *uy, 
*y.

7.2.1

*a from /a/ in all dialects, e.g. NB, 
SB, Si dalan 'way path; means'; PB **dalant (3)
SB, Si ago, NB ago 'charcoal'; PB 
**ago 
NB bagas 'deep; house', Si bagas 
'deep', AnNa bagas 'deep', To bagas 'deep', 
and bagas 'house'; PB **bagas

7.2.2

*i from /i/ in all dialects, e.g. all 
dialects nipit 'dream'; PB **nipit 
SB, Si ribut 'busy, unquiet, unrest, 
commotion', Da ribut 'great stir, commo-
tion; in commotion, rebellion', Ka ribut 
'confused, of thoughts; dissension, 
quarrel'; PB **ribut

PB **ditat 
All dialects laqt 'sky, firmament'; 
PB **laqit 
SB, Si, Da bibir 'lip'; Ka bibir 
id.; PB **bibir 
Lexemes of (CiCi(r,l) structure do not 
occur in Ka except for tili, an Arabic loan. 
Corresponding to this type in other dialects 
Ka has a (Ci)Car(l) structure.

7.2.3

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

7.2.4

*e from /a/ in Ka, /a/ and /o/ in 
Da, /o/ in Si-SB, e.g. SB-Si jolma 'human 
being', NB jolma id., from Skrt. janna 
(from SB-Si /o/ and NB /o/ in the penult it 
is clear that this word came into the 
Batak dialects through other Indonesian 
languages, probably Malay or Javanese);
PB **jolma

SB-Si onom, NB onom 'six';
PB **onom
SB-Si tanom, NB tanom 'to bury';
PB **tanom
Da does not allow /a/ before fin. 
/h/ and in monosyll. lexemes, and has /a/ 
instead. Only Ka confirms reconstruction of 
*e in this position, e.g. Ka tanah 
'ground, land, earth', Da-Si tanah id., 
SB tano id., PB **tanah (for final /h/ 
in SB: see par. 9.3) 
Ka darah, Da-Si daroh, SB daro
b') through vowel assimilation, e.g. Ka puwok, Ka puwog (here a high back V became /o/ under influence of the accent, and subsequently the following pepet assimilated to /o/ as well), To puwok, Si puwok 'navel'; PB **puwok (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: tin/ lown, ðh/dauk, ðk/kalik, kol/kol, kene/kapi, jelah/jelah/jelah, jelah, jembar/jambor, ambat/ambat.

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 guol, 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
SB kastor 'an evil-smelling plant, with roots used as a medicine', Si kastor, Da kastor, Ka kastor id.; PB **kastor (see par. 8.4 for SB-Si /s/: NB, /c/).
In Ka there is a tendency to change /V/- and /Vi/- sequences to /Vw/- and /Vye/- sequences respectively, e.g. the doublets raut/ravet, saip/sayep, lir/liser. I cannot 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 paruam /pa/- + /ruma/ + /on/; To lehon/Ma leken; To tho=the; To baen/Ma baon; To hain/kaen. To kaen comes from (MI) kain.
c) through vowel assimilation: SB, Si degke, NB degke 'fish'. In Si there is also a form degkel [degkey] which means 'meat'. /e/ in the penult of degke developed from *o through regressive assimilation; degke in Si is a borrowing from SB; PB **degke (35)

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 ðykam 'to open, to break open', Si ðykab 'to open', To ðykab 'to open, to get open'; PB **ðykab
NB ðbat, SB-Si ðbat 'remedy, medicine'; PB **ðbat (see also (30))

7.3.2
In Si /o/ developed
a) from *e (see par. 7.2.4);
Ka gelar, name; PB **gelar
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 swara, sora
'sound, voice, tone': from Skrt.
swara; PB ++swara

(c) through vowel variation, e.g. deos/dias,
awo/awo;
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 kahayan, melo from
mallo (both malla and kahayan exist togeth-
er with the contracted forms in SI)
b) through vowel variation, e.g. derder/
darar; darsat/darsat; jeleg/juljq;
c) in borrowings, e.g. degke (see par. 7.3.2) and Si bere, 'to thank (lit.
to give thanks)'; and the inherited form is berel: '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 bere in pa/bere/ax 'a start, a make-weight, an extra! Da
bere 'to give', Ka bere 'to give,
allow'; PB **berey

8 Semivowels
These do not occur in To, and /w/ is
found only in NB and SI.

8.1
In To tY is contracted with a pre-
ceding or following V, and is lost bet-
ween /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 kuyu 'tree, wood'; PB **kuyu
(see also par. 9.3)
To saur in saur matu 'to become
old', Si-AnMa sayur matsu, Da sayurntua
'to become old'; PB **sayur
To bup 'to hold, detain, to lift
up' Da isap 'to put away, to store';
PB **isap; see par. 9.4.2 for /n/
in Da; van der Tuuk also mentions Si
sara', a form which is not given by
Sarajih.
To bau 'block'; other dialects
iyan, i.a.; PB **iyan
To bau 'to hold in one's mouth for
a long while, e.g. a chew of betel nut'
AnMa boym NB bone 'to chew sth. in
a leaf, e.g. fish' Si boyo 'to eat up,finish one's eating'; PB **be(Ny)em.
To tuaq, Si-AnMa tuyeq, 'Solanum
Melongena'; PB **tuyeq

In SB there is a doublet tuyaq
with a cognate toyaq in NB, both with
the meaning 'Solanum Melongena'
(aubergine, eggplant)'; PB **tuyaq

To laeq in tuan laeq 'favourite wife';
von der Tuuk also gives An laeq and Ma laay, but Eggink's dictionary only gives laeq. AccORDING to v.d.T. To aeq 'water' has a cognate
ayok in Si, but this word is not given in Sa-
rangih's dictionary, where we find bah 'water'.

8.2
/w/ occurs in intervocalic position
in NB and SI. The correspondence in SB
is usually ø, and in a few cases To /b/
between /a/ 's and AnMa /o/ after /a/
(with contraction of following a)3. If
we assume that To /b/ represents an
original *b in the sequence *aba that
merged with *w in NB and SI and con-
tacted 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 ø, e.g. To
haban, Si, Da kawan 'cattle';
slave', and AnMa hoo, haban (probably
a To loan) 'herd of white boars, of ele-
phants'; (see also par. 9.3); PB **kaban
To tabar, AnMa taor, NB-Si tawar
'magic remedy, antidote'; PB **tobar
To ak (van der Tuuk gives also
aoak14), Da awak 'waist'; PB **awak
To dan (van der Tuuk also: daoan14
and daban), AnMa dan, dahan (van der
Tuuk) NB dawan 'mushroom'; PB **dawan
To ait, NB awin 'to draw tight, to
draw closer to o.s.'; (see also par.
9.0); PB **awid
To lai-lai, Da lawi 'the long tail-
feathers of a bird'; PB **lavi
To sat 'eye-tooth of humans and
animals', Da savut 'the tusks of a wild
pig'; PB **savut

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 /a/
before V, I reconstruct PB *w; e.g. Ka
uari, SB, SI, Da ari 'day'; NB still has a
semivowel in the compound form ma(u)ari
'sun'; PB **uari
(NB-Si ualuh, SB ualn 'eight';
PB **uuluh
To atoq 'a spider', AnMa atoq, Da
uataq, Si uataq/wataq 'a big and
fiercely stinging wasp'; PB **ataq
There is also NB uull, SB aull '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 dolq, SB dolok, NB doloq
'mountain'; PB **deleg

(36)
(37)
(38)
(39)
(40)
(41)
(42)
(43)
(44)
(45)
(46)
(47)
(48)
(49)
(50)
(51)
(52)
(53)
(54)
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 complementary distribution with /h/ (Ma [\^]) (see par. 3.2, 4.3.2 and 5). /h/ occurs as a phoneme distinctive from /k/ in NB and Si. In To and An /h/ is an allophone of /k/. Corresponding to NB /h/ we find *\^, *\^, *\^ in SB, and *\^, *\^, /h/ in Si. As NB /h/- is in a weak position and is sometimes lost, I reconstruct *\h in PB reconstructions which have NB correspondences with initial /h/- (sometimes alternating with *\^), and *\h in PB reconstructions with a corresponding PAN form that reflects NB /h/- (\=PAN \*\^) cf. \(^{(28)}\) with \*\gu, and \(^{(94)}\) with \*\ukur. There is a regular correspondence of NB /k/- to To, An \(\=h\) /k/, Ma /\^\k/, and Si /k/.

In inherited Si lexemes /k/, /h/- and /h/- which agree with NB /k/ occur in conditioned alternation: /h/ occurs in initial and in intervocalic position, /k/ occurs as final C in a syllable and in the synchronous patterns k\(\nuV\)(C), k\(\nuV(w)y\)\(\nu\)h, and (C)k\(\nuV\)h.\(^{(15)}\) For this correspondence I reconstruct \*\v; examples: SB, Si kutsu [Ma [\u]], NB kutsu 'loud'; PB \*kutsu

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 bati 'part, share'; PB \*bati

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

All dialects jadi 'to become; to be realized'; PB \*jadi

All dialects ajar 'lesson, instruction'; PB \*ajar

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 tulu, Si tulu, NB tulu, 'three'; PB \*tulu To, Si, Ka tiga 'market'; AnMa, Da 'commodity, merchandise'; PB \*tiga

SB, Si, Da urup 'to help, assist', Ka urup 'to do s.th. together (joint action), to assist, to stand by'; PB \*urup

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

SB, NB, Si lupa 'to forget'; PB \*lupa

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.

9.3.1 Voorhoeve (1955:9) considers Si as representing 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 dialects is due to borrowing, and has no bearing 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 occurring 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 vs) 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), (32), (34), (38), (44), (46), (52), (56), (65).

9.4 Palatals occur in all dialects, /j/ everywhere, /c/ in NB and AnMa, /R/ in AnMa; they occur in all dialects in initial and medial position; for the treatment of the vs /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 koqombul/katumbul, sampah, sapah/sgapah, ather/ather, ainiymbu/ siniymbu; Malay borrowings: cakap 'talk, language'; sampah 'stamp', Da tingolom/ siringolom, selebon/selebon, tilaka/tilaka, attuk/attuk, concaq/conqaq, atdahari/ atdahari; Malay borrowings: cakap 'stamp, seal'; cakap 'language', talk; probably from Ml: sangu 'prepared opium'; komin 'mirror'. The Ka borrowing minak aapi 'beef suet' (from Ml minak aapi) 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 /c/ corresponds to NB /c/ before /i/, e.g.: SB, Si nudiq, Ka nuviq, Da kodiq 'cat'; PB **kuviq (75)

SB tiop, Ka oikop, Da oikap 'to hold s.th. in the hand'; PB **oikap (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 vs in NB, whereas this sound merged with /s/ in SB and Si. There is good reason to suppose that a direct influence of Sanskrit on the Batak dialects, rather than through Malay or Javanese, moreover, there are, along with instances of a correspondence To, Si /s/ ~ AnMa /R/ ~ NB /c/ no instances of a correspondence To, Si /s/ ~ AnMa /R/ ~ NB /s/ in lexemes without initial nasal substitution. This suggests the possibility of an original AnMa *c with nasal substitute *ŋ (for /R/ 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 *Nc for the correspondence To, Si /s/ ~ AnMa /R/ ~ NB /c/, because 1) other dialects show no nasal feature for this correspondence; 2) /R/ is secondary in AnMa (see next par.), e.g. NB pisat 'to squeeze s.th., to massage breasts', To, Si pisat 'to pinch, to squeeze, to push', AnMa pisat 'pressed, pushed in'; PB *pi(N)sat (77)

To, Si pisat 'crushed, squeezed flat', Da liqat, AnMa liqat 'squeezed flat'; PB **li(N)sat (78)

To laqo, NB laksa, AnMa laNio '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 laksè.

PB *la*Nio, *laqo/+laqsa (77+78)

To haqo, Ka haqo, AnMa kaNio 'to disturb, to disarrange, make a mess' (I do not have an explanation for AnMa /k/ here); PB **kaNio (79)

To ugor 'to get wet; to put s.th. into water', Si ugor, AnMa apor to sink, to make water, to swallow water'; PB **apor (80)

Ka oqen, AnMa kqen 'tightened, of lips'; one of these forms shows metathesis; PB oqen/kqen or oqen/kqen I reconstruct (81)

For the correspondence NB /c/ ~ Si, SB /s/, e.g. SB, Si raun, NB raun 'poison'; PB raun (82)

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

To aukkalit 'difficult, troublesome', AnMa aukkalit/kukkalit 'knotty (wood)' (no Si or NB cognates); PB aukkalit (83)

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

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

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

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

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

SB sampur 'to mix up; to unite in marriage', Si sampor, Da sampor 'to mix up', Ka sampor 'to have an intimate relationship, of girls and boys; to mix up'; PB **sampor (83)
AnMa /r/ has /y/ as a variant in many lexemes. It also occurs as a nasal substitute of /s/, and corresponds with /n/, /s/, /ns/ or ō (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 /r/ other verbs will have. Van der Tuuk’s hypothesis of an older ō that was maintained in AnMa but changed elsewhere either to /n/ or to /y/ (š 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 ō/r/), thus where other dialects have /y/ (To š), 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. Hāti (‘to sing’; Ml nγañy) and Rāla (‘flame’; Ml nγñay). On the other hand AnMa ponu-ponu ‘turtle’ has no palatal N where PAN peñu has; with To ponu, id.; PB ōpeñu *(84)

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

The AnMa forms with palatal N do not give enough evidence for a PB palatal N, and I reconstruct *(N)N for the correspondence To š ~ AnMa /r/ ō Si, NB /y/, and *(N)C for the correspondence To Si /s/ ō AnMa /r/ ō 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 **Kayat/*Kayat/ *(86)

To so, AnMa oño/oyo, Si oyo, Da (o-)tah, Ka o-tah (/a/ unexplained) ‘urine’; PB **oøyoh *(87)

Examples of doublets with /y/ in AnMa: hañal/hayal, oño/ooyo, hu-róm/ huñom, oña/oyaq, pàñoq/pàภายq, huña-huña/huña-huña, the suffix -/muñu, -/muuyu/. There is also a doublet añam/unjom ‘to found a village’ that does not have doublets in the other dialects; To hasunjom ‘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 moni ‘mouse’ with cognates To, Si moni, NB moni ‘mouse’; PB **moni *(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 critical for the establishment of each proto-phoneme; e.g. SB toru, Si toruk, NB toruh ‘underneath’; PB **toru *(89)

SB torop, NB tóram ‘number, a lot, many; multitude’, Si torob ‘average, sufficient, mediocre (Ml sedang)’; PB **torob *(90)

SB sorot (van der Tuuk) ‘to sting, of insects’, Si sorod, Ka soron ‘sting’; PB **sore*(91)

SB, Si rogit, NB regit ‘mosquito’; PB **regit *(92)

SB, Si ruɣuɣ ‘to loosen, to disengage; to make everything ready’, Ka ruɣuɣ ‘to pour out s.th.’, Da ruɣuɣ golà-golà ‘morning between about 8 and 10 a.m.’; PB **ruɣug *(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 maɣuɣhur ‘to deceive’ (for h- see par. 9.3) **[h]ukur *(94)

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

SB, Si bórqi-n, Da bórqi-n, Ka bórqi ‘night’; PB *bórqi *(96)

(all dialects) lima ‘five’; PB **lima *(97)

SB laɣka, Si, Da laɣkah ‘step, pace, stride’ Ka laɣkah ‘behavior, way of acting’; PB **laɣkah *(98)

(all dialects) li-bag ‘pit, hole; full of holes’; PB **li-bag *(99)

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

SB očit, Si očlid, NB očil ‘to wipe after defecating’; PB **očlid *(101)

To toía, NB teleh ‘to settle a dispute, to part combatants’; PB **teleh *(102)

To arut/alut, AnMa arut, Si alud *(103)

NB alun ‘to massage’; PB **alud *(104)

Ka, Si salal ‘to shine (up) on; shining, glow’, To salal ‘clearly visible, clear’; PB **salal *(105)

SB, Si pógkal ‘hoops around a
9.6

Except for /ŋ/ (which is treated with the palatals in par. 9.4) nasals occur in all dialects in initial, medial and final position. On the basis of their correspondences I reconstruct *m, *n and *ŋ, e.g. (all dialects) mata 'eye'; PB **mata
(909)

PB **manuk
(910)

SB, dalom in 'di dalom laut'; 'to be in the womb, to be pregnant'
(911)

PB **dalom
SB nasog, Ka, Si nasog 'orang utan'; PB **nasog
(912)

(all dialects) mabar 'k. o. large rubber plant'; PB **mabar
(913)

NB, To, Si juma 'rice-field (usually 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
(914)

(all dialects) samun 'to rob'
(915)

PB **samun
SB, Si golom, NB golom 'to seize, take, to grasp with the hand'; PB **golom
(916)

(all dialects) ni, genitive marker; PB **ni
SB, Si, Ka nunut 'lasting, continuous'; PB **nunut
(917)

To panu-pano 'freckles', AnMa, Ka panu, Si panu [panow], Da panu-pano 'white spots on the skin'; PB **panow
(918)

SB boni, Si bonih, NB bonih 'seed'; PB **bonih
(919)

SB, Si dahan, NB dakan 'to cook rice'; PB **dakan
(920)

(all dialects) papan 'board'
(921)

PB **papan
(all dialects) qatag 'to gnaw, to chew, to nibble'; PB **qatag
(922)

SB, Ka qatag 'moth'; PB **qatag
(923)

SB, NB agin 'wind'; PB **agin
(924)

To, eaq 'illness that causes heavy feet' Ka hayaq 'heavy feeling, as after an attack of fever'; PB **hayaq
(925)

(all dialects) timbaq 'to weigh, to consider'; PB **timbaq
(926)

SB, Si bolagaq, NB bolagaq 'to cook'; PB **bolagaq
(927)

sakit, NB sakit 'illness'; PB **sakit
(928)

SB sia, Si siah, NB siah 'nine'; PB **siah
(929)

SB sombu, NB sombu 'appeased, satisfied, quenched (thirst)'; PB **sombu
(930)

PB **suluh
SB, Si suluh 'torch'
(931)

SB, Ka, Si surut 'to go back, to retreat' Da surut 'less; to go back, to retreat'; PB **surut
(932)

PB **asur
SB tasak 'finished, prepared (said of a luluk, a k. o. mushroom)', Si tasak 'dry; done, ripe', NB tasak 'ripe, done'; PB **tasak
(933)

SB mosok, Si mosog, NB mosog 'to burn, be burnt'; PB **mosog
(934)

(all dialects) asor 'bird's nest'; PB **asor
(935)

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

10 Accent

In the southern dialects and in Simalungun the accent is distinctive and its position is syntactically determined (see par. 3.4 and 6.3). Polymorphemic words have the accent on the penultima, irrespective of their syntactical position (except in cases treated in par. 3.4.2 and 6.4.3). In the Northern dialects 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 investigation of SB and Si morphology and lexicon 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 /o/.

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 relationship 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.
III COMPARISON OF PROTO-BATAK WITH PROTO-AUSTRONESEAN

12 Proto-Austronesian phoneme system

Consonants

<table>
<thead>
<tr>
<th>uvuvelar</th>
<th>pal-alveo-retrophic-bilabial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stops</td>
<td>g j/z/Z d D b v1 k s/c C,t T p</td>
</tr>
<tr>
<td>Nasals</td>
<td>q h N,n m</td>
</tr>
<tr>
<td>Liquids</td>
<td>vibrant R r lateral l</td>
</tr>
<tr>
<td>Sibilants</td>
<td>y w/N</td>
</tr>
<tr>
<td>Fricatives</td>
<td>x s</td>
</tr>
<tr>
<td>Laryngeals</td>
<td>q,n,h</td>
</tr>
</tbody>
</table>

Vowels

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

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)
   *bařiW > **bari, (Da bari, AnMa bari 'rotten, musty',
   To bari 'gone bad, spoiled (food)',
   Ka m-bari 'to be cold, of corpse'.

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

3) *-ay and *-ey to PB *-ey, e.g.
   *pahay > **pakey (65)
   *punay > **puney (29)
   *gatey > **gatay (28)
   *macay > **matey (27)

4) *t, *T and *C to PB *t, e.g.
   *maca > **mata (109)
   *eRat > **urat (11)
   *telu > **telu (62)
   *laniC > **laniq (9)
   *tuka > **tukuk, (NB tukuk 'to knock on s.th.', SB, Si tukuk 'beak; to knock on s.th.').

5) *r and *R to PB *r, e.g.
   *kerbow > **kerbow (23)
   *rantaw > **rantow (21)
   *wari > **wari (51)
   *lugar > **lur (107)

6) *d, *D and *Z to PB *d, e.g.
   *Dalan > **dalan (3)
   *ArE > **aar (17)
   *sered > **sered (91)
   *dzilat > **dzilat (8)
   *zadi > **jadi (60)

7) *n, *h and *N to PB *n, e.g.
   *taNen > **tannem (15)
   *(n)[u]pi > **npi (6)
   *peŋu > (**penu) (84)
   *punay > **puney (29)
   *bulaN > **bulaN, ((all dialects) bulan 'moon, month').

The following phonemes were lost:

| ? | e.g. *enem > **enem (14) |
| X | e.g. *apuy > **apuy (24) |
| S | e.g. *ma-busek > **mabuk (13) |
| E | e.g. *eyeh (87) |
| H | e.g. *gunaH > (**h) uma/juma (104) |

W, e.g. *nasu > **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. 
*qkur > **(h)ukur
*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: pe+toto+kən yields pepotokən, 
and po+idak+kən yields peidakən.

3. In loanwords original w becomes /o/, 
and the word gets an additional syllable, e.g. 
To sara-d-ar 'trousers' (v.d.Tuuk; Warneck 
gives saraoal), from Min. sarawal, (Ml. seluar, 
seluar, sarwual; orig. from Persiian); Ml. 
bawang 'onion' becomes To ba-ø-ɑq, 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 -/sə/ is realized as [ca] 
after final C, e.g. dapotca is pronounced 
dapotca.

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

7. This distinction of prefixation be- 
tween 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 dic- 
tionary, see par. 2.

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


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 
tervocalic /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 
KV(kV)(C), kV(w,y)Vh, or (C)Vkh 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 In- 
dian loanwords in Batak dialects which, accord- 
ing to his information, do not occur in other 
Indonesian languages. Many of these words how- 
ever do occur at least in Old or contemporary 
Javanese (like To humba 'water-jar'; To sarpa 
'serpent'; To mintonara 'magical formula'). 

Tideman gives an account of Indian in- 
fluence 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 patto pula 'forty', are 
also found in the Tamil lexicon.

17. The phoneme inventory presented here 
is that of Dempwolf 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 interpre- 
tation 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. Scan- 
dinavian Institute of Asian Studies Mono- 
digraph Series No. 15, Studentliteratur. 
Lund.

Dempwolf, O. 1934—38. Vergleichende Laut- 
lehre des Austronesischen Wortschatzes I, 
II und III. Berlin.

Dyen, I. 1953. The Proto-Malayo-Polynesian 
Laryngeals. Linguistic Society of Ameri- 
ca. Baltimore.

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

Eggink, H.J. 1936. Angkola- en Mandailing- 
Batakseh- Nederlandse Woordenboek. 
Verhandelingen van het Bataviaasch 
Genootschap 72/5. Bandong.

Gonda, J. 1973. Sanskrit in Indonesia. In- 
ternational Academy of Indian Culture, 
New Delhi.

Joustra, M. 1922. Nederlandse-Kamboeg 
Woordenlijst. Uitgaven van het Batakseh 
Instituut nr. 19. Leiden.
Kern, H. 1915. Over de herkomst van eenige
telwoorden in 't Bargoonsch der Zuidin-
dische kooplieden. Verspreide Geschrif-

Nederlands-Bataksche Woordenboek. Pro
Manuscripto in behalf of the missionaries
at the Vicar Apostolic in Medan. Babbe-
rich.

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.

Bahasa Simalungun. Seri bahasa dan bud-
ya Simalungun. Leiden.

Tideman, J. 1936. Hindoe-invloed in Noorde-
ljk 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 Lan-
guage and Anthropology. Leiden.

A Grammar of Toba Batak. KITLV transla-

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

--------------------------------------------- 1872.
't Lampons en zijne Tongvallen. Tijds-
sohrt 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-Deutches
Wörterbuch. Batavia. Reedited in 1977
at the Royal Institute of Language and
Anthropology. Leiden.

Wisman Saragih, J. 1936. Partingkian ni hata
Simalenggoen (Simalungun-Batak Dictiona-
ry). Pamatangraja.

* * *