

PACIFIC LINGUISTICS

Series C - No. 42

SOUTH-EAST ASIAN LINGUISTIC STUDIES

Vol. 2

Nguyen Dang Liem, ed.



Department of Linguistics
Research School of Pacific Studies
THE AUSTRALIAN NATIONAL UNIVERSITY

PACIFIC LINGUISTICS is published through the *Linguistic Circle of Canberra* and consists of four series:

- SERIES A - OCCASIONAL PAPERS
- SERIES B - MONOGRAPHS
- SERIES C - BOOKS
- SERIES D - SPECIAL PUBLICATIONS.

EDITOR: S.A. Wurm.

ASSOCIATE EDITORS: D.C. Laycock, C.L. Voorhoeve, D.T. Tryon, T.E. Dutton.

EDITORIAL ADVISERS:

- | | |
|--|--|
| B. Bender, University of Hawaii | A. Healey, Summer Institute of Linguistics, New Guinea |
| A. Capell, University of Sydney | N.D. Liem, University of Hawaii |
| S. Elbert, University of Hawaii | H. McKaughan, University of Hawaii |
| K. Franklin, Summer Institute of Linguistics, New Guinea | K. Pike, University of Michigan; Summer Institute of Linguistics |
| G. Grace, University of Hawaii | E. Uhlenbeck, University of Leiden |

ALL CORRESPONDENCE concerning PACIFIC LINGUISTICS, including orders and subscriptions, should be addressed to:

The Secretary,
PACIFIC LINGUISTICS,
Department of Linguistics,
School of Pacific Studies,
The Australian National University,
Box 4, P.O.,
Canberra, A.C.T. 2600.
Australia.

Copyright © The Authors.
First published 1976.

The editors are indebted to the Australian National University for help in the production of this series.

This publication was made possible by an initial grant from the Hunter Douglas Fund.

National Library of Australia Card Number and ISBN for complete set of volumes: 0 85883 144 9

National Library of Australia Card Number and ISBN for this volume: 0 85883 143 0

TABLE OF CONTENTS

	<i>Page</i>
VESTIGES OF MORPHOLOGY IN SOME TIBETO-BURMAN LANGUAGES, by Eugénie J.A. Henderson	1
AN EXAMINATION OF THE VOWELS AND FINAL CONSONANTS IN CORRESPONDENCES BETWEEN PRE-ANGKOR AND MODERN KHMER, by Judith M. Jacob	19
THE VALUE OF \bar{I} , I, \bar{U} AND U IN MIDDLE KHMER, by Philip N. Jenner	39
JAH-HUT, AN AUSTROASIATIC LANGUAGE OF MALAYSIA, by Gérard Diffloth	73
THE PHONOLOGICAL BEHAVIOUR OF MALAY PREFIXES WITH A NASAL ENDING, by Sidharta (Sie Ing Djiang)	119
A SUBGROUPING OF 100 PHILIPPINE LANGUAGES, by Teodoro A. Llamzon & Ma. Teresita Martin	141
RECONSIDERING THE NOTION OF FOCUS IN THE DESCRIPTION OF TAGALOG, by Joseph F. Kess	173
THE INTERPRETATION OF POTENTIAL ACTION IN BIKOL VERBS, by Malcolm Warren Mintz	187
GAYO CONSONANT CORRESPONDENCES, by H.L. Shorto	199
THE FUNCTION OF INDONESIAN IN CENTRAL JAVA, by John U. Wolff	219

	<i>Page</i>
FORMOSAN REFLEXES OF PAN NASAL/ORALS, by Paul K. Benedict	237
SOUND SYMBOLISM AND KHASI ADVERBS, by Lilli Rabel-Heymann	253

NOTE: A more detailed Table of Contents will appear
at the beginning of most articles.

EDITORIAL NOTE: Most of the articles included in this volume were received in 1974 or 1975, and theoretical views expressed in them may be at variance with the authors' present views on the subject. This note has been inserted at the specific request of the Editor for this volume on behalf of the authors.

VESTIGES OF MORPHOLOGY IN SOME TIBETO-BURMAN LANGUAGES

Eugénie J.A. Henderson

Introductory Remarks

1. Vestiges of Morphology in Initial Consonants
2. Vestiges of Morphology in Final Consonants
3. Vestiges of Morphology in Tinal Variation

INTRODUCTORY REMARKS

The first point to be made is the distinction between what may be called prosodic and segmental morphology. It seems to me perfectly acceptable that prosodic features such as tone or stress should in themselves be found to operate as morphological devices in the language of any part of the world, and as a member of the so-called London School, I should be inclined to include under 'prosodic' morphology, morphological processes involving the alternation of subsegmental features such as voice or voicelessness, aspiration or absence of aspiration, etc. It seems, however, to be the general view of the majority of the Sino-Tibetan philologists that prosodic morphology must in some sense be regarded as secondary to or derived from what may be called segmental phonology, i.e. that the grammatical use of aspiration, for example, or of tone, is in all probability to be ascribed to the operation of earlier segmental formatives which they have supplanted. I wish to present here some of the vestigial morphological features of Tibeto-Burman languages I have worked with in this latter light, whilst reserving the right to wonder whether our apparent conviction of the primacy of segmental morphology does not derive to a very large degree from our own traditional Indo-European standpoint - a standpoint to some extent challenged by the learning experience of our own children, in whose acquisition of speech tonal features for example, are in the early stages as important as, perhaps more important than, segmental ones.

I propose to exclude from consideration such overt morphological features as prefixes, or verbal particles which could be regarded as morphological elements, in order to concentrate upon phonological features which are an integral part of tonic syllables, and which appear to show signs of having at one time been the expression of live morphological processes. I shall be concerned therefore with

- (1) initial consonants
- (2) final consonants
- (3) tone

1. VESTIGES OF MORPHOLOGY IN INITIAL CONSONANTS

The principal feature that springs to mind in this connection is the well-known though rather limited use in languages like Burmese and Chin of a contrast between unaspirated and aspirated initials to express verbal relationships which may be loosely termed transitive/intransitive, or sometimes causative/noncausative.

Examples from Burmese include such pairs as the following¹:

- | | | | | | | |
|-----|----------|-------------|---------|-----------|-----------|----------|
| (a) | kwe: te | to break | (intr.) | hkwe: te | to break | (trans.) |
| (b) | kya.te | to drop | (intr.) | hkya. te | to drop | (trans.) |
| (c) | pwin. te | to open | (intr.) | hpwin. te | to open | (trans.) |
| (d) | nou:te | to be awake | | hnou: te | to awaken | (trans.) |

Sometimes Burmese spelling shows a similar relationship between pairs of words whose modern spoken form would not lead one to expect it,

- e.g. you' te to be inferior
 hyou' [ʃou?] te to put down

In Tiddim Chin one finds a few similar pairs, e.g.

- | | | | | | | |
|--|-------|-------------------|--|--------|----------------|----------|
| (e) | ˌkia | to fall | | ˌxia | to drop (tr.), | to fell |
| (cp., perhaps, second Burmese example above) | | | | | | |
| (f) | -ka:i | to be suspended | | -xa:i | to hang | (trans.) |
| (g) | -ka:k | to dilate (intr.) | | -xa:k | to open wide | (trans.) |
| (h) | -tu:k | to roll (intr.) | | -xu:k | to roll | (trans.) |
| (j) | ˌpu:k | to fall (intr.) | | ˌphu:k | to fell | |
| (k) | ˌka:ŋ | to rise, raise | | ˌxa:ŋ | to lift | |
| | | oneself | | | | |

Notice the unexpected relationship in (h) between t and x, not the expected th, which in this language is generally accepted as a reflex of earlier *s.

When I had the opportunity to work with a Lushai informant, I was not, unfortunately, looking for such forms. A somewhat cursory examination of Lorrain's dictionary did not produce any likely examples but R.B. Jones has drawn my attention to at least one pair of such forms: tliak² to break (intr.) and thliak² to break (tr.), which sug-

gests that others may exist.

Wolfenden² and Stern³ cite further forms for Sizang, another Northern Chin dialect:

- | | | | | |
|-----|-------|----------------------------|--------|----------------------------|
| (l) | ki:em | to grow less | khi:em | to decrease, make less |
| (m) | kɔm | to assemble, come together | khɔm | to collect, bring together |
| (n) | ka:i | to pull, be suspended | kha:i | to hang up (trans.) |

Scholars (Wolfenden⁴, Pulleyblank⁵) have ascribed this feature to the loss of a former s-prefix, such as exists and has a similar function in written Tibetan. Compare, for example, Tibetan *agyel*, *gyel* to fall, *sgyel* to throw down with examples (b) and (e) above. As La Raw Maran and others⁶ have shown Kachin still has a *ʃa/ʃa* prefix with a causative or transitive function in similar sets of words. It is pertinent to note, furthermore, that in spoken Tibetan the transitive/intransitive relationship is realized as one of absence or presence of aspiration, viz: [ky:gydu:] *he is boiling the water*, but [khy:gy:du:] *the water is boiling* (Sprigg)⁷. Pulleyblank has suggested that this feature might be "an important point from which to start in trying to establish the phonological isoglosses in Tibeto-Burman".⁸

Morphological or quasi-morphological alternation of voiced and voiceless initials is assumed by some scholars for Archaic Chinese, but as far as I am aware there is nothing in Archaic Chinese that corresponds to the s-prefix of Tibetan and Kachin, and the related alternation of aspirated and unaspirated initials in Chin and Burmese. This would therefore seem to be a genuinely Tibeto-Burman grammatical trait. It is possibly significant here that no trace of this trait has so far been reported for Karen. This would appear to support the current view that Karen is to be regarded as Sino-Tibetan but not as Tibeto-Burman (Luce)⁹.

Closer examination of Karen might however show up suggestive initial consonant relationships of other kinds, which are clearly linked to tonal features and possibly also to long vanished grammatical formatives of some kind. In Bwe Karen, for instance, in addition to the not uncommon phonetic variation between voiced and voiceless initials, as in *ci*² and *ji*³, both meaning *to knead*, one sometimes finds alternation between glottalised and non-glottalised initials, as for example *bɛ*¹ *to put, lay, keep*, and *bɛ*³, with the same range of meanings; there is also *da*² *to cut*, beside *da*² *to cut a foothold*; *bwe*¹ = *Bwe Karen*, and *bwe*² = *person*, and many more. One suspects a link too between *kɔ*¹, a prenominal prefix, and two preverbal formatives *gɔ*³ and *khɔ*¹. *kɔ*¹ denotes future time, as in *kɔ*¹*mu*²*nɛ*² *tonight*, *kɔ*¹*mo*¹*hɛ*² *this (coming) evening*, *kɔ*¹*məhɔ*² *tomorrow*, *kɔ*¹*dəhɔ*² *the day after tomorrow*, and also interrogatively *kɔ*¹*lɛ*³ *when?* (of future time, as contrasted with *phɔ*¹*lɛ*³ *when?*

of past time). The two preverbal formatives $g\text{ɔ}^3$ and $kh\text{ɔ}^1$, both referring to future time, are sometimes used interchangeably, but with $g\text{ɔ}^3$ indicating probability rather than certainty, e.g. $kh\text{ɔ}^1g\text{e}^1ph\text{ɔ}^2$ *It will (certainly) fall* as contrasted with $g\text{ɔ}^3g\text{e}^1ph\text{ɔ}^2$ *It will (probably) fall*. Much more work needs to be done on word-families of this kind. Is it fanciful to seek some link here with the alternation of voiced and voiceless initials proposed for Archaic Chinese words, sometimes in free variation, sometimes with linked but systematically differentiated meanings, together with the high and low tone registers associated with them? Bwe Karen and its closest related dialect Geba are exceptional among Karen languages in preserving the ancient distinction between voiced and voiceless stops, and in having a 3-tone system rather than the 5- or 6-tone system common to the dialects which have lost the old voice distinction. Bwe has voiceless unaspirated stops (p , t , k), voiceless aspirated stops (ph , th , kh), plain voiced stops (b , d , g) and voiced glottalised stops (β , d'). The plain voiced stops are associated with the two lower tones (mid and low); all the others, including the glottalised stops, with the two higher tones (high and mid). One might expect therefore that the cognates of pairs like da^2 and da^1 , bwe^1 and bwe^2 , in other Karen dialects would show a difference in tone, with or without an accompanying difference in initial. There is some hint of this in the tonal variation among semantically linked sets of words recorded by R.B. Jones for Palaychi.¹⁰

2. VESTIGES OF MORPHOLOGY IN FINAL CONSONANTS

The comparative rarity of forms showing the vestigial remains of the old s-prefix system in Tibeto-Burman languages is perhaps confirmation of Wolfenden's contention that prefixed forms in these languages are in general older than suffixed forms. It is certainly true that, in the Chin languages at least, alternations of final consonants, such as might be supposed to derive from older suffixed elements, are very much more numerous. These principally concern pronominal forms within the verbal phrase and the shape of verb stems themselves. I have given some account elsewhere of both of these characteristics as they occur in Tiddim Chin¹¹, and so will only summarize them briefly here.

2.1. PRONOMINAL INFLECTIONS

In formal literary Chin there is commonly a pronominal prefix before the verb and a phrase sentence final particle after it, e.g.

(a) $k\check{a}$ -pai \.hi *I go or I went*

The negative particle .kei or the future particle $\text{.di}:\eta$ may be inserted

between the verb and the following phrase-final particle, e.g.

(b) kǎ -pai ʔkei ʔhi I *didn't go*

(c) kǎ -pai ʔdi:ŋ ʔhi I *will go*

In colloquial style, however, the pronominal prefix and the phrase-final particle are omitted, and a pronominal suffix takes their place, viz.

(a) above becomes -pai ʔiŋ

Literary	nǎ -pai ʔhi	<i>you went</i>	becomes colloquial	-pai ʔtɛʔ
Literary	ǎ -pai ʔhi	<i>he went</i>	becomes colloquial	-pai
Literary	ʔ -pai ʔhi	<i>we incl. went</i>	becomes colloquial	-pai ʔhaŋ
Literary	kǎ -pai ʔuʔ ʔhi	<i>we excl. went</i>	becomes colloquial	-pai ʔuŋ
Literary	nǎ -pai ʔuʔ ʔhi	<i>you pl. went</i>	becomes colloquial	-pai ʔuʔ ʔtɛʔ
Literary	ǎ -pai ʔuʔ ʔhi	<i>they went</i>	becomes colloquial	-pai ʔuʔ

So far these all appear to be independent pronominal suffixes of the kind that I excluded from my study at the beginning of this paper. Upon examination of other colloquial verb forms, however, it turns out that certain inflexional elements may perhaps be discerned within some of these forms themselves, as for example the final velar nasal, which is regularly associated with first person forms, either singular or plural.

Compare:

Literary	Colloquial	
kǎ -pai ʔdi:ŋ ʔhi	-pai ʔniŋ	<i>I will go</i>
kǎ -pai ʔdi:ŋ -uʔ ʔhi	-pai ʔnu:ŋ	<i>We (excl.) will go</i>
kǎ -pai ʔkei ʔhi	-pai ʔkeŋ	<i>I didn't go</i>
ʔ -pai ʔkei ʔhi	-pai ʔxaŋ	<i>We (incl.) didn't go</i>

In the colloquial suffixes above there appears to be fusion of the future or negative formative, which is reflected in the first part of the syllable, with the pronominal element in the final consonant. There are a great many other such forms in colloquial usage.

2.2. VERBAL INFLECTIONS

Such accounts as we have of Lushai and of the Northern and Central Chin languages all bear witness to a widespread if rudimentary system of verbal inflection by which the great majority of verbs have at least two, and sometimes three, stems associated with different sets of grammatical contexts. The principal and most regular phonological device used nowadays to differentiate such stems is undoubtedly tonal variation - of which I shall say more below. Tonal variation is often accompanied

by variation in final consonants and sometimes the inflection is expressed by variation in the final consonant (or absence of consonant) alone, without accompanying tone change, as in the following examples from Tiddim Chin:

e.g.	Stem I	Stem II	
(a)	-gɔ:	-gɔ:t	<i>to dry up</i>
(b)	-pua	-puak	<i>to carry</i>
(c)	˩ha:	˩ha:t	<i>to be solid</i>
(d)	˩pha:	˩pha:k	<i>to overtake</i>
(e)	˩ne:	˩ne:k	<i>to eat</i>
(f)	˩la:	˩la:k	<i>to take</i>
(g)	˩ka:	˩ka:t	<i>to be forked</i>
(h)	˩pa:	˩pa:t	<i>to be thin</i>
(j)	_that	_thaʔ	<i>to kill</i>
(k)	_sut	_suʔ	<i>to snatch</i>
(l)	_sat	_saʔ	<i>to jerk</i>
(m)	_tat	_taʔ	<i>to strike against</i>
(n)	_kap	_kaʔ	<i>to cry</i>
(o)	_sak	_saʔ	<i>to be hard</i>
(p)	_pak	_paʔ	<i>to immerse</i>
(q)	_lak	_laʔ	<i>to show</i> etc.

These forms are so varied that I find it difficult to think of a historical solution in terms of, say, suffixation. There seems to be a certain underlying regularity in that all the long open Stem I forms (a) to (h) have a corresponding Stem II with long vowel and final -t or -k, whereas all the short closed Stem I forms (j) to (q) ending in -p, -t, or -k have a Stem II with final glottal stop. L.G. Löffler¹², who has investigated comparable material in Lushai, has pointed out that the set with the long Stem II forms in final -t are always intransitive, those in final -k always transitive, which clearly suggests the possibility of earlier suffixed forms and which could account for the absence of final -p in Stem II verbs of this kind. There are, however, apparent counter-examples, such as ˩xa:, ˩xa:k *to be bitter*, and ˩sia, ˩siat *to spoil*. Caution would seem to be advisable until cognate forms in other Tibeto-Burman languages can be identified. R.B. Jones¹³ cites a number of Lushai forms, verbs and verbal nouns, in which there is alternation of final vowel and final -k. The final -k forms here appear to parallel the Tiddim Chin use of Stem II forms for verbal nouns¹⁴, but do not suggest any regular association with transitive or intransitive verbs. Cp. pe^h *to give*, pek² *gift*; tlu^h *to fall down*, tluk² *fallen*; zu^h *to drink*, zuk² *drunk*; pua^h *to carry on the back*,

puak² *carried*; lua⁴ *to vomit*, luak² *vomit*; su⁴ *to wash (clothes)*, suk² *washed*.

3. VESTIGES OF MORPHOLOGY IN TONAL VARIATION

What may be termed the regular inflection of Tiddim Chin verbs operates as follows:

All verbs whose Stem I has a level or rising tone have a falling tone in Stem II, without consonant change. There is an interesting exception to this 'nonconsonant change' rule in that verbs whose Stem I consists of a syllable closed by a velar nasal regularly have an alveolar nasal in Stem II:

e.g.	Stem I	Stem II	
(a)	˩xa:ŋ	˩xa:n	<i>to lift up</i>
(b)	˩pa:ŋ	˩pa:n	<i>to defend</i>
(c)	˩za:ŋ	˩za:n	<i>to be light</i>
(d)	-tuan	˩tuan	<i>to perch</i>
(e)	˩gaŋ	˩gan	<i>to be profuse etc.</i>

A similar pattern obtains in Lushai and Sizang, and has been reported for Tibetan and some Naga languages.¹⁵ It would seem not unreasonable to suppose that some suffixed element, presumably a dental, might have been at work here. Pulleyblank has indeed suggested that some of the other consonant changes in the Chin verb suggest 'suffixation comparable to Tibetan final -s and da-drag.'¹⁶ These changes are regularly found when the Stem I form has a falling tone already. Stem II then has a falling or low tone plus the homorganic stop corresponding to the continuant final of Stem I.

The vowel of a Stem II of this type is always short and the pitch low level as contrasted with the falling pitch of Stem I. This low level pitch may, however, be regarded as the allotonic variant of the falling tone appropriate to short stopped syllables so that no tone change need be postulated for such forms. Morphologically the relevant features appear to be the shortening of the vowel and the final stop, viz:

(f)	˩kam	_kap	<i>to be dispersed</i>
(g)	˩la:m	_lap	<i>to lift up</i>
(h)	˩ge:m	_gep	<i>to creep up on</i>
(j)	˩i:m	_ip	<i>to keep secret</i>
(k)	˩am	_ap	<i>to be perplexed</i>
(l)	˩la:n	_lat	<i>to gape</i>
(m)	˩man	_mat	<i>to cost</i>
(n)	˩ba:n	_bat	<i>to reach for</i>

(o)	˘pan	_pat	<i>to start</i>
(p)	˘lanŋ	_lat	<i>to appear</i>
(q)	˘naŋ	_nat	<i>to be weak</i>
(r)	˘da:ŋ	_dat	<i>to be pale</i>
(s)	˘baŋ	_bat	<i>to be like</i>
(t)	˘ci:	_ciʔ	<i>to say</i>
(u)	˘hi:	_hiʔ	<i>to be</i>
(v)	˘mu:	_muʔ	<i>to see</i>
(w)	˘gai	_gaiʔ	<i>to consume</i>
(x)	˘ba:l	_balʔ	<i>to be covered with juice</i>
(y)	˘dol	_dolʔ	<i>to be damp etc.</i>

It will be seen that once again the Stem II form of verbs with a velar final in Stem I has the corresponding alveolar. There seems to be strong support here for the hypothesis of a dental suffixal element in Stem II, since a verb form ending in a velar nasal and having a falling tone may always be assumed to be a Stem I form, and never a Stem II form.

Compare such sets as

	Stem I	Stem II	
(a)	-da:ŋ	˘da:n	<i>to be infrequent</i>
(b) but	˘da:ŋ	_dat	<i>to be pale</i>
(c)	-naŋ	˘nan	<i>to defend</i>
(d) but	˘naŋ	_nat	<i>to be weak</i>

If the suffix theory is to hold water, it seems to me that there are four factors to be explained here:

- (i) the falling tone in forms (a) and (c) above,
- (ii) the accompanying change from velar to alveolar nasal in (a) and (c),
- (iii) the further change from nasal to stop in (b) and (d),
- (iv) the shortness of the vowel in syllables affected by (iii) above.

Not being myself a language historian, in the comments that follow I am asking for answers rather than hoping to supply them:

(i) If final -s or -h historically had the effect of inducing a falling pitch, as has, I believe, been supposed by Haudricourt¹⁷ and others, a final suffixed -s might perhaps, as Pulleyblank has already suggested,¹⁸ account quite neatly for the falling tone (or its allotone the low level) in the Stem II of the majority of Tiddim Chin verbs. It is clear from Stern's material that a falling tone is also present in a number of Sizang Stem II verbs also, but in how great a proportion is

not clear from the evidence available. The regrettable absence of tone marking in Lorrain's dictionary makes it impossible to discover how prevalent is the use of the falling tone in this grammatical context in Lushai until more evidence comes to hand.

(ii) The change from velar to alveolar nasal seems also fairly readily attributable to the operation of a suffixed dental element which did not affect stem finals with labial nasals, nor with velar final stops (this latter is perhaps more difficult to understand!):

	Stem I	Stem II	
Compare	(e) ˩xa:ŋ	˩xa:n	<i>to lift up</i>
but	(f) ˩xa:k	˩xa:k	<i>to close</i>

(iii) The change from continuant to stop in verbs with falling tone in Stem I could also, I imagine, be plausibly accounted for by the operation of a final voiceless (i.e. ^{*}s) suffix or by something akin to *da-drag*, which I take it is assumed to represent a dental stop of some kind. The difficulty here as I see it is to reconcile (ii) and (iii), which suggest that two morphological processes must have been at work rather than one. We cannot surely maintain that *-naŋ + ^{*}s → ˩nan*, whereas *˩naŋ + ^{*}s → ˩nat*.¹⁹

Further fodder for morphological speculation is supplied by some of the other derivation processes connected with the Chin verb. Stem II regularly supplies the form for verbal nouns, e.g.

	Stem I	Stem II		Noun
(g)	-la:m	˩la:m	<i>to dance</i>	˩la:m <i>a dance</i>
(h)	-na:k	˩na:k	<i>to breathe</i>	˩na:k <i>nose</i>
(j)	-man	˩mat	<i>to catch</i> (irreg.)	˩mat <i>prisoner</i>

Does older Tibetan and Chinese practice suggest that the same suffix as in the verb might have been at work here? It is perhaps of interest that the Chin nominalising suffix that may be used to form verbal nouns (always based upon the Stem II form of the verb) is *-na:*, which itself may be regarded as a tonally inflected form of *˩na:* *thing, object* (see below).

A further derivation process is that of forming causatives or benefactives from Stem II, e.g.

	Stem I	Stem II	
(k)	-ta:ŋ	˩ta:n	<i>to be bright</i>
Derived form:	˩ta:n	˩tat	<i>to flash a light at s.</i>
(l)	-nam	˩nam	<i>to smell</i> (intr.)
Derived form:	˩nam	˩nap	<i>to smell</i> (tr.)
(m)	˩dim	˩dim	<i>to be full</i>
Derived form:	˩dim	˩dip	<i>to fill</i>

	Stem I	Stem II	
(n)	ˌlam	ˌlam	<i>to earn for oneself</i>
Derived form:	ˌlam	ˌlap	<i>to earn for someone else</i>

All the above verbs, primary or derived, form their second stems perfectly regularly, taking the Stem I as the base; the Stem I of the derived form being the Stem II of the primary form in each case. Occasionally the Stem II of the derived form does not undergo the process whereby the continuant final becomes the corresponding stop, so that we have:

	Stem I	Stem II	
(o)	ˌdan	ˌdan	<i>to be different</i>
but	ˌdan	ˌdan	<i>to differentiate</i>
		(not *ˌdat)	
(p)	ˌdam	ˌdam	<i>to be well</i>
but	ˌdam	ˌdam	<i>to heal</i>
		(not *ˌdap)	

As Pulleyblank has said,²⁰ the complications and irregularities of the Chin verbal system suggest that this system 'stands a better chance of reflecting features of common Tibeto-Burman' than the relatively regular system of tonal inflection found in Tiddim Chin nouns (of which more later). One would be glad, however, to discover traces of this putative verb morphology in other Tibeto-Burman languages. Kachin is said to have traces of it, but so far as I am aware there is nothing comparable in Burmese, except perhaps in the rare cases of semantically related verbal forms differentiated by tone, such as *hcai'/hcaïn. hollow-topped/to be concave*; *ngoun/ngoun. to hold in the mouth/to hold the head down*; etc.²¹ Karen, though more remotely related, has a few tonal features which may perhaps be connected - e.g. the tonal variant forms cited by R.B. Jones for Palaychi.²² Perhaps the tonal variation which is sometimes a part of the Bwe Karen compounding process is also relevant here, e.g. (where ¹ = high level, ² = mid level, ³ = low level):

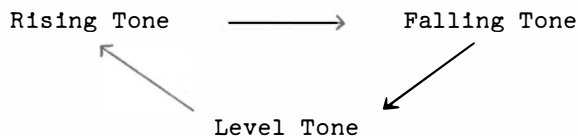
(q)	ɬɔ ²	<i>to speak</i>	beside	ɬɔ ¹ ʃa ²	<i>to tell</i>
(r)	la ²	<i>to descend</i>	beside	la ¹ de ³	<i>to fall</i>
(s)	ca ²	<i>to see</i>	beside	ca ¹ le ²	<i>to search</i>
(t)	ɬa ²	<i>to cut</i>	beside	ɬa ¹ the ¹ pha ²	<i>to cut off</i>

What is particularly striking, however, is the way in which the uses of Chin Stem II forms - as verbal nouns, causatives, in compounding, etc. - resemble those of the derived *chiuh-sheng* forms in classical Chinese, as described by Downer²³. The relationship between the derived *chiuh-sheng* forms and the corresponding basic forms was evidently one of tonal contrast, sometimes associated with an alternation between

voiced and voiceless initials, and possibly in some instances with the loss of a former suffix.²⁴ Downer argues that morphological derivation by tonal variation may date back to Archaic Chinese and is thus a very old feature in the Sino-Tibetan language family. Karlgren's reconstructions suggest an alternation not of tones but of voiced and voiceless final stops. Either way, there is evidence for the very ancient origins of such forms in Chin and allied languages.

3.1. TONAL ALTERNATION IN THE NOUN IN TIDDIM CHIN

For those unfamiliar with the account I have given of this phenomenon elsewhere,²⁵ a brief summary may be of interest here, in case similarities are forthcoming from other languages. Tiddim Chin nouns and pronouns have two alternating forms, the commonest of which I shall refer to as the 'direct' form, the less common as the 'oblique' form. The alternation is tonal, and the forms are mutually predictable. (This is in contrast to the verbal forms, in which Stem II may be predicted from Stem I but not vice versa). The tonal variant found in a given context is grammatically not phonetically determined, and is thus not to be confused with tone sandhi, such as is common in other languages of the family. The tonal relationship between the direct and oblique forms of nouns is illustrated by the diagram below, a direct rising tone implying a falling oblique tone, a direct falling tone implying a level oblique tone, and a direct level tone implying a rising oblique tone:



In nouns of more than one syllable the tonal alternation affects the last syllable only. The oblique form of nouns is found (i) in genitival constructions, (ii) before certain suffixes, (iii) in certain compound nouns, and (iv) is possibly also used in the case of nouns of inanimate or abstract reference to express what appears to be a lexico-semantic rather than a grammatical relationship between the two forms.

Examples will make this clearer:

(i) In genitival constructions

	Direct form	Oblique form of first noun
(a)	˩ha:u ˩gou <i>Haugo</i> (a name)	˩ha:u -gou ˩ʔwi <i>Haugo's dog</i>
(b)	˩vul zǎ -thaŋ <i>Vul Za Thang</i> (a name)	˩vul zǎ ˩thaŋ ˩ʔwi <i>Vul Za Thang's dog</i>

Direct form	Oblique form of first noun
(c) -ga:l ɽtɛ:	-ga:l ɽtɛ: -xuaŋ
<i>The enemies</i>	<i>The enemies' drum</i>

(ii) Before certain suffixes

The oblique form is obligatory before certain postnominal suffixes, such as the masculine and feminine suffixes ɽpa: and ɽnu:, but not before others, such as the pluralising suffix ɽtɛ:.

Compare:

Direct Form	Direct Form + suffix	Oblique Form + suffix
(d) -ga:l	-ga:l ɽtɛ:	ɽga:l ɽpa:
<i>war, enemy</i>	<i>enemies</i>	<i>an enemy</i>

If a noun with a postnominal suffix enters into a construction that requires an oblique form the suffix is treated as the last syllable of the relevant nominal form and is varied tonally in the same way as the last syllable of nouns:

Compare:

Direct	:	Oblique	:	Direct	
(e) -ga:l		ɽtɛ:		-xuaŋ	<i>The enemies' drum</i>
Oblique	:	Oblique	:	Direct	
(f) ɽga:l		ɽpa:		-xuaŋ	<i>The enemy's (singular) drum</i>

In (e) above, -ga:l is the direct form appropriate before the suffix ɽtɛ:, while ɽtɛ: is the oblique form of the suffix ɽtɛ: appropriate to the expression of the genitival relation between -ga:l ɽtɛ: and -xuaŋ. In (f) both ɽga:l and ɽpa: are oblique forms, the first by reason of its position preceding the suffix ɽpa:, and the second by reason of the genitival relation between ɽga:l ɽpa: and -xuaŋ.

(iii) In compounding

A single example in my material - which is all too thin for any but the most tentative generalisations here - suggests that in certain compound nouns consisting of a noun + a verb the oblique form of the noun is used. The example is ɽmi: -hiŋ *human being* which seems clearly to derive from ɽmi: *person* and -hiŋ *to be alive*. In the noun + noun compounds occurring in my material both nouns are in the direct form.

(iv) Lexico-semantic relationships

In one or two interesting cases, all (perhaps fortuitously) referring to temporal expressions, it is possible to suggest very tentatively that we may have direct/oblique tonal alternation with lexical function:

e.g.

	Direct		Oblique	
(h)	˩za:n	<i>night</i>	-za:n	<i>yesterday</i>
(j)	-zi:ŋ	<i>morning</i>	ǎ ˩zi:ŋ	<i>next morning</i>
(k)	˩tu:	<i>now</i>	˩tu: ˩ni:	<i>today</i>
	-ni:	<i>day</i>		

Alternation of this kind defies explanation in terms of the working of some lost segmental formative. Pulleyblank has pointed out that its regularity "suggests that there has been innovation and analogical extension at a comparatively recent date."²⁶ I must confess to a certain unease here in that the innovation seems, as far as has been reported so far, to be confined to Tiddim Chin. Theodore Stern specifically looked for similar behaviour in Sizang, a fairly closely related dialect, but found none. I should be greatly reassured if some similar system were to turn up elsewhere in the family. The nearest similar case seems to be the use of the so-called 'induced creaky tone' tone,²⁷ which is found in some of the same grammatical contexts as Tiddim Chin oblique forms, viz:

In Genitival Constructions

- (l) min: *you* but min. tha: *your son*
- (m) hsain *shop* but hsain. hyin: *owner of the shop*
- (n) hsǎya *teacher* but hsǎya. ka: *teacher's car*
- (o) di lu *this man* but di lu. ka: *this man's car*

Mrs Allott points out that it is possible in such expressions to use the suffix *ye.* after the first nouns, in addition to the use of 'induced creaky tone', and that it has been suggested that the latter may derive historically from a former possessive suffix, such as Tibetan *kyi, gyi, i.* In the earliest Burmese inscription (c. A.D. 1113), however, the most usual way of showing possession was through tonal alternation of the type described above.

Before certain suffixes or particles

'Induced creaky tone' is reported by Allott in the last syllable of expressions suffixed by the 'sentence particle' *-kou*, e.g.

- seitǎna *generosity* but seitǎna. kou māhyi. hpu *(he) is not at all a generous person. etc.*

She reports similar tonal behavior in pronouns, names, titles and kinship terms followed by the 'noun-marker' *-kou* (which she distinguishes from the *-kou* cited in the previous paragraph), or by the 'noun-marker' *-hma.*

In compounding

The first elements in some compound numeral expressions in Burmese are marked by 'induced creaky tone', as are the first elements of many reduplicated expressions, e.g.

(tǎ)hse *ten* but hse. thoun: *thirteen*
 hpyu *white* but mǎhpyu. tǎhpyu *whitish* etc.

This fairly extensive use of a particular tone in Burmese for grammatical purposes, though superficially similar in some respects to the tonal variation reported for the Chin noun, is in fact much more readily explained by the postulation of an earlier suffix than is the Chin phenomenon.

POSTSCRIPT

Since the substance of this paper was first conceived four or five years ago there have been important developments in Sino-Tibetan historical linguistics which prompt me to add two brief comments.

(1) It now seems to be accepted by scholars like Pulleyblank, Bodman and Benedict that *s-prefixes must be reconstructed for Old Chinese, and probably for the Sino-Tibetan proto-language itself. It is also assumed that the causative *s-prefix is an original Sino-Tibetan feature, not confined to Tibeto-Burman as suggested on page 3 above.²⁸

(2) Recent work by Weidert on Lushai confirms the occurrence of tonal alternations in Lushai kinship terms, personal names, etc. which are sufficiently similar to those observed for Tiddim Chin to allay the 'unease' I formerly felt at the apparently quite exceptional behaviour of Tiddim in this respect (see page 13 above).²⁹

N O T E S

1. The Burmese examples in the paper are transcribed according to the system used in Anna J. Allott, 'Grammatical Tone in Modern Spoken Burmese', *Wissenschaftliche Zeitschrift der Karl-Mark Universität Leipzig*, 1967.
2. Stuart N. Wolfenden, *Outline of Tibeto-Burman Linguistic Morphology*, pp.185-6.
3. Theodore Stern, 'A provisional sketch of Sizang (Siyin) Chin', *Asia Major (New Series)*, X, 2, p.251.
4. Op. cit. pp.185 and 199-200.
5. E.G. Pulleyblank, review of Henderson's *Tiddim Chin* in BSOAS, XXIX, Pt 2, p.422.
6. cp. Wolfenden, op. cit. pp.85-6, 199-201; H.F. Hertz, *A Practical handbook of the Kachin or Chingpaw Language*, p.15; Professor La Raw Maran in verbal communication.
7. For this and other examples see R.K. Sprigg, 'Verbal Phrases in Lhasa Tibetan - I', BSOAS XVI, Pt I, pp.155-6.
8. Op. cit. p.422.
9. See G.H. Luce, *Phases of Pre-Pagān Burma: Language and History* (forthcoming).
10. R.B. Jones, *Karen Linguistic Studies*, pp.77-8.

11. Eugénie J.A. Henderson, *Tiddim Chin*, O.U.P. 1965, pp.72-89, and 108-113.
12. In a personal communication.
13. In a personal communication. The translation of the final -k forms in these examples by nouns or past participles in English appears to parallel the Tiddim Chin use of the Stem II form for verbal nouns. See Henderson, *Tiddim Chin*, pp.86-9.
14. See Henderson, *Tiddim Chin*, pp.86-9.
15. Verbal report from N.C. Bodman and J. Matisoff, respectively.
16. Op. cit. p.423.
17. A. Haudricourt, 'De l'origine des tons en vietnamien', *Journal Asiatique*, 1954, pp.80-2.
18. Op. cit. p.423.
19. L.G. Löffler, in a personal communication dated March 1972, says that he would indeed maintain this, 'although I would prefer a dental instead of -s'. He suggests that Stem II is to be derived from 'a voiced (phonetically low) dental final, say *d', which in the example under discussion gave rise to the developments:
1. Stem II *nańd or náńd → *nań' → nàn;
 2. Stem I *nań' → nàń, Stem II *nań'd → *nańd' → *nàn' → nàt.
- The 'stopped nasals' proposed as intermediate stages are attested by Löffler as occurring in Bawm, a Chin language of the Chittagong Hill Tracts.
20. Op. cit. p.422.
21. See Anna J. Allott, op. cit. pp.159-161.
22. Op. cit. pp.77-8.
23. G.B. Downer, 'Derivation by tone-change in Classical Chinese', BSOAS XXII, Pt 2.

24. See A. Haudricourt, 'Comment reconstruire le chinois archaïque', *Linguistics Today*, ed. Martinet and Weinreich, p.244.
25. In Eugénie J.A. Henderson *Tiddim Chin*, pp.69-71.
26. Op. cit., p.422.
27. See Anna J. Allott, op. cit. pp.159-160.
28. See N.C. Bodman, 'Old Chinese s-Clusters, Some Dialect Alternations, and Traces of the Sino-Tibetan s-Causative', paper submitted to Fifth International Conference on Sino-Tibetan Language and Linguistic Studies, Ann Arbor, 1972; E.G. Pulleyblank, 'Some New Hypotheses concerning Word Families in Chinese', *Journal of Chinese Linguistics*, 1, 1, 1973; P.K. Benedict, 'The Chinese s-orgy: further adventures and misadventures', paper submitted to Eighth ILCSTLL, Berkeley, 1975.
29. Alfons Weidert, *Componential Analysis of Lushai Phonology*, Amsterdam, 1975.

AN EXAMINATION OF THE VOWELS AND FINAL CONSONANTS IN CORRESPONDENCES BETWEEN PRE-ANGKOR AND MODERN KHMER

Judith M. Jacob

Preparation of Table I
Summary of the information provided by Table I
PA-MK correspondences and OK Phonology
Appendix of examples with reference to Table I

This study is intended to be complementary to a paper on the correspondences between Old and Modern Khmer initial consonants (Jacob: 197?). Only Pre-Angkor Khmer (PA) is dealt with here, however, and not the Angkor dialect (A). It seemed desirable to limit the investigation to PA, partly because of the inclusion of A, with its different vowel-notations, would cause confusion and partly because the author now has a complete card-index of PA,¹ whereas that for A is still in progress.

The majority of correspondences between Old Khmer (OK) and Modern Khmer (MK) involve the same final consonant at both stages of the language. Two regular exceptions to this are final *r*, which has turned to zero in MK (except in the Battambang area) and final *s*, which has completely merged with the final aspirate in MK. An example of each final consonant of PA with its MK correspondence is given below:

Final consonant	PA orthography	MK pronunciation	Meaning
k	sruk	/srok/	<i>inhabited area, district</i>
ñ	sroñ	/sroŋ/	<i>bathe</i>
c	roc	/rò:c/ ²	<i>wave</i>
ñ	añ	/ʔaŋ/	<i>I</i>
t	ket	/kaet/	<i>wax, be born</i>
n	mān	/mì:ən/	<i>have</i>

Final consonant	PA orthography	MK pronunciation	Meaning
p	cap	/cap/	<i>seize</i>
m	psam	/phsɔm/	<i>unite</i>
y	toy	/daoy/	<i>along</i>
r - zero	dār	/tì:ə/	<i>claim v.</i>
l	jmol	/chmò:l/	<i>male</i>
v	jāhv	/cì:əv/	<i>barter</i>
s	thās	/tha:h/	<i>tray</i>
h	vrah	/prèəh/	<i>holy; god</i>
Zero	lvā	/lvì:ə/	<i>fig</i>

These regular correspondences between 15 PA and 13 MK final consonants were taken as the basis of the comparison. A table was made, using over 200 well-established correspondences, so as to find out their frequency and versatility of occurrence with vowels and arrange them in order. This is the order in which they are presented in Table 1, except that h is placed next to s for easy comparison.

Up to this point, the enquiry had been made with only established vocabulary. When the validity of certain empty slots in the table was tested by reference to some less certain correspondences, however, it became apparent that a much wider variety of vowels and final consonants was probably in operation. The PA card-index was therefore searched for additional probable correspondences, as much in order to avoid claiming empty slots mistakenly as to try to fill more slots. This led to the preparation of Table 1, which forms the basis of this study.

PREPARATION OF TABLE 1

The vocabulary used comprises:

ESTABLISHED CORRESPONDENCES. These consist of (i) correspondences established by the occurrence of the PA form in a grammatical context with clear meaning and (ii) correspondences established by analogy, the PA form being in a context which clearly indicates the kind of word to be expected. In a place name, for example, *vrai forest* and *stuk pool* are regularly followed by names of trees. Thus, a word such as *ramteñ*, which on K.430 follows *vrai*, and which also has a form which suggests the correspondence, is held to be an established correspondence for MK /rùmde:ŋ/ *amomum galanga* (zingibéracées).

NAMES. Although, as has just been stated, words occurring in place-names may be held to be established when the context makes their nature clear, other words occurring either as place names or personal names

or parts of such names may be held to be probable but not established correspondences. Cambodian names, even now, are often names of animals, plants or objects; or they may be descriptive of physical or mental characteristics. Where such words are used as material for the present study the PA word is followed by 'N.' in the Appendix and, if the vowel-symbol in the table depends entirely on examples which are names, it is followed by an asterisk.

In identifying correspondences of the two categories just described, certain regular relationships between PA and MK initial clusters were regarded as being dependable. These are:

(i) The relationship between a MK base and a PA form consisting of base with name-prefix, kN. This prefix, which consists of initial k, neutral vowel a, as junction, and nasal consonant, was tentatively included in an analysis of OK affixes (Jacob:1963,pp.69-70). This now seems well substantiated by the preponderance of names beginning with kN in contrast to other prefixes and by information supplied by Mr. Kuoch Hak Srea, that a kN name-prefix is used as a term of endearment before girls' names still. Some examples are given below with a reference to the number of an inscription on which they occur and the MK equivalent:

kañcañ 562	/caŋ/ <i>malaria</i>	kandai 561	/tèy/ <i>cloth bag</i>
kañjāhv 1030	/cì:əv/ <i>barter</i>	kansuc 582	/soc/ <i>N. of insect</i>
kantrT 748	/trɨy/ <i>fish</i>	kan-a 808	/ʔɔ:/ <i>flute</i>
kanden 926	/tìən/ <i>candle</i>	kan-ek 18	/ʔæk/ <i>noisy device attached to a kite</i>

(ii) The relationship between a MK base and a PA form consisting of base with infix -an-. This correspondence was mentioned before (Jacob: 197?) but more examples have now been found, of which some are given here:

kanmoy 38	/kmu:əy/ <i>nephew, niece</i>	canmat 126	/chmat/ <i>milk (cow)</i>
kanmau 357	/khmau/ <i>black, dark</i>	ransi 134	/rəsɨy/ <i>bamboo</i>
tuñnot 9	/tñaot/ <i>sugar-palm</i>	suññay 9	/snāy/ <i>salvadora capitulata</i>
danhum 424	/thùm/ <i>scented, smell</i>	anlik 480	/ʔolɨk/ <i>melon</i>

(iii) The relationship in which the PA form has the 'wrong' consonant as the first consonant of an initial sequence, i.e. surd where one expects sonorant or vice versa. This type of relationship was accounted for as being due to loss of voicing in this pre-consonantal context and examples were given (Jacob:197?).

Table 1 represents the result of examining over 550 correspondences. Down the left side are arranged all MK vowels, the pairs of vowels produced on the two registers being kept together in accordance with

the written symbol with which they are now spelt, since the written tradition is held to represent a development which has taken place from a time when the surd and sonant initial consonants genuinely represented surd and sonant initial represent vowels diverging according to register. In each slot is placed one or more vowel-symbols, representing PA spellings found in the correspondences. An asterisk indicates that the only examples available are names. For symbols without an asterisk, established correspondences are offered. The Appendix gives as far as possible two examples for each vowel-symbol in the table, reading from left to right.

Notes with reference to Table 1

(i) It is held that the vowel occurring before y , v , c and \tilde{n} and having in those contexts, in MK, the realisations $ay/\grave{e}y$, $au/\grave{u}u$, $ac/\grave{e}c$ and $ap/\grave{e}p$, according to register, may best be regarded as being a plus final consonant. Strictly according to the conventions of the present table, their representations on the second register would be: $\grave{e}y$, $\grave{e}v$, $\grave{e}c$ and $\grave{e}p$.

(ii) Short e/\grave{e} and i in MK occur, in native words, apart from \grave{e} in $\grave{e}p$ and $\grave{e}y$ mentioned in (i), only before h . Since MK long $e:$ and $i:$ cannot occur before h , the few examples of MK $eh, \grave{e}h$ and ih are entered along the $e:, \grave{e}:$ and $i:$ lines to save space.

SUMMARY OF THE INFORMATION PROVIDED BY TABLE 1

1. USE OF THE PA VOWEL-SYMBOLS

(i) Consistency. It will be observed that in the case of two MK vowel-pairs, PA correspondences were represented consistently by one vowel-symbol with all final consonants. Thus PA vowels \sim MK $ao/\grave{o}:$ were exclusively represented by o and PA vowels \sim MK $ae/\grave{e}:$ were exclusively represented by e . Two other pairs were almost as consistent: PA vowels \sim MK $i:\grave{a}$ were exclusively written with \bar{a} , (though a few PA words of which the vowel \sim MK partner vowel $a:$ were written with a) and PA vowels \sim MK $ae/\grave{e}:$ were all represented by e , with the exception of $kamrat\bar{a}\tilde{n}$ (written $kamrate\tilde{n}$ in the A period, however) and one non-established word.

The short vowel-symbols, i and u , were fairly consistently used to write respectively the vowels \sim MK y/\grave{u} and o/\grave{u} . In each case, however, some confusion of spelling with the more open vowels, e on the one hand and o on the other, took place. Open i and u also represent vowels \sim MK $y/\grave{i}:$ (open) and $o:v/\grave{u}:$. The extent to which i and u occur apart from these uses is minimal.

While \bar{a} , i and u were chiefly used to write vowels \sim one MK vowel-pair, a , o and e were used to write vowels corresponding with a variety of MK vowel-pairs. Table 2 summarises these uses.

(ii) Register difference. PA vowels \sim MK vowels of one register are sometimes represented by one vowel while correspondences with the MK partner vowel on the other register are not. It seems that this may indicate the beginning of a differentiation between the vowels of the two registers. The details are therefore summarised here:

(i) As was mentioned above, a occurs as well as \bar{a} to represent PA vowels \sim MK $a:$, but not $\dot{i}:\bar{e}$.

(ii) u occurs alongside o or a to represent the vowel \sim MK $\dot{u}:\bar{e}$, but not \bar{o} .

(iii) The experimental $va, v\bar{a}$ were used in the PA period to write the diphthong \sim MK $\dot{u}:\bar{e}$ in second register words only.

(iv) The experimental ya, ye were used in the PA period to write the diphthong \sim MK $i\bar{e}$ in first register words, while $ya, y\bar{a}$ occur in words \sim MK second register words.

2. MK VOWELS AND FINAL CONSONANTS FOR WHICH PA HAS NO CORRESPONDENCES

(i) The first observation which must be made is that MK $w\bar{e}/\dot{u}:\bar{e}$ and \bar{e} : have no entries in the Table. To the best of the writer's knowledge, no correspondence for any of these vowels occurred in PA or A. The first occurrences which are datable are probably the following, all Thai loanwords, found in the Middle Khmer inscriptions (Inscriptions Modernes d'Angkor): $m\dot{u}:\bar{e}\eta$ on No.8 (1625 A.D.)l. 19; $kr\dot{u}:\bar{e}\eta$ and $(r\dot{u}:\bar{e})-r\dot{u}:\bar{e}\eta$ on No.12 (1638 A.D.), ll. 16 and 19. $l\dot{v}:\bar{n}(-l\bar{e}\eta)$ occurs on No.38 (1701 A.D.),l.26.

MK \dot{u} : is represented by only one correspondence in PA, $gi, gui, \sim /k\dot{u}:/$ that *is, is*. In MK, although it is not widely distributed, \dot{u} : occurs before \dot{n}, t, p and h .

(ii) An examination of the MK vowels and final consonants which are not represented in PA correspondences showed that the chief lack of correspondence occurred in connection with MK forms having the vowels $\bar{a}o/\bar{o}:\bar{e}, o:\bar{u}:\bar{e}$, and $u:\bar{e}/\dot{u}:\bar{e}$ and having palatal final consonants, c, \bar{n} or y . It happens that many PA words ending in $o\bar{n}$ seem to have no correspondence with MK.

3. CORRESPONDENCES FOR WHICH TABLE 1 DOES NOT PROVIDE A SLOT

Although generally speaking the correspondence between 15 PA and 13 MK final consonants is regular and reliable, there are some exceptions:

(i) In spite of regular correspondences between PA and MK \dot{n} and PA

and MK ñ, as in *tiñ* - /dɿŋ/ *know*, *duñ* - /tèŋ/ *buy*, etc., there is an instance of ñ going to ɿ: *tmiñ, tmñ* 557 (cf. MK /dɿŋ/ *pluck strings*) *player of stringed instrument*. MK /vèŋ/ *back, again* also goes back to an old form with velar nasal final consonant. When this is considered in conjunction with the variation between ñ and ñ̄ in the final consonant of several OK titles (*tāñ* and *tāñ̄*, *mratāñ* and *mratāñ̄*, *steñ* and *steñ̄*), it seems probable that we are dealing here with dialectal differences.

(ii) One PA word is written both with and without final r. Thus we have *camkā* 426, 115 and *camkār* 664 ~ MK /cɔmkɑ:/ *plantation, market-garden*. This may be a dialectal variation or a mistake. In Middle Khmer poetry r rhymed with l (Jacob:1966,232); this would suggest that at least it was not absent in the pronunciation of speakers in the earlier PA and A periods. One or two words which have in MK final l were written with r in PA:

kantor N. 1030 /kɔndol/ *rat, mouse*
sampor N. 910 /sɔmbol/ *complexion*
knur N. 557 /khnol/ *jack-fruit*

(iii) Written s and h as final consonants, now both realised as the aspirate, were also rhymed together in Middle Khmer poetry. It is possible that already there was a merging of pronunciation of the two final consonants at the Middle Khmer stage and that it was after that that s was used in writing to indicate certain long vowels preceding it, while h indicated a short vowel preceding it. In any case, certain correspondences between PA s and MK written h seem acceptable. M. Cœdès identified PA *pas* with MK /bɔh/, written *poh*, for example, in interpreting PA *pamas* as *grinder*; and some names suggest such a correspondence, e.g. *trases* 765 /trəseh/, written *traseh* N. *of a kind of woodpecker* and *kañcus* 956 /kɔpcoh/, written *kañcuh* N. *of a small fish with two spears*.

PA-MK CORRESPONDENCES AND OK PHONOLOGY

The question of what vowels were actually operating at the phonological level in PA times now faces us. In an earlier study (Jacob: 1960) the writer attempted to analyse OK phonology from a synchronic viewpoint. She would, however, be the first to admit that more progress has been made by diachronic methods. The combined vowel phonology of PA and A has been admirably tackled by Professor Sakamoto in a series of four articles (Sakamoto:1970-4). He uses two methods to analyse his material. One method is the examination of the spelling of the vowels in individual words corresponding with a given MK vowel, to show where there was confusion and where not. The confused spellings, taken in

conjunction with the MK corresponding vowel, show that a comparable distinction was already in operation but that it was a distinction for which symbols were lacking. The second method is the examination of the initial and final consonants of correspondences to show that where OK vowels, written with one symbol, correspond with MK words involving more than one MK vowel (quite apart, that is, from the question of MK divergence according to register), if no explanation of the divergence can be found in the initial or final consonants of the OK words, and if no correlation can be seen between particular vowels and particular final consonants, then one is at liberty to suppose that OK had the same distinctions as MK and that spelling ambiguities were tolerated.

Sakamoto's use of the first method was highly successful in sorting out the *a* and *ā* vowels so as to establish the phonemes which he writes **v* and **a*. It was also successful in separating his **o* (∪ *ao/ò:*), **o* (∪ *o:/ù:*) and **uo* (∪ *u:ə/ù:ə*). The second method was used to establish that **ε* (∪ *ae/è:*) and **ə* (∪ *aə/ÿ:*) were separate phonemically from **e* (∪ *e:/è:*). Here Sakamoto did not refer to initial consonants but his implication is that they provide no explanation of the MK differences. We have, for example, PA *tem, tek* ∪ MK /*daəm, daek/*; PA *ceñ, cer* ∪ MK /*caeŋ, ce:/*; PA *vek, ver* ∪ MK /*vè:k, vè:/*; PA *pre, preñ* ∪ MK /*praə, pre:ŋ/*.

In some parts of his analysis, however, it seems that Sakamoto did not apply his methods where he might have done. In 'e de khmer ancien' he sets out to use the confused or non-confused spellings of individual words to define phonemes written with *e*. Having found no confused spellings for PA *tamre, ampel, et, vañe, ber* or *pi, ti piñ, siñ*, he deduces that the A spelling *ya*, which all these words later have, represented a fusion of two distinct PA sounds, **i* and **e*. In fact he had an example of confused spelling among his few words, in the word *tamre*, which was spelt two ways in PA times, with *r* as well as with *e*, and he knew that *vñe* had later spellings *vñi* and *vñya*. He could therefore have allowed these cases of confused spellings to assign the *tamre* set to **i*. Similarly, in 'i, T, ya, yā de khmer ancien', p. 498, he concludes, in spite of the confused A spellings *i, T, ya* of **i* and *ya, yā* of **iə*, that the same diphthong phonetically was heard in the pronunciation of these two phonologically distinct vowels, **i* and **iə*, and that they were somehow distinguished again later so as to yield two separate MK vowels. It still seems to the present writer more satisfactory to assume that there was at least a difference in A between **ia* (spelt *ya* and *yā*) and previously spelt *e, ya, ye* and *yā* in PA) and **i* (spelt *i, T* and *ya* and previously spelt *i, e* in PA). The distinction made (Jacob:1961,359) between the short and long A diphthongs *iə, ĩə* still seems therefore to be valid.

With most of Sakamoto's conclusions, however, the writer heartily agrees. His second method might now be applied further to the analysis of *a and *v. In paragraph 6 of 'a and ā de khmer ancien' Sakamoto examines his examples with reference to their final consonants to see whether short and long pairs may be substantiated in each case and decides that his twelve instances in which a PA vowel corresponds with both a short and a long vowel before the same consonant in MK are not enough to make a satisfactory conclusion. However, PA examples in the Appendix to this study indicate the occurrence of:

PA *a ~ both MK a: and MK a before k ṅ, m, p, l, s, y, t, c, ñ.

PA *a ~ both MK ì:ə and MK èə~ḍə before ṅ, n, m, p, y, v.

PA *v ~ both MK ɔ: and MK ɔ before k, ṅ, m, p.

PA *v ~ both MK ò: and MK ùə before k, ṅ, n,

Pre-vocalic consonants may - in one case only! - assist the conclusion. PA soṅ 493 /sɔ:ŋ/ *pay back* may be compared with PA saṅ 137 /sɔŋ/ *build*.

AN EXAMINATION OF THE VOWELS AND FINAL CONSONANTS IN CORRESPONDENCE
BETWEEN PRE-ANGKOR AND MODERN KHMER

APPENDIX OF EXAMPLES WITH REFERENCE TO TABLE 1

As far as possible, two examples are given for each PA vowel-symbol entered on the table, reading from left to right along each line. When possible, different words have been chosen to illustrate different spellings of vowels corresponding with one MK vowel. N. = name. Pl.N. = place name. Numerals give a reference to an inscription. MK pronunciation is given between sloping lines.

- a:k cāk N. 24 /ca:k/ corypha pilearia. skāk N. 109 /ska:k/ *shelf, rack.*
- a:ŋ kantrāñ N. 748 /tra:ŋ/ ficus pilosa. tanlāñ N. 74 /thla:ŋ/ *large earthenware cooking pot.*
- a: kralā 557 /krəla:/ *court.* pkā 21 /phka:/ *flower.*
- a:m hām N. 129 /ha:m/ *forbid.*
- a:p c-āp N. 559 /chʔa:p/ *smelling of fish.* ^asāp N. 877 /sa:p/ *woven basket for catching small fish.* tarap 56 /dəra:p/ *all the way, always.*
- a:l kantāl 79 /konda:l/ *middle.* āl N. 357 /ʔa:l/ *be hasty.* ampall 389 /ʔɔmba:l/ *all.*
- a:h thās 505 /tha:h/ *tray.*¹ kmās N. 11 /khma:h/ *bashful.*
- a:y tañhvāy 9 /tɔŋva:y/ *offering.* svāy 134 /sva:y/ *mango.*
- a: skār N. 76 /ska:/ *weasel.* cpar 562 /cba:/ *garden.* cmar 424 /chma:/ *fine, small.*
- a:t tmāt N. 49 /tma:t/ *vulture.*
- a:c trāc 726 /tra:c/ N. *of large tree of dipterocarpus family, producer of excellent oil.* tralāc 51 /trəla:c/ N. *of a variety of gourd.*
- a:p tmāñ 956 (cf. /tba:p/ *weave v.*) *weaver.* kravāñ N. 561 /krəva:p/ *cardamum.* tpāñ, tpañ N. 18,66 /tba:p/ *weaving.*
- ì:əŋ jāñ N. 24 /cì:əŋ/ *artisan.* vāñ N. 149 /pì:əŋ/ *sterculia foetida.*
- ì:ə nā 79 /nì:ə/ *at a time when, at, with regard to.* ^amā 1029 /mì:ə/ *uncle.*

- ì:ən m̄ān 154 /mì:ən/ *have*.
 ì:əm marām N. 149 /mrì:əm/ *finger*. amvām Pl.N. 451 /pì:əm/ *place where stream runs into river or river into sea*.
 ì:əp prāp N. 562 /prì:əp/ *pigeon*.
 ì:əl gvāḷ 689 /khvì:əl/ *herd v*.
 ì:əh mās 388 /mì:əh/ *gold*.
 ì:əy kajāy N. 149 /khcì:əy/ *Kaempferia pandurata*.
 ì:ə dār 493 /tì:ə/ *claim*. dmār 79 (cf. /tì:ə/ *claim*) *claimant*.
 ì:əv jāhv 79 /cì:əv/ *buy, barter*. jamnāhv 811 (cf. /cì:əv/ *barter*) *barter n*.
- ak canlakk 79 /cōmlak/ *sculpted object*. prak 388 /prak/ *silver*.
 cpāk Pl.N. 76 /cbak/ *place where roots are exposed*.
- aṅ pnañ 155 (cf. /baṅ/ *screen v.*) *screen n*. prañ 24 /praṅ/ *dry (of weather, rice-growing)*. lpāñ 155 /rəbaṅ³/ *shelter, screen*. grāñ 877 /kraṅ/ *hill on flat plain*.
- an tlann 561 /thlan/ *cobra*. santan N. 137 /sōndan/ *Garcinia Merguensis*.
 cān N. 66 /can/ *sandalwood*.
- am cam 438 /cam/ *guard*. pram 607 /pram/ *five*. tām 689 /dam/ *cook v*.
 lkām N. 560 /rəkam³/ N. *of a tree with very long branches and strong thorns*.
- ap cap 44 /cap/ *seize*. kracap 1029 /krəcap/ *edible water-creeper*.
 snap Pl.N. 689 /snap/ *thicket*.
- al cal N. 138 /cal/ *jealous*.
- ah cas,^a cas 493,726 /cah/ *old*. panlas 137 /bōnlah/ *exchange*.
- ah kanlah 416 /kōnlah/ *half*. klah 726 /khlah/ *some*.
- ay tñai 38 /thṅay/ *day, sun*. tlai 133 /thlay/ *expensive, costing*.
- at kan-at N. 134 /kʔat/ *tadpole*. sñat N. 423 /sṅat/ *quiet*. chvāt 341 (cf. /chvat/ in /chvat-chviəl/ and /chvat-chvaeṅ/ *criss-cross*) *mark out*.
 sāt N. 18 /sat/ *drift*.
- ac srac 555 /srac/ *achieved, finished*. stac N. 560 /sdac/ *king; pre-verbal particle in royal vocabulary*. kāk 790 /kac/ *cut off*.
 hāc N. 138 /hac/ *restlessly looking for something new and different*.
- au krau 18 /krau/ *outside*. cau 51 /cau/ *grandchild*.
- aṅ añ 561 /ʔaṅ/ I. tarañ 341 /traṅ/ *small flat grass-covered area*.
 kantāñ N. 689 /kōndaṅ/ *curly*. krāñ N. 557 /kraṅ/ *unwilling*.
- èək ^anak 422 /nèək/ *person*. vnāk 21 (cf. /pèək/ *be attached, wear*) *adornment, attachment*.
- èeṅ glañ 505 /khlèeṅ/ *storehouse*. ramañ 129 /rəmèeṅ/ *roe-buck*.
 plāñ 1029 /phlèeṅ/ *thatching grass*. kandāñ N. 11 /kōntèeṅ/ *part of N. of three plants*.

- ɔ̄ən jamnan 561 (cf. /c̄ɔ̄ən/ tread on) causeway.
 ɔ̄əm ramam 557 (cf. /r̄ɔ̄əm/ dance v.) dancer. nam 137 /n̄ɔ̄əm/ lead, bring.
 ɔ̄əp rañap N. 557 /r̄əŋɔ̄əp/ calmed down. ramñap N. 66 /r̄umŋɔ̄əp/ appease,
 silence v.tr.; kill.
 èəh vrah passim /pr̄èəh/ holy;god. hvah 502 /v̄èəh/ cut open.
 èy (strictly
 èəy) bhai 582 cf. /m̄əphèy/ twenty. vrai 18 /pr̄èy/ forest. vanli 561
 /p̄ùənlièy/ Zingiber cassumunar.
 ɔ̄ət mat N. 66 /m̄ɔ̄ət/ mouth, edge.
 èc (strictly
 èəc) luc 38 /l̄èc/ sink, set; west. kandec N. 904 /k̄ontèc/ 1. scabies
 2. part of name of plant which stings when touched. vlac N. 11
 /phl̄èc/ forget.
 ỳu (strictly
 èəv) dau 76 /t̄ỳu/ go. tvau N. 54 /thp̄ỳu/ sorghum millet. nu,nuv 424,
 939 /n̄ỳu/ and, with.
 èŋ (strictly
 èəŋ) duñ 424 /t̄èŋ/ buy.

 ɔk kak N. 129 /k̄ɔk/ N. of a rush grass. kantok 557 /k̄ondɔk/ Cyatula
 genicolata (Nyctogynées). campok 438 /c̄ombɔk/ Buchanla fastigiata.
 ɔŋ sañ 137 /s̄ɔŋ/ build. snañ 137 (cf. /s̄ɔŋ/ build) foundation.
 sroñ 124 /sr̄ɔŋ/ bathe.
 ɔm cam 561 /c̄ɔm/ as in /riəp c̄ɔm/ make ready, provide. psam 127
 /phs̄ɔm/ unite.
 ɔp tap 388 /d̄ɔp/ ten. trap 21 /tr̄ɔp/ egg-plant.
 ɔl tnal 493 /thn̄ɔl/ main road. kañval N. 22 /k̄əŋvɔl/ trouble. cañol
 N. 138 /chŋ̄ɔl/ surprised.
 ɔh klas 505 /k̄hl̄ɔh/ parasol. tras N. 502 /tr̄ɔh/ N. of tree which
 grows in wet places; medic.
 ɔh samtoh N. 8 /s̄ɔmd̄ɔh/ spittle. soh N. 24 /s̄ɔh/ exhausted.

 ̀ək lak 22 /l̄̀ək/ sell. caruk 124 /cr̄̀ək/ pickle.
 ̀əŋ kamvañ 421 /k̄əmp̄̀əŋ/ waterfront. grañ 155 /kr̄̀əŋ/ N. of tree
 with brittle wood. yuñ N. 79 /ȳ̀əŋ/ gleaming. camdoñ N. 424
 /c̄əmt̄̀əŋ/ adolescent.
 ̀ən pandan 726 /b̄ənt̄̀əŋ/ tame v. tr. lahvan N. 137 /lv̄̀əŋ/ graceful.
 ̀əl candal 877 /c̄ənt̄̀əl/ support, stand n. rahval N. 926 /r̄əv̄̀əl/
 busy. pradul Pl.N. 557 /pr̄ət̄̀əl/ opposite.
 ̀əh tnas 877 (cf. /thn̄̀əh/ object contrived with branches used to
 channel fish into the place where they are to be caught) means of
 directing water in gutter (?). kvas N. 24 /k̄hp̄̀əh/ tall, high.
 kavos N. 910 /k̄hp̄̀əh/ tall, high.

- ùəh jmah 557 /chmùəh/ *name*. vlah 561 /phlùəh/ *double*.
 ùət rat 66 /rùət/ *run*
- ɔ:k clok N. 138 /chlɔ:k/ *N. of animal of weasel family*. sampok N. 562 /sɔmbɔ:k/ *bark*.
 ɔ:ŋ kañ 21 /kɔ:ŋ/ *bracelet, anklet*. coñ 341 /cɔ:ŋ/ *put together, tie*.
 phoñ 38 /phɔ:ŋ/ *all, too*.
 ɔ: ka 137 /kɔ:/ *construct*. ta passim /dɔ:/ *the one which, the*.
 tmo 21 /thmɔ:/ *stone*. so 79 /sɔ:/ *white*.
 ɔ:m kan-am N. 648 /kʔɔ:m/ *pitcher*. kraham N. 41 /krəhɔ:m/ *red*.
 ɔ:p ckop 44 (cf. /kɔ:p, prəkɔ:p/ *equipped with*) *involved (in debt)*.
 ɔ: tampar N. 155 /dɔmbɔ:/ *foursome, 4*
 ɔ:t trat N. 22 /trɔ:t/ *hurrying, though weary, in quick spurts*.
- ɔ:k mok 341 /mɔ:k/ *come*. svok 124 /spɔ:k/ *tray*.
 ɔ:ŋ cgoñ N. 137 /chkɔ:ŋ/ *imperfect*.
 ɔ: lño 124 /lɲɔ:/ *sesamum*. kambho 438 /kɔmphɔ:/ *N. of variety of fish, barbuis macrolepidotus*.
 ɔ:n mon N. 149 /mɔ:n/ *mulberry*.
 ɔ:m kadam N. 424 /khtɔ:m/ *hut*. sagom N. 357 /skɔ:m/ *thin (of human physique)*.
 ɔ:p karap 21 /krɔ:p/ *cover n.*
 ɔ: bhar N. 664 /phɔ:/ *tell a lie*.
- ao sno 904 /snao/ *sesbania paludosa*.
 aom karom 137 /kraom/ *below*. camnom 44 (cf. /caom/ *enclose, enclosure*).
 aol sramol N. 18 /srəmaol/ *shadow*.
 aoy oy passim /ʔaoy/ *give*. toy 590 /daoy/ *along, by*.
 ao kantor N. 1030 /kɔndaɔ/ *mouse, rat*. Alternative pronunciation in MK: /kɔndol/.
 aot tuñnot 9 /tñaot/ *sugar-palm*. sot 79 /saot/ *in addition*.
 aoc kmoc Pl.N. /khmaoc/ *ghost*.
- ɔ:ŋ moñ N. 138 /mɔ:ŋ/ *N. of area between Byo and Pursat*. yoñ N. 451 /yɔ:ŋ/ *draw water in bucket at end of rope*.
 ɔ:l jmol 127 /chmɔ:l/ *male*. yol Pl.N. 134 /yɔ:l/ *swing*.
 ɔ:c roc 451 /rɔ:c/ *period of waning moon*. samvoc N. 816 /sɔmpɔ:c/ *civet-cat*.
- o:ŋ toñ 424 /do:ŋ/ *coconut*. antoñ, antvoñ 124 (cf. /ʔɔndo:ŋ/ *well n.*) *beehive, container for oil*.
 o:n kon, kun 66 /ko:n/ *child*. paon 76 /phʔo:n/ *younger sibling*.

- o:m smom N. 127 /sɔ:m/ *beggar*.
- o: canhor, canhvar 904, 341 /cɔŋho:/ *stream*. vnur 341 /phno:/ *mound*.
 hvar 726 /ho:/ *flow*.
- o:t vnot N. 422 /phno:t/ *birthmark*. slot N. 940 /slo:t/ *good-natured*.
- o:c kroc 757 /kro:c/ *citrous fruit*. ptoc 726 (cf. /do:c/ *like, as*)
equivalent.
- o:v plu 76 /phlo:v/ *way, path*. sru, srū 424, 726 /sro:v/ *paddy*.
- ù:k dok, dvak 44, 341 /tù:k/ *boat*.
- ù:ŋ tmoñ 415 (cf. /tù:ŋ/ *beat drum*) *drummer*.
- ù: ru, rū 341, 664 /rù:/ *like, as*.
- ù:n jon, jvan 30, 74 /cù:n/ *offer*. jamnon 689 /cùmnù:n/ *offering*.
- ù: ñarñor, ñrañor 689, 124 /rəŋù:/ *syrup*.
- u:ə tao, taor N. 559, 149 /tʰu:ə/ *terminalia tree*.
- u:ən kalmon 124 /krəmu:ən/ *wax*. pon 90 /bu:ən/ *four*.
- u:əh pos 44 /bu:əh/ *enter the religious life*.
- u:əy kanmoy 38 /kmu:əy/ *nephew, niece*. troy N. 137 /tru:əy/ *new leaf*.
- ù:ək vnok, vnvak 134, 7 (cf. /pù:ək/ *group*).
- ù:ə mo N. 24 /mù:ə/ *wild rice*.
- ù:əy moy 137 /mù:əy/ *one*.
- ù:ət kamdot N. 109 /tù:ət/ *great-grandparent*. vot N. 149 /pù:ət/
join forces, help at work.
- ù:əc lvāc 341 /lù:əc/ *steal*.
- ok samruk 21 (cf. /sɔmrok/ *sink deep in*) *repoussé*. sruk 79 /srok/
inhabited area, district. samlok 22 (cf. /sɔmlɔ:-sɔmlɔk/
variety of cooked dishes) *cooking*.
- oŋ kamluñ 44 (cf. /kɔmlɔŋ/ *group, place*) *inside*. kuruñ 388 /kroŋ/
king.
- om kñum *passim* /khɔm/ *servant*. kləmun 24 /krəmun/ *maiden*. kam
 724 /kom/ *do not*. kantrom N. 711 /trom/ *N. of kind of egret*.
 som 939 /som/ *ask for*. tvam 22 /dom/ *piece*.
- oh cuh 557 /coh/ *go down, put down in writing*. uh N. 24 /ʰoh/
firewood. kantvoh N. 134 /doh/ *grow*.
- oy canhoy 877 (cf. /cɔmhoy/ *steamer; steamed food*) *container for*
burning in making perfume. sa-uy N. 28 /sʰoy/ *putrid*. uy N. 562
 /ʰoy/ *little basket*.
- ot kaupt N. 561 /kɔmbot/ *having lost a limb*.

- ùk duk 904 /tùk/ *put, reserve, keep.* vuk N. 8 /pùk/ *beard.*
 ùŋ ruñ 76 (cf. /rùŋ/ in /rùŋ-rùəŋ/ *grand, splendid*) *big.*
 ùn mun N. 24 /mùn/ *spotty.*
 ùm gmum 124 /khumùm/ *bee.* danhum 424 /thùm/ *scented.* paŋgam 44
 /bɔŋkùm/ *greet respectfully.* vnam 44 /phnùm/ *hill.* vom 154
 /pùm/ *not.* klavom N. 357 /krəpùm/ *bud.*
 ùp dap 38 /tùp/ *block, impede.* damnap 689 /tùmnùp/ *dam, barricade.*
 ùl jul N. 137 /cùl/ *mend.*
 ùh noh 79 /nùh/ *that.* loh,lvāh 561,341 /lùh/ *as far as.*
 ùy duy N. 24 /tì:tùy/ *owl.*
 ùt mut 451 (cf. /mùt/ *cut, pierce*) *harm.* vut N. 11 /pùt/ *pretend.*
- aæk t-ek N. 149 /tʰaæk/ 1. *N. of a creeper.* 2. *hiccough.*
 aəŋ taken N. 357 /thkaəŋ/ *lofty.*
 aə pre 561 /praə/ *use, order.* pamre 561 /bɔmraə/ *servant.*
 aən camren 451 /cɔmraən/ *increase, prosper.*
 aəm tem 560 /daəm/ *tree.* phem 79 /phaəm/ *with young.*
 aəh pares Pl.N. 9 /praəh/ *deer.*
 aəy ley 557 /laəy/ *at all.* Trey Pl.N. 1029 /traəy/ *the far side.*
 aət kan-et N. 155 /ʔaət/ *strain neck to see.* ket 44 /kaət/ *wax (of moon); be born.*
- ỳ:ŋ jeñ 79 /cỳ:ŋ/ *foot, lower part (of hill, etc.).* vleñ 877
 /phlỳ:ŋ/ *fire.*
 ỳ: jhe 66 /chỳ:/ *wood.* tve 127 /thvỳ:/ *do.*
 ỳ:m kamvem N. 109 /khpỳ:m/ *despise.*
 ỳ:p dep 726 /tỳ:p/ *then.*
 ỳ:l vrel N. 163 /prỳ:l/ *careless.*
 ỳ: der 149 (cf. /cɔmtỳ:/ *middle-sized*) *adolescent, middle-sized (?).*
- aek tek 388 /daek/ *iron.* santek 689 /sɔndaek/ *bean.* cacāk (cf.
 /caek/) *divided (?).*
 aəŋ kamrateñ 400 /kùmdaəŋ/ *lord.* prahveñ 424 /prəvaəŋ/ *length.*
 kammratāñ 561 /kùmdaəŋ/ *lord.*
 ae plē 561 (cf. /phlae/ *fruit*) *produce, revenue.* sre passim /srae/
ricefield.
 aen pen N. 877 /baen/ *thresh.*
 aem pa-em N. 127 /phʔaem/ *sweet.*
 ael tel 66 /dael/ *which, who.*
 ae tañker N. 134 /dɔŋkae/ *tick, acarus.* her 44 (cf. /hae/ *go in procession*) *flow.*
 aev cehv N. 560 /caev/ *row a boat.* slev N. 357 /slaev/ *having a squint.*

- è:k ahvek 124 /vè:k/ *ladle*. tvek N. 480 /thpè:k/ *bald*.
 è:ŋ reñ 426 /rè:ŋ/ *weave, plait*.
 è: vave 562 /pəpè:/ *goat*.
- iæk canlek, canlyak 561,7 /sɔ̃mliæk/ *clothing*.
 iəŋ camreñ 557 (cf. /criəŋ/ *sing*, /cɔ̃mriəŋ/ *song*) *singer*. kanteñ
 N., Pl.N. 562,76 /kɔ̃ndiəŋ/ *commelina communis*.
 iəm hyam N. 109 /hiəm/ *have a discharge*.
 iəp tkep N. 8 /thkiəp/ *pinch, nip*.
 iəl trel N. 563 /triəl/ *N. of a creeper*. kryel 134 /kriəl/ *crane n.*
 iə kantyar N. 755 /kɔ̃ndiə/ *white ant.* yer Pl.N. 421 /ʔiə in
 ʔiə-məmiə/ *go carefully along an edge*.
- iəŋ deñ N. 149 /tiəŋ/ *exact*.
 iən den 388 /tiən/ *candle*. rmmen N. 424 (cf. /riən/ *learn*) *student*.
 iəp tgyap N. 904 /thkiəp/ *pinch, nip*.
 iəh chdyās 723 (cf. /cɔ̃ntiəh-cɔ̃ntè:ŋ/) *hampere*.
 iəv kandehv N. 808 /tiəv/ *blue jay*. vñau 561 /phpiəv/ *guest*.
- e:k trasek Pl.N. 560 /trəse:k/ *peltophorum dasyrachis*.
 e:ŋ kanmeñ, kanmin 49,423 /kme:ŋ/ *young*. pareñ 451 /pre:ŋ/ *oil*.
 e:h ces N. 137 /ce:h/ *cotton thread*.
 eh ceh N. 137 /ceh/ *know how to*. treh N. 149 /treh/ *pluck stringed
instrument*.
 e: cer 38 (cf. /ce:/ *long, as time proceeds*) *go, transgress*.
 cmer 502 (cf. /ce:/ *long, as time proceeds*) *transgressor*.
 tmir 137 (cf. /de:/ *sew*) *one who sews, attaches*. sir 133
 /səse:/ *write*.
- è: ge *passim* /kè:/ *person*. danle 904 /tuənlè:/ *river, Tonle Sap*.
 è:p vep N. 357 /pè:p/ *stick out lower lip*.
 è:h dves N. 607 /thvè:h/ *careless*.
 èh veh N. 76 /vèh/ *slip off, avoid an encounter*.
 è: hvera 127 /vè:/ *help in turn, form a rota*.
- ɤk sliik 560 /sliɤk/ *400*. anlik N. 480 /ʔolɤk/ *melon*.
 ɤŋ ktiñ 451 (cf. /kɔ̃mɤŋ/ *owing*) *pay (back)*. piñ 557 /bɤŋ/ *lake*.
 ɤn sin 24 /sɤn/ *further, too*.
 ɤp criip 726 /cɤɤp/ *detach*.
 ɤl ampel 940 /ʔɔ̃mbɤl/ *salt*.
 ɤh is 562 /ʔɤh, ʔɔ̃h/ *all*.

- ry krapì 11 /krəbry/ *buffalo*. pi 388 /bry/ *three*. tmT Pl.N. 493
 /thmry/ *new*. kmT 11 /khmry/ *reclaim*. tamre, tamrr 21, 388
 /dɔmry/ *elephant*. tr 25 /try/ *fish*. camnya 341 /cɔmry/
food, pasture.
- rt kampit N. 562 /kambyt/ *knife*. spit 137 (cf. /syt/ *pour*) *libation*.
 samrat 30 /sɔmryt/ *husk, clean (rice)*.
- ùk dik 137 /tùk/ *water*.
- ùŋ cdiñ 134 /stùŋ/ *river*. kajiñ N. 648 /khcùŋ/ *lamprey, numb-fish*.
- ùn uden 389 /ʔo:tùn/ *container for fish sauce*.⁴ tvin N. 726
 /thpùn/ *blindfold*. vin N. 562 /pùn/ *topmost, best*.
- ùm dnem 560 /thnùm/ *yoke, pair*.
- ùl valvel 877 /pəpùl/ *taper-holder*. amvil 129 /ʔɔmpùl/ *tamarind*.
- ùh vis N. 11 /pùh/ *venom, poison*.
- ùh sgih N. 133 /skùh/ *slow*.
- ùt kamvit N. 155 /pùt/ *true*. jit N. 76 /cùt/ *near*.
- ì:ŋ vreñ 790 /prì:ŋ/ *Syzygium*. kanden N. 8 /khtì:ŋ/ *wild
 buffalo*.
- ì: amvi 79 /ʔɔmpì:/ *from*. ^aji 451 /cì:/ *grandparent, ancestor*.
 vñe 134 /phñì:/ *flower*. ye 562 /rì:/ *female (?)*. laye N. 66
 /lñì:/ *douroucouli monkey (?)*.
- ìh neh 388 /nìh/ *this*. gnih 555 *id.*
- ì: ber 388 /pì:/ *two*.
- ù: gi,gui 561 (both) /kù:/ *is, that is*.

TABLE 1

USES OF THE PA VOWEL-SYMBOLS, IN RELATION TO THE MK VOWELS WITH WHICH THEY CORRESPONDED, BEFORE ALL FINAL CONSONANTS

MK vowels	PA-MK final consonants															
	k	ŋ	Zero	n	m	p	l	s-h	h	y	r	Zero	t	c	v	ŋ
a:	ā*	ā*	ā		ā*	ā* a	ā a	ā		ā	ā* a	ā*	ā			ā a*
ɨ:ə		ā*	ā	ā	ā*	ā*	ā	ā		ā*	ā				ā	
a	a ā*	a ā		a ā*	a ā	a ā*	a*	a	a	ai			a* ā	a ā	au	a ā*
əə.əə	a ā	a ā		a	a	a*		a		ai i			a*	u e* a*	au u†	u
ɔ	a* o	a o		a	a	a o*	a	o*								
ùə	a u	a u* o*		a		a u*	a* o*	a					a			
ɔ:	o*	a o	a o	a*	o								a*	a*		
ɔ:	o	o*	o	o*	a* o*	a							a*			
ao			o	o		o*			o			o*	o	o*		
ò:		o*				o								o		
o:		o* vo*		o* u*	o*							o* va* u*	o*	o*	u* u†	
ù:	o va	o	u ũ	o va								o				
u:ə			o*	o				o		o						
ù:ə	o va		o*							o			o*	vā		
o	u o*	u		u a o va				u vo*	u* o				u*			
ù	u	u		u*	u a o	a	u*	o vā*	u*				u			
aə	e*	e*	e	e	e			e*		e			e			
ɨ:		e	e	e*	e	e*							e*			
ae	e ā*	e ā	e	e*	e*		e						e		e*	
è:	e	e	e													
iə	e ya	e		ya*	ye	e* ye							ya* ye*			
lə		e*	e		ya*		yā								e* au†	
e:	e*	e i					e*	e*					e i*			
è:			e			e*	e*	e*					e			
ɨ	i	i		i	i	e	i		i Te r ya				i a			
ù	i	i		e i*	e		e i	i*	i*				i*			
ɨ:		e	i e					e i					e			
ɨ:			i u i													
wə																
ùə																

† Final v usually not written in PA following these vowels.

TABLE 2
SUMMARY OF THE MK CORRESPONDENCES WITH PA VOWEL-SYMBOLS

PA vowel-symbol	MK correspondences		
	Usual	Unusual	Exceptional
\bar{a} :	a:/ì:ə	a/ĕə~ðə	ae
a	a/ĕə ðə	a:	
a, o	ɔ/ùə ɔ:/ò:	o/ù	
o	ao/ò:		o/ù
o, u, ū, va	o:/ù:		
o, va, vā	u:ə/ù:ə		
e	e:/è: ae/ÿ: ae/ĕ:	ɣ/ù ì:	
e.ya, ye, yā	iə/ìə		
i (closed) i (open)	ɣ/ù ɣy/ì:	ì (+h)	e:(+r)
u (closed) u (open)	o/ù o:v/ù:	ùə(+k, ñ, l)	
ai	ay/èy		
au	au/ÿu		

AN EXAMINATION OF THE VOWELS AND FINAL CONSONANTS IN CORRESPONDENCES
BETWEEN PRE-ANGKOR AND MODERN KHMER

N O T E S

1. This includes the texts published by G. Cœdès plus the texts of K. 1029 and 1030 kindly supplied by Monsieur C. Jacques of the Ecole Pratique des Hautes Etudes, Paris.
2. The accent indicates second register.
3. r and l are not in opposition in this pre-consonantal context.
4. I am grateful to Mr. Kuoch Hak Srea for suggesting this correspondence.

BIBLIOGRAPHY

JACOB, J.M.

- 1960 'The structure of the word in Old Khmer'. *BSOAS*. 23:351-8.
- 1963 'Prefixation and infixation in Old Mon, Old Khmer and Modern Khmer'. In *Linguistic comparison in South East Asia and the Pacific* (ed.) H.L. Shorto. London. 62-70.
- 1966 'Some features of Khmer versification'. In *In Memory of J.R. Firth*. London. 227-41.
- 197? 'Some problems arising from the orthography of consonants on the Khmer inscriptions'. To appear in *Austroasiatic Studies*, University of Hawaii.

SAKAMOTO, Y.

- 1970 'e de khmer ancien'. *Journal of Asian and African Studies*. 3:28-43. Tokyo.
- 1970 'i,ī,ya,yā de khmer ancien'. *Tonan Ajia Kenkyu. Southeast Asian Studies*. 7:492-503.
- 1971 'Sur quelques voyelles de khmer ancien'. *Journal of Asian and African Studies*. 4:53-73. Tokyo.
- 1974 'a and ā de khmer ancien'. *Journal of Asian and African Studies*. 7:75-100. Tokyo.

THE VALUE OF \bar{I} , I , \bar{U} AND U IN MIDDLE KHMER

Philip N. Jenner

This is the fourth of a series of studies on the vowel and consonant systems of Middle Khmer, tentatively assigned to the period from the 14th to the 18th centuries. The first paper in the series¹ reported the results of an analysis of eighteen short metrical texts and showed how these can be placed in quasi-chronological sequence on the basis of their rhymes. This was followed by a description of the changing distribution of final $-ra$ and $-la$ ² and by an examination of the graphemes au and ai ,³ both using the same corpus of 5,164 rhymes as had been developed for the first paper. We turn now to investigate the graphemes \bar{I} , i , \bar{U} , and u as these are found in the same material. Future studies will treat vowels occurring before the final palatals, the development of modern $/o\bar{e}/$, and the grapheme e in the same body of texts. A sketch of the vocalism, including information on the Middle Khmer system, has been given elsewhere.⁴

The analytical procedure followed here is essentially the same as before. The entire corpus is first sifted, all rhymes involving orthographic \bar{I} , i , \bar{U} , and u being separated out. The rhymes in question are then typed and weighed from the point of view of their relevance to the problem of determining their value during the Middle Khmer period. The modern value of the four graphemes is taken as a known quantity, and the modern regularized orthography is preferred over that originally used in the seventeen undated $cp\bar{a}'pa$ texts. The members of each rhyming pair are compared and inferences are drawn from their orthographic shape, from their modern value, from the pre-modern pronunciation of loanwords present, and from known systematic changes in the vocalism. These inferences are finally tested against the collected rhymes.

While many of the rhymes used have been invalidated by later phono-

logical changes, it is assumed that all of the rhymes indicated by the meter of each text were, with very few exceptions, valid at the time of composition.

In isolating the evidential from the nonevidential portion of the assembled data, it is necessary to draw a distinction between 'perfect' and 'imperfect' rhymes. Perfect \bar{T} rhymes are those in which both members of a rhyming pair show graphic \bar{T} in their relevant syllables. Perfect i , \bar{u} , and u rhymes are similarly those in which both members have the same orthographic vowel. Imperfect rhymes are those in which only one member of a rhyme contains graphic \bar{T} , i , \bar{u} or u , which rhymes with some other orthographic form. Exceptions are made in the case of certain allographs, to be mentioned as they are introduced. The reason for this distinction, as well as examples of both classes of rhyme, will be given shortly. In our group of texts perfect rhymes account for 62.4% of all \bar{T} , i , \bar{u} and u rhymes, imperfect rhymes accounting for the remaining 37.6%.

Table 1 shows that the total corpus contains 366 perfect and imperfect \bar{T} rhymes and 299 perfect and imperfect i rhymes. The \bar{T} rhymes have an average frequency of 6.45% and individual frequencies which rise and fall erratically between a peak of 13.82% and a low of 0.79%. Maxima are registered for the *Cpā'pa sṛṭ*, the '*Ariyasatthā*, and the *Ker(i) kāla*, minima for the *Kūna cau lpæka* (A) and the *Bākyā cā'sa*. The i rhymes show an average frequency of 5.64% and individual frequencies which, while roughly mirroring the profile of \bar{T} rhymes, range narrowly between 8.3% and 4.00%.

Table 2 shows that the corpus contains 217 perfect and imperfect \bar{u} rhymes and 243 perfect and imperfect u rhymes. The \bar{u} rhymes have an average frequency of 4.23% and individual frequencies which, though less extreme than those of \bar{T} rhymes, rise and fall at random between a peak of 7.57% and a low of 2.08%. Maxima are registered for the *Cpā'pa dūnmāna kūna*, the '*Ariyasatthā*, and the *Kūna cau lpæka* (A), minima for the *Hai mahājana* (I), the *Krama*, and the *IMA*. The i rhymes have an average frequency of 4.72% and individual frequencies ranging from a high of 10.07% down to zero. Maxima are registered for the *Prusa*, the *Bākyā cā'sa*, and the *Hai mahājana* (I), minima for the *Krama*, the *Vidhūrapandita*, and the *Kūna cau*.

At issue in this paper, therefore, are 1,124 rhymes representing 21.8% of the total corpus. It remains to be seen how many of these are usable in our investigation.

TABLE 1
GENERAL FREQUENCIES OF Ī AND i RHYMES

	total rhymes	Ī rhymes	i rhymes
<i>Ker(tī) kāla</i>	105	11 10.48%	7 6.66%
<i>Kūna cau</i>	328	22 6.71%	14 4.26%
<i>Rājaneti</i>	214	20 9.35%	13 6.07%
<i>Krama</i>	154	6 3.90%	7 4.54%
<i>Hai mahājana (I)</i>	337	29 8.61%	20 5.93%
<i>IMA 38</i>	493	39 7.91%	28 5.67%
<i>Kūna cau lṛæka (A)</i>	297	9 3.03%	14 4.71%
<i>Vidhūrapandita</i>	225	14 6.22%	9 4.00%
<i>Pantām pitā</i>	295	9 3.05%	19 6.44%
<i>Kūna cau lṛæka (B)</i>	127	1 0.79%	6 4.72%
<i>Trīneti</i>	380	22 5.79%	19 5.00%
<i>Dūnmāna khlwna</i>	175	11 6.29%	10 5.71%
<i>Bākya cā'sa</i>	97	1 1.03%	4 4.12%
<i>Hai mahājana (II)</i>	324	23 7.10%	20 6.17%
<i>Pantām 'ū buka</i>	291	10 3.44%	14 4.81%
<i>Dūnmāna kūna</i>	251	22 8.76%	21 8.36%
<i>Srī</i>	579	80 13.82%	42 7.25%
<i>Prusa</i>	298	16 5.37%	19 6.37%
<i>'Ariyasatthā</i>	194	21 10.82%	12 6.18%
	5,164	366 6.45%	298 5.63%

TABLE 2
GENERAL FREQUENCIES OF \bar{u} AND u RHYMES

	total rhymes	\bar{u} rhymes	u rhymes
<i>Ker(tī) kāla</i>	105	6 5.71%	5 4.76%
<i>Kūna cau</i>	328	10 3.05%	6 1.83%
<i>Rājaneti</i>	214	8 3.74%	10 4.67%
<i>Krama</i>	154	4 2.60%	0
<i>Hai mahājana (I)</i>	337	7 2.08%	22 6.53%
<i>IMA 38</i>	493	14 2.84%	13 2.64%
<i>Kūna cau lṛæka (A)</i>	297	19 6.40%	16 5.39%
<i>Vidhūrapandita</i>	225	7 3.11%	3 1.33%
<i>Pantām pitā</i>	295	18 6.10%	15 5.08%
<i>Kūna cau lṛæka (B)</i>	127	4 3.15%	5 3.94%
<i>Trīneti</i>	380	15 3.95%	15 3.95%
<i>Dūnmāna khlwna</i>	175	5 2.86%	11 6.29%
<i>Bākya cā'sa</i>	97	3 3.10%	7 7.22%
<i>Hai mahājana (II)</i>	324	10 3.09%	20 6.17%
<i>Pantam 'ū buka</i>	291	18 6.19%	15 5.15%
<i>Dūnmāna kūna</i>	251	19 7.57%	11 4.38%
<i>Srī</i>	579	25 4.32%	29 5.01%
<i>Prusa</i>	298	12 4.03%	30 10.07%
<i>'Ariyasatthā</i>	194	13 6.70%	10 5.15%
	5,164	217 4.23%	243 4.72%

In Table 3 we see that perfect \bar{T} rhymes number 313 and perfect i rhymes number 127. The \bar{T} rhymes, with an average frequency of 5.52%, have individual frequencies ranging from 11.9% to 0.8%. Maxima are registered for the *Sr̄*, the '*Ariyasatthā*' and the *Ker(ti) kāla*, as before, while minima are registered for the *Kūna cau lpœka* (B), the *Bākya cā'sa*, and the *Kūna cau lpœka* (B). The i rhymes, with an average frequency of 2.47%, have individual frequencies ranging from only 5.2% to 0.8%. The highest are registered for the *Dūnmāna kūna*, the *Pantām pitā*, and the '*Ariyasatthā*', the lowest for the *IMA*, the *Bākya cā'sa*, and the *Vidhūrapandita*. In the case of neither \bar{T} rhymes nor i rhymes do we have reason to assume any correlation between these frequencies and the relative age of the texts.

TABLE 3
PERFECT \bar{i} AND i RHYMES

	total rhymes	\bar{i} rhymes	i rhymes
<i>Ker(ti) kāla</i>	105	9 8.6%	2 1.9%
<i>Kūna cau</i>	328	21 6.4%	5 1.5%
<i>Rājaneti</i>	214	16 7.5%	6 2.8%
<i>Krama</i>	154	6 3.9%	0
<i>Hai mahājana</i> (I)	337	21 6.2%	8 2.4%
<i>IMA 38</i>	493	35 7.1%	4 0.8%
<i>Kūna cau lpœka</i> (A)	297	9 3.0%	10 3.4%
<i>Vidhūrapandita</i>	225	13 5.7%	3 1.3%
<i>Pantām pitā</i>	295	9 3.1%	13 4.4%
<i>Kūna cau lpœka</i> (B)	127	1 0.8%	4 3.1%
<i>Tr̄neti</i>	380	19 5.0%	7 1.8%
<i>Dūnmāna khlwna</i>	175	9 5.1%	4 2.3%
<i>Bākya cā'sa</i>	97	1 1.0%	1 1.0%
<i>Hai mahājana</i> (II)	324	15 4.6%	10 3.1%
<i>Pantām 'ū buka</i>	291	9 3.1%	10 3.4%
<i>Dūnmāna kūna</i>	251	19 7.6%	13 5.2%
<i>Sr̄</i>	579	69 11.9%	11 1.9%
<i>Prusa</i>	298	12 4.0%	9 3.0%
<i>'Ariyasatthā</i>	194	20 10.3%	7 3.6%
	5,164	313 5.52%	127 2.47%

In Table 4 we see that perfect \bar{u} rhymes number 138 while perfect u rhymes number 124. The \bar{u} rhymes show an average frequency of 2.71% and individual frequencies ranging from 5.7% to 0.8%. Maxima are seen for the *Ker(tī) kāla*, the *Kūna cau lpæka* (A) and *Pantām pitā*, and the *Pantām 'ū buka*, minima for the *Ḍūnmāna kūna*, the *Bākya cā'sa*, and the *Hai mahājana* (I and II). The u rhymes show an average frequency of 2.32% and individual frequencies ranging from 4.7% down to 0.6%. Peaks are registered for the *Rājaneti* and *Prusa*, the *Pantām pitā* and the *Kūna cau lpæka* (A), lows for the *IMA*, the *Ḍūnmāna khḷwna*, and the *Kūna cau*. In neither case, again, can any correlation be discerned between these frequencies and the age of the texts.

TABLE 4
PERFECT \bar{u} AND u RHYMES

	total rhymes	\bar{u} rhymes	u rhymes
<i>Ker(tī) kāla</i>	105	6 5.7%	2 1.9%
<i>Kūna cau</i>	328	10 3.0%	4 1.2%
<i>Rājaneti</i>	214	7 3.3%	10 4.7%
<i>Krama</i>	154	4 2.6%	0
<i>Hai mahājana</i> (I)	337	4 1.2%	9 2.7%
<i>IMA 38</i>	493	12 2.4%	3 0.6%
<i>Kūna cau lpæka</i> (A)	297	13 4.4%	12 4.0%
<i>Vidhūrapandita</i>	225	5 2.2%	0
<i>Pantām pitā</i>	295	13 4.4%	12 4.1%
<i>Kūna cau lpæka</i> (B)	127	2 1.6%	2 1.6%
<i>Trāneti</i>	380	13 3.4%	6 1.6%
<i>Ḍūnmāna khḷwna</i>	175	3 1.7%	2 1.1%
<i>Bākya cā'sa</i>	97	1 1.0%	2 2.1%
<i>Hai mahājana</i> (II)	324	4 1.2%	7 2.2%
<i>Pantām 'ū buka</i>	291	12 4.1%	12 4.1%
<i>Ḍūnmāna kūna</i>	251	2 0.8%	5 2.0%
<i>Srī</i>	579	12 2.1%	17 2.9%
<i>Prusa</i>	298	8 2.7%	14 4.7%
<i>'Ariyasatthā</i>	194	7 3.6%	5 2.6%
	5,164	138 2.71%	124 2.32%

Perfect rhymes thus number 702 and account for 62.4% of all Ī, i, ū, and u rhymes in the corpus.

We address ourselves now to the question of what these perfect rhymes have to contribute to a solution of the problem under consideration. Since the phonological value of Ī, i, ū, and u is unknown in Middle Khmer but known in modern Khmer, it can be said that the perfect rhymes in our eighteen texts are divisible into three types according as their members have High Register (HR) or Low Register (LR) reflexes in the modern language.

In one of these types, both members of the rhyme presuppose HR reflexes, e.g. jĪ /cii/ : gambĪ /kumpii/ (R 33fe);⁵ dharanĪ /thəɔɔɔnii/ : rddhĪ /rýtthii/ (HMI 75bc); gamnita /kumnit/ : gita /kit/ (KC 68de); 'āditya /qaatit/ : jita /cit/ (HMII 66d/67b); gūra /kuur/ : dadūra /ttuur/ (KK 3bc), grū /kruu/ : 'abhirū(ha) /qaphiruu/ (A 32ab); buka /puk/ : duka /tuk/ (KC 29de); gūna /kun/ : muna /mun/ (T 67ab).

In a second type, both members of the rhyme presuppose LR reflexes, e.g. ktĪ /kdəj/ : krakrĪ /krakrəj/ (KC 20ef); pĪ /bəj/ : sratĪ /sradəj/ (P 9lab); phtita /pdət/ : citta /cət/ (R 32ce); tamrih /tamrəh/ : trih /trəh/ (T 19de); kūna /kòon/ : p'ūna /pqòon/ (KK 14bc); 'āsrūva /qaasròow/ : kūva /kòow/ (S 176c/177a); 'usa /qoh/ : khusa /khoh/ (HMI 40ab); cuka /cok/ : sukha /sok/ (PP 54g/55c).

In the third type of rhyme one member is reflected by a HR form, the other by a LR form, e.g. dhūĪT /thuulii/ : tĪ /dəj/ (KK 18bc); sālĪT /saaləj/ : kutumbĪT(ka) /kdompii/ (V 50ab); tica /təc/ : dranica /trənic/ (KC 4g/5c); gita /kit/ : citta /cət/ (S 215a); madhūra /mathuur/ : cūla /còol/ (R 48g/49c); trūva /tròow/ : byū(ha) /pjuu/ (KCLA 8g/9c); sam'uya /samqoj/ : ruya /ruj/ (KC 20ab); manussa /mnuh/ : prusa /proh/ (P 3ab).

In considering these examples we note at once the close apparent correspondence between the four graphemes and their HR realizations: Ī /ii/ [i:], i /i/ [ɪ], ū /uu/ [u:], u /u/ [ʊ]. On the other hand, we cannot fail to recognize that the same graphemes have LR realizations and that these differ appreciably from their HR counterparts. We are warranted in assuming from their orthographic representation that the HR nuclei have undergone little or no change from their pre-modern shapes. We are warranted in assuming from the difference between the HR nuclei and their LR counterparts that the latter have developed from their premodern forms by several types of lowering: i /əj/ [ɣ:j], i /ə/ [ə̃], ū /oo(w)/ [o:~ɣ:w], u /o/ [õ]. Both of these assumptions accord well with what we know of the development of the modern vocalism.⁶ Briefly, the register of each nucleus is governed by its former environment, originally voiced initials yielding HR nuclei, originally

voiceless initials yielding LR nuclei. The circumstance that the former environment is, with very few exceptions, shown by the conservative writing system is an incidental convenience. The main point here is that by knowing the register of a nucleus we can reconstruct its pre-modern equivalence: HR /ii, i, uu, u/ and LR /èəj, è, òo(w), o/ are alike from pre-modern /ii, i, uu, u/. In this light, the meaning of the perfect rhymes given above becomes clearer. While the HR : HR rhymes undergo no change, the LR : LR rhymes appear as kt̄ /kʔdii/ : krakr̄ /krəkrii/, p̄ /ʔbii/ : srat̄ /srəʔdii/; phtita /pʔdit/ : citta /cit/, tamrih /təmrih/ : trih /trih/; kūna /kuun/ : p'ūna /pquun/, 'āsrūva /qaasruuw/ : kūva /kuuw/; and 'usa /quh/ : khusa /khuh/, cuka /cuk/ : sukha /suk/. The discrepant HR : LR rhymes, moreover, are now heard as they were at the time of composition: dhūīT̄ /thuulii/ : t̄ /ʔdii/, sālīT̄ /saalii/ : kutumbT̄(ka) /kʔdumbii/; tica /tic/ : dranica /trɔnic/, gita /kit/ : citta /cit/; madhūra /madhuur/ : cūla /cuul/, trūva /truuw/ : byū(ha) /bjuu/; sam'uya /səmquj/ : ruya /ruj/, manussa /mnuh/ : prusa /pruh/. In and by themselves, however, our perfect rhymes tell us nothing regarding the pre-modern value of the graphemes in question. Were it not for (a) the fact that we are concerned with rhymes, (b) our assumption that these were almost entirely valid at the time of composition, and (c) their orthographical form, we should be on very infirm ground indeed. What we have been doing up to this point, it may be said, is comparable to weighing English rhymes (*soot* : *moot*, *cough* : *rough*) which are spelled similarly but are pronounced differently in modern English and may or may not have rhymed in Middle English. In short, we need other evidence. We must consequently look to imperfect rhymes and see what confirmation or what new conclusions can be had from them.

The four tables given hereafter show that imperfect rhymes number 422, a figure representing 37.5% of our working corpus, and have an average frequency of 2.0%.

Table 5 shows that imperfect T rhymes number 53 and have an average frequency of only 0.9%. Of these 33 (62.3%) are rhymes in which one member has orthographic T and the other has ai in the relevant syllable. Of the remainder 6 (11.3%) are T : e rhymes and 5 (9.4%) are T : ʔ rhymes, while in 9 others (17%) T rhymes with some other graphic vowel. It should be noted in particular that the individual frequencies for the eighteen texts rise and fall unpredictably between a maximum of 2.5% and a minimum of 0.3%, and show no pattern of increase or decrease correlatable with the age of the texts.

TABLE 5
IMPERFECT ʈ RHYMES

	total rhymes	a i	e	ʈ	other	total	
<i>Ker(ṭi) kāla</i>	105	0	1	1	0	2	1.9%
<i>Kūna cau</i>	328	0	1	0	0	1	0.3%
<i>Rājaneti</i>	214	1	1	1	1	4	1.9%
<i>Krama</i>	154	0	0	0	0	0	
<i>Hai mahājana (I)</i>	337	7	0	0	1	8	2.4%
<i>IMA 38</i>	493	2	1	0	1	4	0.8%
<i>Kūna cau lṣaka (A)</i>	297	0	0	0	0	0	
<i>Vidhūrapandita</i>	225	1	0	0	0	1	0.4%
<i>Pantām pitā</i>	295	0	0	0	0	0	
<i>Kūna cau lṣaka (B)</i>	127	0	0	0	0	0	
<i>Trīneti</i>	380	0	0	2	1	3	0.8%
<i>Dūnmāna khḷwṇa</i>	175	2	0	0	0	2	1.1%
<i>Bākya cā'sa</i>	97	0	0	0	0	0	
<i>Hai mahājana (II)</i>	324	8	0	0	0	8	2.5%
<i>Pantām 'ū buka</i>	291	0	0	0	1	1	0.3%
<i>Dūnmāna kūna</i>	251	3	0	0	0	3	1.2%
<i>Shī</i>	579	7	0	0	4	11	1.9%
<i>Prusa</i>	298	1	2	1	0	4	1.3%
<i>'Ariyasatthā</i>	194	1	0	0	0	1	0.5%
	5,164	33	6	5	9	53	0.9%

In 'The Value of *au* and *ai* in Middle Khmer' 25 $\bar{T} : ai$ rhymes were listed in full and discussed. These may be exemplified here by $sm\bar{a}rat\bar{T} /smaard\bar{e}j/ : thlai /tlaj/$ (HMI 66bc), $khc\bar{T} /kc\bar{e}j/ : nai /n\bar{e}j/$ (DKh 4bc), $metr\bar{T} /m\bar{e}etr\bar{e}j/ : kansai /kansaj/$ (DKn 86ab), and $pt\bar{T} /pd\bar{e}j/ : tadai /daat\bar{e}j/$ (S 22ab).⁷ To these we may now add eight rhymes between \bar{T} and $\bar{a}ya$, an allograph of *ai*:

HMI	45cb	$sm\bar{a}rat\bar{T} /smaard\bar{e}j/$:	$sans\bar{a}ya /san\bar{s}aj/$ ⁷
	45cd	$sm\bar{a}rat\bar{T} /smaard\bar{e}j/$:	$vin\bar{a}ya /win\bar{e}j/$
V	51b/50d	$kt\bar{T} /kd\bar{e}j/$:	$jalasr\bar{a}ya /cul\bar{s}raj/$
HMII	79d/80b	$p\bar{a}ram\bar{T} /baar\bar{c}m\bar{e}j/$:	$bh\bar{a}ya /ph\bar{e}j/$
DKn	46ab	$metr\bar{T} /m\bar{e}etr\bar{e}j/$:	$pr\bar{a}sr\bar{a}ya /praasraj/$
S	61c/62a	$sri /sr\bar{e}j/$:	$hard\bar{a}ya /hr\bar{y}t\bar{e}j/$
	62ba	$srat\bar{T} /srad\bar{e}j/$:	$hard\bar{a}ya /hr\bar{y}t\bar{e}j/$
A	41c/40g	$kt\bar{T} /kd\bar{e}j/$:	$-a\bar{l}\bar{a}ya /-aalaj/$

It has already been shown that the \bar{T} member of the 25 $\bar{T} : ai$ rhymes was in every case reflected by a modern LR syllable in $/-\bar{e}j/$. This observation finds unqualified confirmation in the eight $\bar{T} : \bar{a}ya$ rhymes just cited. On the other hand, in the case of the $\bar{T} : ai$ rhymes previously reported it was found that 17 out of the 25 *ai* members were reflected by modern HR forms in $/-\bar{e}j/$ while eight corresponded to modern LR forms in $/-aj/$. The eight $\bar{a}ya$ members of the rhymes just given are similarly ambivalent, corresponding to four HR and four LR syllables in the modern language. It can therefore be said that, of the 33 $ai \sim \bar{a}ya$ members in question here 21 or 63.6% reflect the HR while 12 or 36.4% reflect the LR. The inequality of these percentages is not enough to justify assumption of any clearly defined pattern. Hence we can only repeat what was said earlier. It is clear that if *ai* (and $\bar{a}ya$) had originally represented [aj] these various rhymes would not have been possible. Moreover, from the fact that we find *ai* and $\bar{a}ya$ members which presuppose different registers in rhyme with \bar{T} members presupposing only the LR we can only suppose that the LR reflex of Middle Khmer /ii/ (\bar{T}) had developed before registral contrast was fully established for Middle Khmer $/\bar{e}j/$ ($ai \sim \bar{a}ya$). In all 33 cases, therefore, \bar{T} must have had its modern LR value of $/-\bar{e}j/$ while *ai* and $\bar{a}ya$ must have had their pre-modern value of $/-aj/$. The eight rhymes just given might hence be rewritten as follows in pre-modern as well as modern terms:

$sm\bar{a}rat\bar{T} /smaar\bar{d}ii \succ smaard\bar{e}j/$	$sans\bar{a}ya /s\bar{o}ns\bar{e}j \succ san\bar{s}aj/$ ⁸
$sm\bar{a}rat\bar{T} /smaar\bar{d}ii \succ smaard\bar{e}j/$	$vin\bar{a}ya /win\bar{e}j \succ win\bar{e}j/$
$kt\bar{T} /k\bar{d}ii \succ kd\bar{e}j/$	$jalasr\bar{a}ya /j\bar{c}l\bar{s}r\bar{e}j \succ cul\bar{s}raj/$
$p\bar{a}ram\bar{T} /?baar\bar{c}mii \succ baar\bar{c}m\bar{e}j/$	$bh\bar{a}ya /bh\bar{e}j \quad ph\bar{e}j/$

metrT /meetrii > méetrèəj/	prāsrāya /praasrəj > praasraj/
srT /srii > srèəj/	hardāya /hrytəj > hrytáj/
sraT /srəʔdii > sradèəj/	hardāya /hrytəj > hrytáj/
ktT /kʔdii > kdèəj/	-ālāya /-aaləj > -aalaj/

These examples are enough to show that rhymes of the type in question could have been possible only during a limited period in the transition from the vocalism of Old Khmer to that of the modern language.

Leaving T : āya (ai) rhymes and going on to the remaining imperfect T rhymes in the corpus, we find that Table 5 registers six rhymes in which orthographic T is paired with e. Upon examining these we are struck by the fact that the T members of five are reflected by modern HR forms:

KK	15/16b	bTra /piir/	:	ker(ti) /kèer/ ⁷
KC	6ed	bTra /piir/	:	ker(ti) /kèer/
R	23ab	jhnānTsa /cnfi niih/	:	cacesa /ccèeh/
P	19ab	raTna /rliiŋ/	:	dheña /théəŋ/
	25ba	raTna /rliiŋ/	:	dheña /théəŋ/

Regarding the e member of these rhymes, we are reminded in the first place that modern HR /éé/ is often articulated as [ɛ:] while modern LR /èè/ is often [e:]. It seems unlikely that the authors of the *Ker(ti) kāla* and the *Kūna cau* ever intended rhyme between /ii/ and [e:] [ɛ:]. We are reminded in the second place that *ker(ti) fame, glory* is a modification of Sanskrit kṛti, in conformity with a fairly common correspondence.⁹ In the two cases above we can only suppose that it was still [ki:r] at the time of composition. In the late *Cpā'pa prusa*, on the other hand, the twofold occurrence of *raTna : dheña* is good evidence that *dheña* was already being pronounced [thɛ:ŋ] instead of pre-modern [dhe:ŋ] and consequently that Middle Khmer /ee/ (e) had by the date of this text already split into its present registral shapes. As for the form *cacesa* in the above rhymes, we must assume either that the rhyme in which it occurs was never valid or, as is more probable, that the contemporary orthography must have been **cacTsa*. The sixth and last of our T : e rhymes is also the most instructive. This is:

IMA	105a/104c	ktT /kdèəj/	:	leya /laəj/ ⁷
-----	-----------	-------------	---	--------------------------

in which both members are reflected by modern LR forms. The latter member has the Middle Khmer orthography of modern *ɬəya*. Rewriting the rhyme in pre-modern as well as modern terms makes it clear that it could have been valid only during a limited period, namely after registral contrast had been established for original /ii/ but before contrast had been established for original /əə/:

ktT /kʔdii > kdèəj/	:	ɬəya /ləəj > laəj/ ⁸
---------------------	---	---------------------------------

Table 5 also shows five $\bar{T} : \bar{T}$ rhymes. In reality only three of these have long \bar{T} in the member rhyming with \bar{T} . These are:

KK	4cb	krap \bar{T} /krab \bar{e} ej/	:	bh \bar{i} \bar{T} /pl \bar{y} y/ ⁷
\bar{T}	36ec	metr \bar{T} /m \bar{e} etr \bar{e} ej/	:	ban \bar{i} \bar{T} /pun \bar{l} y/ ⁷
	44ec	metr \bar{T} /m \bar{e} etr \bar{e} ej/	:	ban \bar{i} \bar{T} /pun \bar{l} y/ ⁷

It would be rash to draw conclusions from so few examples, which are consequently reserved for a forthcoming study on \bar{T} and \bar{i} . Of the two remaining rhymes, one has orthographic long \bar{T} pronounced [i], in rhyme with short \bar{i} , and hence reassignable to \bar{i} :

P	lab	brahmag \bar{T} ti /prum \bar{m} okit/	:	br \bar{i} ddha /pr \bar{y} t/ ⁷
---	-----	--	---	---

This conforms with the modern standard language, in which contrast is easily lost between HR /i/ and HR /y/. The other rhyme is

R	34ab	jañj \bar{T} na /cunci \bar{i} ŋ/	:	t \bar{i} na /d \bar{y} ŋ/ ⁷
---	------	---------------------------------------	---	---

where the first term (*balance, scales*) is from Old Khmer janjyañ /j \bar{e} n \bar{j} i \bar{e} ŋ/¹⁰ while the second (*to know*) is from Old Khmer tyañ /ti \bar{e} ŋ ~ $\bar{?}$ di \bar{e} ŋ/ and Middle Khmer (e.g., IMA 16b) tyiña / $\bar{?}$ dli \bar{e} ŋ > $\bar{?}$ d \bar{y} ŋ/. Notwithstanding the modern orthography given above, it seems clear that the two rhymed when the *Rājaneti* was composed.

Tabel 5 shows, finally, nine rhymes in which \bar{T} is paired with other orthographic vowels. These are:

R	12ba	s \bar{T} la /s \bar{e} l/	:	kh \bar{j} ila /k \bar{c} il/ ⁷
HMI	87ab	h \bar{T} na /hli \bar{e} ŋ/	:	bh \bar{l} l \bar{e} na /pl \bar{f} i \bar{e} ŋ/
IMA	28ab	byira /pilr/	:	ramila /rmil/
\bar{T}	5de	mantr \bar{T} /muntr \bar{e} ej/	:	bi- /pi-/
PUB	2ed	'v \bar{T} /qaw \bar{e} ej/	:	h \bar{e} ya /ha \bar{e} ej/
S	193a	pt \bar{T} /pd \bar{e} ej/	:	n \bar{a} ya /ŋ \bar{f} i \bar{e} ŋ/
	199a	pt \bar{T} /pd \bar{e} ej/	:	n \bar{a} ya /ŋ \bar{f} i \bar{e} ŋ/
	202a	pt \bar{T} /pd \bar{e} ej/	:	n \bar{a} ya /ŋ \bar{f} i \bar{e} ŋ/
	204a	pt \bar{T} /pd \bar{e} ej/	:	n \bar{a} ya /ŋ \bar{f} i \bar{e} ŋ/

The first of these shows in its prior member the same long \bar{T} pronounced [i] as has already been seen in the case of brahmag \bar{T} ti : br \bar{i} ddha, and is hence to be reassigned to short \bar{i} . The second rhyme with bh \bar{l} l \bar{e} na (= ph \bar{l} l \bar{e} na /pl \bar{l} i \bar{e} ŋ/) *rain* is acoustically admissible, and will be treated in a future study on orthographic \bar{i} \bar{e} . In the third, the form byira corresponds to modern b \bar{T} ra. The latter member of this and the fourth rhyme shows short \bar{i} /i/, which apparently varied freely in these forms with /ii/.¹¹ The fifth rhyme conforms exactly with the kt \bar{T} : leya cited above from IMA 38 and confirms the inferences made from its occurrence. From its development,

'vT /wii > wəəj/ : həya /həəj > haəəj/,⁸

one can see that the new LR reflex of Middle Khmer /ii/ appeared before the LR reflex of /əə/. The four identical rhymes from the Cpā'pa sṛṭ occupy optional positions in the metrical scheme. While it is possible that their constituents may not have been meant to rhyme, it is probable that they were so intended. By the time of this late text the prior member had certainly developed from /pʔdii/ to /pdəəj/. The change /ŋaaj > ŋéəj > ŋíəj/ is established but not yet dated in relation to other changes. The important thing here, however, is that the syllable nucleus before final /j/ typically undergoes in modern standard Khmer a shift of stress and length from [i:ə] to [ie:] and even [je:] which would make it more than ordinarily congruent with /pdəəj/. We can therefore suppose that the rhyme in question was as admissible as the ḥṭña : bhliəña mentioned above.

We turn now to imperfect i rhymes. Table 6 shows, first of all, a preponderance (84 items = 49.1%) of rhymes in which one member of the pair has graphic iya. These are somewhat awkward to classify inasmuch as iya, while written with short i, is an allograph of long \bar{I} . As such it occurs paired either with itself or with \bar{I} in all but four cases. Like \bar{I} , iya is reflected by modern forms of both registers. The bulk of the rhymes in which it occurs are hence divisible into three types according as their members show HR or LR reflexes. As before, in one type both members presuppose HR reflexes, e.g. 'indriya /qəntrii/ : byādhṭ /pjíiəthii/ (KC 50ba) and muḥhni /muhni/ : medriya /méetrii/ (IMA 126ce). In a second type, both members presuppose LR reflexes, e.g. lokiya /lóokəəj/ : pṭ /bəəj/ (K 9c/9g) and lokiya /lóokəəj/ : 'apriya /qaprəəj/ (K 9ce). In the third type of rhyme one member is reflected by a HR form, the other by a LR form, e.g. 'indriya /qəntrii/ : pṭ /pdəəj/ (KK 2lcb) and 'apriya /qaprəəj/ : lajṭ /lacci/ (KC 19g/20c). It can be seen from this much that rhymes of this type are to be reassigned to perfect \bar{I} rhymes. Like the latter, they tell us nothing of the pre-modern value of \bar{I} or i. It is nonetheless of interest to note that the ratio of the three types of rhyme just described is decidedly uneven. In 59 cases (70.2%) both members correspond to modern HR forms. In the remaining eighteen cases (21.4%) the two members have different registers.

It was mentioned above that there are four exceptions to this pattern. In these graphic iya is in rhyme with graphic ai and its āya allograph. The rhymes in question are:

HMI	25bc	-tirthiya /-dèerthəəj/	:	bhāya /phéj/ ⁷
IMA	25cb	-dhivatiya /-thipdəəj/	:	didaiya /tiitáj/

TABLE 6
IMPERFECT | RHYMES

	total rhymes	iya	e	ä	ı	other	total	
<i>Ker(tı) k̄ala</i>	105	2	2	0	1	0	5	4.8%
<i>Kūna cau</i>	328	7	0	2	0	0	9	2.7%
<i>Rājaneti</i>	214	3	0	0	3	1	7	3.3%
<i>Krama</i>	154	6	1	0	0	0	7	4.5%
<i>Hai mahājana (I)</i>	337	5	2	1	2	2	12	3.6%
<i>IMA 38</i>	493	15	7	2	0	1	25	5.1%
<i>Kūna cau lpøka (A)</i>	297	2	0	2	0	0	4	1.3%
<i>Vidhūrapandita</i>	225	4	1	0	1	0	6	2.7%
<i>Pantām pitā</i>	295	2	2	2	0	0	6	2.0%
<i>Kūna cau lpøka (B)</i>	127	0	1	0	1	0	2	1.6%
<i>Trīneti</i>	380	9	1	1	0	0	11	2.9%
<i>Dūnmāna kh̄lwna</i>	175	5	1	0	0	0	6	3.4%
<i>Bākya cā'sa</i>	97	0	0	0	3	0	3	3.1%
<i>Hai mahājana (II)</i>	324	0	8	2	0	0	10	3.1%
<i>Pantām 'ū buka</i>	291	2	0	2	0	0	4	1.4%
<i>Dūnmāna kūna</i>	251	3	3	0	2	0	8	3.2%
<i>Sr̄t</i>	579	12	6	11	0	2	31	5.4%
<i>Prusa</i>	298	3	3	3	1	0	10	3.4%
<i>'Ariyasatthā</i>	194	4	1	0	0	0	5	2.6%
	5,164	84	39	28	14	6	171	3.2%

DKh	20ab	lokiya /lókəəj/	:	ksāya /ksaj/
DKn	75c/76a	lokiya /lókəəj/	:	'ai /qaj/

These four are to be reassigned to the imperfect ai (āya) : T which have been treated elsewhere.¹²

The next largest group on Table 6 consists of 39 i : e rhymes, representing 22.7% of all imperfect i rhymes. These are of sufficient interest to warrant being listed in full:

KK	16cb	sarila /sarəl/	:	ker(ti) /kèer/ ⁷
	20ab	rih /rih/	:	ceh /cèh/
K	5ba	khjila /kcil/	:	sarasera /səarsèer/
HMI	55ab	kipca /kəc/	:	bhleca /pléec/
	73d/74b	jamnih /cumnih/	:	radeh /rtéh/
IMA	3d/4b	-gitta /-kit/	:	sarbvejña /sarpéec/
	18ab	tica /təc/	:	bejña /péec/
	39ba	rih /rih/	:	nēh /néh/
	57ab	kica /kəc/	:	sreca /srèec/
	119fe	miña /mip/	:	pambeña /bampéer/
	140ab	samliña /samlèeŋ/	:	phsiña /psèeŋ/
	150ab	viña /wip/	:	beña /péer/
V	46bc	miña /mip/	:	ceña /cèer/
PP	21ec	tamrih /tamrèh/	:	camneh /camnèh/
	21ef	tamrih /tamrèh/	:	neh /néh/
KCLB	35ba	tamrih /tamrèh/	:	ceh /cèh/
T	79fe	pantica /bantəc/	:	greca /kréec/
DKh	5ba	diña /tip/	:	ceña /cèer/
HMII	8ab	tica /təc/	:	bhleca /pléec/
	10d/11b	tica /təc/	:	geca /kéec/
	11dc	banlica /punlic/	:	pambhleca /bampléec/
	51d/52b	viña /wip/	:	jreña /créer/
	57d/58b	viña /wip/	:	ceña /cèer/
	58cb	viña /wip/	:	ceña /cèer/
	60d/61b	viña /wip/	:	'amneña /qamnéer/
	78cb	lica /lic/	:	steca /sdac/ (sic)
DKn	71c/72a	viña /wip/	:	'amneña /qamnéer/
	72ba	viña /wip/	:	'amneña /qamnéer/
	91a/90c	rih /rih/	:	neh /néh/
S	49a	kipca /kəc/	:	beca(na) /péec/
	77ab	viña /wip/	:	mneña /mnéer/
	162c/163a	viña /wip/	:	pambeña /bampéer/
	163ba	viña /wip/	:	pambeña /bampéer/
	164a	tih /tèh/	:	ceh /cèh/
	221ba	kipca /kəc/	:	beca(na) /péec/

P	12bc	miña /mip/	:	ceña /cèep/
	16b/15d	viña /wip/	:	ceña /cèep/
	16bc	viña /wip/	:	ceña /cèep/
A	19ab	pantica /bantəc/	:	bhleca /pléec/

One can see at a glance that these rhymes fall into two main groups. They comprise, in addition to a residue of three rhymes, eight rhymes (20.5%) in which the *i* and *e* members occur before the visarga (/~h/) and 28 rhymes (71.8%) in which the *i* and *e* members occur before a palatal final, in 12 cases /-c/, in 16 cases /-p/. This majority includes the defective rhyme cited above from IMA 38 brahmagitta (now brahmagīti) : sarbvejña (now sarbejña), about which there is little that can be said. For the rest it is enough to mention that in modern standard Khmer the palatal stop and nasal have, when final, the effect not only of typically inducing a palatal glide before their main articulation but also of definitely muddling the quality and the quantity of the preceding syllable nucleus. In impressionistic terms the vowels of miña (formerly meña) and ceña are indistinguishable; length contrast between them is lost, while both are centralized and lowered to a degree difficult to define without acoustic analysis. Thus the pronunciation of stecca (formerly stā¹ca) recommended by the semi-official *Vacanānukrama khmēra* is /sdac/ [sʔdɑvč], as above.¹³ There is no reason not to attribute these same phenomena to Middle Khmer, since they offer the simplest justification for the rhymes in question.

The eight rhymes with visarga can be explained in terms of divergent development. The form riḥ (formerly reḥ) occurs three times, and changed from Middle Khmer /reh/ to early modern HR /réh/ [rɪh], after which its orthography was adjusted. The form ceḥ and its derivative camneḥ account for another three occurrences, and change from Middle Khmer /ceh/ and /cəmneh/ to modern LR /cèh/ [čɛh] and /camnèh/ [čəmneh]. Hence the rhyme riḥ : ceḥ (KK 20ab) would have sounded as [reh : čɛh] at the time of composition. The form neḥ (formerly neḥ or nēḥ) also occurs three times; this changed from Middle Khmer /neh/ to modern HR /néh/ [nɪh], without modification of the orthography. The rhyme riḥ : neḥ (IMA 39ba, DKn 91a/90c) would therefore have sounded as [reh : neh] at the time of composition. The form tamriḥ (formerly tamreḥ) also occurs three times, and underwent the same change as riḥ. Hence tamriḥ : neḥ (PP 21ef), tamriḥ : ceḥ (KCLB 35 ba), and tamriḥ : camneḥ (PP 21ec) would all have been in rhyme at the time of composition. The form jamniḥ /jəmniḥ, cumniḥ/ has undergone no change in its main syllable, and can still rhyme with radeḥ /rdeh, rtéh/ [r³tɪh] cart. The form tiḥ apparently changed from Middle Khmer /tih/ to early modern

LR /tèh/, whence modern /tèh/; presumably it could have rhymed with *ceh* (S 164a) only before the development of the modern standard language.

Our three-item residue consists of two rhymes with liquid finals, *sarila* /saril > sarəl/ : *ker(ti)* /kir > kèer/ (KK 16cb)⁸ and *khjila* /kʃii > kcil/ : *sarasera* /sərsir > sarsèer/ (K 5ba),¹⁴ and one rhyme which appears to have undergone little but orthographical change: *samliña* (now *samteña voice*) and *phsiña* (now *phseña to be different*) at IMA 140ab, which seem to have rhymed as /səmleɲ/ : /pseɲ/ and hence to be reassignable to *e*.

The next largest group on Table 6 consists of 29 rhymes in which *i* is paired with an \bar{a} or *a* type graphic vowel other than *ai* ~ *āya*. These represent 16.9% of all imperfect *i* rhymes, and fall into several classes. One of these comprises seven rhymes (six from the late *Cpā'pa* $\Delta\bar{a}\bar{z}$, the other from the still later *Cpā'pa prusa*) with palatal finals, exemplified by *miña* /mip/ : *sam⁴ā¹ña* /samlap/ (S 141c/142a) and *tica* /təc/ : *tā¹ca* /dac/ (P 32ba). These may be grouped with *i* : *e* rhymes with palatal finals, already discussed, and on identical grounds. One is tempted to dismiss the remaining rhymes as simply defective, but closer examination suggests that this may not be the case with the majority. Indeed, some of these rhymes may constitute the best evidence available for the early bifurcation of Middle Khmer /i/ into HR /i/ and LR /ə/ [e^ʔ ~ ə^ʔ]. In particular, four rhymes (including one duplication) consist of loans from Sanskrit or Pāli, the *a* member of which may have been given its Indic (= learned) value of [ə] instead of its Khmer value of [ɔ:],¹⁵ though the actual pronunciation in Cambodia may well have been [a^ˆ] – that is to say, a low central vowel raised toward [ʌ]. These are:

IMA 8dc	<i>sthita</i> /sthit > sthət/ ⁸ <i>bodhisatva</i> /boodhisət > póothisat/
T 63de	<i>citta</i> /cit > cət/ <i>pranipāta(na)</i> /prənipət > pranəbat/
HMII 17ab	'ānita /qaanit > qaanət/ <i>satva</i> /sət > sat/
HMII 62ab	'ānita /qaanit > qaanət/ <i>satva</i> /sət > sat/

More will be said regarding this type of change in a separate study of the low vowels of Middle Khmer. For the moment, let it be noted that two other rhymes containing Indic loans cannot be justified on the same basis as the four just given, inasmuch as their *i* members never descended to the emergent LR: *bita* /bit > pit/ : *rata(na)* /rət > rat >

roet/ (KC 57ab) and gita /git > kit/ : -ksatra /-ksət > -ksat/ (HMI 29d/30b). With these may be grouped gita /git > kit/ : mā¹ta /mat > moət/ (S 75c/76a), which does not so far as is known contain an Indic loan. Regarding these three it seems best to withhold judgment. On the other hand, in the following eight cases (including two duplications) the rhyme may have pivoted on a raised variety of /a/ in imitation of, or at least set off by, the learned pronunciations just seen:

KC	10ab	citta /cit > cət/ pā ¹ ta /ʔbat > bat/
IMA	123de	'issa /qih > qəh > qah/ (now 'a sa) hnāssa /nah/ (now nā ¹ sa)
KCLA	56ab	pañhina /ʔbəŋhin > baŋhən/ khjā ¹ na /kʃan > kcoən/
PP	56ab	pañhina /ʔbəŋhin > baŋhən/ khjā ¹ na /kʃan > kcoən/
PUB	56ab	pañhina /ʔbəŋhin > baŋhən/ khjā ¹ na /kʃan > kcoən/
S	76ba	citta /cit > cət/ mā ¹ ta /mat > moət/
S	169c/170a	citta /cit > cət/ sñā ¹ ta /sɲat/
P	56cd	cina /cin > cən/ dā ¹ na /dan > toən/

In two rhymes modern editions of our texts show the orthographic form 'a¹ta /qət > qat/ *to withstand*, which the original manuscripts must have represented as 'ita /qit > qət/:

S	67c/68a	citta /cit > cət/ 'ita /qit > qət/
P	62ab	jita /ʃit > cit/ 'ita /qit > qət/

This change is well attested, being already met with in the form 'issa (now 'a¹sa) just cited from IMA 123d; it is also seen in the free variation in the modern language between 'ita /qət/ and 'a¹ta /qat/ *to lack, be missing*.¹⁶ Even this variation, however, cannot justify the 'ita /qit > qət/ : khnāta /knaat/ found first in the *Kūna cau lpxəka* (A) (24ed) and again in the *Pañtām pitā* and the *Pañtām 'ū buka* (also 24ed). These three occurrences, together with sina /sin > sən/ : da¹na /dɔn > tun/ (S 79c/80a), must be regarded as defective.

The next group of imperfect i rhymes comprises 14 cases in which orthographic i is paired with ɨ, representing HR /y/ [w^v] and LR /y/ [y^v ~ ə] in modern standard Khmer. These appear to pose no problems.

Contrast is easily lost between their modern realizations on both registers, and there is no reason to assume contrast was formerly greater than today. The modern orthography in fact tolerates a number of alternant forms such as *kritya* /krèt/ and *kritya* /krýt/ while the dictionary of Guesdon, published just as the modern orthography was being promulgated, records hundreds of alternants such as *niña* /niŋ/ and *n+ña* /nýŋ/ *with*. This small group of rhymes may therefore be exemplified by *gita* /git > kit/ : *ñañita* /ŋŋyt > ŋŋýt/ (HMI 22ab, P 22ba), *smita* /smit > smèt/ : *samriddhi* /sòmryt > sòmryt/ (V 49ba), and *sucarita* /sucarit > socarèt/ : *briddha* /pryt > prýt/ (KK 1ba).

Lastly, Table 6 shows a residue of six imperfect rhymes in which i is paired with 'other' orthographic vowels. Three of these include the anomalous form *sTla* /səl/, previously discussed and reassigned to short i : *khjila* /kʃil > kcil/ : *sTla* /sil > səl/ (R 12 ab), *'ampila* /qəmʔbil > qambəl/ : *sTla* /sil > səl/ (HMI 77b/76d), and *babila* /ppil/ : *sTla* /sil > səl/ (HMI 77cd). A fourth, *pica* /ʔbic > bəc/ : *cuca* /cuc > coc/ (S 208a), is probably referable to the muddling attendant upon palatal finals, also previously discussed. The other two may well have been defective: *'isa* /qih > qah/ (now *'a'sa*) : *dosa* /dooh > tóoh/ (IMA 47a) and *citta* /cit > cèt/ : *chuta* /chut > chot/ (S 216a).

Table 7 shows that imperfect \bar{u} rhymes number 79 and have an average frequency of 1.5%. Of these 35 (44.3%) are rhymes in which orthographic \bar{u} is paired with w , while 33 (41.7%) have \bar{u} paired with au . Of the remainder six (7.6%) have \bar{u} paired with o , while five (6.3%) make up a small residue in which \bar{u} is paired with other orthographic vowels. It is worthy of note that imperfect \bar{u} rhymes are not found in the first, second, and fourth of our texts, that they are of minimal frequency in the third, fifth, sixth, eighth, and eleventh texts, and that they reach a peak frequency of 6.8% only in the fairly late *Dūnmāna kūna*. However, apart from this striking maximum, the rise and fall of frequencies for the other texts follows no recognizable pattern, and the most that can be said is that imperfect \bar{u} rhymes are less common in the early texts than in the later texts.

The 35 $\bar{u} : w$ rhymes are reflected in the modern language by ten HR: HR rhymes, sixteen LR : LR rhymes, and nine HR : LR rhymes. These are exemplified by

KCLA	9ce	by \bar{u} (ha) /pjuu/ ⁷	gwra /kúuær/
DKn	21	kūna /kòon/	sñwna /sŋùuæn/
HMI	48d/49b	tp \bar{u} ña /tbòŋ/	lwña /lúuəŋ/

TABLE 7
IMPERFECT ū RHYMES

	total rhymes	w	au	o	other	total
<i>Ker(tī) kāla</i>	105	0	0	0	0	0
<i>Kūna cau</i>	328	0	0	0	0	0
<i>Rājaneti</i>	214	0	0	1	0	1 0.5%
<i>Krama</i>	154	0	0	0	0	0
<i>Hai mahājana (I)</i>	337	0	2	1	0	3 0.9%
<i>IMA 38</i>	493	2	0	0	0	2 0.4%
<i>Kūna cau lṛaeka (A)</i>	297	3	2	0	1	6 2.0%
<i>Vidhūrapandita</i>	225	0	0	1	1	2 0.8%
<i>Pantām pitā</i>	295	3	2	0	0	5 1.7%
<i>Kūna cau lṛaeka (B)</i>	127	0	2	0	0	2 1.6%
<i>Trāneti</i>	380	0	2	0	0	2 0.5%
<i>Dūnmāna khḷwṇa</i>	175	0	1	0	1	2 1.1%
<i>Bākya cā'sa</i>	97	0	1	0	1	2 2.1%
<i>Hai mahājana (II)</i>	324	3	2	1	0	6 1.9%
<i>Pantām 'ū buka</i>	291	3	2	0	1	6 2.1%
<i>Dūnmāna kūna</i>	251	10	6	1	0	17 6.8%
<i>Srī</i>	579	8	4	1	0	13 2.2%
<i>Prusa</i>	298	3	1	0	0	4 1.3%
<i>'Ariyasatthā</i>	194	0	6	0	0	6 3.1%
	5,164	35	33	6	5	79 1.5%

In weighing the significance of these rhymes it is worth noting in the first place that the seventy lexical items they include show a decidedly limited range of finals: on the one hand, 38 items with /-n/ and four with /-ŋ/, on the other hand, 23 items with /-r/ and five with zero final. It is well known that in modern Khmer a final dental nasal or liquid may induce before itself a neutral phonetic glide: *jūna* /cuun/ [č̥u:ⁿ], *kūna* /kòon/ [ko:ⁿ]. It may be mentioned in the second place that the distribution of these seventy items according to their modern register is uneven. HR ū /uu/ and w /úuə/ account for 29 items (41.4%): two in *IMA 38*, six in the *Kūna cau l̥pæka* (A), four in the *Hai mahājana* (II), four in the *Dūnmāna kūna*, and one in the *Sr̥*. LR ū /òo/ and w /ùuə/ thus account for 41 items (58.6%), but are clustered in the later texts: two in *IMA 38* and two in the *Hai mahājana* (II) but sixteen in the *Dūnmāna kūna*, fifteen in the *Sr̥*, and six in the *Prusa*. Although it can be assumed that this concentration of LR forms (23 ū, 17 w items) is fortuitous, it would appear to constitute particularly cogent proof that, as late as the time of the *Cpā'pa prusa*, Middle Khmer /uu/ (ū) had not yet bifurcated into modern HR /uu/ and LR /òo(w)/ or, perhaps it should be said, had not yet developed fixed registral contrast. Rhymes on the order of *dhūra* /dhuur > thuur/ : *gwra* /guuər > kúuər/ (KCLA 9 fe) are not difficult to justify once we accept the possibility of [dhu:ⁿr]. On the other hand, while a rhyme foreshadowing LR ū /òo/ from an early text such as *tr̥sūra* (Sanskrit *triśūla*) *tr̥isuur* > *tr̥əjsòor* : *gumnwra* /gumnuuər > kumnúuər/ (*IMA 26d/27b*) may not surprise us, a similar rhyme from a late text such as *p'ūna* /pquun > p̥òon/ : *khlwna* /kluuən > kl̥uuən/ (*DKn 56c/57a*, P 45ab) is difficult to accept as normal, albeit imperfect, unless we see LR /òo/ as not yet fully established.

However, this reminds us of our next largest group of imperfect ū rhymes, namely the 33 in which ū is paired with au. These have already been listed in full and discussed in sufficient detail in an earlier paper,¹⁷ where it was shown that most of the au items in question presuppose modern HR /éw/ while most of the ū items presuppose modern LR -ūva /òow/ [ɣ:w]. It was also shown that this type of rhyme increases in frequency, from earlier to later texts, with the diphthongization and lowering of -ūva from [w:w] to [ɣ:w] as the emergent LR became fixed. This observation would seem to modify somewhat, or at least place in better perspective, what has just been said regarding ū : w rhymes.

The next largest group of imperfect ū rhymes comprises six ū : o rhymes, as follows:

R 16de grū /kruu/ : bola /póol/¹⁸

HMI 74ab	dūka /tuuk/	:	goka /kóok/
V 5bc	drūña /truu /	:	camkoña /camkaaon/
HMII 16cb	santūka /sandòok/	:	thoka /thaaok/
DKn 33	kūna /kòon/	:	lamdona /lumtóon/
S 185c/186a	cambūka /campuuk/	:	soka /saaok/

The first thing we observe in these rhymes is that two of the \bar{u} members (santūka and kūna) are on the LR and three of the \bar{o} members (camkoña, thoka, and soka) are on the LR. This observation does not appear to be helpful, however. These rhymes are to be compared with the $\bar{T} : e$ rhymes already discussed. In particular we are reminded that, like the modern registral reflexes of Middle Khmer /ee/, those of /oo/ may differ considerably in tongue-height, HR /óo/ being often articulated as [o:], LR /òo/ being often [ɔ:]. Moreover, in the orthographical forms samliña : phsiña (now samteña : phseña), already cited from IMA 38, we have a good indication that Middle Khmer /ee/ was heard as a lower-high [ɪ:] at the opening of the 18th century notwithstanding the fact that both of the forms in question were to develop LR nuclei in the modern language. This qualifies what was argued earlier on the basis of the repeated rhyme ralīña : dheña. The vowel of dheña may well have been [ɪ:] but not, after all, because it was evolving into the HR reflex of /ee/. It may have originally been on this level. In the case of $\bar{T} : e$ rhymes general conclusions, even of a provisional kind, were ruled out by what was found regarding the form ker(ti), which significantly reduced the number of rhymes remaining. In the case of the present $\bar{u} : \bar{o}$ rhymes we may have sufficient reason to postulate that Middle Khmer /oo/ was characteristically articulated on the lower-high level as [o:]. Particularly if the thesis is accepted that registral contrast for Middle Khmer /uu/ was not fully established by the time of the *Cpā'pa prusa*, we are unable to force a justification of three of the six rhymes listed above by juggling the registral and prerregistral forms of the orthographic vowels. Specifically, drūña /druuŋ > truuŋ/ : camkoña /cɔmkooŋ > camkaaon/, santūka /sɔn?duuk > sandòok/ : thoka /thook > thaaok/, and cambūka /cɔmbuuk > campuuk/ : soka /sook > saaok/ are plausible only if we interpret the pre-modern /oo/ as [o:]; and if we admit this, we provide the rationale for all six rhymes including grū /gruu > kruu/ : bola /bool > póol/, dūka /duuk > tuuk/ : goka /gook > kóok/, and kūna /kuun (> kòon)/ : lamdona /lɔmdoon > lumtóon/. All six rhymes, in other words, paired [u:] with [o:].

Our five-item residue comprises three $\bar{u} : a$ (a') rhymes and two $\bar{u} : u$ rhymes, as follows:

KCLA	23fe	tūña /tòŋ/	:	daña /tsoŋ/
DKh	25d/26b	kūna /kòon/	:	la ¹ na /lun/
PUB	47ef	mūla /muul/	:	ya ¹ la /jul/
V	39ba	pradūsta /pratuuh/	:	khusa /khoh/
B	1lab	yūra /juur/	:	dur- /tur-/

Although it was mentioned above that LR /ò/ is often [ɔ:] in modern standard Khmer, we cannot assume that the first of these rhymes sounded originally as [tɔ:ŋ] : [dɔ:ŋ]. To do so would be to contradict our assumption that registral contrast for Middle Khmer /uu/ was not fully established as late as the time of the *Cpā'pa prusa*, an assumption which has proved useful thus far. Modern *daña*, *ya¹la*,¹⁹ and probably also *la¹na* belong to an important group of lexical items which have *o* in Old Khmer but undergo lowering (e.g. *phoñ* /phooŋ/ > *phaña*/phaaŋ/ *in company with, together*), and should hence to reassigned to the *ū* : *o* rhymes previously discussed. The original rhymes were almost certainly *tūña* /tuuŋ/ : *doña* /doŋ/ [dɔ:ŋ], *kūna* /kuun/ : *lona* /loon/ [lɔ:n], and *mūla* /muul/ : *yola* /jool/ [jɔ:l]. In the case of *pradūsta* : *khusa* the two syllable nuclei have the same phonetic length but likewise show different degrees of openness. Despite its phonemic form, *pradūsta* /pratuuh/ *to fault* reverts to a short vowel (Sanskrit *praduṣṭa*) under the influence of its final /-h/, while *khusa* /khuk > khoh/ had not yet manifested the LR. The rhyme was therefore [prəduh] : [khoh]. In the last rhyme *yūra* /juur/ *to be long in time* is paired with the first syllable of *durjana*²⁰ /turjun/ *wicked person*. The vowel of this syllable, with the gradual loss of final /-r/, had probably undergone compensatory lengthening by the time of the *Bākya cā'sa*, though it is difficult to say whether the /-r/ had entirely disappeared. With this reservation, the rhyme sounded as [ju:(r)] : [du:(r) > tu:].

We turn now to imperfect *u* rhymes, the types and distribution of which are laid out in Table 8.

The first and largest group of imperfect *u* rhymes comprises 98 items (82.3% of all such rhymes) in which *u* is paired with shortened *a*. This is the one type of rhyme which cannot be satisfactorily analyzed from the date provided by the modern orthography. There was considerable vacillation between *u* and *a* in Middle Khmer times while original /ɔ/ [ɔ^v] after voiced initials was being raised to modern [ɔ] in the formation of the HR.²¹ As a result, continuity with the Old Khmer orthographic form of the items in question here was lost. As often as not the modern regularized orthography has fixed upon one symbol or the other in a wholly arbitrary manner. Pending closer examination of the manuscripts, therefore, we shall describe the data on hand without drawing any firm conclusions.

TABLE 8
IMPERFECT u RHYMES

	total rhymes	ă	o	other	total
<i>Ker(ti) kāla</i>	105	3	0	0	3 2.9%
<i>Kūna cau</i>	328	2	0	0	2 0.6%
<i>Rājaneti</i>	214	0	0	0	0
<i>Krama</i>	154	0	0	0	0
<i>Hai mahājana (I)</i>	337	11	2	0	13 3.9%
<i>IMA 38</i>	493	10	0	0	10 2.0%
<i>Kūna cau lpæka (A)</i>	297	4	0	0	4 1.3%
<i>Vidhūrapandita</i>	225	1	1	1	3 1.3%
<i>Paṅtām pitā</i>	295	3	0	0	3 1.0%
<i>Kūna cau lpæka (B)</i>	127	3	0	0	3 2.4%
<i>Trīneti</i>	380	9	0	0	9 2.4%
<i>Dūnmāna khlwna</i>	175	6	3	0	9 5.1%
<i>Bākya cā'sa</i>	97	4	0	1	5 5.2%
<i>Hai mahājana (II)</i>	324	11	1	1	13 4.0%
<i>Paṅtām 'ū buka</i>	291	3	0	0	3 1.0%
<i>Dūnmāna kūna</i>	251	4	1	1	6 2.4%
<i>Srī</i>	579	8	2	2	12 2.1%
<i>Prusa</i>	298	11	4	1	16 5.4%
<i>'Ariyasatthā</i>	194	5	0	0	5 2.6%
	5,164	98	14	7	119 2.4%

What is not shown on Table 8 is the frequency of u : a rhymes for each text. This averages 2.0% precisely. Individual averages below this point are far from what we should expect were the orthography unadjusted. They run, quite erratically, from the eighth text (the *Vidhūrapandita*, 0.4%) to the second (0.6%) and on to the fifteenth and ninth (both 1.0%), the seventh (1.3%), seventeenth (1.4%), and the sixteenth (1.6%). Above this same point averages start from the thirteenth text (the *Bākya cā'sa*, 4.1%) and drop to the eighteenth (3.7%), the fourteenth and twelfth (both 3.4%), the fifth (3.3%), the first (2.9%), the nineteenth (2.6%), and the eleventh and tenth (both 2.4%); the IMA stands right at 2.0%. The random nature of these frequencies clearly rules out any correlation with the date of the texts.

As before, it is convenient to classify u : a rhymes according to the register of their modern reflexes. In this way we obtain four groups, as follows:

HR : HR rhymes number 30, and are exemplified by the following:

HMI	68ba	rum /rum/ ⁸ dham /dhom > thum/
IMA	143ab	guna /gun > kun/ lana /lon > lun/
KCLA	49g/50c	muna /mun/ bandha /bon > pun/
V	68ba	jum /jum > cum/ dham /dhom > thum/
T	12ba	manussa /mnuh/ yasa /joh > juh/
DKh	27b/26d	buka /buk > puk/ naraka /nərok > naruk/
HMII	5ab	guna /gun > kun/ dāmna ¹ na /dōmṅon > tumṅun/
P	11bc	lupa /lup/ gra ¹ pa /grōp > krup/
A	24de	buta /but > put/ rama ¹ ta /rmot > rmut/

While one naturally supposes that these rhymes were devised after the /o > u/ shift and were meant to sound as [o] : [o], it is not altogether certain that this was the case. For in the next group of rhymes we meet evidence that [o] and [ɔ] could rhyme.

HR : LR rhymes number 8, as follows:

HMI	92ba	buka /buk > puk/ 'ākra ¹ ka /qaakrōk > qaakrak/
-----	------	---

IMA	87a/96c	ñuña /ruŋ/ paña /ʔbɔŋ > baŋ/
IMA	87ab	nuña /ruŋ/ camnaña /cɔmnoŋ > camnaŋ/
IMA	94a/93c	ghluña /gluŋ > klūŋ/ thlaña /tloŋ > tlaŋ/
IMA	94ab	ghluña /gluŋ > klūŋ/ 'aŋga /qɔŋ > qaŋ/
IMA	97c/98a	mukha /muk/ 'ākṛaka /qaakrɔk > qaakrak/
B	2cd	ghmum /gmum > kmum/ samñam /sɔmŋɔm > samŋam/
P	21bc	buña /buŋ > puŋ/ 'aŋga /qɔŋ > qaŋ/

This small group of forms is puzzling, but may contain invaluable indices. Four out of the eight rhymes in question come from IMA 38, and are therefore given here in the original orthography. The rhyme /muk/ (Sanskrit *mukha face, front*): /qaakrɔk/, providing it is not defective, suggests that [ɔ] and [ɔ̄] were in rhyme at the time of composition - possibly because the /ɔ > u/ shift was well under way by 1702 A.D. If this possibility is disallowed, we have two alternatives. On the one hand, we can assume that our eight u members were originally a forms representing Middle Khmer short /ɔ/, raised to [ɔ̄] and respelled accordingly before the date of composition. On the other hand, we can suppose that the relevant vowel of the a members for some reason passed through an intermediate [ɔ̄] before being lowered to /a/ [ɔ̄] for the emergent LR. Of these two possibilities the first is simple and in keeping with what is already known of Middle Khmer while the second is quite unlikely. Both, however, would disqualify /muk/ : /qaakrɔk/, which should not be done without firmer evidence.

LR : HR rhymes number 37, and are exemplified by the following:

KK	35ab	ṭula /ʔdul > dol/ ya ¹ la /jɔl > jul/
KC	14fe	knuña /knuŋ > knoŋ/ mama ¹ ña /mmoŋ > mmuŋ/
HMI	76dc	parisuddha /ʔbɔrisut > barəsot/ samba ¹ ta /sɔmbɔt > sampuṭ/
IMA	143ed	suña /suŋ > soŋ/ 'aŋga /qɔŋ > qaŋ/
KCLA	40fe	phlum /plum > plom/ pragam /prɔgɔm > prakum/

KCLB	8d/9b	kantura /kənʔdul > kandoi/ ya ⁱ la /jɔl > jul/
T	69ab	khusa /khuh > khoh/ ra ^l sa /rɔh > ruh/
DKh	16ab	cuka /cuk > cok/ lala ^l ka /llɔk > lluk/
HMI	55ab	punya /ʔbun > bon/ ja ⁱ na /jɔn > cun/
DKn	64c/65a	sukha /suk > sok/ lāmaka /laamɔk > līiəmuk/
S	192a/191c	chuta /chut > chot/ ma ^l ta /mɔt > mut/
P	51cb	prusa /pruh > proh/ yasa /jɔh > juh/
A	35de	suṅa /suŋ > soŋ/ la ^l ṅa /lɔŋ > luŋ/

These rhymes are simply an amplification of the HR : HR rhymes already given, inasmuch as they occur in texts composed before Middle Khmer /u/ split into modern HR /u/ and LR /o/. The qualification previously expressed applies here as well.

LR : LR rhymes number 22, and are illustrated by the following:²²

KK	6ba	kum /kum > kom/ pārambha /ʔbaarɔm > baaram/
HMI	1ba	prusa /pruh > proh/ sappurasa /sɔpʔburɔh > sapborah/
IMA	98ba	sukkha /suk > sok/ 'ākraka /qaakrɔk > qaakrak/
T	62ed	pamruṅa /ʔbɔmruŋ > bamroŋ/ phca ^l ṅa /pɔŋ > pɔŋ/
DKh	51ba	prusa /pruh > proh/ rapa ^l sa /rʔbɔh > rbah/
DKn	43ab	suṅa /suŋ > soŋ/ 'aṅga /qɔŋ > qɔŋ/
S	192ab	chuta /chut > chot/ 'a ^l ta /qɔt > qat/
A	5de	cuṅa /cuŋ > coŋ/ ca ^l ṅa /cɔŋ > cɔŋ/

These are to be added to the eight HR : LR rhymes given above. The presence of /suk/ (Sanskrit sukha) and /pruh/ (Sanskrit puruṣa) tends to legitimate the /muk/ : /qaakrɔk/ cited previously as a clue that [o] and [ɔ] rhymed at this period. Indeed, it seems more certain than

before that the a members of these 30 rhymes did not have [o] in the relevant syllable.

The next largest group of imperfect u rhymes given on Table 8 comprises 14 rhymes in which u is paired with o. It is worth noting that the u members of these rhymes are divided, in modern terms, between six HR and eight LR items while the o members are divided between seven HR and seven LR items. HR : HR rhymes are four in number, LR : LR rhymes, five, and rhymes of mixed register five. The rhymes are:

HMI	84ab	manussa /mnuh/ ⁸ smoh /smoh > smaoh/
HMI	95cd	khusa /khuh > khoh/ smoh /smoh > smaoh/
V	52ab	luh /luh/ noh /noh > nôh/
DKh	30ba	cuh /cuh > coh/ noh /noh > nôh/
DKh	32d/33b	dhluh /dluh > tluh/ noh /noh > noh/
DKh	33cb	khduh /kduh > ktuh/ noh /noh > nôh/
HMI	51ba	dhluh /dluh > tluh/ coh /coh > caoh/
DKn	49ba	cuh /cuh > coh/ smoh /smoh > smaoh/
S	199c/200a	susa /suh > soh/ kra'oh /krəqoh > kraqaoh/
S	200ba	susa /suh > soh/ kra'oh /krəqoh > kraqaoh/
P	22b/21d	juh /j̃uh > cuh/ noh /noh > nôh/
P	26d/27b	'usa /quh > qoh/ joh /j̃oh > cōh
P	27cb	khusa /khuh > khoh/ joh /j̃oh > cōh/
P	57b/56d	cuh /cuh > coh/ khnoh /knoh > knaoh/

It cannot escape notice that all 28 members of these rhymes have final /-h/. In the case of the u members this latter is represented by visarga in eight items, by final -sa in six items; in the case of all 14 o members it is represented by visarga, the shortening effect of which permits rhyme with u. It seems clear in this light that we have

here an extension of the ū : o rhymes previously discussed with the difference that, while the o items were meant to be articulated on the higher-mid level or higher, [o ~ o], as before, the u items were articulated on the lower higher level: manussa /mnuh/ [m^ənoh], luh /luh/ [loh], khusa /khuh/ [khoh], cuh /cuh/ [čoh]. It may be worth observing that nine of the above rhymes are probably still valid: four of the rhymes with modern /nóh/ and the five rhymes in which /-oh/ [-o^h] is paired with /-aoh/ [-əh]. The five rhymes which are no longer valid are, on the one hand, /mnuh/ [m^ənoh] : /smaoh/ [sməh] and /tluh/ [t^hloh] : /caoh/ [čəh] and, on the other, /coh/ [čo^h] : /nóh/ [nóh], /qoh/ [qo^h] : /cóh/ [čoh], and /khoh/ [kho^h] : /cóh/ [čoh].

The last group of imperfect u rhymes shown in Table 8 comprises seven rhymes in which u is paired with other graphic vowels. These are:

V	39ab	khusa /khuh > khoh/ ⁸ pradūsta /prəduuh > pratuuh/
B	11ba	dur- /dur- > tur-/ yūra /juur/
HMII	80ab	lupa /lup/ rā'pa /rap/
DKn	39	'aṅguya /qəŋguj > qəŋkuj/ kantəya /kəntəej > kantaəej/
S	208a	cuca /cuc > coc/ pica /?bic > bəc/
S	216a	chuta /chut > chot/ citta /cit > cət/
P	34ab	trum /trum > trom/ tām /?dam > dam/

The first two of these have already been discussed under imperfect ū rhymes, while *cuca* : *pica* is adequately explained by the palatal finals treated under imperfect i rhymes. The remaining four rhymes must be regarded as defective, and are probably the work of a careless copyist.

The main conclusions to which the present study leads may be re-stated as follows:

1. Imperfect T : āya rhymes demonstrate prosodic congruence between LR /əej/ (T) and preregstral /ej/ (āya) and at the same time show that the registral differentiation of Middle Khmer /ii/ (T) antedated that of Middle Khmer /ej/ (āya ~ ai).

2. Registral differentiation of Middle Khmer /əə/ (e > ə) was established after that for Middle Khmer /ii/ (T), and was not fixed as late as 1702 A.D.

3. It is probable that registral contrast between modern HR /*é*e/ and LR /*è*e/ (e) had appeared by the date of the *Cpā'pa pusa*.

4. It is probable that Middle Khmer /*aa*/ (*ā*) had bifurcated into HR /*é*ee > *í*iə/ and LR /*aa*/ by the time of the *Cpā'pa sūl*, if not much earlier.

5. As recently as the time of the *Cpā'pa pusa* the registral reflexes of Middle Khmer /*uu*/ (*ū*) were not fixed.

6. The preregistral articulatory level of /*ee*/ and /*oo*/ (e, o) may have been characteristically lower-high [ɪ:] and [o:], or may have ranged between lower-high and higher-mid, [ɪ: ~ e:] and [o: ~ u:].

N O T E S

1. Philip N. Jenner, 'The Relative Dating of Some Khmer Cpā'pa,' to appear in *Austroasiatic Studies*, edited by Philip N. Jenner, Laurence C. Thompson, and Stanley Starosta. Oceanic Linguistics Special Publications, No. 13 (Honolulu: The University Press of Hawaii, 1976). The texts in question comprise 17 undated works of the cpā'pa /cbap/ genre plus No. 38 of the *Inscriptions modernes d'Angkor (IMA 38)*, known as the *Grande inscription d'Angkor Vat (K. 301)* and dated śaka 1623 (= A.D. 1702 or 1701). The cpā'pa are homiletic works of surpassing interest from the linguistic as well as the cultural and literary points of view; IMA 38 is a devotional text no less interesting, included to provide one reference point in the chronological sequence. The titles of the cpā'pa are given on the accompanying tables. For the benefit of readers who have not seen my earlier papers it must be explained that there are two distinct texts known as the Cpā'pa hai mahājana and that the Cpā'pa kuna cau lpæka is a composite work, its first part consisting of a version of the Pantām pitā (itself a prototype of the Pantām 'ū buka), its second part consisting of a version of the Bākya cā'sa.
2. Philip N. Jenner, 'The Final Liquids of Middle Khmer,' to appear in a forthcoming issue of *Zeitschrift für Phonetik*.
3. Philip N. Jenner, 'The Value of au and ai in Middle Khmer,' to appear in *South-East Asian Linguistic Studies*, edited by Nguyen Dang Liem. Pacific Linguistics, Series C - No. 31 (Canberra: The Australian National University, 1974).
4. Philip N. Jenner, 'The Development of the Registers in Standard Khmer,' in *South-East Asian Linguistic Studies*, 47-60.

5. Glosses are not furnished for the forms cited, inasmuch as we are concerned with a phonological problem. The examples are given first in an Indianist transliteration and again, between slants, in phonemic transcription. In the notation of modern standard Khmer the acute (´) marks HR, the grave (`) marks LR, for 14 otherwise ambivalent nuclei.

6. See 'The Development of the Registers...'

7. All forms save those cited from IMA 38 are given in the modern regularized orthography. Unless otherwise indicated, all phonemic transcriptions represent the modern realization, for the reason already given.

8. The transliterations continue to show the modern orthography, but the first phonemic form represents Middle Khmer while the second represents modern Khmer. The symbols /ʔb, ʔd/ are used for the unitary imploded voiced labial and dental stops of Middle Khmer, which contrast with exploded /b, d/. This contrast is lost in modern Khmer, and /b, d/ are to be construed as implosive.

9. The parentheses in *ker(ti)*, now /kèer/, correspond to the *danda-ghāta*, a diacritic 'canceling' final written syllables. The Sanskrit form alone warrants our assuming that the earliest Khmer pronunciation of this item was [ki:r]. Perhaps under the influence of the orthographic -rt- sequence, this must have begun to be altered to [kir] by the date of the *Cpā'pa ker(ti) kāla*. For this text we can postulate /kiir/, rhyming as here with /piir/, and also /kir/ or even /kil/, rhyming with /saril/ (KK 16cb). Both pronunciations were possible during the early Middle Khmer period but with the progressive loss of final /-r/, with intolerance of short nuclei in open syllables, and with the development of the registers the original nucleus had to be lowered. In open syllables Middle Khmer /ii/ appears to have first dropped to higher-mid front [e:] and then to have undergone the typical centralization seen in such syllables as pī /ʔbii > bəəj/; in syllables closed by final /-r/, however, this change was arrested before centralization set in. That the same change could occur when the original vowel was short is shown by modern *sira* /sèer/ *head*, < Pāli *sira*, and its alternants *sira(sa)* /sèer/ (< Sanskrit *śiras*) and *sirsa* ~ *sir(sa)* /sèer/ (< older Sanskrit *śīrṣa*), for which see VK II: 1352b, 1353a, 1353b. Note finally that the VK (I: 39b) sanctions *kir(ti)* /kèer/ as a variant of *ker(ti)* and attributes both to a nonexistent Sanskrit *kirti.

10. See Saveros Pou and Philip N. Jenner, 'Some Chinese Loanwords in Khmer,' in *JOS*, XI (1973).1: 45, item 148.
11. At least it can be noted that *ramila* /rmil/ *to glance at* is an /r-/ derivative of an allomorph of *mæla* /mæəl/ *to look at* and may have had a long as well as a short vowel. The syllable *bi-* in the next rhyme is the initial of *bicāraṇā* (Sanskrit and Pāli *vicāraṇa*) *investigation* and like other loans with the same prefix is commonly articulated [pi?] or even [pi:] rather than [pɪ?].
12. See 'The Value of *au* and *ai* in Middle Khmer.'
13. *VK*, II: 1437b.
14. For the disparate liquid finals see 'The Final Liquids of Middle Khmer.'
15. W.A. Allen, *Phonetics in Ancient India* (London: Oxford University Press, 1961), 57-61, notably 58 and note 4.
16. *VK*, II: 1591b, 1772a; Joseph Guesdon, *Dictionnaire cambodgien-français* (Paris: Plon, 1930), I: 10a, 51a.
17. See 'The Value of *au* and *ai* in Middle Khmer,' especially Table 4, which includes two rhymes in *-ova* and *-o* from *IMA 38* which are not taken into account here.
18. Rhymes of this type have been discussed in 'The Final Liquids of Middle Khmer.'
19. The form *yola* /jool/ occurs in *IMA 12*, dated in correspondence with 1628 A.D.; see Saveros Lewitz, 'Inscriptions modernes d'Angkor 10, 11, 12, 13, 14, 15, 16a, 16b, et 16c,' in *BEFE0*, LIX (1972): 221-49, in particular 226. Most of the forty-odd Middle Khmer inscriptions which use the word at all have the orthographic form *yala*, with the length of the syllable nucleus unmodified by a diacritic. However, in *IMA 38*, the only metrical text in this corpus, the same form is consistently in rhyme with short syllables, e.g. '*akusala* /qakosaI/ (97a/96c), *tala* /daI/ (97ab), *saṃṇala* /samnaI/ (110c/109g), *mandala* /mundul/ (129ec), *kravala* /krawal/ (129ef). It is not without interest that modern editions of the earlier *Kūna cau lpæka* and *Paṅtām pitā* have *cūla* /còol/ *to approach, enter* where the *Paṅtām 'ū buka* has *ya¹la* /jul/

to see. In view of what has been said, we are led to assume that *ya¹la* is the original, and is roughly contemporary with *IMA 12*.

20. This form with short *u* is given in two of my editions of the *Bākya cā'sa* while that of the Institut Bouddhique gives *dūrjana*. The *VK* admits only the former.

21. See 'The Development of the Registers...'

22. Table 8 specifies 98 items whereas only 97 have been accounted for here. One defective rhyme has been omitted.

23. In my transcription, the nongeminate /*o*/ of /*smoh*/ [*smoh*] represents a short allophone of Middle Khmer /*oo*/ before visarga as opposed to final *-sa*, while the /*ao*/ of modern /*smaoh*/ [*smoh*] represents a short allophone of /*aao*/ in the same environment.

JAH-HUT, AN AUSTROASIATIC LANGUAGE OF MALAYSIA

G. Diffloth

Foreword

Introduction

1. Previous literature
2. Affiliation
3. Social situation
4. Syntax
5. Morphology
6. Phonology
7. Remarks on Jah Hut and Malay

FOREWORD

The present account of Jah Hut is based on several field trips I made to Malaysia in 1966, 1968, 1969, 1971 and 1973.

Each time, my work on Jah Hut was only a diversion from my main occupation: the study of Semai, Jah Hut's fairly distant, but closest relative, which will be the subject of a larger description. The total amount of direct contact I had with Jah Hut speakers is about one month, with long intervals between sessions for sorting out, comparing and thinking about the data. Most of my observations, but not all, were made by direct questioning, either in Malay or in my own hesitating Jah Hut. As my questions were not always meaningful to the Jah Hut, the answers were not always consistent. In such cases, I had to modify the question, or to rely on my experience with Semai, Temiar, Cheq Wong and Semelai to guess what the answer was likely to mean. Since my questions were also intended for comparison with Semai, I may have unwittingly introduced a pro-Semai bias in my description; for this, I alone am responsible since the Jah Hut do not know Semai. But as the two languages are related, I felt such a bias could be more

revealing than any other, if bias there must be. In such short time as I spent, I did not record much text, nor observe or participate in many spontaneous Jah Hut conversations, although these approaches would have been much more fruitful and reliable than direct questioning.

Published literature on related languages was helpful, up to a point; of the fifteen or so Aslian languages of Malaya, only Temiar, Semai, Kentaq Bong and Jahai have brief grammatical sketches; the present description of Jah Hut grammar, in spite of its many gaps and defects, is more comprehensive. As for the few other languages of the Mon-Khmer family for which we have full grammars: Khmer, Khasi, Palaung, Nicobarese, Chrau, Sre, they are too distantly related to Jah Hut to serve as guides. So nearly every single statement made here represents a step in the unknown, and a possible error.

In 1971, I was supported for travel expenses by the American Council of Learned Societies, and in 1973 I was supported by a grant from the National Science Foundation to the University of Chicago. At all stages of my research, the Jabatan Hal Ehwai Orang Asli (Department of Aboriginal Affairs, Kuala Lumpur) has been very helpful in granting me permissions to visit aboriginal areas. In 1971 and 1973, the Jabatan Perpaduan Negara (Department of National Unity) kindly gave me permission to conduct research in Malaysia. To all the Jah Hut I met, I am indebted for their hospitality and cordial assistance. This work is intended for their benefit.

INTRODUCTION

To the world outside, both in Malaysia and beyond, the Jah Hut people are practically unknown. Their language has never been written nor described, and the total amount of vocabulary printed in word lists probably does not exceed two hundred items. Yet, a study of the language will help us to understand certain problems and to raise new, more interesting questions. For instance, what does the presence of a Mon-Khmer language so far south in the Malay Peninsula mean for the linguistic history of South East Asia; or, does Jah Hut present, in the typology of ergative constructions, a kind of system which was so far unrecognized. These questions, and others, were in the back of my mind while I was studying Jah Hut, and have limited my queries.

1. PREVIOUS LITERATURE

The name Jah Hut does not appear at all in Skeat and Blagden's monumental work (W.W. Skeat, C.O. Blagden - 1906), but, by studying the Comparative Vocabulary (vol. 2, part IV) it is possible to ascertain that some groups which they included in their 'Eastern Sakai

cluster' must have spoken Jah Hut. However, within this 'cluster', one also finds groups who did not speak Jah Hut but Semaq Beri, a rather distant relative of Jah Hut. Semaq Beri is a South Aslian (and hence also Mon-Khmer) language, closely related to Mah Meri and Semelai (Benjamin, 1973a), whereas Jah Hut is a Central Aslian (or Senoic) language, whose closest relatives are Semai and Temiar. Judging from the Vocabulary, Jah Hut proper seems to correspond to what Skeat and Blagden called the 'Inner Subgroup of Eastern Sakai'. Vocabulary entries preceded by the labels: Sak. Guai, Krau Ket., Krau Tem., and Kerdau are clearly Jah Hut, while certain words entered as being Krau Em., U. Cher., and U. Tem., seem to have a Jah Hut origin.

R.J. Wilkinson (1926) rejected Blagden's inclusion of Jah Hut (which he calls 'Krau Sakai') in a special division of 'Eastern Sakai'. He collected fifty to sixty 200-word vocabularies of Aslian languages and had access to a 'very full vocabulary of the Krau dialect' collected by A.J. Sturrock, a District Officer in Temerloh (unpublished, unseen). Yet, he would not decide on the position of Jah Hut, and preferred to mention it in a chapter entitled 'Mixed and doubtful tribes' saying that it showed features of both 'Central Sakai' (i.e. Semai) and 'Jakun' (i.e. Semelai), without discussing any specific example.

Peter Williams-Hunt (1952) placed Jah Hut squarely within the Senoi branch together with Semai and Temiar; he referred to it either as 'South Eastern Senoi' or properly, as 'Jah Hut'. Unfortunately he also included Mah Meri in the Senoi branch, and used the term Jakun interchangeably with Semelai.

Robert Dentan (1964) had first-hand information on the Jah Hut language and noted the shortcomings of Williams-Hunt's classification but did not propose one of his own.

Finally, Geoffrey Benjamin (1973a) presented a full-fledged classification of Aslian languages with probable separation dates based on lexicostatistical techniques. Although we have reservations about glottochronology and the dates proposed, we agree with his language subgrouping, having reached ourselves very similar results by different methods. His subclassification of Senoic is essentially the one presented here, and the position of Jah Hut within Senoic which he proposes is confirmed by our study.

2. AFFILIATION

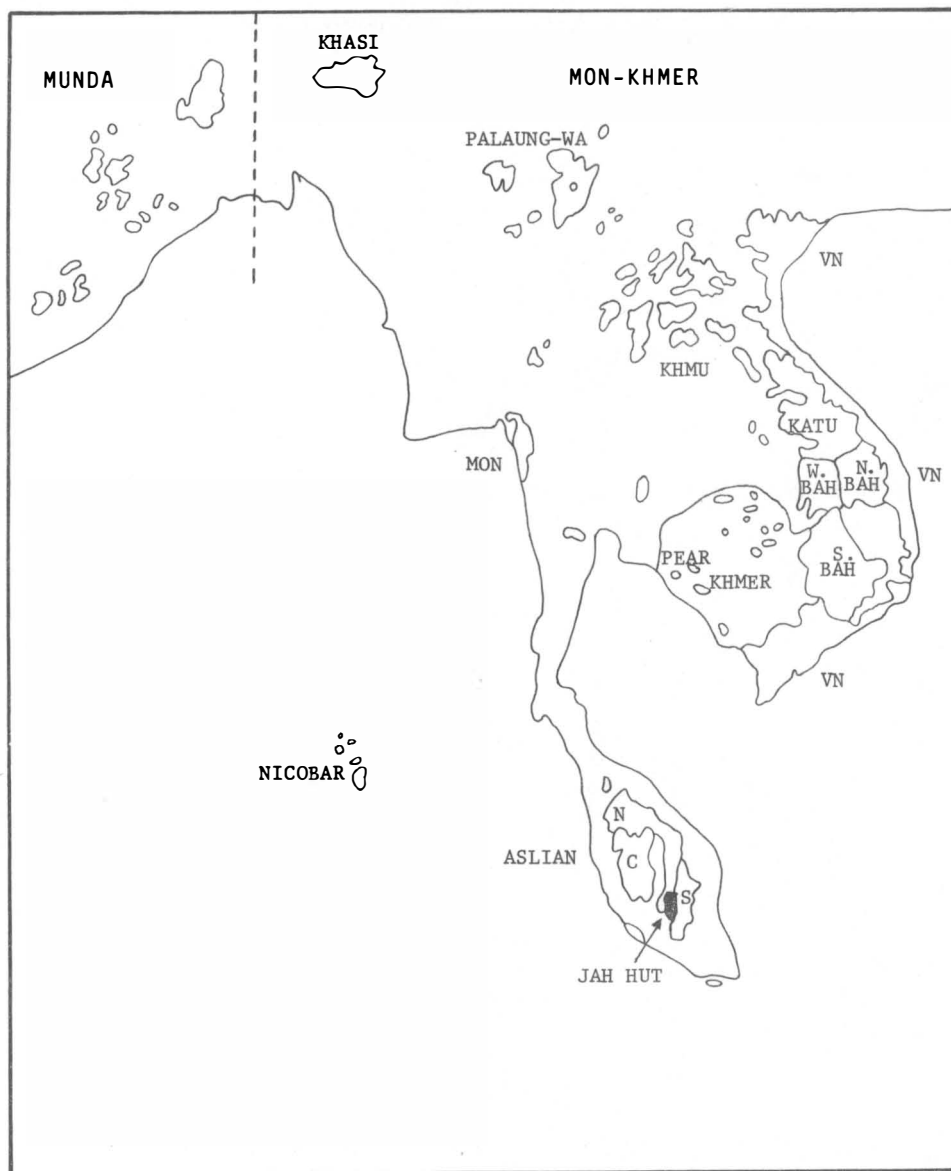
Jah Hut is an Austroasiatic language, and belongs specifically to the Mon-Khmer branch, but due to numerous internal changes, and to a good number of Malay borrowings, this fact is not immediately evident. Perhaps the most important change which has obscured the affiliation

of Jah Hut has been the loss of contrast between long and short vowels, a contrast which has been preserved in other Senoic languages (Semai, Temiar), and is found in most of the Mon-Khmer family. The only remnants of this contrast which survive in Jah Hut today are the diphthongs /ye, wə, wo, wɔ, wa, wɛ/ which correspond to Senoic long vowels; if these diphthongs are analyzed as clusters of consonant plus vowel, Jah Hut only has one vowel quantity, like North Aslian and Malay, but other analyses are possible (see 6.2.1.3.). As for the Lexicon, there is a tendency to use many Malay borrowings, especially when speaking to non-Jah Huts, but Benjamin's lexicostatistics (Benjamin, 1973a) show that the basic lexicon of Jah Hut remains Senoic, and therefore Mon-Khmer.

In spite of this obscured situation, there remains enough evidence of a structural nature to demonstrate the Mon-Khmer membership of Jah Hut in a fairly rigorous fashion. It is not our purpose to do that here, but one bit of evidence is worth mentioning, however briefly: the existence in Jah Hut of final palatal obstruents (/c/ and /ɲ/). These two finals are found in every single branch of Mon-Khmer and in Munda languages; they are to be reconstructed for Proto-Austroasiatic. Austronesian languages never have them, except when they were borrowed from Mon-Khmer, as in the Chamic languages of Viet Nam. They occur after any Jah Hut vowel, in words that have regular sound correspondences with Mon-Khmer cognates in Aslian, Nicobar, and all over continental South East Asia. Only a few examples will be given here:¹

- excrement* (SB: B161, D114): Jah Hut /ʔec/
 Khmer: ʔac, Proto-North-Bahnaric (238): *ʔc, Chrau: /ǎc/
 Pear: ich, Khasi: /ʔec/, Nancowry Nicobar: /ʔǎc/.
- to harvest, to pluck* (Pin. K40): Jah Hut: /kɛc/
 Khmer: /kac/, Proto-South-Bahnaric (353): *kac, Pear: khach,
 Khasi: /kʰɛc/, Nancowry Nicobar: /kéc/.
- meat, flesh* (SB: F170): Jah Hut: /sɛc/
 Khmer: /sac/, Proto-North-Bahnaric (247): *sǎc, Proto-East-Katuic (340): *sâJ, Old Mon: sac, /sɔc/ (*fruit*).
- ghost* (SB: G18): Jah Hut: /kmɔc/
 Khmer: /kʰmaoc/ (*corpse, ghost*), Proto-South-Bahnaric (p.29):
 *kɔʔmôc (*grave*), Pear: khmuch (*corpse*), Pacoh: /kumuuy?/.
- to weave, to plait* (Pin. 301): Jah Hut: /tap/
 Khmer: bɔnda:ɲ/ (spell.: paŋa:ɲ), Proto-North-Bahnaric (94):
 *tañ, Proto-South-Bahnaric (335): *tañ, Proto-East-Katuic (640):
 *taañ, Khasi: /tʰa:ɲ/, Pear: thanh, Nancowry Nicobar: /tǎɲ/.
 The Aslian words given in SB:P126 actually mean *to braid* and

MAP 1: AUSTRO-ASIATIC LANGUAGE GROUPS



represent a different etymon, but the Mainland-Mon-Khmer words quoted are indeed cognate with Jah Hut /tap/.

-to ask (SB: A165): Jah Hut: /smaŋ/

Old Mon: smāñ, /smaŋ/, Khmu: /maap/, Khamet: /maap/,

Riang-Lang: -maŋ, Lawa: /hmaiñ/.

-termite² (SB: A110): Jah Hut: /gruŋ/~/druŋ/

Theng: druñ, Riang (White Striped): pruiñ\, (Black): priñ\,

Khasi: krúin, Central Nicobar: daòin, Bahnar: groñ (*a small worm, causes tooth cavities*).

The morphology of Jah Hut would also provide systematic evidence for including the language in the Mon-Khmer family (see 5.1.2.4. and 6.2.1.3.).

Among the Aslian languages of Malaysia, Jah Hut belongs to the Central branch, also called the Senoic branch. This can be shown by way of elimination: Jah Hut does not have the aspirated stops which are typical of South Aslian: Semelai /t^hi/, Jah Hut /tiŋ/ *hand*; nor does it exhibit the change of Proto-Aslian *a: to /e/ and /i/ which is a North Aslian innovation: Cheq Wong /kle?/, Jah Hut /kla?/, Proto-Aslian *kla:? *tiger*.

It is difficult to find positive evidence for assigning Jah Hut to Senoic, besides that of Lexicostatistics³. The reason may be that Jah Hut has separated from Proto-Senoic very early, soon after the three branches of Aslian themselves separated, so that few innovations could take place in all Senoic and only there.

A comparison of Proto-Senoic (hereafter PSc) vowels illustrates the position of Jah Hut: leaving aside PSc *u:, there are three PSc long vocalic nuclei in the back region: PSc *ua, *uə, and *ɔ:. The three North-Senoic languages Temiar, Lanoh and Semnam all share the innovation of having merged PSc *ua and *ɔ: in favor of /ɔ:/, with Semnam later losing vowel length contrasts. On the other hand, Proto-Semai has merged PSc *ua and *uə in favor of /o:/, with a subsequent change of all /o:/'s to /ə:/'s in North-East Semai. Jah Hut did not undergo any merger of these proto-vowels and is thus the only Senoic language to have kept this older three-way distinction: PSc *ua became Jah Hut /wɛ/ before palatals, /wa/ before alveolars, and /wɔ/ elsewhere, whereas PSc *uə became Jah Hut /wo/ or /wə/ depending on the dialect, and PSc *ɔ: became Jah Hut /ɔ/.

meaning:	<i>finger nail</i>	<i>dog</i>	<i>Ipoh poison</i>
proto Senoic:	*c(n)ruas	*cuə?	*dɔ:k
Semnam	cnyɔs	cwo?	dɔk
Temiar (SW)	cɛnrɔ:s	cwɔ?	dɔ:k
Temiar (NE)	cɛnrɔ:s	cwə?	dɔ:g
Semai (NW)	cŋrɔ:s	co:?	dɔ:k
Semai (NE)	cnrɛ:s	cə:?	dɔ:k
Jah Hut (1)	crwɛs	cwo?	dɔk
Jah Hut (2)	crwɛs	cwə?	dɔk

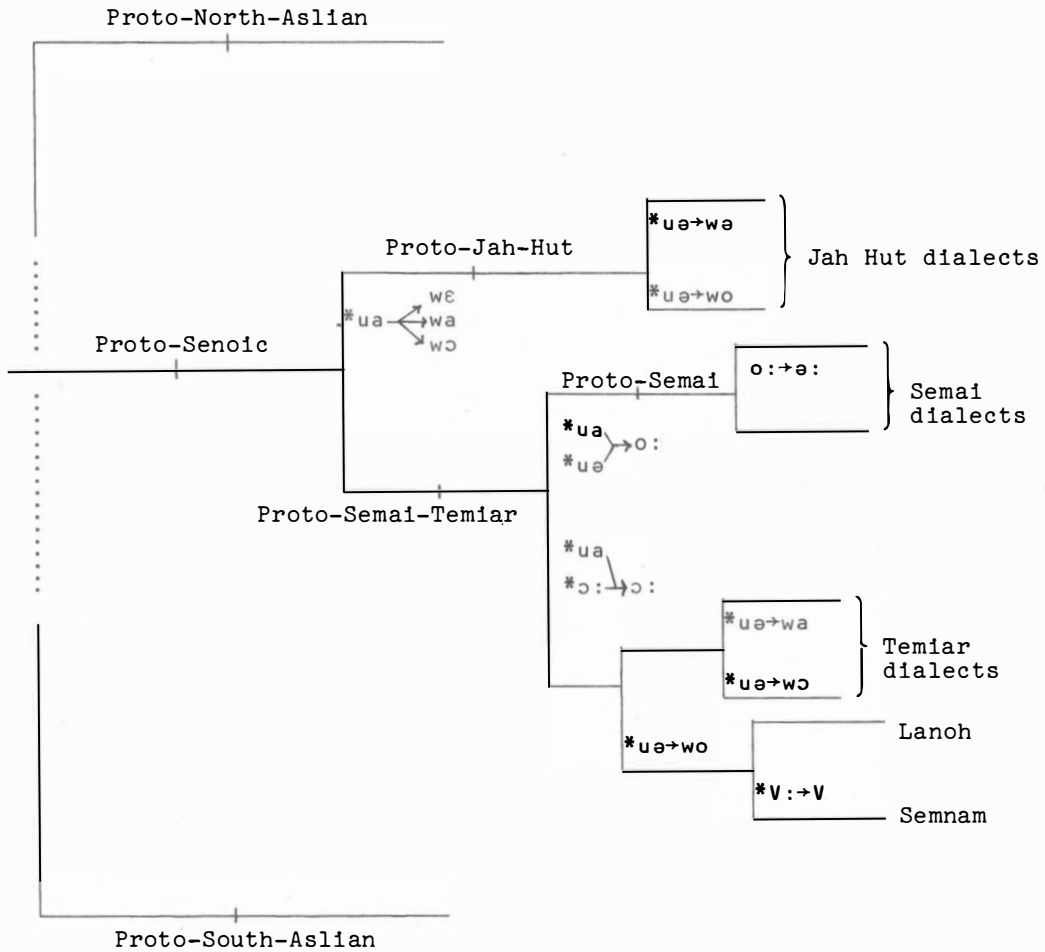
As for North and South Aslian, they have kept distinct reflexes of these proto-vowels, but with different realisations, as shown below.

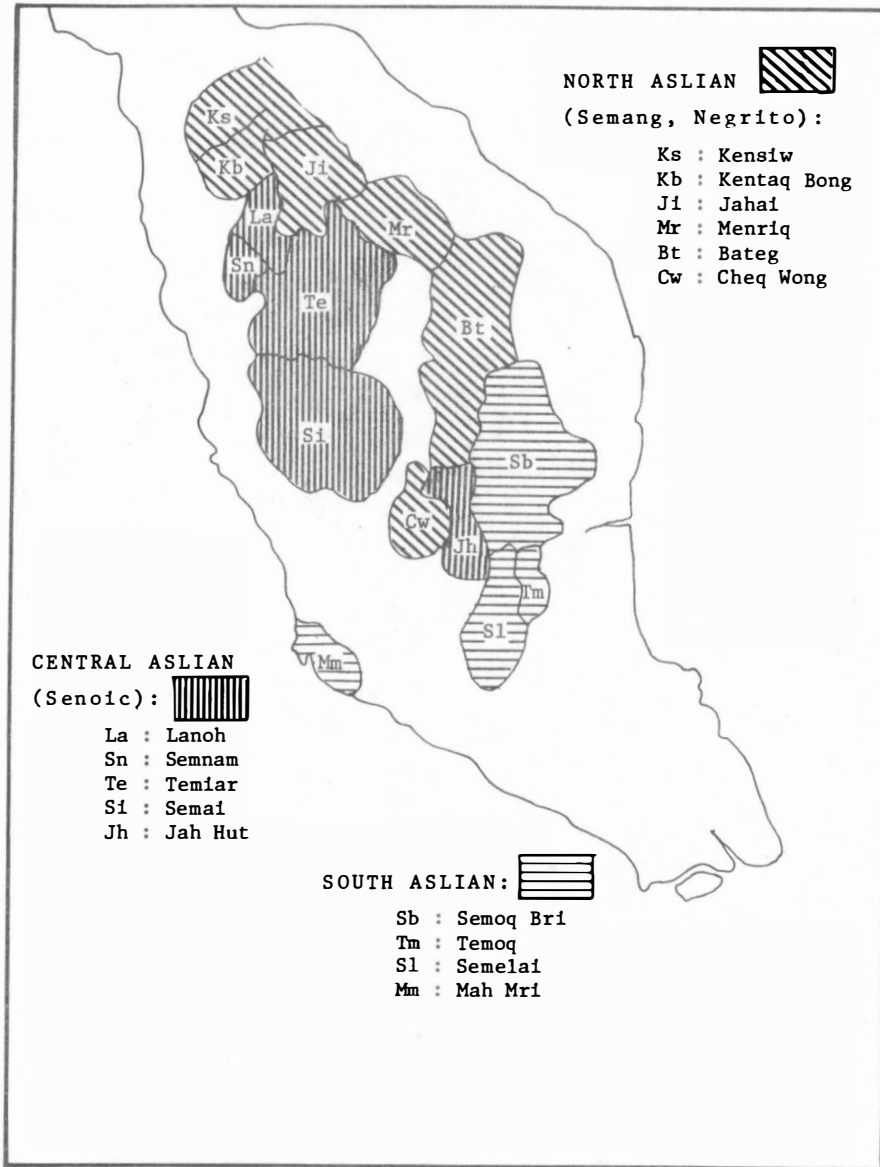
Proto Senoic	*ua	*uə	*ɔ:
North Aslian	a	ɛ	ɔ
South Aslian	o	ɔ	u

It can be seen that the modern Jah Hut reflexes are easier to relate to the modern Semai and Temiar vowel qualities than to those of either North or South Aslian.

The position of Jah Hut can be represented with the following tree diagram:

Historical position of Jah Hut in the Aslian branch of Mon-Khmer:





3. SOCIAL SITUATION

Geographically, the Jah Hut occupy a peculiar position, being cut off from their nearest relatives, the Semai, by a distance of over fifty miles, in contact with distant linguistic cousins on all sides, the Cheq Wong, the Bateg, the Semoq Beri, the Semelai, and flanked by Temuans who speak an Austronesian language, on the South-West. Speakers of Malay, also an Austronesian language, form part of the wider environment.

Such a remarkable position implies interaction between these groups, and one is not surprised to find a number of North Aslian (i.e. Semang, or Negrito) and South Aslian features in Jah Hut, both in phonology and lexicon. The very ethnonym, Jah Hut, is revealing in this respect: in /jah hət/, /jah/ means *people* and /hət/ means *no*. The neighbouring Cheq Wong, who speak a North Aslian language, distinguish all ethnic groups they know by the word for *no*, or *don't want*, or *there isn't* used in their respective languages: In Cheq Wong, the Jah Hut are called /bri? hət/ (*people no*) *the no* (hət) *people*, the Temuans are /bri? nɛp/, some Bateg are /bri? nɔŋ/, different Semai groups are /bri? pɛʔ/ and /bri? tɔʔ/, all *no people*. The Jah Hut seem to have adopted the idea, but only for designating themselves.

Jah Hut speakers are not numerous (1300 to 1700 according to Dentan, 1964), and they live in very small settlements, traditionally composed of half a dozen or less houses near a swidden field. These settlements move every few years, but the general location of the Jah Hut remains the same: a hill tract area, ten to twenty miles deep along the right bank of the Pahang river, between Jerantut and Temerloh.

Dialect variation in Jah Hut is not very great; having rapidly sampled most of the Jah Hut speaking area, except for the more remote parts of the Krau valley, I have found only one systematic dialect difference: in the village of Paya Mendoi (lower Krau valley) the reflex of Proto-Senoic *uə is /wo/, while it is /wə/ elsewhere. There are other differences associated with this: in the /wo/ dialect, the first person pronoun is /ʔihãʔ/, but /ʔihãh/ in the /wə/ dialects. There is also a great deal of variation in the whole Jah Hut area in the pronunciation of initial *cl-, *jl- and *sl- clusters: while they are intact in Paya Mendoi, they change to /tɪ-/ , /dɪ-/ or /gɪ-/ , and /hɪ-/ respectively, in other dialects. Ex. Paya Mendoi: /jlep/ *red*, elsewhere: /dlep/ or /glep/. There are even individual speech mannerisms: three men in Kyol village (and only three) are famous for saying /ʔihãh/ instead of /ʔihãh/ for *I* and /ʔimãh/ instead of /ʔimãh/ for *you* Sg. In microsocieties like those of the Jah Hut where individualism is appreciated, language change can be a matter of deliberate

personal creation. Such societies, and such values, may have been common-place in the Austroasiatic past.

4. SYNTAX

What we know of Jah Hut syntax confirms the Senoic character of the language; most statements made in this section, including those about ergativity, have precise equivalents in Semai.

4.1. INDEXICAL WORDS

Jah Hut has a large number of words which are not used in statements, but indexically (Jakobson, 1966). In a statement, the speaker asserts the truth of what he says; indexical meanings, on the other hand, have little to do with truth, but accompany the state of mind of the speaker and indicate it. There are at least three categories of indexical words in Jah Hut: Verb clitics, Exclamations, and Expressives.

4.1.1. Verb Clitics

A verb clitic can be added immediately after a verb, before any complement or any other word. It indicates the attitude or the role of the speaker in the given speech situation; for instance, /meh/ accompanies a gentle protest:

- (1) /ʔihãh ʔakən meh/ *I, not-want, Clitic:*
'I don't want (to do it)...What do you think!'

or a mild command:

- (2) /pɔcyɛk meh ʔiwã? he? dɔh/ *put-to-sleep, Clitic, child, our, this:*
'Why don't you put our child to sleep, dear!'

This clitic is very similar in meaning to the Malay particle *lah*; in fact, *lah* itself is often added to /meh/ in Jah Hut:

- (3) /ʔihãh ʔakən meh lah/ *same meaning as (1).*

The clitic /meh/ can also be used at the end of a Sentence (see Predicate clitics 4.2.1.3.); and it may be placed just after an Auxiliary and after the Negative /hət/. In this last case it signals the attitude of the speaker with regard to the Negation (for example that it is obvious, not worth arguing over). Thus, Auxiliaries and the Negative function in some ways like verbs.

Another verb clitic is /bəh/, the question marker, also placed immediately after the verb:

- (4) /yɔn hət yɔn ca? bəh ka? dɔh/: *you, not, you, eat, Clitic, fish, this:*
 'Won't you eat this fish?'

4.1.2. Exclamations

Exclamations are sufficient in themselves to identify the emotion which they accompany, so, they may be used alone. They may also be followed by complete sentences. For example, /ʔes/ in the following dialogue between speaker A and speaker B:

- (5) A: /ʔimāh cip we? tuy/: *you, go, to, there:*
 'Go over there!'
 B: /ʔes, hāh ʔakən meh lah sbap hāh bhec/: *exclamation, I, not-want, clitic, clitic, because, I, afraid:*
 'Forget it! I don't want (to go), I am scared.'

Another exclamation is /tɛʔ/, *enough said, let's go!* Exclamations and Verb clitics are invariable.

4.1.3. Expressives

Expressives, on the other hand, have a derivational morphology of sorts. They are also extremely numerous. An Expressive signals the presence of certain sensations in the speaker. These sensations may be due to activities the subject is bodily undergoing (e.g.: loss of balance) or simply observing (e.g.: visual pattern). Expressives are generally iconic: that is, there are elements in common between the sensation signalled and the sensation produced by uttering and hearing the Expressive.

/lpuŋ/: 'sound of heavy fruit falling on the ground'. Ex: /dyɛw/: 'loss of balance (due to oscillations of a bridge or flexing the knees)'

The Expressive may be repeated, indicating that the sensation is repeated in time: /dyɛw dyɛw/, a morphological (or syntactic) process which is also iconic.

It may take certain prefixes indicating plurality:

/dyɔp/: 'visual impressions of a wave rolling'

/slaʔdyɔp/: 'impressions due to several waves at the same time'

This /slaʔ-/ or /hlaʔ-/ prefix contains an /-l-/ infix which cannot always be isolated but indicates intensity or great numbers: /saʔbyɛr/: 'visual impression of dishevelled hair', /slaʔbyɛr/: 'id., with hair longer and more plentiful'.

There is a good deal of individual variation in the meanings and forms of Expressives, e.g. some prefer /dɛw/ to /dyɛw/ *loss of balance*.

Expressives are used either alone, or, like Exclamations, at the beginning of a sentence:

- (6) /hla?yaŋ, mna? ntaŋ ?əh nin/: Expressive, *big, ear, his, there*:
'(large ears!) his ears are large',

or directly preceding a Noun phrase, like a Stative verb:

- (7) /gihuŋ mat ?əh nin/: Expressive, *eye, his, there*:
'(caved in!) are his eyes',
- (8) /sla?dew sla?dew n?cip jah nin dɛ?/: plural-losing-balance,
plural-losing-balance, *the-act-of-walking, people, those, now*:
'those people walk like drunks'.

4.2. STATEMENTS

Statements are assertions of truth. There are three types of statements: Equational sentences, Stative sentences and Process sentences.

Equational sentences consist of two Noun Phrases where the second normally constitutes the predicate: /NP₁ - NP₂/: 'NP₁ is a NP₂'. It is possible to have the predicate first but it must then be followed by a predicate clitic like /meh/:

- (9) /tɛl cnu?ɔŋ jah meh dɔh/: *trace, act-of-making-fire, people, clitic, this*:
'these are traces of fire making (of some people)'.

Jah Hut does not have an overt Copula corresponding to English *is a*, but it has an overt Negative Copula: /?iwɔŋ/: *not to be* distinct from the ordinary Negative /hət/.

- (10) /?ihāh ?iwɔŋ jah cina?/: *I, NOT-BE, person, China*:
'I am not a Chinese'.

Note that the predicate, here: /jah cina?/, can optionally be preceded by the Verb phrase particle /na?/:

- (11) /?ihāh ?iwɔŋ na? jah cina?/: (same meaning). This particle is homophonous with the Contemplated Aspect particle /na?/ which is borrowed from Malay (written *hendak* or 'nak), and with several Jah Hut prepositions (see Ergative Construction).

Stative sentences normally have the order: Predicate-Topic:

- (12) /mna? kɔy māh/: *big, head, your*:
'your head is big'.

The reverse order: Topic-Predicate is also possible when the Topic is newly introduced in the conversation.

Process sentences have the order: Topic-Predicate:

- (13) /yəh kdi? kay syē?/: *he, stay, in, house:*
'he stayed at home'

The other order, with Predicate, or at least Verb, in first position is also possible (see 4.2.4.).

4.2.1. Predicate

The Predicate consists of a Verb Group followed by Complements, and optionally followed by a Predicate Clitic.

4.2.1.1. The Verb Group

The Verb Group consists maximally of one or more Auxiliaries, a Personal Prefix, a Verb, and a Verb Clitic, in that order.

Among Auxiliaries we find: /mɛ?/ or /na?/ 'Contemplated Aspect', /dah/ 'Perfective' (cf. Malay sudah).

- (14) /?ihāh na? cip cwɔm kyɛy/: *I, Contemp.-Asp., go, dig, tuber:*
'I am going (somewhere) to dig up tubers'.

Personal Prefixes agree in Person, Number, and Respectability with the Agent or the Experiencer, depending on the type of verb; note for instance /yəh/ '3rd Pers.' and /hāh/ '1st Pers.' in the following sentences:

- (15) /cwə? yəh m?mɔs/: *dog, 3rd-Pers., growl:*
'the dog growls'
- (16) /?iwā? nin hət yəh srə?/: *child, this, Negative, 3rd-Pers., know:*
'this child does not know'
- (17) /?ihāh hət hāh srə?/: *I, Negative, 1st-Pers., know:*
'I don't know'

As in (16) and (17), the Personal Prefix is very commonly used after the Negative /hət/. This is true even if the Experiencer is a Pronoun, as in (17).

These Personal Prefixes are phonologically reduced forms of the Personal Pronouns with which they may agree.

	Personal Pronouns	Personal Prefixes
<i>I</i>	?ihāh (ʌ?ihā?)	hāh (ʌhā?)
<i>We (Exclusive)</i>	?ibɔ?	bɔ?
<i>We (Inclusive)</i>	?ihɛ?	hɛ?
<i>You (Sg. Casual)</i>	?imāh	māh

	Personal Pronouns	Personal Prefixes
<i>You</i> (Sg. Respectful)	?ihɨ?	hi?
<i>You</i> (Plural)	yɔn	yɔn
<i>He</i>	yəh	yəh
<i>They</i>	?igən	gən

It is worth noting that the Personal Prefixes are identical to the Possessive Adjectives: /häh/, *my*, /mäh/, *your*, etc..., with the exception of the third person singular: /?əh/, *his* (for other uses of /?əh/ see: Predicate Clitics; see also 5.2.1.2. regarding the Prefix ?i-).

Not all Verbs can have Personal Prefixes: for some of them, which we can call Process Verbs, the Personal Prefix is used, but is optional; for others, which we can call Stative Verbs, the Personal Prefix cannot be used at all. Process Verbs and Stative Verbs constitute the two major subclasses of Verbs.

Examples of Stative Verbs:

- (18) /plə? nin kdək/: *fruit, that, BITTER:*
'that fruit is bitter'
- (19) /?ihäh dah k?ot/: *I, Perfective, EXHAUSTED:*
'I am exhausted'

4.2.1.2. Complements

Complements other than the Direct Object are introduced by Prepositions, of which there is only a small number:

-/ge?/, also /gwe?/ or /we?/ *to, with movement*

- (20) /b?jɔk ge? syĕ? pay/: *move, TO, house, near:*
'move to a new house'

- 21) /yəh cruh ge? tɛ?/: *he, fall, TO, earth:*
'he fell to the ground'

-/kay/ *towards, with movement*

- (22) /?imäh ?ageh pɲca? kay jah ?asiŋ/: *you, give, food, TO(wards), people, other:*

'you gave food to other people'

- (23) /?imäh sməp kay yɔn/: *I, ask, TO(wards), you:*
'I asked you'

- (24) /yəh dlɛh kay ?ihäh/: *he look, TO(wards), me:*
'he looked in my direction'

-/na?/ *at, with movement*

- (25) /yəh ŋʔŋək na? dɔh/: *he, sit, AT, this:*
'he is sitting here'

-/lam/ *inside, with or without movement*

- (26) /yəh crɔh lam tow/: *he, fall, IN, water:*
'he fell into water' (cf. Mal. dalam *inside*)

-/han/ *with (Instrumental or Associative)*

- (27) /ʔihāh na? cɔp rap tuy han bulus/: *I, will, stab, boar, that,*
WITH, spear:
'I'll stab that wild boar with a spear'

It is not always possible to translate a Jah Hut sentence with a parallel one in English because the Verb and its accompanying Preposition do not always divide up the total meaning in the same way in both languages:

- (28) /ɲcɛm ge? tuy/: *near, TO, there:*
'It is near there'
- (29) /bilit kay nhɔʔ/: *wrap, TOWARDS, tree:*
'wrap (it) around the tree'
- (30) /yəh krpuh han krpa?/: *it, open, WITH, wings:*
'It opened its wings'

Some Prepositions can also be used without a main Verb, and function as the Verb itself:

- (31) /ge? pat ʔimāh nin dɔh/: *TO, where, you, here, just-now:*
'Where are you going?'
- (32) /ʔihi? dɔh na? we? pat ʔɔh/: *you, just-now, will, TO, where,*
Clitic:
'Where do you want to go?'

Only one Complement does not require any Preposition: the Direct Object. While in many cases the Direct Object construction serves to convey the meaning 'target' of the action:

- (33) /səh ba?/: *pound, RICE:*
'to pound rice'

or the meaning 'prey' of the action:

- (34) /wəh tow/: *drink, WATER:*
'to drink water'

it also has many other meanings:

- (35) /cwə? ju! jah/: *dog, bark, PEOPLE:*
'the dog barks at people'
- (36) /yəh bhec ?ihāh/: *he, afraid, I:*
'he fears me'
- (37) /?ihāh na? kɔy/: *I, sick, HEAD:*
'I have a headache'
- (38) /?imāh c?cyek syē? hāh/: *you, sleep, HOUSE, MY:*
'you are coming to sleep at my house'
- (39) /bey-kɔ? mpāc nina? hāh/: *don't, step, MAT, MY:*
'don't step on my sleeping mat'.

4.2.1.3. Predicate Clitic

There is a variety of Predicate Clitics which are not absolutely required by rule, but are nevertheless extremely common. Jah Hut speakers feel that without a Predicate Clitic a sentence is not 'wrong' but is not 'full' either. We know practically nothing about the semantics of these clitics.

-/ŋɛc/ *only (?)*

- (40) /?ihāh mrpah ŋɛc/: *I, restless, ONLY:*
'I am just restless'

-/nan/ or /nan/ *for a while*

- (41) /hɛ? bra?do? na? dɔh nan/: *we, stop, at, here, FOR-A-WHILE:*
'we'll stop here for a while'.

4.2.2. Noun Phrase

In the Noun Phrase, be it Subject or Complement, the head-Noun can be preceded only by a Quantifier; it may be followed by Adjectives, Possessives and Demonstratives, in that order.

4.2.2.1. Quantifiers

The Quantifier is either a Quantity Modifier, e.g. /ba?lɔ?/ *many*, /mrəm/ *how many?*, or a Numeral, e.g. /ni?wey ~ niwey/ *one*.

Only Count-Nouns can occur with a Quantifier; these are either Nouns having an inherent durational meaning, e.g. /nahun/ *year*, /kntɔ?/ *day*, /dɔy/ *night* (as a time span) or Nouns used as standards of measurement: /del/ *three dimensional object*. Mass Nouns cannot occur with Quantifiers. In case they need to be counted this must be done mentioning the standard of measurement being used: ex. /nar del syē?/ *two, three-dimensional-objects, house, (=two houses)*. This construction

is very similar to the Malay use of so-called 'numeral classifiers' and has the same word order: Number-Classifier-Noun.

4.2.2.2. Possessives, Adjectives and Relatives

Possessives and Adjectives are directly added after the head Noun without any grammatical marker. This is also true of Relative Clauses.

- (42) /ʔidəh pləʔ kəm kəʔ caʔ/: [*this*]_{NP} [(*fruit*)_N (*CAN, WE, EAT*)_{Rel}]_{NP}:
'*this is a fruit which we can eat*'.

In this example, the Relative Clause /kəm kəʔ caʔ/ is directly added to the head Noun /pləʔ/. Sometimes the Relative Particle /yaŋ/, a borrowing from Malay, is inserted between the head Noun and the Relative Clause:

- (43) /ʔidəh syɛʔ hãh yaŋ kʔnar/: [*this*]_{NP} [(*house, my*)_N (*Rel. Part, second*)_{Rel}]_{NP}:
'*this is my second house*'

4.2.2.3. Deictics

Definiteness in Noun Phrases is indicated by Possessives: /syɛʔ hãh/ *my house*, /syɛʔ ʔəh/ *his house*, or by Deictics, of which there are five:

- /dəh/ *this here*, (near speaker)
/nin/ *that there*, (near addressee, or not too far from speaker)
/tuy/ *that out there*, (far from both speaker and addressee)
/teh/ *that up*, (higher than speaker)
/reh/ *that down*, (lower than speaker)

The combination of a Possessive and a Deictic is possible, with the Possessive first:

- (44) /ʔiwãʔ hɛʔ dəh/: *child, our, this*:
'*our child here*'

The Possessive also precedes Relative Clauses (see ex. 43) and the Deictic always comes at the end of the Noun Phrase.

4.2.3. Temporal and other Clitics

A number of clitics indicating punctual times are added to Noun Phrases, Verb Phrases or full Sentences. Several of them are translated into Malay as meaning *tadi*, *just now*: /dyeh/, /baʔ/, /dəh ~ deʔ/, others appear to be generally deictic: /rəʔ/ *this*. When used with a Noun Phrase, they exclude the Deictics mentioned above.

- (45) /bila? j?jət ?inin, yəh dyeh pr?du?/: *when, startled, clitic, he, JUST-NOW, run:*
'When thus startled he runs away'
- (46) /tapi? hət yəh dləh kay ?ihāh rə?/: *but, not, he, look, at, me, Clitic:*
'but he did not look at me'
- (47) /bila? dapat na? yəh ?inin ba?, tr?wə?/: *when, get, by, he, this, Clitic, feverish:*
'When he gets it, he becomes feverish'⁴

We may also mention here the Clitic /?əh/ which occurs in phrase final position when the sentence contains a question word like /pat/ *Where ?*:

- (48) /cwəm kyey pat ?əh/: *dig, tubers, where, Clitic:*
'Where shall (we) dig up tubers?'
- (49) /?ihi? də? na? wə? pat ?əh/: *you, just-now, will, towards, where, Clitic:*
'Where are you going?'

This /?əh/ Clitic being the bound form of the Third Person Pronoun, it may be possible to interpret (49) as meaning 'You will go somewhere, Where is IT?', with 'it' (/?əh/) referring to the omitted indefinite 'somewhere'.

4.2.4. Permutations

We have outlined, so far, only those constructions which we considered basic. From this order of constituents, it is possible, by permutation, to obtain other constructions.

4.2.4.1. Subject permutation

In both types of Declarative Sentences (NP-NP and NP-VP) it is possible to place the Predicate in front. There seems to be no obvious change of meaning in this permutation.

- (50) /?inin meh bəs kra?bə?/: *that, Clitic, SPIRIT, FOOD MIXING*
'the food-mixing taboo spirit is like that' (with the Subject, /bəs kra?bə?/ at the end of the Sentence).
- (51) /gmac bəh plə? nin/: *good, Interrog., FRUIT, THIS:*
'Is that fruit good to eat?'

4.2.4.2. *Ergative construction*

With a large number of Process Verbs, the Agent, if placed after the Verb, must have the Preposition /na?/, which also has the meanings 'from' and 'at', as well as other functions (see ex. 11, 14).

- (52) /brce? meh na? ?imāh ra?wā? dɔh/: *delouse*, Clitic, AGENTIVE, you, infant, this:
'delouse this child, won't you'

The Preposition /na?/ can be called Agentive because it is restricted to Animate or personified Nouns, and is used only with certain Sentences whose Subject is, semantically, an Agent. In such Sentences, there is a choice to place the Agent in post-verbal position and mark it with /na?/, forming what will be called an Agentive Subject, or to place it before the Verb, without any marking, like other kinds of Subjects.

Transitive Sentences can have Agentive Subjects:

- (53) /cu?oŋ na? nah ?əh ka? nin dɛ?/: *cook*, Agentive, who, Clitic, fish, that, just-now:
'Who just cooked that fish?'

This remains true even in semantically transitive Sentences whose Object is not overtly mentioned:

- (54) /k?ku? na? ?ihāh/: *vomit*, Agentive, I:
'I was vomiting (something)'

Certain Intransitive Sentences also may have Agentive Subjects, if they mention, or imply, a direction toward of from something other than the Subject himself:

- (55) /jwɔŋ na? ?ihāh/: *stand-up*, Agentive, I:
'I stood up'
- (56) /yok na? ?ihāh meh/: *return*, Agentive, I, Clitic:
'I just went back'

But Intransitive and Directionless actions like to shiver, to be sleeping, cannot have Agentive Subjects:

- (57) *(tr?wə? na? ?ihāh): *(I am shivering)
- (58) *(c?cyek na? ?ihāh): *(I am sleeping)

Note that Agentivity is not simply determined by the Verb, but by the meaning of the whole sentence: the same verb /c?cyek/ which cannot have an Agentive Subject in (58), can have one in (63) because a Direction is meant. A very similar situation is found in Semai (Diffloth, 1974).

It is difficult to find out what nuance of meaning is introduced by having the Agent placed in post-verbal position. Since it is especially common in answers to 'Who?' questions, we may assume that it is the preferred position when the Agent represents 'new information' (Chafe, 1970):

(59) Speaker A: /*ṛah səh baʔ*/: *who, pound, rice*:
'*Who pounded the rice?*'

(60) Speaker B: /*səh naʔ ʔihäh*/: *pound, Agentive, I*:
'*I pounded it*'

The reasons for calling such constructions 'Ergative' are given below (4.2.4.3.).

4.2.4.3. Complement permutation

Any Complement can be permuted to pre-verbal position. If there is a Case-marking preposition it remains with the Noun Phrase; the position of the Subject is independent of such permutations.

Locative permutation:

(61) /*naʔ dɔh məh ʔimäh krʔdiʔ*/: *AT, HERE, Clitic, you, stay*:
'*You'll stay here, won't you*'

Object permutation:

(62) /*cyək nin dah yəh caʔ*/: *BANANA, THAT, Completive, he, eat*:
'*he already ate that banana*'

Direction permutation:

(63) /*syëʔ häh dɔh cʔcyək məh naʔ ʔimäh*/: *HOUSE, MY, HERE, sleep, Clitic, Agentive, you*:
'*Come to my house to sleep, won't you*'

This last example shows that it is possible to have both Complement permutation and an Agentive Subject in the same Sentence. The two are independent of each other: to sentence (62) where the Object is permuted and the Agent, being pre-verbal, has no Agentive marker /*naʔ*/, corresponds sentence (64):

(64) /*cyək nin dah caʔ naʔ ʔihäh*/: *banana, this, Completive, eat, Agentive, I*:
'*I already ate the banana*',
with a permuted Object, and an Agentive Subject.

To sum up, in Intransitive and Complementless sentences like (51), it is possible to have a Subject without Case marker, regardless of its position in the sentence. The Object of transitive sentences

behaves in exactly the same way: no Case marker, regardless of position.

In transitive sentences, or sentences with Complements, the Agent receives a special Case marker /na?/, which is lost only if the Agent is in pre-verbal position.

This amounts to saying that sentences (52-56), (60), (63) and (64) are instances of Ergative constructions of a new kind (see Golab, 1969, and Silverstein, 1973, on the typology of ergativity).

It might be objected that sentences like (64) represent a Passive construction in Jah Hut. Such an analysis can be rejected on three accounts: first, as we showed above, the placement of the Subject and the Object in relation to the Verb are independently variable, second, when the Agent and the Object exchange positions, the Verb does not undergo any morphological change which would indicate a change of Voice, third and more crucial, when the Verb has a Personal Prefix, it always agrees with the Agent, whatever its position in the sentence is, and regardless of the position of the Object:

- (65) /cyək nin dah hãh ca? na? ?ihãh/: *banana, that, Completive, 1st-Person-Prefix, eat, Agentive, I:*
'I already ate the banana'

If (65) was a Passive, the permuted Object '*banana*' would become a superficial Subject, and the Verb would agree with it; we would have the Personal Prefix /yøh/ (3rd-Pers.) instead of /hãh/.

Jah Hut Ergativity is not simply a formal syntactic device, devoid of meaning: being restricted to actions with a Direction or a Complement located outside the Agent, it overtly manifests an important semantic category of the language.

A similar system is also found, and easier to see in Semai, but Temiar lacks any sort of Agentive Subject; whatever its history in Senoic may be, Jah Hut and Semai are the only Mon-Khmer languages in which Ergativity has been found so far.

4.2.5. Dependent Clauses

Dependent clauses, introduced by Conjunctions, are usually placed before the main clause. The last word of the dependent clause is often a Deictic Pronoun which appears to have no semantic relation with either the main Verb or the Verb of the dependent clause; its sole function seems to consist in marking the end of the dependent clause:

- (66) /bila? j?jət ?inin yøh dyəh pr?du?/: *when, startled, pronoun, he, just-now, run:*
'When he is startled he runs away'

Thus all Deictic words have boundary marking functions: Deictics mark the end of Noun Phrases (see ex. 44) and Deictic Pronouns mark the end of Dependent Clauses.

There are two Jah Hut constructions to express the notion 'to do S_1 in order that S_2 ': the Purposive and the Prospective.

4.2.5.1. Purposive

The Purposive construction indicates two consecutive actions carried out by the same Agent, the first action being a prerequisite for the second: 'Agent does VP_1 so that HE can do VP_2 '. The Agent, being the same in both clauses, is deleted in the second, dependent clause:

(67) /ʔihāh naʔ cip cwəm kyey/: I, will, go, DIG, TUBER:
'I want to go (in order to) dig up some tubers'.

(68) /yəh səh prah naʔ bwat rəm/: he, pound, perah, WILL, MAKE, FERMENTED-PASTE:
'he pounded perah nuts⁵ in order to make fermented perah paste'

4.2.5.2. Prospective

The prospective construction indicates an action carried out by someone on an object (or a person) so that this object (or person) reaches a certain state: 'Agent does V to Object so that Object is Stative'.

The dependent clause is introduced by the Prospective particle /ʔεʔ/, and the Object of the main verb is not repeated in the dependent clause:

(69) /tin nəm ʔεʔ cmaʔ/: sharpen, knife, Prospective, SHARP:
'sharpen the knife (so it becomes) sharp'

Sometimes the Object is not mentioned at all and the construction Verb-ʔεʔ-Adj. becomes a compound verbal unit expressing both the Action and its Result:

(70) /bəʔ ʔεʔ lajuʔ/: throw, Prospective, far:
'throw (it) far'

5. MORPHOLOGY

The Senoic branch of Mon-Khmer is unique in the family in having a relatively rich and productive morphology, and Jah Hut seems to have preserved some archaic patterns lost in the rest of Senoic. These two facts combine to make Jah Hut one of the most interesting Mon-Khmer language to study, if morphology is of any help in linguistic pre-

/mna?/ to be big	/pmna?/ to enlarge sth.
/bhec/ to be afraid	/prbhec/ to frighten
/tlas/ to escape	/pnlas/ to release

Syntactically, the main difference between Causatives and the corresponding (Active) Verbs lies in the Causatives having an additional Agent, the Causer. Because of this, the Case assignments associated with the Active Verb are modified as the Verb becomes Causative.

If the Active Verb is intransitive, its Subject becomes the Direct Object of the Causative; for instance, the Direct Object /?iwā?/ in ex. 2 would be the Subject of the Active Verb /cyək/ to sleep.

If the Active Verb is transitive, its Direct Object remains Direct Object of the Causative, while its Agent becomes an Indirect Object, marked with the particle /kay/; for instance, from:

(73) /ra?wā? dɔh dah yəh ca? cyək nin/: CHILD, THIS, Completive, 3rd-Pers., EAT, banana, that:
'this child ate that banana' one derives:

(74) /?ihāh pɔca? cyək nin kay ra?wā? dɔh/: I, Caus.-EAT, banana, that, TO, CHILD, THIS:
'I fed that banana to this child'

In all cases, the Causer is, semantically, an Agent, and, like all Agents, can be placed after the Verb with the Agentive particle to form ergative constructions:

(75) /pcyək na? pəh ?əh/: Caus.-sleep, Agentive, who, Clitic:
'Who put (him) to sleep?'

There are other possibilities: a Verb like /bhec/ to fear has a Direct Object which is semantically the origin of the fear:

(76) /kuciŋ bhec cwə?/: cat, fear, DOG:
'the cat fears the dog'

In the Causative, this 'origin-object' becomes the Subject:

(77) /cwə? prbhec kuciŋ/: DOG, Caus.-fear, cat:
'the dog frightens the cat'

thus embodying the idea that this 'origin' is semantically the same as the 'Causer' of the fear.

5.1.1.3. Superlative

Stative Verbs generally have a derived form with a Superlative meaning and the same syntactic behavior as the original Stative Verb:

/num/ ripe	/ra?num/ very ripe
/hlək/ heavy	/sra?lək/ very heavy

5.1.2. Derived Noun Forms

5.1.2.1. Action nominalisation

Nearly every Verb has a corresponding Gerund Noun form. With Process Verbs, the Gerund refers to the fact that the action took place, or to the way it did; with Stative Verbs it indicates the State.

Ex. /ca?/ to eat /n?ca?/ the act of eating
 /bhec/ to be afraid /bnahec/ the feeling of fear

The gerund may be followed by all the complements that accompany the underlying verb, the whole construction functioning like a Noun Phrase. The former Subject of the underlying Verb becomes the possessive of that Noun Phrase and is, therefore, placed at the end:

(78) /lambat n?cip ?əh/: slow, Gerund-walk, HIS:
 'his walking is slow'

5.1.2.2. Agent nominalisation

Most Verbs have a corresponding Noun with the meaning: 'the one who does V', often with the added connotation of an habitual or excessive action:

/lyep/ to plait palm leaves /mlayep/ one who plaits
 /cyek/ to sleep /m?cyek/ one who usually sleeps
 /ca?/ to eat /m?ca?/ one who eats too much

Sometimes, upon questioning, Jah Huts will say that these forms with initial /m/ also have another use similar to that of the Malay mēng-forms, which they take to mean 'to be in the process of doing something'. Whether this is actually true of the Jah Hut language or is only a commonly held stereotype, I am not sure: it is not always easy to parse Jah Hut sentences unambiguously. There are many cases however where the /m-/ forms are clearly Agent Nouns:

(79) /?ihāh ?iwɔŋ na? jah m?ca?/: I, not-to-be, Preposition,
 person, Agent-EAT:
 'I am not a big eater'

where /jah m?ca?/ is a Noun-Noun construction just like /jah cina?/ in example (11), an exact parallel to example (79).

5.1.2.3. Object nominalisation

Transitive Verbs have a corresponding Noun designating 'the thing which is V-ed'.

Ex. /ca?/ to eat /pɔca?/ food
 /phɔm/ to breathe /pɔŋ?hɔm/ breath

This affix is no longer productive at present, and it is not always

easy to disambiguate it from other Nominalised Verb forms: whereas /pɾcaʔ/ is not ambiguous, as a Noun, /pɾʔhɔm/ can also mean: *the act of breathing*.

5.1.2.4. *Instrument nominalisation*

Some verbs yield a Noun form meaning 'the object with which V is done'

Ex. /kʔrɔʔ/ *to intoxicate fish* /knrɔʔ/ *the root used for intoxicating fish*
 /tlɔs/ *to knock fruits down* /tnalɔs/ *pole for knocking fruits*

Here again, such forms are often, but not always, homonymous with other Nominalisations, especially Action Nominalisation.

The phonological forms of the various types of Nominalisations found in Jah Hut indicate a basic two-way contrast between forms with an /m/ for Agent Nominalisation, and forms with an /n/ for all others. The same contrast, expressed by the same sounds /m/ and /n/ is found in the Nicobarese languages, (Radhakrishnan, 1970), a group of Austro-Asiatic languages whose exact relationship to Mon-Khmer and Munda has not been determined yet.

5.2. NOUN MORPHOLOGY

The morphology used with basic Nouns is less productive than that used with basic Verbs, and does not form a neat system. The meaning elements expressed in Noun morphology are usually detailed and specific. Only two such affixes have a more systematic meaning: the Quantifier infix and the Verbaliser.

5.2.1. Derived Noun Forms

5.2.1.1. *Quantified nouns*

As we saw earlier (section 4.2.2.), the majority of Jah Hut Nouns are Mass Nouns, and only a few Noun sub-classes can be used directly with a Numeral or a Quantifier. But it is possible to create Count Nouns from Mass Nouns by morphological derivation:

/kɔʔ/ *day light* /kntɔʔ/ *day (unit of time)*

This is certainly an ancient morphological pattern as it is found in all branches of Aslian; but in Jah Hut, it is being displaced by a curious pattern of suppletion; the Malay equivalent of most duration Nouns is known to the Jah Hut, for instance, /hariʔ/ *day* (Mal. hari) is equivalent to /kntɔʔ/, and /malam/ *night* is equivalent to /dɔy/; in Numerals, Jah Hut uses Mon-Khmer words only for *one* (/niʔ/~niʔwey/), and *two* (/nar/) (Diffloth, to appear) while Malay borrowings, e.g. /tigaʔ/ *three* are used for higher figures; in a construction, if the

last numeral is 'one' or 'two', the Mon-Khmer Count Noun is used, otherwise, the equivalent Malay one: /ni? kntɔʔ/ *one day*, /nar kntɔʔ/ *two days*, but: /tigaʔ hariʔ/ *three days*, and similarly: /niʔ dɔy/ *one night*, /nar dɔy/ *two nights*, but /tigaʔ malam/ *three nights*. Thus there is agreement between Noun and Numeral with regard to the 'original' vs. 'borrowed' distinction. A similar case is found in Theng (Maspero, 1955) with Thai borrowings (see mak entry). As there are only two remaining original numerals, original quantified Nouns are being phased out and replaced by equivalent Malay count nouns.

5.2.1.2. Existential nouns

Several small classes of Nouns have a prefix /ʔi-/ which can still be isolated in most cases.

With kinship terms, /ʔi-/ indicates a reference term as opposed to the bare root which is a term of address:

/ʔitaʔ/	<i>someone's grand father</i>	/taʔ/	<i>grand pa!</i>
/ʔiʔɛm/	<i>someone's elder brother</i>	/ʔɛm/	<i>elder brother!</i>
/ʔidɛh/	<i>someone's parent's yo. sis.</i>	/dɛh/	<i>parent's yo. sis.!</i>
/ʔiwãʔ/	<i>someone's offspring (parents usually address their own children by name, but see /raʔwãʔ/ child, infant below).</i>		

The prefix /ʔi-/ is also used to form pronouns from Possessives and personal prefixes, e.g. /ʔihãh/ *I*, /ʔibɔʔ/ *We Excl.* (see section 4.2.1.1.), and from deictics, e.g. /ʔidɔh/ *this one here* (see section 4.2.2.2.).

In all cases, the /ʔi-/ prefix asserts the existence of the entity to which it is added. This prefix is historically related to the Semai Definite Article /ʔi/ which is also a third person Possessive, and to the Temiar particle /ʔi/, a Subject marker (Benjamin, 1973b).

5.2.1.3. Expressive nouns

A number of animal names contain a recurrent /-lʔ-/ sequence which cannot be isolated as a morpheme in present-day Jah Hut:

/klʔbak/	<i>butterfly</i>
/hlʔdɛʔ/	<i>cockroach</i>
/klʔjɛh/	<i>a small bird sp.</i>

This appears to be a remnant of a Proto Senoic infix /-l-/ which probably meant 'step by step' and was used to derive Expressives from Stative Verbs. We saw (section 4.1.3.) that an /-l-/ infix can still be isolated in some Jah Hut Expressives, and the very same /-lʔ-/ sequence can be found, but not isolated, in many other Jah Hut Expressives: /klʔpɔr/ *big mouth!*, /blʔhir/ *blue-green* etc. It is probably not a coincidence that the animal names where /-lʔ-/ is found designate

animals with rapid jerky movements. These names thus appear to be former Expressives used to describe the 'step by step' movements of these animals. This is clearer in Semai where every animal species has a large number of nicknames referring to movements, habits and appearance, many of which are drawn from the grammatical class of Expressives.

5.2.1.4. Reduplicated nouns

Many Jah Hut Nouns have the same phonological structure as reduplicated Verb forms;

/tntyɛn/ *bridge*

/sɔŋʔwiŋ/ *skies*

/cʔcek/ *house lizard* (cf. Mal. *chichak*)

/tntwɔŋ/ *a bird sp.*

Since the corresponding monosyllabic forms are not found in Jah Hut today, these forms are not analysable; but such roots are often found in related Senoic languages, e.g. Semai /swi:k/ *skies*. In any event, the Reduplication affix only applies to Verbs; but as the phonology of these Nouns follows exactly the phonological pattern of reduplication in Verbs, they will be made to undergo the same phonological rules, even though they do not contain a morpheme 'Reduplication'.

5.2.1.5. Superlative nouns

The superlative /raʔ/ affix (cf. section 5.1.1.3.) is another case of a Verb morpheme unexpectedly found in some Nouns; for instance, the word /raʔwãʔ/ *child, infant* (cf. /ʔiwãʔ/ section 5.2.1.2.). The Nouns 'woman' and 'man (male human)' also contain this affix:

/kraʔkəŋ/ *woman* from /kəŋ/ *female*

/kraʔkɔŋ/ *man* from /kɔŋ/ *male* (e.g. /ʔiwãʔ kɔŋ/ *son*)

The prefixed reduplication of the initial /k/ in these two words does not fit any regular morphological pattern, but see section 6.2. The term superlative Noun was suggested by the overt similarity to Superlative Verbs, but it might also be justified semantically, e.g. a woman is female par excellence.

5.2.2. Derived Verb Forms

Verbs may also be derived, sporadically, from basic Nouns:

/cɛʔ/ *louse* /brcɛʔ/~/brɔcɛʔ/ *to delouse someone*

/sek/ *rotan* /brsek/ *to look for rotan*

/rudɔŋ/ *friend* /brudɔŋ/ *to accompany*

The similarity of this pattern to the Malay *bĕr*-Noun: 'to have Noun' construction, in both form and function is quite striking. However, the Malay prefix is more productive. In Jah Hut, the meanings of derived Verb forms are as varied as 'to take Noun from someone', 'to get Noun', 'to take someone as Noun', each Noun producing a different sort of Verb; such information has to be included in the Lexicon, as in the case of other idiomatic compounds, and yet the phonological forms are as regular as in any normal morphological pattern.

6. PHONOLOGY

6.1. ROOT STRUCTURE

Jah Hut, like other Mon-Khmer languages, has prefixes and infixes but no suffixes; the end of the word is therefore unaffected by morphophonemic alternations, and usually constitutes the root; it is also the only part of the word to receive stress. In describing Jah Hut phonology, it is therefore appropriate to start from the end of the word and move backwards to the initial; rhyming dictionaries are a must for all Mon-Khmer languages, and are more informative than initially ordered dictionaries of the traditional European type.

6.1.1. Finals

All Jah Hut words and roots end in one, and only one consonant. When Malay words ending in a vowel are borrowed, they receive a final glottal stop: /tuhaʔ/ *old* (Mal. tua), /saʔluʔ/v/slamuʔ/ *always* (Mal. selalu).

All Jah Hut consonants, except voiced stops, can be used as finals:

Jah Hut Finals:	p	t	c	k	ʔ
	m	n	ŋ	ŋ	
	w	r, l	s, y		h

Besides /-uw/, /-uw/ and /-iy/, which are excluded, there seem to be very few restrictions between the final consonant and the preceding vowel. This makes Jah Hut a very useful language for reconstructing Proto-Mon-Khmer finals as even /-is/ and /-es/ are preserved, while only /-ih/ and /-eh/ are normally found in other languages of the family.

There are only four Velar and Post-Velar finals in Jah Hut, but they represent nearly half of the vocabulary, almost as much as the eleven other finals. The glottal stop is mostly responsible for this imbalance.

6.1.2. Vocalic Nuclei

A Vocalic Nucleus consists of a simple vowel or a diphthong; there are twelve diphthongs which may become vocalic nuclei: /yɛ/ /wo/ /wə/ /wa/ /wɛ/ /wɔ/ and their nasalised counterparts. Other types of diphthongs function as sequences of Semiconsonant + Vowel; thus, /cyɛk/ *to sleep* has only one initial consonant and a /yɛ/ vocalic nucleus, whereas /cyək/ *banana* has two initial consonants and a single /ə/ vowel. The reasons for distinguishing two kinds of diphthongs are morphophonemic (see section 6.2.1.3.).

Simple vowels form a three by three system:

i	u	u
e	ə	o
ɛ	a	ɔ

For many speakers, /u/ and /ə/ are not distinguished, but others do maintain this old contrast consistently; e.g. /sʔup/ *sweat* vs. /grʔəp/ *to burp*, /kbus/ *to be dead* vs. /bəs/ *to throw away*.

All vocalic nuclei have nasalized counterparts, but they are not very frequent (around 6% in Lexicon frequency).

Ex: /cɔr/ *to burn a swidden* vs. /cɔ̃r/ *a squirrel* sp. (*Sundasciurus Lowii*), /sec/ *flesh* vs. /ʔisɛ̃c/ *a bird* sp. (*spider hunter*).

Nasal vowels can occur before every final consonant, and after every consonant except /g-/. There seems to be no simple way of predicting their occurrence, even by setting up appropriate nasal consonants in 'underlying' phonological representations.

The great frequency of nasal vowels in Expressives suggests in many cases the presence of a separate morphological element. But it also appears that Expressives do not have affixes in the traditional sense (Diffloth, 1973).

6.1.3. Initials

In considering the sounds which precede vocalic nuclei, we can distinguish three types of roots: Simple roots, with only one initial consonant, complex roots, with two initial consonants, and disyllabic roots with an unstressed syllable, the 'Minor', preceding the stressed, final, 'Major' syllable.

6.1.3.1. Simple roots

Every Jah Hut consonant can occur as the initial of a simple root; this includes all the consonants which can be final, plus a full series of Voiced stops:

b d j g

Ex: /got/ *hungry*, /gum/ *to winnow*, /jɔŋ/ *foot*, /jɔn/ *to send*, /den/ *bamboo*, /duʔ/ *to run away*, /bam/ *mouth piece of blowpipe*, /baʔ/ *rice (padi)*.

6.1.3.2. *Complex roots*

At the beginning of complex roots, most combinations of two consonants can occur, with the following restrictions:

- No clusters of two identical consonants.

There are apparent counter-examples: /kkēr/ *brush-tailed porcupine* (*Atherurus macrourus*), /ʔʔak/ *crow* (*Corvus* sp.); but they are actually Reduplicated Nouns which have lost the infix /-ʔ-/ normal in such forms. /kkēr/ has a free variant /kʔkēr/ (see also Cheq Wong /krkēr/ *brush-tailed porcupine*), as for /ʔʔak/, the normal reduplicated noun form should be /ʔʔʔak/ ([ʔəʔʔak]), which is automatically simplified to [ʔəʔak]: /ʔʔak/ by a phonetic rule.

- No clusters of homorganic stops. This rule excludes the following clusters: gk-, kg-, jc-, cj-, dt-, td-, bp-, pb- in addition to the ones already excluded by the rule above. There are two apparent exceptions in our data: /bput/ *to blow* which probably contains a b-prefix (see Semai /pu:t/ *to blow*) and /tduh/ *evening* which is a Malay word perhaps not used in everyday Jah Hut. Non-homorganic Stop clusters are not restricted: /tkak/ *palate*, /dkaŋ/ *bamboo rat*, /tgoh/ *solid*, /pkaŋ/ *a spirit*, /bku/ *grey*, /bgok/ *goitre* etc...

- No clusters of homorganic Stop plus Nasal. This rule excludes: kŋ-, gŋ-, cŋ-, jŋ-, tŋ-, dŋ-, pŋ-, mŋ-. But non-homorganic clusters are not restricted: /dŋɔy/ *straight*, /kneʔ/ *rat*, /jnes/ *arm*, /bnum/ *high mountain*, /kmat/ *gizzard*, /cmaʔ/ *sharp*, /tmoʔ/ *stone* etc...

- Clusters of Nasal plus Stop must be homorganic. Ex: /mpäc/ *to step on*, /ntaŋ/ *ear*, /pɕem/ *near*, /ndum/ *ripe* (v/num/). There is one apparent exception: /mcək/ *Yellow-throated marten* (*Martes flavigula*). The glottal stop does not function as a Stop in this rule since it has no homorganic Nasal: /mʔun/ *comfortable to sit in*, /nʔɔs/ *firewood* (historically derived from the Jah Hut word for *fire*: /ʔɔs/).

- With two Stops, combinations of Dentals and Palatals are excluded: this rule further prohibits: ct-, cd-, jt-, jd-, tc-, tj-, dc-, dj-.

This rule also extends to clusters with -s- in second position: ts-, ds-, cs-, js-, are excluded.

A further extension of this rule could also explain the instability of cl-, jl-, and sl- initial clusters; even ty- clusters are unstable and alternate freely with cy-: /cyek/ *to sleep* is sometimes pronounced /tyek/; (the t- is historical and attested in other Mon-Khmer languages: Semnam /tek/ *to sleep*, Khmer /ʔeic/, (spell.: te:k) *to lie down*).

- No clusters of Liquids. This restriction can be seen as another extension of the rule just above.

- The initial of a cluster cannot be a Semi-Vowel (w,y), nor a

Laryngeal (h,ʔ). The only exceptions are clusters of h plus Voiced continuants (hm-, hn-, hp-, hŋ-, hr-, hl-); the first four are unstable, with the initial h- freely disappearing: /hŋɛm/~/ɲɛm/ *knife*, /hməs/~/məs/ *to sniff out*; the last, hl-, varies freely with sl-: /hlaʔ/~/slaʔ/ *leaf* (Proto-Senoic and Proto-Mon-Khmer *slaʔ?), /hlay/~/slay/ *swidden, ladang* (Proto-Senoic *sla:y, Khmer /sla:y/ *fallow land*), while the hr- cluster could be analysed as a single unit (see section 7.).

There are other clusters that would appear to be impossible from our limited collection; some may simply be very rare possibilities e.g. clusters with initial Nasals, others may represent true restrictions on the language, other still, accidental gaps. Somewhere among these possibilities lies the historical explanation for the appearance of Nasal Vowels.

6.1.3.3. Disyllabic roots

A number of roots contain more than two segments before the main Vowel. These include roots with three initial consonants, the second of which is vocalic (a Nasal or a Liquid): /smpaʔ/ *durian*, /grteʔ/ *a tick*, /pɿʔeŋ/ *lukewarm*. Such roots may contain obsolete infixes, but they are no longer analysable in contemporary Jah Hut. Other roots contain a true vowel after the initial consonant. These are, for the most part, borrowings from Malay: /suraʔ/ *to sing* (Mal. *suara voice*), but some have no known Malay source, and have Semai and Temiar cognates which are also disyllabic (see: Diffloth 1973b):

/makɔʔ/ *pregnant* (Semai, Temiar: /makɔ:ʔ/)

/kabok/ *Monitor lizard* (Semai: kabuk/, Temiar: /kabug/)

/baŋkeŋ/ *a bird sp. woodpecker?* (Semai: /maŋki:k/, Temiar: /maŋke:k/)

6.2. MORPHOPHONEMICS

6.2.1. Reduplication

Reduplication is a productive process in Verbs (see section 5.1.1.1.); in Reduplicated Nouns, it is not a morphological process at all (see section 5.2.1.4.); and yet, the two follow exactly the same complicated phonological rules.

Reduplication takes several forms depending on the root type: for complex roots, it simply consists of 'Final infixation'; for simple roots it is better described in two steps: first, 'Initial copying', then 'Final infixation'.

6.2.1.1. Initial copying

Reduplicated forms of simple C_iV C_f roots always contain two occur-

rences of the initial consonant C_i :

/jɔl/ → /jʔjɔl/ to be barking
 /rɛp/ → /rŋʔrɛp/ to be gnawing
 /həy/ → /hiʔhəy/ to be waking up
 /cip/ → /cʔcip/ to be walking

the following rule is therefore needed: $C_i V C_f \rightarrow C_i C_i V C_f$

It creates non-existing intermediate forms: jjɔl, rrɛp, etc...which are similar to complex roots in having initial clusters, and are thus suited for the application of the next, 'Final infixation' rule.

6.2.1.2. Final infixation

Reduplicated forms of complex roots show that, during Reduplication, something is inserted between the two consonants of the initial cluster:

/jleɥ/ → /jʔleɥ/ to be seeing
 /kɫɛŋ/ → /kŋʔɫɛŋ/ to be speaking
 /cwɔm/ → /cŋʔwɔm/ to be digging
 /syɔc/ → /sʔyɔc/ to be whistling

In a first approximation, this 'something' is most conveniently described as a copy of the final consonant, followed by a glottal stop:

$C_i C_m V C_f \rightarrow C_i - C_f ? - C_m V C_f$

This rule produces intermediate forms like: j-hʔ-leɥ, k-ŋʔ-ɫɛŋ, c-mʔ-wɔm, s-cʔ-yɔc; it also applies to the outputs of the 'Initial copying' rule, to produce new intermediate forms:

jjɔl → j-lʔ-jɔl
 rrɛp → r-rʔ-rɛp
 hhəy → h-yʔ-həy
 ccip → c-pʔ-cip

Several adjustments are necessary to produce the actually observed forms:

- Clusters consisting of C-yʔ- are syllabified as Ciʔ-:
 h-yʔ-həy → /hiʔhəy/ to be waking up
 n-yʔ-wey → /niʔwey/ one

One would expect C-wʔ- clusters to be syllabified as Cuʔ- initials, but no clear example has been found yet.

- Clusters consisting of $C_i - N ? - C_m$ where N is 'any Nasal', are simplified to: $C_i \eta ? C_m$, unless C_m is a Stop:

c-mʔ-wɔm → /cŋʔwɔm/ to be digging ([cəŋʔwɔm])
 r-rʔ-rɛp → /rŋʔrɛp/ to be gnawing ([rəŋʔrɛp])
 k-ŋʔ-ɫɛŋ → /kŋʔɫɛŋ/ to be speaking ([kəŋʔɫɛŋ])

- Finally, all remaining clusters created by the 'Final infixation' rule lose the consonant preceding the glottal stop:

j-h?-leh → /j?leh/ to be seeing ([ji?leh])
 s-c?-yoc → /s?yoc/ to be whistling ([si?yoc])
 j-l?-jəl → /j?jəl/ to be barking ([ji?jəl])
 c-p?-cip → /c?cip/ to be walking ([ci?cip])

this rule also affects clusters where C_m is a Stop, and the infix is -N?-:

t-n?-tin → /t?tin/ to sharpen
 g-m?-gum → /g?gum/ to winnow (I heard once a /gm?gum/ variant).

There are a few forms with an extra, optional, adjustment rule, one that deletes the infix -?-:

k-r?-kēr → /k?kēr/~/kkēr/ brush tailed porcupine
 t-ŋ?-twŋ → tŋtwŋ → /tntwŋ/ a bird sp.: large racket-tailed drongo (*Dissemurus paradiseus*)
 s-ŋ?-wiŋ → /sŋ?wiŋ/~/sŋwiŋ/ skies

this rule is apparently found only among Reduplicated Nouns, but it may be spreading in some dialects; e.g. in Kuala Krau in casual speech.

Finally, disyllabic roots reduplicate by having 'Final infixation' after the first vowel:

/sura?/ → /su?ra?/ to be singing
 /mati?/ → /ma?ti?/ to be dying

but as both examples are Malay borrowings, there may be some other pattern.

6.2.1.3. Diphthongs in reduplication

The operation of reduplication rules shows that the sequences /ye, wo, wə, wɔ, wa, wɛ/ may be considered as single vocalic nuclei, whereas other sequences of semi-consonant plus vowel, e.g. /yɔ/ or /wi/, function as if they contained one consonant and one vowel. Thus, the root /cyɛk/ to sleep is reduplicated, not as a complex CCVC root, but as a simple CVC one:

/cyɛk/ → ccyɛk → c-k?-cyɛk → /c?cyɛk/ to be sleeping ([ci?cyɛk])
 (there is no */c?yɛk/ parallel to /s?yoc/ or /sŋ?wiŋ/)

The explanation for this is historical: Jah Hut /ye/ comes from a Proto-Senoic diphthong *iə, Jh. /wo/ and /wə/ come from *uə, and Jh. /wɔ/, /wa/ and /wɛ/ come from *ua; all three units *iə, *uə, *ua function in Proto-Senoic, Semai and Temiar as single vocalic nuclei, and their reflexes still do in Jah Hut today.

However, as w and y are pronounced in present-day Jah Hut as semi-consonants, Reduplication creates very long consonantal clusters:

/jwɔŋ/ → j-ŋ?-jwɔŋ

these can be reduced by the usual adjustment rules:

j-ŋ?-jwɔŋ + /jʔjwɔŋ/ *to be standing* ([jiʔjwɔŋ])
 but also, optionally, by an early deletion of the infix -ʔ- mentioned
 above:

j-ŋ?-jwɔŋ + j-ŋ-jwɔŋ + /jŋjwɔŋ/ ([jiŋjwɔŋ]) or /ŋŋjwɔŋ/ ([ŋiŋjwɔŋ]).

There are also a few cases where a /wɔ/, for instance, is treated
 as containing a consonant, in spite of its historical vocalic origins:

/cwɔm/ *to dig* (Semai /co:p/) + /cŋʔwɔm/ *to be digging*.

Reduplication itself is a very ancient process: all Senoic languages
 have it, and both North and South Aslian have similar processes. In
 non-Aslian Mon-Khmer, 'Initial copying' is a prominent feature of Khmer
 morphology (see 'Prefix /R-/', Jenner, 1969, p.63 ff.), and 'Final
 infixation' has a strikingly close parallel in Nancowry Nicobar (see
 'Root duplication', Radhakrishnan, 1970, p.149 ff.).

6.2.2. Affixation of n

Action nominalisations (section 5.1.2.1.), and Quantified Nouns
 (section 5.2.1.1.) have affixes with a variety of forms all containing
 an /n/.

Here again, Simple roots must be distinguished from others; in
 simple roots, nʔ- is prefixed:

/coy/ + /nʔcoy/ *act of gutting* ([niʔcoy])

/səh/ + /nʔsəh/ *act of pounding* ([nəʔsəh])

/cip/ + /nʔcip/ *act of walking* ([niʔcip])

but in complex roots, -nʔ- is infix after the first consonant:

/tlɔs/ + /tnʔtlɔs/ *act of knocking fruits* ([tənəʔtlɔs])

/jkət/ + /jnʔkət/ *act of tying* ([jənəʔkət])

and we also find that roots with /ye, wɔ/ etc. are treated as simple
 CVC roots:

/cyek/ + /nʔcyek/ *act of sleeping* ([niʔcyek]) (not *(cnʔyek))

/cwɔm/ + /nʔcwɔm/ *act of digging* ([niʔcwɔm]) (not *(cnʔwɔm))

Disyllabic roots simply infix an -n- after the first consonant:

/bilit/ + /bnilit/ *act of wrapping*

/cuʔɔŋ/ + /cnuʔɔŋ/ *act of cooking*

but there are initial consonants which do not allow infixation of
 nasals and simply have an n- prefix:

/rʔoh/ + /nrʔoh/ *act of sweating* ([nəroʔoh])

/mpäc/ + /nʔmpäc/ *act of stepping on*

/ʔagan/ + /nʔagan/ *goodness*

/lajuʔ/ + /nlajuʔ/ *distance*

/hawac/ + /nhawac/ *stinginess*

These examples suggest that this affix was originally a simple n-
 prefix and not an infix. The glottal stop in forms with an nʔ affix

is probably the remnant of an application of the 'Final infixation' rule. There are a few forms which can only be explained in this fashion:

/həy/ → */niʔhəy/* *act of waking up* must be derived from *n-yʔ-həy*
(cf. */hiʔhəy/* *to wake up* from *h-yʔ-həy*)

Such forms are rare and limited to -y finals; roots with final Nasals do not have a *nŋʔ* affix, as an application of the 'Final infixation' rule would produce:

/pləm/ → */pnʔləm/* *day after tomorrow* and not **(pŋŋʔləm)*.

6.2.3. Affixation of m

Not enough is known about this process to propose rules for it, but the following examples suggest that it is similar to 'Affixation of n':

/caʔ/ → */mʔcaʔ/* *eater*

/cyək/ → */mʔcyək/* *sleep*

/ʔudot/ → */mʔudot/* *smoker*

This Agent nominalisation affix must be distinguished from the m-'Progressive' affix which nasalizes initial stops according to the Malay pattern;

/cuŋaŋ/ → */mpuŋaŋ/* (*teeth*) *are protruding*

but otherwise contains a single m-:

/ʔudot/ → */mʔudot/* *be smoking* (the Malay rule would have produced **(mŋudot)*).

6.2.4. Miscellanea

There is a rich variety of affixed forms about which little is known at the moment; these include many forms with an -a- infix:

/crəŋ/ *long* → */cnarəŋ/* *length*

/sŋɛc/ *cold* → */snaŋɛc/* *the cold*

/hlək/ *heavy* → */snaʔlək/* *weight*

/rŋap/ *red* → */nraŋap/* *redness*

/bhec/ *afraid* → */bahec/* *afraid (?)*

/lyɛp/ *to weave* → */nlayɛp/* *act of weaving*, */mlayɛp/* *be weaving*

Patterns in Semai and Temiar morphology suggest that this -a- infix could be a separate morpheme, but its meaning in Jah Hut is not apparent so far.

Causative morphology also contains unknown elements: the regular affix is either p- or pr-:

/caʔ/ → */prcaʔ/* *to feed*

/cyək/ → */pcyək/* *to put to sleep*

but there are also Causatives in tr- and kr-:

/hus/ (*clothes*) *loosen* + /trhus/ *to undress*

/luy/ *to be inside* + /krluy/ *to put inside*

6.3. PHONETICS

The notation in // used in the present work is a fairly abstract one, but, together with phonetic rules, is sufficient to predict phonetic details of the words represented. Some of these phonetic rules are given below.

The long consonant clusters at the beginning of many Jah Hut words would be unpronounceable without syllabification rules which insert vocalic segments in appropriate places;

/pnʔlum/ + [pəneʔlum]

Without trying to be exhaustive, one can say that:

- in C_1C_2V - initials (V being a vocalic nucleus), a vocalic segment is inserted between C_1 and C_2 , unless C_2 is a liquid or a semi-vowel,
- in $C_1C_2C_3V$ - initials, two vocalic segments are inserted ($+C_1vC_2vC_3V$ -), unless C_2 is a glottal stop, in which case only one segment is inserted, after C_1 ($+C_1v?C_3V$ -).

The quality of these epenthetic vocalic segments is subject to the following rules:

6.3.1. Suprasegmental laryngeals

When a laryngeal (h or ʔ) immediately precedes a major vowel, the preceding epenthetic vowel takes on the quality of the major vowel:

/jʔaŋ/ *bone* + [jaʔaŋ]

/nhʕʔ/ *tree* + [nʕhʕʔ]

/sʔit/ *rotten smell* + [sʔitʔ]

the major vowel thus seems to be anticipated by the epenthetic vowel. Actually, from the point of view of articulation, this 'anticipation' is only a notational illusion: Laryngeals and Vowels are articulated independently and can be superposed in time: there is only one articulatory gesture for the vowel, not two separate and identical ones; the laryngeal intervenes at some point during the execution of the vocalic gesture; these laryngeals are in fact suprasegmental. But phonologically, they function as 'main consonants' (C_m) in our description of root structures and morphophonemics.

6.3.2. Epenthetic high vowels

A similar sort of 'anticipation' occurs in some $C_1v?C_3V$ - initials: if C_3 is a palatal, v takes on the quality [i], if C_3 is a labial,

v takes on the [u] quality:

- /nʔcaʔ/ → [niʔcaʔ] *act of eating*
 /pʔɲar/ → [piʔɲar] *to be noisy*
 /sntʔyɔ̃l/ → [səntiʔyɔ̃l] *unhealthy (way of walking)*
 /pʔbar/ → [puʔbar] *to be two*
 /wʔwec/ → [wuʔwec] *to be climbing*
 /trʔwəʔ/ → [truʔwəʔ] *to be feverish*

here again, 'anticipation' is only illusory; the [i] and [u] segments are part of the articulatory transition from a vocalic segment to a palatal or labial consonant, the suprasegmental glottal stop intervening late in the middle of the transition, without affecting it.

In C_1C_2V - initials where C_2 is a palatal or a labial, the epenthetic vowel is not quite so high (nor front or back) as in $C_1vʔC_3V$ - initials:

- /tbuɪ/ → [təbuɪ] *a bee*
 /pcəh/ → [pəcəh] *to leave food*

the reason seems to be that in C_1C_2V - initials, the epenthetic vowel is not syllabic, and often hardly audible, whereas in $C_1vʔC_3V$ initials the epenthetic vowel is syllabic.

Other epenthetic vowels have the neutral quality [ə].

6.3.3. Decomposed final stops

Jah Hut final Stops are checked, or unreleased, as in most languages of the Southern Far East. However, when final stops are preceded by nasal vowels, or by vowels preceded by Nasals, they are decomposed into two phonetic segments; the first is a Nasal homorganic with the Stop, the second is a glottal stop:

- /ŋɔk/ → [ŋɔ̃ŋʔ] *to sit*
 /hlaʔɲac/ → [hlaʔɲäpʔ] *to be shy*
 /ʔisēc/ → [ʔisēpʔ] *a bird sp; (spider hunter)*
 /mat/ → [mānʔ] *eye*
 /sʔīt/ → [sīʔīnʔ] *rotten smell*
 /ʔaŋʔhliēp/ → [ʔaŋʔhliēmʔ] *sound of breathing, of the creation of the world*

The final glottal stop preserves the checked character of the stop and maintains the contrast with final nasals.

This could be seen as an indication that all Jah Hut final stops, regardless of what precedes, have a glottal closure in addition to the oral one, a feature which would not surprise Mundaists; but experimental evidence is wanting.

7. REMARKS ON JAH HUT AND MALAY

Speakers of Aslian languages have been in contact with speakers of Austronesian languages for several centuries. Bilingualism across these two language families is not rare, and, probably, many present-day speakers of Austronesian languages like Temuan, Belandas and Jakun are descendants of Aslianophones. It is perhaps through such people that a good deal of linguistic interaction took place. So little is known about Temuan, Belandas, Jakun, and even Malay dialects, that we can only speculate. But two bits of evidence will illustrate the problem.

Modern Malay has lost intervocalic /h/ at an early date. The /h/ is sometimes found in the orthography, e.g. *mahu want*, sometimes not, e.g. *tiang house pole*. Temuan has preserved these h's: Temuan /tihan/ *house pole*. Jah Hut has borrowed the word, not in the Modern Malay form, but in the more archaic Temuan form: Jah Hut /tihan/ *house pole*, and this is not an isolated case; but Jah Hut could also have borrowed it from Malay when the /h/ was still pronounced.

There are other surprising h's in Jah Hut, especially in front of /r/: Jh. /hraket/ *raft* (Mal. *rakit*), Jh. /hrbus/ *to boil* (Mal. *rebus*). Since Semai and Temiar have similar occurrences, the explanation must be sought in the past. Temuan, and colloquial Malay dialects in Malaysia, generally have a voiced velar fricative [ɣ] for /r/, some dialects even have a uvular fricative, and many simply have a breathy [ʁ], especially in final position. Semai and Temiar, on the other hand, and most probably Proto-Senoic, have a distinctly trilled alveolar /r/. In modern Jah Hut, the /r/ is an alveolar approximant, articulated without friction with the tip of the tongue. In order to explain Jah Hut /hraket/ and /hrbus/ one would need to go back in time when Jah Hut still had a trilled /r/ and borrowed Malay or Temuan words with a velar [ɣ] or a post-velar; the sequence /h/ plus /r/ would be a good analysis, in Jah Hut terms, of that unfamiliar sound. Or was it Temuan which introduced the /h/ for similar reasons?

N O T E S

1. In this comparative vocabulary, the abbreviation 'Pin.' followed by a Number refers to the entry in Pinnow 1959, 'SB' followed by a letter and a number refers to the Vocabulary in Vol. II of Skeat and Blagden, 1906. Proto-North-Bahnaric words are from Smith, 1972; Proto-South-Bahnaric words are from Blood, 1966; Proto-East-Katuic words are from Thomas, 1967. The numbers in parentheses for the last three sources refer to entries in their vocabulary lists. Chrau words are from Thomas, 1971; Bahnar from Guilleminet 1959; Pear from Morizon, 1936; Nancowry Nicobar from Radhakrishnan, 1970; Central Nicobar from Man 1889; Riang from Luce 1965; Old Mon and Riang-Lang from Shorto, 1971; Theng from Maspero, 1955; Khmu from Smalley, 1961; Khamet and Lawa from Mitani, 1965. Khasi, Khmer and Jah Hut words are my own recordings.
2. The Khasi cognate is from N. Singh, 1906; the Jah Hut, Bahnar and Theng forms show that the initial stop must have been voiced in Proto-Mon-Khmer. If so, here is a case of the Khasi *g + k innovation which Haudricourt had expected but not found a good example of in 1965.
3. The Jah Hut retention rates (Benjamin, 1973a) are higher with Semai (38-40%) than with the other Senoic languages. If one assigns this to borrowings, as Benjamin does, the average retention rate between Jah Hut and the rest of Senoic would be around 27%, as compared to an average 25% with North Aslian and 24% with South Aslian.
4. Examples 45-46-47 were given to me by Duncan Holaday who did original and inspiring anthropological fieldwork among the Jah Hut in 1969.
5. *Elateriospermum* sp.

6. The plant, a *Derris* sp., is called /jnu?/ in Jah Hut (Mal. tuba jenu).

7. In Malay, such animals often have fully reduplicated names: kupu-kupu *butterfly*, anai-anai *termite*.

BIBLIOGRAPHY

BENJAMIN, G.

1973a 'Austroasiatic subgroupings and prehistory in the Malay Peninsula'. *Austroasiatic Studies, Oceanic Linguistics*, Special Publication No.13. Hawaii. University of Hawaii Press. In press.

1973b 'An outline of Temiar grammar'. *Austroasiatic Studies, Oceanic Linguistics*, Special Publication No.13. Hawaii. University of Hawaii Press. In press.

BLOOD, H.F.

1966 *A Reconstruction of Proto-Mnong*. Summer Institute of Linguistics. Grand Forks. University of North Dakota.

CHAFE, W-L.

1970 *Meaning and the Structure of Language*. University of Chicago Press.

DENTAN, R.K.

1964 'Senoï-Semang'. *Ethnic Groups of Mainland South-East Asia* 176-86. New Haven. HRAF Press.

DIFFLOTH, G.

1973a 'Expressives in Semai'. *Austroasiatic Studies, Oceanic Linguistics*, Special Publication No.13. Hawaii. University of Hawaii Press. In press.

1973b 'Minor syllable vocalism in Senoic Languages'. *Austroasiatic Studies, Oceanic Linguistics*, Special Publication No.13. Hawaii. University of Hawaii Press. In press.

DIFFLOTH, G.

- 1974 'Body moves in Semai and in French'. *Papers from the 10th regional meeting*. Chicago. Chicago Linguistic Society.
- Forth- 'Mon-Khmer numerals in Aslian languages'. *Linguistics*.
coming In press.

GOLAB, Z.

- 1969 'Subject as a linguistic category'. *General Linguistics* 9.

GUILLEMINET, P.

- 1959 *Dictionnaire Bahnar-Français*. Tome I. Paris. Ecole Française d'Extrême-Orient.
- 1963 *Dictionnaire Behnar-Français*. Tome II. Paris. Ecole Française d'Extrême-Orient.

HAUDRICOURT, A.

- 1965 'Les mutations consonantiques des occlusives initiales en Mon-Khmer'. *Bulletin de la Société de Linguistique de Paris* 60:160-72.

JAKOBSON, R.

- 1966 'Quest for the essence of language'. *Diogenes* 51. Montreal.

JENNER, P.N.

- 1969 *Affixation in Modern Khmer*. Ph.D. Dissertation. University of Hawaii.

LUCE, G.H.

- 1965 'Danaw, a dying Austroasiatic language'. In: G.B. Milner and E.J.A. Henderson (eds.) *Indo-Pacific Linguistic Studies*, Pt.1:98-129. Amsterdam. North-Holland Publishing Co.

MAN, E.H.

- 1889 *Dictionary of the Central Nicobarese Language*. London.

MASPERO, H.

- 1955 'Matériaux pour l'étude de la langue T'eng'. *Bulletin de l'Ecole Française d'Extrême-Orient* 47:457-507.

MITANI, Y.

- 1965 *Descriptive and Comparative Study of the Khamet Phonology.*
Kyoto University. Tonan Ajia Kenkyu. 3/3:22-51.

MORIZON, R.

- 1936 *Essai sur le dialecte des populations Pears des Cardamomes.*
Paris. Etudes Internationales.

PINNOW, H.J.

- 1959 *Versuch einer historischen lautlehre der Kharia-Sprache.*
Wiesbaden. O. Harrassowitz.

RADHAKRISHNAN, R.

- 1970 *A preliminary descriptive analysis of Nancowry.* Ph.D.
Dissertation. University of Chicago.

SHORTO, H.L.

- 1971 *A Dictionary of the Mon inscriptions.* Oxford University
Press.

SILVERSTEIN, M.

- 1973 *Hierarchy of features and ergativity.* Paper read at the
January 1973 meeting of the Chicago Linguistic Society.

SINGH, N.

- 1906 *Khasi-English Dictionary.* Shillong. Eastern Bengal and
Assain Secretariat Press.

SKEAT, W.W. and C.O. BLAGDEN

- 1906 *Pagan Races of the Malay Peninsula.* 2 Vols. London.
Macmillan.

SMALLEY, W.A.

- 1961 *Outline of Khmu Structure.* American Oriental Essays, No.2.
Baltimore. American Oriental Society.

SMITH, K.D.

- 1972 *A phonological reconstruction of Proto-North-Bahnaric.*
Santa Ana, California. Summer Institute of Linguistics.

THOMAS, D.D.

- 1971 *Chrau Grammar, Oceanic Linguistics,* Special Publication
No.7. University of Hawaii Press.

THOMAS, D.M.

- 1967 *A phonological reconstruction of Proto-East-Katuic.* M.A. Thesis. Grand Forks. University of North Dakota.

WILKINSON, R.J.

- 1926 *Papers on Malay Subjects. Supplement: The Aboriginal Tribes.* Kuala Lumpur. Federated Malay States Government Press.

WILLIAMS-HUNT, P.D.R.

- 1950 *An introduction to the Malayan Aborigines.* Kuala Lumpur. Government Press.

THE PHONOLOGICAL BEHAVIOR OF MALAY PREFIXES WITH A NASAL ENDING

Sidharta (Sie Ing Djiang)

1.1. This paper is an attempt, within the framework of generative phonology, to set up the rules governing the behavior of a certain class of Malay¹ prefixes, namely those which end in a nasal consonant. There are two of them, /məN/ and /pəN/ (where the morphophonemic symbol N stands for 'nasal consonant'), and both behave in a similar manner: the nasal consonant becomes homorganic with the following sound (that is, the initial sound of the stem) and in certain cases both the nasal consonant of the prefix and the initial consonant of the stem undergo further regular changes as exemplified in Tables 1 and 2 below.

TABLE 1

initial sound	stem	preceded by /məN/	plus stem re- duplication	preceded by /pəN/
1	2	3	4	5
1. /p/	(a) pindah <i>move</i>	məindah <i>to move</i>	məindahindah <i>to spread, infect</i>	pəindahan ² <i>transfer</i>
	(b) pəduli <i>heed</i>	məpədulikan ³ <i>to heed</i>		
2. /b/	bəsar ⁴ <i>big</i>	məbesar <i>to get bigger</i>	məbesarbesar ⁵ <i>to exaggerate</i>	pəbesar <i>prominent person</i>
3. /t/	(a) tari <i>dance</i>	mənari <i>to dance</i>	mənarinari <i>to dance for joy</i>	pənari <i>dancer</i>
	(b) tərjəmah <i>translate</i>	mətərjəmahkan ⁶ <i>to translate</i>		pəntərjəmah <i>translator</i>
4. /d/	doron <i>push</i>	mədoron <i>to push</i>	mədoron ⁷ doron <i>keep pushing</i>	pəndoron <i>incentive</i>
5. /k/	(a) kirə ⁷ <i>count</i>	məņirə <i>to count</i>	məņirəņirə <i>to estimate</i>	pəņirə <i>enumerator</i>

Table 1 (cont.)

initial sound	stem	preceded by /mƏN/	plus stem re-duplication	preceded by /pƏN/
1	2	3	4	5
5. /k/	(b) kritik ⁸ <i>criticism</i>	mənkritik <i>to criticize</i>		pənkritik <i>critic</i>
6. /g/	goso? ⁹ <i>rub</i>	məngoso? <i>to scour</i>	məngoso?goso? <i>to polish</i>	pəngoso? <i>scrubber</i>
7. /f/	fitnah <i>slander</i>	məmfitnah <i>to slander</i>		pəmfitnah ¹⁰ <i>defamation</i>
8. /v/	vito ¹¹ <i>veto</i>	məmvito <i>to veto</i>		
9. /θ/	θabet ¹² <i>authentic</i>	mənθabetkan <i>to authenticate (a tradition)</i>		
10. /ð/	ðarab ¹³ <i>multiply</i>	mənðarab <i>to multiply (numbers)</i>		pənðarab <i>multiplier</i>
11. /s/	(a) sapu <i>broom</i>	mənapu <i>to sweep</i>	mənapunapu <i>to wipe</i>	pənapu <i>sweeper</i>
	(b) sah <i>legal</i>	mənsahkan ¹⁴ <i>to legalize</i>		pənsahan ¹⁵ <i>legalization</i>
12. /z/	ziarah <i>pilgrimage</i>	mənziarahi ¹⁶ <i>to visit a holy place</i>		pənziarah <i>pilgrim</i>
13. /ʃ/	šarah <i>lecture</i>	məņšarahkan ¹⁷ <i>to lecture in</i>		pəņšarah <i>lecturer</i>
14. /x/	xianat <i>betray</i>	məņxianati ¹⁸ <i>to betray</i>		pəņxianat <i>traitor</i>
15. /ɣ/	ɣəb <i>invisible</i>	məņɣəbkan ¹⁹ <i>to make invisible</i>		
16. /č/	čuri <i>steal</i>	məņčuri ²⁰ <i>to steal</i>	məņčuričuri <i>to be stealthy</i>	pəņčuri <i>thief</i>
17. /j/	jilat <i>lick</i>	məņjilat ²¹ <i>to lick</i>	məņjilatjilat <i>to spread (fire)</i>	pəņjilat <i>flatterer</i>
18. /m/	masa? ²² <i>cook</i>	məmasa? <i>to cook</i>	məmasa?masa? <i>prepare a feast</i>	pəmasa? <i>cook, chef</i>
19. /n/	nanti <i>wait</i>	mənanti <i>to wait</i>	mənantinanti <i>to await eagerly</i>	pənanti <i>receptionist</i>
20. /ŋ/	ŋəpi <i>sing</i>	məŋəpi <i>to sing</i>	məŋəpiŋəpi <i>sing all the time</i>	pəŋəpi <i>singer</i>
21. /ŋ/	ŋəŋə ²³ <i>gape</i>	məŋəŋə <i>to gape</i>		
22. /l/	lompat <i>jump</i>	məlompat <i>to jump</i>	məlompatlompat <i>keep jumping</i>	pəlompat <i>jumper</i>
23. /r/	rasə ²⁴ <i>taste</i>	mərasə <i>to taste</i>	mərasərasə <i>to touch, feel</i>	pərasə <i>sensitive</i>
24. /h/	harap <i>hope</i>	məņharap <i>to hope</i>		pəņharapan ²⁵ <i>hope</i>
25. /ʔ/	ʔibarət <i>comparison</i>	məņʔibarətkan ²⁶ <i>to compare, liken</i>		

Table 1 (cont.)

initial sound	stem	preceded by /məN/	plus stem re-duplication	preceded by /pəN/
1	2	3	4	5
26. /y/	yaken <i>convinced</i>	məyakenkan ²⁷ <i>to convince</i>		
27. /w/	warnə <i>colour</i>	məwarnəkan ²⁸ <i>to colour</i>		pəwarnə <i>pigment</i>
28. /i/	intay <i>spy</i>	məŋintay <i>to spy on</i>		pəŋintay <i>spy</i>
29. /e/	ekor <i>tail</i>	məŋekori ²⁹ <i>to tail</i>		pəŋekor <i>follower</i>
30. /ə/	əmbərə ³⁰ <i>roam</i>	məŋəmbərə <i>to roam</i>		pəŋəmbərə <i>wanderer</i>
31. /a/	aŋgo? ³¹ <i>nod</i>	məŋaŋgo? <i>to nod</i>	məŋaŋgo?aŋgo? <i>to nod repeatedly</i>	pəŋaŋgo? <i>bob-stay</i>
32. /o/	olah <i>process</i>	məŋolah <i>to process</i>		pəŋolah <i>manufacturer</i>
33. /u/	undaŋ <i>enact</i>	məŋundaŋkan ³² <i>to enact a law</i>		pəŋundaŋundaŋ <i>legislator</i>

TABLE 2

stem	preceded by /məN/ and /pəN/	preceded by /məN/ and /təN/	preceded by /pəN/ and /pəN/
1. hati <i>heart</i>	məmpərhəhatikan ³³ <i>to pay attention to</i>		pəmpərhəhati <i>observer</i>
2. čəŋaŋ <i>surprise</i>		məntərčəŋaŋkan ³⁴ <i>to astonish</i>	

The rules to be set up must be able to account for all the forms contained in the above tables.

1.2. Within the theoretical framework proposed by Chomsky and Halle (1968)³⁵ the sounds of Malay can be described in terms of fifteen binary features as in Table 3.

With regard to Table 3 the following remarks are in order. Since Malay has no syllabic consonants, the feature [+ syllabic] is sufficient to distinguish the vowels from the consonants (including the laryngeal glides and the semi-vowels), which are all [- syllabic]. The vowels are redundantly [+ sonorant, - consonantal]. For the [+ low] vowel the feature [- high] is redundant by reason of universal redundancy.

1.3. Returning to the examples in Tables 1 and 2, the behavior of the nasal ending of the prefix and the changes in the stems can be described as follows.

(i) Except in examples 1(a), 3(a), 5(a), 11(a), 18 - 23, and 26 - 27 in Table 1, the final nasal consonant of the prefix becomes homorganic with the initial sounds of the following stem.

(ii) In the case of items 18 - 23 and 26 - 27, the nasal consonant of the prefix is deleted when followed by stems beginning with /m, n, ŋ, ŋ, l, r, ɣ, w/, that is, when followed by segments having the features [- syllabic, + sonorant].

(iii) In examples 1(a), 3(a), 5(a) and 11(a), which are representative of the majority of Malay stems beginning with /p, t, k, s/, the initial consonants of the stems undergo a further change through assimilation to the preceding nasal by becoming /m, n, ŋ, ŋ/ respectively, while the final nasal consonant of the prefix is deleted. These changes do not occur in a class of stems beginning with /p, t, k, s/ represented by examples 1(b), 3(b), 5(b) and 11(b), comprising mostly unassimilated borrowings from foreign languages.

(iv) Column 4 of Table 1 shows that the second member of reduplicated stems undergoes the same change as the one occurring in the first member, if any.

(v) Although they begin with /p/ and /t/ respectively, the prefixes /pər/ and /tər/ are not subject to the changes described under (iii) above when preceded by the prefix /məN/, but /pər/ is regular when preceded by /pəN/. This is shown in Table 2.

2.1. Based on the description of Malay sounds in Table 3, the following rules can be set up to account for the changes described in section 1.3.

The changes in 1.3. (i), in which the nasal consonant of the prefix becomes homorganic with the initial segment of the stem can be handled by a rule with multiple variables involving the features 'anterior' and 'coronal'. Given the following table of the values of the features 'anterior' and 'coronal' for all sounds of Malay:

TABLE 4

	p	b	t	d	k	g	f	v	θ	ð	s	z	ʃ	x	γ	č	ǰ	m	n	ɲ	ŋ	l	r	h	ʔ	y	w	i	e	ə	a	o	u		
anterior	+	+	+	+	-	-	+	+	+	+	+	+	-	-	-	-	-	+	+	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	
coronal	-	-	+	+	-	-	-	-	-	+	+	+	+	+	-	-	+	+	-	+	+	-	+	+	-	-	-	-	-	-	-	-	-	-	-

the following rule can be set up:

$$(1) \begin{bmatrix} - & \text{syll} \\ + & \text{nasal} \end{bmatrix} \longrightarrow \begin{bmatrix} \alpha & \text{ant} \\ \beta & \text{cor} \end{bmatrix} / \text{---} + \begin{cases} \begin{bmatrix} - & \text{syll} \\ - & \text{son} \\ \alpha & \text{ant} \\ \beta & \text{cor} \end{bmatrix} & (a) \\ \begin{bmatrix} + & \text{syll} \\ \alpha & \text{ant} \\ \beta & \text{cor} \end{bmatrix} & (b) \end{cases}$$

This rule states that the final nasal consonant of the prefix (the + sign between the segments is the morpheme boundary) becomes homorganic with the initial segments of the following stems if they are obstruents (which have the features [- syllabic, - sonorant]) (rule 1(a)) or vowels (which are [+ syllabic]) (rule 1(b)), that is, the nasal consonant is realized as [m] ([+ anterior, - coronal]) when followed by /p, b, f, v/ all of which are [+ anterior, - coronal], as [n] ([+ anterior, + coronal]) when followed by /t, d, θ, ð, s, z/ (all of them [+ anterior, + coronal]), as [ɲ] ([- anterior, + coronal]) when followed by /ʃ, x, γ/ (all of them [- anterior, + coronal]), and as [ŋ] ([- anterior, - coronal]) when preceding /k, g, x, γ, h, ʔ/ or the vowels /i, e, ə, a, o, u/ all of which have the features [- anterior, - coronal]. Thus rule 1 accounts for all the examples in columns 3 and 5 of Table 1, with the exception of 1(a), 3(a), 5(a), 11(a), 18 - 23, and 26 - 27.

2.2. The items in columns 3 and 5 of examples 18 - 23 and 26 - 27 can be accounted for by setting up rule 2.

$$(2) \begin{bmatrix} - & \text{syll} \\ + & \text{nasal} \end{bmatrix} \longrightarrow \emptyset / \text{---} + \begin{bmatrix} - & \text{syll} \\ + & \text{son} \end{bmatrix}$$

which states that the final nasal consonant of the prefix is deleted when preceding sonorant consonants, that is, the nasals /m, n, ɲ, ŋ/, the liquids /l, r/ and the semivowels /y, w/ all of which have the features [- syllabic, + sonorant]. This accounts for the changes described in section 1.3. (ii). Rules 1 and 2 can be collapsed into one rule:

$$(3) \begin{bmatrix} - \text{syll} \\ + \text{nasal} \end{bmatrix} \longrightarrow \left\{ \begin{array}{l} \begin{bmatrix} \alpha \text{ ant} \\ \beta \text{ cor} \end{bmatrix} / \text{---} + \begin{bmatrix} - \text{syll} \\ - \text{son} \\ \alpha \text{ ant} \\ \beta \text{ cor} \end{bmatrix} \quad (a) \\ \begin{bmatrix} + \text{syll} \\ \alpha \text{ ant} \\ \beta \text{ cor} \end{bmatrix} \quad (b) \\ \emptyset / \text{---} + \begin{bmatrix} - \text{syll} \\ + \text{son} \end{bmatrix} \quad (c) \end{array} \right.$$

2.3. Two further rules are needed to account for the fact that the vast majority of Malay stems undergo a further change as stated in section 1.3. (iii):

$$(4) \begin{bmatrix} - \text{son} \\ - \text{del.rel.} \\ \alpha \text{ ant} \\ \beta \text{ cor} \\ - \text{voice} \end{bmatrix} \longrightarrow \begin{bmatrix} + \text{nasal} \\ \alpha \text{ ant} \\ \beta \text{ cor} \end{bmatrix} / \begin{bmatrix} - \text{syll} \\ + \text{nasal} \end{bmatrix} + \text{---}$$

$$(5) \begin{bmatrix} - \text{son} \\ + \text{strid} \\ + \text{ant} \\ + \text{cor} \\ - \text{voice} \end{bmatrix} \longrightarrow \begin{bmatrix} + \text{nasal} \\ - \text{ant} \end{bmatrix} / \begin{bmatrix} - \text{syll} \\ + \text{nasal} \end{bmatrix} + \text{---}$$

Rule 4 states that /p, t, k/ in the stem assimilate to the preceding nasal in the prefix by becoming their respective homorganic nasals, namely /m, n, ŋ/, while by rule 5 an initial /s/ in the stem assimilates to the preceding nasal by becoming /ɾ/. Both rules can be collapsed into:

$$(6) \left. \begin{array}{l} \begin{bmatrix} - \text{son} \\ - \text{del.rel.} \\ \alpha \text{ ant} \\ \beta \text{ cor} \\ - \text{voice} \end{bmatrix} \longrightarrow \begin{bmatrix} + \text{nasal} \\ \alpha \text{ ant} \\ \beta \text{ cor} \end{bmatrix} \quad (a) \\ \begin{bmatrix} - \text{son} \\ + \text{strid} \\ + \text{ant} \\ + \text{cor} \\ - \text{voice} \end{bmatrix} \longrightarrow \begin{bmatrix} + \text{nasal} \\ - \text{ant} \end{bmatrix} \quad (b) \end{array} \right\} / \begin{bmatrix} - \text{syll} \\ + \text{nasal} \end{bmatrix} + \text{---}$$

Since the output of rule 6 meets the condition of rule 3(c) the final nasal consonant of the prefix is deleted. Thus the application of rules 6 and 3(c) accounts for examples 1(a), 3(a), 5(a) and 11(a) in Table 1.

As mentioned earlier the vast majority of Malay stems beginning with /p, t, k, s/ behave in the manner of the above examples, that is, they undergo rule 6. On the other hand, a small class of stems beginning with /p, t, k, s/, consisting of unassimilated borrowings from foreign languages represented in Table 1 by examples 1(b), 3(b), 5(b) and 11(b), do not undergo rule 6.

This can be handled by a minus rule feature, that is, their lexical entries are specified [- rule 6], or rather [- rule 7] (see following section), for example:

$$\left[\begin{array}{l} \text{pəduli} \\ + \text{STEM} \\ - \text{rule 7} \\ \vdots \\ \vdots \end{array} \right]^{36}, \quad \left[\begin{array}{l} \text{tərjəmah} \\ + \text{STEM} \\ - \text{rule 7} \\ \vdots \\ \vdots \end{array} \right], \quad \left[\begin{array}{l} \text{kritik} \\ + \text{STEM} \\ - \text{rule 7} \\ \vdots \\ \vdots \end{array} \right], \quad \left[\begin{array}{l} \text{sah} \\ + \text{STEM} \\ - \text{rule 7} \\ \vdots \\ \vdots \end{array} \right], \text{ etc.}$$

2.4. In order to account for the changes in reduplicated stems described in section 1.3. (iv), rule 6 must be changed as follows.

$$(7) \left. \begin{array}{l} \left[\begin{array}{l} - \text{son} \\ - \text{del.rel.} \\ \alpha \text{ ant} \\ \beta \text{ cor} \\ - \text{voice} \end{array} \right] \longrightarrow \left[\begin{array}{l} + \text{nasal} \\ \alpha \text{ ant} \\ \beta \text{ cor} \end{array} \right] \\ \left[\begin{array}{l} - \text{son} \\ + \text{strid} \\ + \text{ant} \\ + \text{cor} \\ - \text{voice} \end{array} \right] \longrightarrow \left[\begin{array}{l} + \text{nasal} \\ - \text{ant} \end{array} \right] \end{array} \right\} / \left[\begin{array}{l} - \text{syll} \\ + \text{nasal} \end{array} \right] + \left[\text{STEM}[\text{---}]X_i \right]_{\text{STEM}} +$$

$$\left(\left[\text{STEM}[\text{---}]X_j \right]_{\text{STEM}} \right) \quad (a)$$

$$(b)$$

Conditions: $X_i = X_j$, $X_i \neq \emptyset$, $X_j \neq \emptyset$

Rule 7 (in which the variables X_i and X_j stand for nonnull segment sequences) states that in reduplicated stems preceded by /məN/ the second member undergoes the same change as the first, if any. This accounts for all forms in column 4 of Table 1.

2.5. As stated in section 1.3. (v) the prefixes /pər/ and /tər/ do not undergo rule 7 when preceded by /məN/, but /pər/ is regular when preceded by /pəN/³⁷. This can be accounted for by rule 8.

$$(8) \left. \begin{array}{l} \left[\begin{array}{l} \text{pər} \\ + \text{PREFIX} \end{array} \right] \\ \left[\begin{array}{l} \text{tər} \\ + \text{PREFIX} \end{array} \right] \end{array} \right\} \longrightarrow [- \text{rule 7}] / \left[\begin{array}{l} \text{məN} \\ + \text{PREFIX} \end{array} \right] + \text{---}$$

Being an exception, rule 8, which accounts for the examples in Table 2, must apply before rules 3 and 7, if those examples are to be derived correctly.

2.6. From the above it would seem that, in addition to a minus rule feature in the lexical entries of a small class of stems, only three rules, namely rule 8, rule 3 and rule 7, in that order, are needed to account for the following facts of the Malay language:

(a) The final consonant of the prefixes /məN/ and /pəN/ becomes homorganic with the initial segment of the following stem (including vowels), except when it precedes segments with the specifications [-syllabic, +sonorant], in which case it is deleted. In other words, the final nasal consonant of the prefix is realized as [m] before /p, b, f, v/, as [n] before /t, d, θ, ð, s, z/, as [ŋ] before /ʒ, ʃ, ʃ/ and as [ŋ] before /k, g, x, ɣ, h, ʔ/ and the vowels /i, e, ə, a, o, u/, but deleted before /m, n, ŋ, ŋ, l, r, ɣ, w/, that is, before nasals, liquids and semivowels. As can be seen from the above, these changes also involve the so-called 'secondary consonants' of Malay (Maris 1966: 144 ff), namely /f, v, θ, ð, z, ʒ, x, ɣ/, which are borrowed from foreign languages, mainly from Arabic and English. Malay has only one native fricative phoneme: /s/.

(b) Native Malay stems beginning with /p, t, k, s/ undergo a further change. In stems beginning with /p, t, k/ the initial stops assimilate to the preceding nasal by becoming their homorganic nasals, whereas in those beginning with /s/ the initial fricative assimilates to the preceding nasal by becoming /p/. The final nasal consonant of the prefix is then deleted since they are followed by a nasal (see (a) above). A small class of stems beginning with /p, t, k, s/, however, do not undergo this further change. This class includes only borrowings from foreign languages that have not been assimilated (represented by examples 1(b), 3(b), 5(b) and 11(b) in Table 1). Loan words which have been assimilated, such as /palsu/ *false* (from Portuguese), /tayp/ *type* (spelled taip, from English), /kontrol/ *control* (from English) and /səkolah/ *school* (from Portuguese)³⁸ behave like native Malay stems:

məməlsu to *counterfeit*
 mənayp to *typewrite*
 məŋontrol to *control*
 məpəkolahkan to *enrol (somebody) in a school*

The following simplified derivations of some examples representative of the above changes (cf. Tables 1 and 2) prove the adequacy of the rules set up so far.

- (1) /p/ Stem: pindah
 # məN + pindah #
 # məm + pindah # (rule 3(a))
 # məm + mindah # (rule 7(a))
 # mə + mindah # (rule 3(c))
 məmindah³⁹
 Stem: pəduli [- rule 7]
 # məN + pəduli + kan #
 # məm + pəduli + kan # (rule 3 (a))
 məmpədulikan

- (ii) /t/ Stem: tari
 # məN + tari #
 # mən + tari # (rule 3(a))
 # mən + nari # (rule 7(a))
 # mə + nari # (rule 3(c))
 mənari
 Stem: tərjəmah [- rule 7]
 # məN + tərjəmah + kan #
 # mən + tərjəmah + kan # (rule 3(a))
 məntərjəmahkan
- (iii) /k/ Stem: kira
 # məN + kira #
 # məŋ + kira # (rule 3(a))
 # məŋ + ŋira # (rule 7(a))
 # mə + ŋira # (rule 3(c))
 məŋira
 Stem: kritik [- rule 7]
 # məN + kritik #
 # məŋ + kritik # (rule 3(a))
 məŋkritik
- (iv) /s/ Stem: sapu
 # məN + sapu #
 # mən + sapu # (rule 3(a))
 # mən + ŋapu # (rule 7(b))
 # mə + ŋapu # (rule 3(c))
 məŋapu
 Stem: sah [- rule 7]
 # məN + sah + kan #
 # mən + sah + kan # (rule 3(a))
 mənsahkan
- (v) /j/ Stem: jilat
 # məN + jilat #
 # məŋ + jilat # (rule 3(a))
 məŋjilat
- (vi) /i/ Stem: intay
 # məN + intay #
 # məŋ + intay # (rule 3(b))
 məŋintay

- (vii) /m/ Stem: masa?
 # məN + masa? #
 # mə + masa? # (rule 3(c))
 məmasa?
- (viii) /l/ Stem: lompat
 # məN + lompat #
 # mə + lompat # (rule 3(c))
 məlompat
- (ix) /w/ Stem: warnə
 # məN + warnə + kan #
 # mə + warnə + kan # (rule 3(c))
 məwarnəkan
- (x) Stem: hati
 # məN + pər + hati + kan #
 # məN + $\left[\begin{array}{l} \text{pər} \\ \text{- rule 7} \end{array} \right]$ + hati + kan # (rule 8)
 # məm + pər + hati + kan # (rule 3(a))
 məmpərhatikan
- (xi) Stem: čəŋaŋ
 # məN + tər + čəŋaŋ + kan #
 # məN + $\left[\begin{array}{l} \text{tər} \\ \text{- rule 7} \end{array} \right]$ + čəŋaŋ + kan # (rule 8)
 # mən + tər + čəŋaŋ + kan # (rule 3(a))
 məntərčəŋaŋkan

3.1. Although, as shown above, the three rules set up so far, do generate the relevant forms, there is some question as to their naturalness.

It is very common for a nasal consonant to become homorganic with the following consonant as, for instance, in Yoruba (see Schane 1973:51). However, rule 3 states that the nasal ending of the prefix becomes homorganic only with a following obstruent or vowel, whereas it is deleted when preceding a nasal, liquid or semivowel. It is much more natural for the nasal ending of the prefix to become homorganic - that is, to totally assimilate - to a following nasal than to be dropped. If it is assumed that this is indeed the case in Malay, it will be necessary to find a well-motivated rule to account for the deletion of the nasal ending of the prefix. As a matter of fact there is a quite general rule in Malay which can account for it. Consider the following facts. There is a prefix /bər/ which, when followed by the stem /rumah/ *house*, would give *[bərrumah] *possessing a house*. Actually the

resulting form is [bərumah]. In other words, geminate consonants are degeminated.⁴⁰ This can be expressed by rule 9 (where C stands for 'consonant', that is, sounds having the feature [- syllabic]):

$$(9) C_i \longrightarrow \emptyset / \text{---} + C_j$$

Condition: $C_i = C_j$

Rule 9 also provides an explanation for another phenomenon. When /bər/ precedes the stem /layar/ *sail* the resulting form with some speakers is [bər^hlayar] *to sail*, which is regular, but with others it is [bəl^hayar]. The latter case cannot be accounted for by simply stating that in the case of some speakers [r] is deleted if /bər/ is prefixed to /layar/. The form [bəl^hayar] results from the application of a dissimilation rule present in the phonology of some speakers which specifies the non-lateral liquid of the prefix with the opposite value of the non-lateral liquid in the stem, that is, the final [r] of the prefix becomes [l] which is [+ lateral], giving the form # bəl + layar #, which in turn becomes # bə + layar # by rule 9.⁴¹

The dissimilation rule is also operative in such changes as /bər/ + /aʃar/ *study* + # bəl + aʃar # *to study*. Now that rule 9 is available, rule 3(c) can be dropped and rules 3(a) and (b) replaced by a quite general rule:

$$(10) \begin{bmatrix} - & \text{syll} \\ + & \text{nasal} \end{bmatrix} \longrightarrow \begin{bmatrix} \alpha & \text{ant} \\ \beta & \text{cor} \end{bmatrix} / \text{---} + \begin{bmatrix} \alpha & \text{ant} \\ \beta & \text{cor} \end{bmatrix}$$

This rule states that the final nasal consonant of /mən/ and /pən/ is realized as a nasal homorganic with the initial segment of the following stem.

Now rules 8, 10, 7 and 9, in that order, will account for all the examples in Tables 1 and 2, except for the items containing stems beginning with a liquid or a semivowel.

3.2. In order to account for the latter two further rules are needed:

$$(11) \begin{bmatrix} - & \text{syll} \\ + & \text{nasal} \end{bmatrix} \longrightarrow \begin{bmatrix} - & \text{nasal} \\ \alpha & \text{lateral} \end{bmatrix} / \text{---} + \begin{bmatrix} - & \text{syll} \\ + & \text{son} \\ - & \text{nasal} \\ \alpha & \text{lateral} \end{bmatrix}$$

$$(12) \begin{bmatrix} - & \text{syll} \\ + & \text{nasal} \end{bmatrix} \longrightarrow \begin{bmatrix} - & \text{cons} \\ \alpha & \text{back} \end{bmatrix} / \text{---} + \begin{bmatrix} - & \text{syll} \\ + & \text{son} \\ - & \text{cons} \\ \alpha & \text{back} \end{bmatrix}$$

Rule 11 states that the homorganic nasal of the prefix assimilates to the initial liquid of the stem (a common occurrence, cf. Schane 1973: 55), and rule 12 that the nasal assimilates to the initial semivowel of the stem. The application of rule 9 will automatically degeminate

the resulting geminate liquids and semivowels.

Rules 10, 11 and 12 can be collapsed into rule 13.

$$(13) \begin{bmatrix} - \text{syll} \\ + \text{nasal} \end{bmatrix} \longrightarrow \left\{ \begin{array}{l} \begin{bmatrix} \alpha \text{ ant} \\ \beta \text{ cor} \end{bmatrix} / \text{---} + \begin{bmatrix} \alpha \text{ ant} \\ \beta \text{ cor} \end{bmatrix} \quad (a) \\ \begin{bmatrix} - \text{nasal} \\ \alpha \text{ lateral} \end{bmatrix} / \text{---} + \begin{bmatrix} - \text{syll} \\ + \text{son} \\ - \text{nasal} \\ \alpha \text{ lateral} \end{bmatrix} \quad (b) \\ \begin{bmatrix} - \text{cons} \\ \alpha \text{ back} \end{bmatrix} / \text{---} + \begin{bmatrix} - \text{syll} \\ + \text{son} \\ - \text{cons} \\ \alpha \text{ back} \end{bmatrix} \quad (c) \end{array} \right.$$

3.3. The application of rules 8, 13, 7 and 9, in that order, will account for all forms in Tables 1 and 2. The addition of rule 9, a general rule which is needed anyhow for independent reasons, does not add to the complexity of the phonological description of Malay, although the replacement of rule 3 by rule 13 admittedly has been made at some cost from the viewpoint of the simplicity metric.

It is, however, the contention of this paper that the above sequence of rules accounts for the facts of Malay under discussion in a more natural way. More specifically, the rules assert that the changes exemplified in Tables 1 and 2 are the result of the following processes:

- (a) The oppositions amongst nasal consonants in final position in the two existing Malay prefixes with nasal ending (/mən/ and /pən/) are neutralized, that is, the nasal consonant is realized as a nasal homorganic with the initial segment of the stem. In the case of stems beginning with a nasal consonant this means that the nasal ending of the prefix totally assimilates to the stem initial (rule 13(a)).
- (b) Subsequently the homorganic nasal undergoes assimilation when followed by a liquid (rule 13(b)) or a semivowel (rule 13(c)).
- (c) Native Malay stems and assimilated borrowed ones beginning with /p, t, k, s/ undergo a further change. Stem initial /p, t, k/ assimilate to the preceding nasal by becoming their homorganic nasals [m, n, ŋ] (rule 7(a)), whereas stem initial /s/ becomes [ɲ] (rule 7(b)). In the latter case the preceding [n] becomes homorganic with the stem initial [ɲ] as a result of rule 13(a). Non-assimilated stems beginning with /p, t, k, s/ are prevented from being affected by rule 7 by a minus rule feature, that is, they have the feature [- rule 7] in their lexical entries.
- (d) The resulting geminate consonants are degeminated by rule 9.

(e) The application of rule 8 (which is ordered prior to the above rules) and rule 13(a) results in the forms exemplified in Table 2.

A few illustrative simplified derivations are given below for the purpose of comparison with those given in section 2.6. above.

- (i) /p/ Stem: pindah
 # məN + pindah #
 # məm + pindah # (rule 13(a))
 # məm + mindah # (rule 7(a))
 # mə + mindah # (rule 9)
 məmindah
 Stem: pəduli [- rule 7]
 # məN + pəduli + kan #
 # məm + pəduli + kan # (rule 13(a))
 məmpədulikan
- (ii) /s/ Stem: sapu
 # məN + sapu #
 # mən + sapu # (rule 13(a))
 # mən + ɲapu # (rule 7(b))
 # məɲ + ɲapu # (rule 13(a))
 # mə + ɲapu # (rule 9)
 məɲapu
 Stem: sah [- rule 7]
 # məN + sah + kan #
 # mən + sah + kan # (rule 13(a))
 mənsahkan
- (iii) /m/ Stem: masa?
 # məN + masa? #
 # məm + masa? # (rule 13(a))
 # mə + masa? # (rule 9)
 məmasa?
- (iv) /l/ Stem: lompat
 # məN + lompat #
 # mən + lompat # (rule 13(a))
 # məl + lompat # (rule 13(b))
 # mə + lompat # (rule 9)
 məlompat
- (v) /w/ Stem: warnə
 # məN + warnə + kan #
 # məɲ + warnə + kan # (rule 13(a))
 # məw + warnə + kan # (rule 13(c))

mə + warnə + kan # (rule 9)

məwarnəkan

(vi)

Stem: hati

məN + pər + hati + kan

məN + $\left[\begin{array}{l} \text{pər} \\ - \text{rule 7} \end{array} \right] + \text{hati} + \text{kan} \# \text{ (rule 8)}$

məm + pər + hati + kan # (rule 13(a))

məmpərhatikan

4.1. As in the case with most rules, the rules governing the behaviour of the prefixes under discussion have their exceptions. In this section and the next the discussion will concern an apparent exception and the handling of a genuine one.

The apparent exception involves monosyllabic stems, which constitute a very small minority in Malay, where the vast majority of stems are disyllabic.

A majority of speakers have, instead of [mənsahkan] (see section 3.3. (ii), the form [məŋəsahkan]. Another case in point is [məŋəbom] to *bomb*, derived from the monosyllabic stem /bom/ *bomb*, instead of [məmbom], which also exists. For these speakers the following rule applies.

$$(14) \quad \emptyset \longrightarrow \left[\begin{array}{l} + \text{syll} \\ - \text{high} \\ - \text{low} \\ + \text{back} \\ - \text{round} \end{array} \right] / \left\{ \begin{array}{l} \left[\begin{array}{l} \text{məN} \\ + \text{PREFIX} \end{array} \right] \\ \left[\begin{array}{l} \text{pəN} \\ + \text{PREFIX} \end{array} \right] \end{array} \right\} + \text{---} \left[\begin{array}{l} S_1^1 \\ + \text{STEM} \end{array} \right]$$

Rule 14 states that monosyllabic stems (for the symbol S for SYLLABLE, see Harms 1968: 117) are made disyllabic by adding a schwa in front of them when preceded by /məN/ or /pəN/. This results in disyllabic stems with an initial vowel, which regularly undergo rule 13.

4.2. The real exception involves the following cases. The prefix /pəN/ added to the stem /lihat/ *see* gives [pəlihat] *one who sees, a seer*, which is regular. However, when /pəN/ and the suffix /an/ are added to the stem the resulting form is not the expected *[pəlihatan] but [pəŋlihatan] *sight, perception*. There exists also a literary form [pəŋlihat] meaning *vision*. Furthermore, there is the form [pəŋlipor] as in *pəŋlipur lara comfort, diversion* in the speech of some speakers where others have the regular [pəlipor]. These exceptional forms can be handled by the use of minus and plus rule features in the following rules.

- (15) $\begin{bmatrix} \text{p}\text{ə}\text{N} \\ + \text{PREFIX} \end{bmatrix} \longrightarrow \begin{bmatrix} - \text{rule 13} \\ + \text{rule 16} \end{bmatrix} / \text{---} + \begin{bmatrix} \text{STEM} \\ + \text{rule 16} \end{bmatrix} \begin{bmatrix} - \text{syll} \\ - \text{nasal} \\ + \text{lateral} \end{bmatrix}$
- (16) $\begin{bmatrix} - \text{syll} \\ + \text{nasal} \end{bmatrix} \longrightarrow \begin{bmatrix} - \text{ant} \\ - \text{cor} \end{bmatrix} / \left[\begin{array}{l} \begin{bmatrix} - \text{son} \\ - \text{del. rel.} \\ + \text{ant} \\ - \text{cor} \\ - \text{voice} \end{bmatrix} \begin{bmatrix} + \text{syll} \\ - \text{high} \\ - \text{low} \\ + \text{back} \\ - \text{round} \end{bmatrix} \begin{bmatrix} \text{---} \end{bmatrix} \\ + \text{PREFIX} \\ - \text{rule 13} \\ + \text{rule 16} \end{array} \right]$
- + $\begin{bmatrix} \text{STEM} \\ + \text{rule 16} \end{bmatrix} \begin{bmatrix} - \text{syll} \\ - \text{nasal} \\ + \text{lateral} \end{bmatrix}$

Rule 15 states that the final nasal consonant of the prefix /pəN/ does not become homorganic with, or assimilated to, the initial consonant of the stem (that is, has the feature [- rule 13]) if the stem begins with /l/ and has the feature [+ rule 16]. Instead the prefix undergoes rule 16 (that is, acquires the feature [+ rule 16]), which states that /pəN/ is realized as [pəŋ] when followed by stems beginning with /l/ and having the feature [+ rule 16]. The two rules imply that the exceptional behavior of /pəN/ only involves a very limited number of stems beginning with a lateral liquid.

In order to generate the correct forms, in their lexical entries the stems of the words [pəŋlihatan] *sight*, *perception*, [pəŋlihat] *vision* and [pəŋlipor] *comfort*, *diversion* are specified [+ rule 16]. The stem of the word [pəlihat] *seer* and that of [pəlipor] *comfort*, *diversion* (for speakers who have the latter form) do not have the feature [+ rule 16].

5. In conclusion, some remarks must be made regarding the ordering of the rules. As has been demonstrated above, the vast majority of cases involving the two Malay prefixes with a final nasal consonant can be accounted for by a set of four rules, namely rules 8, 13, 7 and 9, in that order. Rule 14, for those speakers that have it, must apply prior to the above set of rules. Since rules 15 and 16 account for exceptions, they also must precede the same set. In other words, the ordering of the rules is: (i) rule 15, (ii) rule 16, (iii) rule 14, (iv) rule 8, (v) rule 13, (vi) rule 7 and (vii) rule 9. Rules (i) - (vii), in addition to the minus rule feature discussed at the end of section 2.3, account for all the changes involving the prefixes /məN/ and /pəN/ in Malay.

N O T E S

1. Malay refers here to Standard Malay, that is, the variety of Malay which is used in formal gatherings, on radio and television, and taught in schools in Malaysia and Singapore. For a description of the sounds of Standard Malay, see Maris 1966.
2. The addition of a suffix (in this case /an/) will in no way affect the arguments set forth in this paper.
3. Plus suffix /kan/.
4. The non-lateral liquid /r/ is realized as \emptyset in word final position by many speakers in Malaysia and Singapore (Maris 1966: 126 note 1).
5. Plus suffix /kan/.
6. Plus suffix /kan/.
7. The underlying form is /k \acute{r} a/; /a/ is realized as [ə] in certain positions by a rule which is not relevant to the subject of this paper.
8. The final /k/ of this recent borrowing is not realized as [ʔ] as in the following example. See note 9.
9. The underlying form is /g \acute{o} sok/; morpheme final /k/ is realized as [ʔ] by a rule not relevant to the discussion.
10. Plus suffix /an/.
11. Spelled veto. Speakers who have not been exposed to English replace /v/ by /f/.

12. Spelled *thabit*. Speakers not familiar with Arabic replace /θ/ by /s/. The underlying form is /θabit/; the opposition between /i/ and /e/ is neutralized in closed syllables in morpheme final position where only [e] occurs, except in unassimilated borrowed forms such as [kritik] (see example 5 above).

13. Spelled *dharab* or *dzarab*. A foreign sound borrowed from Arabic, /ð/ is replaced with /d/ by some speakers. Since in native Malay words the opposition between voiced and unvoiced for stops (in addition to the fact that there is only one native fricative in Malay: unvoiced /s/) is neutralized in syllabic final position (where only the unvoiced variety occurs), the final /b/ is often replaced by /p/. No affricates occur in syllabic final position.

14. Plus suffix /kan/.

15. Plus suffix /an/.

16. Plus suffix /i/.

17. Plus suffix /kan/. The word is spelled *mensyarahkan*.

18. Plus suffix /i/. The stem is spelled *khianat*. Some speakers substitute /k/ for /x/, in which case the form is [məŋkianati].

19. Plus suffix /kan/. For [b] in the stem, see note 13. The stem is spelled *ghaib*. With many speakers /ɣ/ is replaced by /g/.

20. Spelled *mencuri*.

21. Spelled *menjilat*.

22. For the final [ʔ] in the stem, see note 9.

23. See note 7.

24. See note 7.

25. Plus suffix /an/.

26. Plus suffix /kan/. The stem begins with /ʔ/ (the noncontinuant laryngeal glide), which is substituted by Malays for Arabic /ʕ/ (the

voiced pharyngeal fricative). It must, therefore, be represented in the underlying form. In the speech of many speakers any vowel in word initial position is realized with [ʔ] preceding it (Maris 1966: 102 note 1), but in the latter case [ʔ] is non-phonemic and need not be represented in the underlying form.

27. Plus suffix /kan/. For [e] in the stem, see note 12.

28. Plus suffix /kan/. For [ə] in the stem, see note 7.

29. Plus suffix /i/. When preceding suffixes beginning with a vowel, [r] must be pronounced. Cf. note 4.

30. For the final [ə] in the stem, see note 7.

31. For the final [ʔ] in the stem, see note 9.

32. Plus suffix /kan/.

33. Plus suffix /kan/.

34. Plus suffix /kan/.

35. Except for the feature 'vocalic', which has been replaced by 'syllabic'. For a different description of the sounds of Malay in terms of the Jakobsonian distinctive features, see Abas 1971:131.

36. For convenience a sequence of unit symbols is used here to represent the stem morpheme, but it should be kept in mind that all such sequences in this paper should be interpreted as bundles of specified features.

37. As far as I know *pemerhati* *observer* (see column 3 of Table 2) is the only word in Malay in which /pəN/ is followed by /pər/. In Indonesian there is the word *pemersatu unifier* (= /pəN/ + /pər/ + *satu one*).

38. For the origin of loan words, see Winstedt 1963.

39. The word - and morpheme boundaries (# and + respectively) are eliminated by a general rule which is irrelevant to the discussion.

40. Another example is: /tər/ + /rasə/ *feel* → #tər + rasə # → # tə + rasə # *felt*.

41. The deletion of [r] in the prefix also takes place in such forms as [bəkərjə] which results from the application of another rule which states that /bər/ is reduced to /bə/ before stems of which the first syllable is of the shape /Cər/ (where C stands for 'consonant'). Since this rule applies to other prefixes ending in /r/ as well, it can be put in the following form:

$$\left[\begin{array}{l} - \text{syll} \\ - \text{nasal} \\ - \text{lateral} \end{array} \right] \longrightarrow \emptyset / C_i \left[\begin{array}{l} + \text{syll} \\ - \text{high} \\ - \text{low} \\ + \text{back} \\ - \text{round} \end{array} \right] \text{---} + C_j \left[\begin{array}{l} + \text{syll} \\ - \text{high} \\ - \text{low} \\ + \text{back} \\ - \text{round} \end{array} \right] \left[\begin{array}{l} - \text{syll} \\ - \text{nasal} \\ - \text{lateral} \end{array} \right] C_k$$

The rule accounts for such forms as: /bər/ + /kərjə/ *work* → # bə + kərjə # *to work*; /pər/ + /kərjə/ → # pə + kərjə # *worker*; /pər/ + /sərtə/ *along with* → # pə + sərtə # *participant*, etc. The addition of C_k in the rule is to prevent it from applying to such stems as /kəras/ *hard*, where /r/ belongs to the second syllable of the stem. The prefixing of /bər/ to /kəras/ results, regularly, in # bər + kəras # *to be obstinate*. The deletion of /r/ expressed by the above rule is due to the fact that in Malay $C_i \text{ər} C_j \text{ər}$ is not a preferred syllable sequence.

BIBLIOGRAPHY

ABAS, Lufti

- 1971 *Linguistik deskriptif dan nahu bahasa Melayu*. Kuala Lumpur.
Dewan Bahasa dan Pustaka.

CHOMSKY, Noam and Morris Halle

- 1968 *The sound patterns of English*. New York. Harper and Row.

HARMS, Robert T.

- 1968 *Introduction to phonological theory*. Englewood Cliffs,
New Jersey. Prentice-Hall.

MARIS, Yunus

- 1966 *The Malay sound system*. Mimeographed. Kuala Lumpur.
University of Malaya.

SCHANE, Sanford A.

- 1973 *Generative Phonology*. Englewood Cliffs, New Jersey.
Prentice-Hall.

WINSTEDT, R.O.

- 1963 *An unabridged Malay-English dictionary*, fifth edition
enlarged. Kuala Lumpur. Marican and Sons.

A SUBGROUPING OF 100 PHILIPPINE LANGUAGES

Teodoro A. Llamzon and Ma. Teresita Martin

- 1.0 Introduction
- 2.0 The Languages of the Study Group
- 3.0 Inventory of ESI's
- 4.0 Subgrouping Hypothesis
- 5.0 Glottochronology
- 6.0 Conclusion

1.0 INTRODUCTION

The subgrouping of the so-called 'Philippine Languages' dates back to at least Wilhelm R. von Humboldt (1836), Wilhelm P. Schmidt (1926), Hendrik Kern (1887) and Otto Dempwolff (1934). These men attempted to relate these languages to the austronesian language family and cited one or two of them as examples of the type of languages which belonged to this subgroup. A much larger number (59) of these languages were included by Isidore Dyen (1965) in his lexicostatistical classification of some 371 speech communities belonging to the Austronesian language family.

Within the Philippine subfamily itself, efforts to classify the various daughter languages were made preliminarily by Harold C. Conklin (1952, 1955). These were then followed by two lexicostatistical classifications: one by Robert B. Fox, Willis E. Sibley and Fred Eggan (1954), and the other by David Thomas and Alan Healey (1962). Douglas Chretien (1951) made a classification of 21 Philippine languages according to the distribution of 1,904 morphemes; and Teodoro A. Llamzon (1969) subgrouped nine of these languages according to their exclusively shared innovations.

Dyen's study identified two large subgroups within the Philippine

language family: the 'Sulic Hesion', which included 28 speech communities together with five major languages (Tagalog, Cebuano, Ilonggo, Bikol, and Pampango), and the 'Cordilleran Hesion', which included all the other languages (together with Ilocano) except nine (Maranao, Casiguran, Yakan, Baler, Tiruray, Dusun, Murutic Subfamily, Bilic Subfamily, and Ivatan), which stood in coordinate relationship with the two groups mentioned above. He left Ilonggot unclassified.

Thomas and Healey's work posited an early (ca. 700 B.C.) triple split of the 'Philippine Stock' into a 'Northern Philippine Family', a 'Southern Philippine Family', and 'Pangasinan'. These three groups together lay in coordinate relationship with Ivatan, Ilonggot, and Baler Dumagat, and formed the 'Philippine Superstock', which broke away from the 'Southern Mindanao Family' and the 'Malay Stock' around 1300 B.C.

Chretien's method of classifying 21 Philippine languages on the basis of the distribution of 1,904 morphemes yielded three divisions: the 'Luzon Sequence', the 'Macro-Bisayan Group', and the 'Mindanao-Sulu Group'. This last group was linked to the Macro-Bisayan Group by Tausug, while the Macro-Bisayan Group was linked to the Luzon Sequence by Bicol.

This study of 100 Philippine languages is, I hope, another effort towards completing the subgrouping of all the Philippine languages. It differs from the above mentioned studies in two respects, namely: in the number of languages included in the study group, and in the method used to classify these languages.

The classification of the present set of 100 languages is based on their exclusively shared innovations (ESI) as the principles which underlie the method were explained by Brugmann (1884). The details of the procedures for determining probable ESI's and setting up a Stammbaum on the basis of these ESI's were elaborated by Llamzon (1969). The use of glotto-chronology, as suggested by George W. Grace (1964), has been restricted only to the calculation of time depth of separation between the languages of the study group.

2.0 THE LANGUAGES OF THE STUDY GROUP

This subgrouping includes 100 Philippine languages, the data for which were obtained from three main sources: Lawrence A. Reid's PHILIPPINE MINOR LANGUAGES (PML), the files of the Summer Institute of Linguistics in the Philippines (SIL), and the files of the Ateneo de Manila University Language Centre (ALC).¹ The following list gives the number assigned to each of these 100 languages, their names, location, and the sources from which data on them were obtained:

NO.	LANGUAGE	LOCATION	S O U R C E S		
			Phonology	Morphology	Syntax
1.	Agta	Central Cagayan Valley	PML	PML	SIL
2.	Atta	Pamplona, Cagayan	PML	PML	SIL
3.	Balangaw	Botak, Natonin, Mt. Province	PML	PML	SIL
4.	Batak	Tagnipa, Palawan Is.	PML	PML	ALC
5.	Bilaan (Koronodal)	Koronodal Valley, Cotabato	PML	PML	SIL
6.	Bilaan (Sarangani)	Southern Davao	PML	PML	ALC
7.	Binukid	Northern Bukidnon	PML	PML	ALC
8.	Bontoc	Bontoc, Mt. Province	PML	PML	ALC
9.	Dmagat	Casiguran, Quezon	PML	PML	SIL
10.	Gaddang	Butigui, Mt. Province	PML	PML	SIL
11.	Amganad (Ifugao)	Amganad, Ifugao	PML	PML	SIL
12.	Batad (Ifugao)	Batad, Ifugao	PML	PML	SIL
13.	Bayninan (Ifugao)	Bayninan, Ifugao	PML	PML	ALC
14.	Ilonggot	Kakidugen, Nueva Vizcaya	PML	PML	ALC
15.	Inibaloi	Kabayan, Benguet	PML	PML	SIL
16.	Isneg	Kabugao, Kalinga-Apayao	PML	PML	SIL
17.	Itbayaten	Itbayat Is., Batanes	PML	PML	ALC
18.	Itneg	Ba'ay Valley, Abra	PML	PML	SIL
19.	Ivatan	Basco, Batanes	PML	PML	SIL
20.	Kalagan	Digos, S. Davao	PML	PML	SIL
21.	Kalinga	Guinaang, Kalinga-Apayao	PML	PML	SIL
22.	Kallahan (Kayapa)	Kayapa, Nueva Vizcaya	PML	PML	ALC
23.	Kallahan (Keleyqiq)	Antipolo, Ifugao	PML	PML	SIL
24.	Kankanay	Sagada, Mt. Province	PML	PML	SIL
25.	Mamanua	Santiago, Agusan	PML	PML	ALC
26.	Ata (Manobo)	Mansalinao, N. Davao	PML	PML	SIL
27.	Dibabawon (Manobo)	Asuncion, N. Cavao	PML	PML	SIL
28.	Ilianen (Manobo)	Kibudtungan, Cotabato	PML	PML	SIL
29.	Kalamansig (Manobo)	Kalamansig, Cotabato	PML	PML	SIL
30.	Sarangani (Manobo)	Jose Abad Santos, Cotabato	PML	PML	SIL
31.	Tigwa (Manobo)	San Fernando, Bukidnon	PML	PML	SIL
32.	Western Bukidnon (Manobo)	West Bukidnon	PML	PML	SIL
33.	Mansala	Pantukan, N. Davao	PML	PML	SIL
34.	Siasi	Siasi, Sulu	PML	PML	SIL
35.	Sambal	Botolan, Zambales	PML	PML	SIL
36.	Sangil	Sarangani Is., Davao	PML	PML	ALC
37.	Sangir	Tabukang, Davao	PML	PML	SIL
38.	Sindangan (Subanon)	Sindangan, N. Zamboanga	PML	PML	SIL
39.	Siocon (Subanon)	Siocon S. Zamboanga	PML	PML	SIL

NO.	LANGUAGE	LOCATION	S O U R C E S		
			Phonology	Morphology	Syntax
40.	T'boli	Eduards, Cotabato	PML	PML	SIL
41.	Aborlan (Tagbanwa)	Aborlan, Palawan	PML	PML	ALC
42.	Kalamian (Tagbanwa)	Kalamian Is., Palawan	PML	PML	SIL
43.	Tausug	Jolo, Sulu	PML	PML	ALC
44.	Tagalog	Manila City	ALC	ALC	ALC
45.	Cebuano	Cebu City	ALC	ALC	ALC
46.	Hiligaynon	Iloilo City	ALC	ALC	ALC
47.	Waray	Catbalogan, Samar	ALC	ALC	ALC
48.	Ilocano	San Fernando, LaUnion	ALC	ALC	ALC
49.	Bicol	Naga, Camarines Sur	ALC	ALC	ALC
50.	Pampango	San Fernando City	ALC	ALC	ALC
51.	Pangasinan	Dagupan City	ALC	ALC	ALC
52.	Tagakaolo	Malita, S. Davao	ALC	ALC	ALC
53.	Dabaweano	Davao City	ALC	ALC	ALC
54.	Isamal	Samal Is., Davao	ALC	ALC	ALC
55.	Bagobo (Tagabawaq)	Kidapawan, Cotabato	ALC	ALC	ALC
56.	Yakan	Lamitan, Basilan Is.	ALC	ALC	ALC
57.	Simunul	Bongao, Sulu	ALC	ALC	ALC
58.	Sibutu	Tawi-Tawi, Sulu	ALC	ALC	ALC
59.	Kapul	Kapul Is.	ALC	ALC	ALC
60.	Palun Mapun	Cagayan de Sulu Is.	ALC	ALC	ALC
61.	Maranao	Marawi City	ALC	ALC	ALC
62.	Magindanaon	Cotabato City	ALC	ALC	ALC
63.	Tasaday	Tasaday, Cotabato	ALC	ALC	ALC
64.	Kiniray'a	Sta Barbara, Iloilo	ALC	ALC	ALC
65.	Masbateño	Masbate Is.	ALC	ALC	ALC
66.	Sorsogonon	Sorsogon, Sorsogon	ALC	ALC	ALC
67.	Cuyunen	Cuyu Is., Palawan	ALC	ALC	ALC
68.	Butuanon	Butuan, Agusan	ALC	ALC	ALC
69.	Hanunoo	Mansalay, Or. Mindoro	ALC	ALC	ALC
70.	Isinay	Dupax, Neuva Vizcaya	ALC	ALC	ALC
71.	Itawes	Enrile, Cagayan	ALC	ALC	ALC
72.	Ibanag	Tuguegarao, Cagayan	ALC	ALC	ALC
73.	Yogad	Echague, Isabela	ALC	ALC	ALC
74.	Malaweg	Rizal, Cagayan	ALC	ALC	ALC
75.	Aklanon	Kalibo, Aklan	ALC	ALC	ALC
76.	Capiznon	Roxas City	ALC	ALC	ALC
77.	Cagayancillo	Cagayan Is.	ALC	ALC	ALC
78.	Inunhan	Looc, Tablas Is.	ALC	ALC	ALC

NO.	LANGUAGES	LOCATION	S O U R C E S		
			Phonology	Morphology	Syntax
79.	Romblonon	Romblon, Romblon	ALC	ALC	ALC
80.	Hamtikon	San Jose, Antique	ALC	ALC	ALC
81.	Higaonon	Pusilaw, Agusan	ALC	ALC	ALC
82.	Alangan	Baco, Or. Mindoro	ALC	ALC	ALC
83.	Datagnon	Bulalakaw, Or. Mindoro	ALC	ALC	ALC
84.	Bulalakaw	Bulalakaw, Or. Mindoro	ALC	ALC	ALC
85.	Iraya (Mangyan)	San Teodoro, Or. Mindoro	ALC	ALC	ALC
86.	Tagaydan	Paitan, Or. Mindoro	ALC	ALC	ALC
87.	Buhid	Bongabon, Or. Mindoro	ALC	ALC	ALC
88.	Tadyawan	Victoria, Or. Mindoro	ALC	ALC	ALC
89.	Tiruray	Upi, Cotabato	ALC	ALC	ALC
90.	Kamayu	Tagu, N. Davao	ALC	ALC	ALC
91.	Mandaya	Mati, Or. Davao	ALC	ALC	ALC
92.	Agutaynon	Agutaya Is., Palawan	ALC	ALC	ALC
93.	Iraya (Cagayan)	Cabagan, Isabela	ALC	ALC	ALC
94.	Talaud	N. Celebes	ALC	ALC	ALC
95.	Bol'anon	Talibon, Bohol	ALC	ALC	ALC
96.	Bantoanon	Banton Is.	ALC	ALC	ALC
97.	Yami	Botel Tobago Is., RC	ALC	ALC	ALC
98.	Zamboangeño	Zamboanga City	ALC	ALC	ALC
99.	Caviteño	Cavite City	ALC	ALC	ALC
100.	Jaun-Jaun	Surigao del Norte	ALC	ALC	ALC
101.	Palawano	Quezon, Palawan	ALC	ALC	ALC

The control languages used in this study were primarily those included in Otto Dempwolff's work (1934). Data on the pronominal forms of the Formosan languages, however, found at the ALC, as well as those of the Micronesian languages in A. Thalheimer (1908) were also used. Finally, the languages included in J.A.L. Brandes (1884) and H. Kern (1887) were also employed.

3.0 INVENTORY OF ESI'S

The ESI's on which the subgrouping hypothesis in 4.0 is based are the following:

I - PHONOLOGICAL

These ESI's consist of vowels, Diphthongs, Labials, Apicals and Dorsals. The laryngeals have not been included.

1. Vowels:

a. *e, u > u:	Cebuano	Hiligaynon	Tausug
	Waray	Kiniray'a	Capiznon
	Aklanon	Masbateño	Dabaweño
	Isamal	Siasi	Simunul
	Sibutu	Palun Mapun	Higaonon
	Hanunoo	Datagnon	Romblonon
	Buhid	Tagaydan	Bulalakaw
	Jaun-Jaun	Kamayu	Sorsogonon

2. Diphthongs:

a. *ay, uy > uy:	Ifugao (Amganad)	Ifugao (Batad)
	Ifugao (Bayninan)	Kallahan (Keleyqiq)
b. *aw, ew > ew:	Manobo (Dibabawon)	Manobo (Ilianen)
	Manobo (Kalamansig)	Yakan
c. *ew, iw > ew:	Balangaw	Bontoc
	Kallahan (Keleyqiq)	Inibaloi
	Kallahan (Kayapa)	

3. Consonants:

a. *l, -r > l:	Itawes	Malaweg
b. *r-, -r, l-, -l-:	Masbateño	Hanunoo
c. *r, l > l:	Bulalakaw	Datagnon
d. *l, -D- > l:	Mamanua	Manobo (Ata)
	Tasaday	Manobo (Dibabawon)
	Manobo (Tigwa)	Manobo (Sarangani)
e. *l, r-, -r-, -D- > l:	Inunhan	Butuanon
	Cagayancillo	
f. *l, -j-, -D > l:	Subanon (Siocon)	Subanon (Sindangan)
	Mansaka	
g. *l-, -l-, R-, -R, -R ₁ , -R ₁ -, R ₂ -, -R ₂ > l:	Itneg	Inibaloi
h. *l, r, R, R ₁ , R ₂ -, -R ₂ -, -D- > l:	Bontoc	Ifugao (Batad)
		Ifugao (Amganad)
i. *r-, -j-, R, R ₁ , R ₂ -, -R ₂ -, l, D-, -D-, -Z- > l:	Bilaan (Koronodal)	Bilaan (Sarangani)
j. *r, l, -j-, -Z-, -z-, -D- > l:	Tagalog	Cebuano
		Hiligaynon

- k. *R, R₁, R₂, -R₃-, -R₃, g-, -g- > g:
- | | | |
|------------|--------------|-----------|
| Cuyunen | Aklanon | Jaun-Jaun |
| Kamayu | Bulalakaw | Inunhan |
| Kiniray'a | Capiznon | Higaonon |
| Masbateño | Butuanon | Datagnon |
| Sorsogonon | Cagayancillo | |
- l. *R, R₁, R₂, g > g:
- | | | |
|--------------------|--------|------------|
| Tagbanwa (Aborlan) | Tausug | Waray |
| Cebuano | Bicol | Hiligaynon |
- m. *j, R₂, -R₁-, -R₁, -R-, -R > g:
- | | | |
|------|------|--|
| Agta | Atta | |
|------|------|--|
- n. *-R, -R₂, -j-, -j > g:
- | | | |
|---------|--------|--|
| Kalinga | Bontoc | |
|---------|--------|--|
- o. *-R₂, -j-, -j > g:
- | | | |
|---------------------|-------------------|------------------|
| Kallahan (Keleyqiq) | Ifugao (Batad) | |
| Ifugao (Bayninan) | Kallahan (Kayapa) | Ifugao (Amganad) |
- p. *-R-, -R, R₁, -R₂-, -R₂, R₃, g > g:
- | | | |
|---------|-------|--|
| Hanunoo | Buhid | |
|---------|-------|--|
- q. *-j-, -j, R, R₁, R₂, R₃-, g-, -g- > g:
- | | | |
|--------|-------|--|
| Ibanag | Yogad | |
|--------|-------|--|
- r. *R-, -R-, -R, R₂, R₃, g-, -g- > g:
- | | | |
|-------------|-------|--|
| Magindanaon | Kapul | |
|-------------|-------|--|
- s. *-R, R₁-, -R₁-, -g-, g- > g:
- | | | |
|---------|-------------|--|
| Simunul | Palun Mapun | |
|---------|-------------|--|
- t. *R, R₁, R₂, -R₃-, -R₃, g > g:
- | | | |
|-----------|----------|--------|
| Tagakaolo | Dabaweño | Isamal |
|-----------|----------|--------|
- u. *d, D-, -D, z-, Z- > d:
- | | | |
|------------|---------------------|----------|
| Kalagan | Batak | Pampango |
| Pangasinan | Subanon (Sindangan) | Mansaka |
| Sangil | Subanon (Siocon) | Tausug |
- v. *d-, -d, D-, -D, z-, Z- > d:
- | | | |
|------------------|---------------------------|--|
| Tagalog | Cebuano | |
| Manobo (Ilianen) | Manobo (Western Bukidnon) | |
- w. *d, D-, -D, z-, Z- > d:
- | | | |
|------------|-------|-------|
| Haligaynon | Waray | Bicol |
|------------|-------|-------|
- x. *Z-, D-, z- > d:
- | | | |
|--------|--------|--|
| Itawes | Ibanag | |
|--------|--------|--|
- y. *d, D-, -D, Z, z > d:
- | | | |
|----------------|---------------------|--------------------|
| Manobo (Ata) | Manobo (Dibabawon) | |
| Manobo (Tigwa) | Manobo (Kalamansig) | Manobo (Sarangani) |

- z. *d, D, z, Z > d: Kallahan (Keleyqiq) Bontoc
 Balangaw Kallahan (Kayapa) Kankanay
 Isneg Ifugao (Bayninan) Ifugao (Amganad)
 Ifugao (Batad) Itneg
- aa. *d-, -d, D-, -D, Z-, z-, j- > d:
 Kiniray'a Masbateño Sorsogonon
- bb. *-d-, -d, D-, -D > d: Itbayaten Ivatan
- cc. *d, -D, -z, Z- > d: Bilaan (Koronodal) Bilaan (Sarangani)
- dd. *d, D, Z-, z-, j-, -j, > d:
 Tagakaolo Aklanon Cagayancillo
 Dabaweño Capiznon Inunhan
 Isamal Romblonon
- ee. *d, D-, -D, Z-, z-, j-, -j > d:
 Butuanon Hanunoo Datagnon
 Bulalakaw Mandaya
- ff. *-y-, -R-, -R₁-, -R₂- > y:
 Itbayaten Sambal Iraya (Mangyan)
 Ivatan Pampango
- gg. *Metathesis of s and t in a syllable:
 Ifugao (Amganad) Bontoc Isneg
 Ifugao (Bayninan) Kankanay Gaddang
 Ifugao (Batad) Kalinga Ibanag
 Kallahan (Keleyqiq) Malaweg Itawes
 Itneg Ilocano Yogad
 Sangil Balangaw Atta
 Sangir Agta

II - PERSONAL PRONOUNS

1. *^caku^c (Demp.), ?akú? (Dyen) I
- a.
- b. *yaken:² Ivatan Itbayaten Yami
- c. *siak:
 Agta Manobo (Dibabawon)
 Sangir Binukid
 Manobo (Tigwa) Ilocano
 Ilonggot Manobo (Western Bukidnon)
 Pangasinan Isneg
 Sangil Magindanaon
 Higaonon

- d. *sikaak: Inibaloi Kallahan (Keleyqiq) Kallahan (Kayapa)
- e. *sikanak: Gaddang Iraya (Cagayan)
- f. *aken: Manobo (Kalamansig) Subanon (Siocon) Tasaday
- g. *siaken: Balangaw Ifugao (Amganad) Kalinga
 Kankanay Ifugao (Batad) Manobo (Ilianen)
 Bagobo Ifugao (Bayninan) Ibanag
 Yogad Maranao Magindanaon
 Bontoc
2. *kav (Demp.), kaw (Dyen) you (Sing.)
- a. *sikaw: Atta Manobo (Western Bukidnon)
 Binukid Manobo (Tigwa)
 Manobo (Ata) Ibanag
 Higaonon
- b. *ikaw: Tagalog Palawano Cebuano Isamal
 Kiniray'a Kalagan Aklanon Sorsogonon
 Datagnon Mansaka Itawes Tagakaolo
 Capiznon Haligaynon Malaweg Jaun-Jaun
 Cuyunen Waray Butuanon Hamtikon
 Kamayu Sangil Masbateño Romblonon
 Mandaya Sangir Dabaweño Inunhan
- c. *sika: Balangaw Bilaan (Koronodal) Bontoc
 Kankanay Bilaan (Sarangani) Pampango
 Isinay Gaddang Ifugao (Amganad)
 Ilonggot Bicol Ifugao (Batad)
 Yogad Tagbanwa (Aborlan) Ifugao (Bayninan)
 Ilocano Pangasinan Kalinga
 Sambal Subanon (Siocon) Maranao
 Magindanaon
- d. *sikamu: Inibaloi Kallahan (Keleyqiq) Kallahan (Kayapa)
- e. *siku: Mamanua Dumagat (Casiguran) Agta
- f. *sikuna: Manobo (Ilianen) Manobo (Sarangani)
 Manobo (Kalamansig) Bagobo
- g. *kaqaw: Kapul Palun Mapun Sibutu
 Siasi Yakan Cagayancillo
- h. *kawu: Hanunoo Iraya (Mangyan)

- b. *qikami: Agta Sangir Gaddang
Pampango Itawes Sangil
Malaweg
- c. *dikami: Itneg Isneg Ifugao (Amganad)
Kalinga Kankanay Ifugao (Batad)
Bontoc Ilocano Ifugao (Bayninan)
Isinay
- d. *yamen: Itbayaten Ivatan Yami
Tagbanwa (Kalamian) Iraya (Mangyan)
5. *kita^c (Demp.), kítà? (Dyen) *we* (inclusive)
- a. *sikita: Manobo (Ilianen) Sangir Yogad
Manobo (Ata) Inibaloi Higaonon
Manobo (Tigwa) Bagobo
- b. *sikitayu: Kallahan (Keleyqiq) Kallahan (Kayapa) Pangasinan
- c. *sikitam: Agta Dumagat (Casiguran)
- d. *ditayu: Ilocano Itneg
- e. *ditaku: Ifugao (Amganad) Bontoc Balangaw
Ifugao (Bayninan) Kankanay Isinay
Ifugao (Batad) Kalinga
- f. *sikitadun: Kalagan Tagakaolo Manobo (Sarangani)
- g. *sitamu: Sambal Pampango Ibanag
Iraya (Cagayan)
- h. *kitabí: Palun Mapun Yakan Simunul Siasi
- i. *sikitamu: Maranao Magindanaon
- j. *kitam: Sibutu Kapul Alangan
Siasi Tagaydan
6. *kamu^c (Demp.), kamú? (Dyen) *you* (plural)
- a. *sikamu: Bilaan (Koronodal) Agta Yogad
Bilaan (Sarangani)
- b. *sikam: Dumagat (Casiguran) Iraya (Cagayan) Yogad
- c. *sikayu: Atta Pampango
Inibaloi Pangasinan
Gaddang Manobo (Ilianen)
Kallahan (Keleyqiq) Manobo (Ata)
Manobo (Tigwa) Kallahan (Kayapa)
Manobo (Western Bukidnon) Manobo (Dibabawon)
Manobo (Sarangani)

- d. *ikayu: Malaweg Itawes
- e. *dikayu: Itneg Ilocano
Isneg Kalinga
Bontoc Isinay
Ifugao (Bayninan) Ifugao (Amganad)
Ifugao (Batad) Kankanay
- f. *dikayu: Kapul Palun Mapun Sibutu
Siasi Yakan Simunul
7. *t'ida^c (Demp.) *they*
- a. *isiDa: Sangil Sangir
- b. *sikaDa: Kallahan (Keleyqiq) Inibaloi Pangasinan
Kallahan (Kayapa) Iraya (Mangyan)
- c. *sikanDa: Manobo (Ata) Manobo (Tigwa)
Manobo (Western Bukidnon) Manobo (Ilianen)
Manobo (Dibabawon) Manobo (Sarangani)
- d. *kagDa: Manobo (Kalamansig) Tasaday
- e. *aggiDa: Isneg Atta Malaweg
- f. *diDa: Balangaw Bontoc Ifugao (Amganad)
Kankanay Itneg Ifugao (Bayninan)
Isinay Kalinga Ifugao (Batad)
- g. *siDan: Maranao Isamal Higaonon
Magindanaon Tagakaolo Mandaya
Sibutu Kamayu
- h. *sinDa: Bicol Romblonon Mastateño
Aklanon Kiniray'a Hamtikon
Inunhan

III - NUMERALS

1. *^cet'a^c (Demp.), ?esa?/?isa? (Dyen) *one*
- a. *takday: Agta Atta Ibanag
- b. *satu: Bilaan (Koronodal) T'boli Bilaan (Sarangani)
- c. *sebeka: Manobo (Ilianen) Manobo (Tigwa)
Manobo (Western Bukidnon)
- d. *sakey: Inibaloi Kallahan (Keleyqiq) Kallahan (Kayapa)
- e. *sala: Subanon (Siocon) Subanon (Sindangan)
Aklanon Bol'anon

- f. *usaiq: Alangan Tagaydan
- g. *issa: Bagobo Malaweg Siasi
Simunul
- h. *isara: Datagnon Bulalakaw Hamtikon
2. *duva^c (Demp.), DewS₃a? (Dyen) *two*
- a. *lewe: Bilaan (Koronodal) T'boli
Bilaan (Sarangani) Tiruray
- b. *DaDawe: Binukid Tagalog
Ivatan Ata
Manobo (Dibabawon) Manobo (Ilianen)
Manobo (Tigwa) Manobo (Western Bukidnon)
Sangil Sangir
Tagbanwa (Kalamian) Agutaynon
Bulalakaw Cagayancillo
Cuyunen Higaonon
Iraya (Mangyan) Kiniray'a
Hamtikon Datagnon
Talaud Yami
- c. *duwaiq: Alangan Tagaydan
- d. *duha: Itbayaten Tausug Bantoanon
Capiznon Cebuano Hiligaynon
Jaun-Jaun Mandaya Masbateño
Romblonon Waray
- e. *eduwaq: Dumagat (Casiguran) Gaddang Yogad
Pampango
3. *telu^c (Demp.), telu (Dyen) *three*
- a. *tatelu: Binukid Sambal Dabawefio
Ivatan Sangil Higaonon
Atta Sangir Hiligaynon
Manobo (Dibabawon) Aklan Iraya (Mangyan)
Manobo (Ilianen) Bantoanon Kiniray'a
Manobo (Tigwa) Bulalakaw Hamtikon
Cuyunen Manobo (Western Bukidnon)
Tagalog Datagnon Yami
- b. *tellu: Bagobo Agta Atta
Gaddang Isneg Kallahan (Keleyqiq)
Cagayancillo Siasi Kallahan (Kayapa)
Ibanag Ilocano Iraya (Cagayan)

Itawes	Itneg	Magindanaon
Maranao	Pangasinan	Kapul
Sibutu	Simunul	Yakan
Yogad		

4. *^cə(m)pat (Demp.), x₂epate (Dyen) *four*
- a. *heqepat: Atta Manobo (Tigwa) Manobo (Western Bukidnon)
Binukid
- b. *pat: Bilaan (Koronodal) Magindanaon Bogobo
Bilaan (Sarangani) Maranao
- c. *mpat: Palun Mapun Sibutu Siasi
Simunul T'boli
- d. *qapqat: Yogad Aklanon Ivatan
Agta Atta Cagayancillo
Gaddang Isneg Binukid
Itbayaten Malaweg Talaud
Bantoanon Ibanag Iraya (Cagayan)
Itawes
- e. *apat: Tagalog Bol'anon Hiligaynon
Bicol Bulalakaw Iraya (Mangyan)
Pangasinan Pampango Cuyunen
Kiniray'a Hanunoo Yami
- f. *ampat: Kapul Yakan
- g. *epat: Tasaday Tiruray
5. *lima^c (Demp.), lⁱma? (Dyen) *five*
- a. *lalima: Binukid Ivatan Atta
Manobo (Ilianen) Manobo (Western Bukidnon)
Manobo (Tigwa) Higaonon
- b. *limaq: Isneg Dumagat (Casiguran)
Kallahan (Keleyqiq) T'boli
Tagbanwa (Kalamian) Tausug
6. *^cənəm (Demp.), ?eneme (Dyen) *six*
- a. *qennem: Agta Inibaloi Yogad
Atta Ibanag Iraya (Cagayan)
Siasi Simunul Itawes
Yakan Malaweg

- b. *anem: Gaddang Hiligaynon Datagnon
 Ivatan Iraya (Mangyan) Tagalog
 Sambal Jaun-Jaun Tasaday
 Bulalakaw Kiniray'a Yami
 Cagayancillo Pangasinan Cuyunen
 Pampango
- c. *nem: Bilaan (Koronodal) Maranao Palun Mapun
 Bilaan (Sarangani) T'boli Sibutu
 Magindanaon
- d. *heqenem: Manobo (Western Bukidnon) Manobo (Tigwa)
 Manobo (Ata)
- e. *anqem: Itbayaten Alangan Bantoanon
7. *pitu^c (Demp.), pltu? (dyen) *seven*
- a. *pituq: Dumagat (Casiguran) Agutaynon
 Itbayaten Sibutu
 Kallahan (Keleyqiq) Simunul
 Siasi Yakan
 T'boli Palun Mapun
 Tagbanwa (Kalamian) Tausug
- b. *pittu: Isneg Bagobo
- c. *papitu: Ivatan Manobo (Ata)
8. *valu^c (Demp.), w₂alu? (Dyen) *eight*
- a. *waluq: Dumagat (Casiguran) Yakan
 Itbayaten Tausug
 Kallahan (Keleyqiq) T'boli
 Kankanay Tagbanwa (Kalamian)
 Siasi Agutaynon
 Kapul Sibutu
- b. *wawalu: Ivatan Manobo (Ata)
- c. *wau: Manobo (Dibabawon) Sangil Isamal
 Palun Mapun Yami
9. *t'iva^c (Demp.), s₂iwa? (Dyen) *nine*
- a. *siam: Bicol Ifugao (Amganad) Subanon (Sindangan)
 Buhid Ifugao (Batad) Ilocano
 Butuanon Ifugao (Bayninan) Iraya (Cagayan)
 Bulalakaw Inibaloi Iraya (Mangyan)
 Cagayancillo Isneg Isamal
 Capiznon Itbayaten Isinay

	Cebuano	Ivatan	Itawes
	Cuyunen	Kalagan	Jaun-Jaun
	Dabaweño	Itneg	Kiniray'a
	Hanunoo	Kallahan (Keleyqiq)	Hamtikon
	Higaonon	Kallahan (Kayapa)	Malaweg
	Hiligaynon	Kalinga	Kamaya
	Ibanag	Kankanay	Mansaka
	Agta	Mamanua	Masbateño
	Atta	Manobo (Tigwa)	Palawano
	Balangaw	Manobo (Sarangani)	Palun Mapun
	Batak	Manobo (Dibabawon)	Pangasinan
	Bilaan (Koronodal)	Manobo (Ata)	Pampango
	Bilaan (Sarangani)	Tagakaolo	Datagnon
	Binukid	Bontoc	Romblonon
	Dumagat (Casiguran)	Siasi	Kapul
	Simunul	Sambal	Sibutu
	Tadyawan	Sorsogonon	Tagalog
	Tagaydan	Waray	T'boli
	Tausug	Yami	Yogad
	Yakan		
b. *siyew:	Tiruray	Maranao	Magindanaon
	Tasaday	Sangil	Manobo (Ilianen)
	Manobo (Kalamansig)	Manobo (Western Bukidnon)	
	Talaud	Sangir	Bagobo
10. *pulu ^c (Demp.), puluqe (Dyen) <i>ten</i>			
a. *mapulu:	Agta	Itawes	Ibanag
	Atta	Sangil	Malaweg
	Bol'anon	Talaud	Iraya (Cagayan)
	Sambal		
b. *simpulu:	Balangaw	Kalinga	Ifugao (Amganad)
	Ifugao (Batad)		
c. *sampuluq:	Binukid	Bicol	Kallahan (Kayapa)
	Kalagan	Cagayancillo	Pangasinan
	Manobo (Ata)	Cuyunen	Mansaka
	Bantoanon	Kallahan (Keleyqiq)	Higaonon
	Yakan	Inbaloi	Mandaya
	Manobo (Sarangani)		
d. *sanpuluq:	Batak	Agutaynon	Tagbanwa (Kalamian)
	Tagbanwa (Aborlan)		

- e. *sepuluq: Manobo (Ilianen) Tasaday
 Bilaan (Sarangani) T'boli
 Manobo (Kalamansig) Bilaan (Koronodal)
 Subanon (Sindangan) Manobo (Western Bukidnon)
- f. *sampuq: Alangan Siasi Manobo (Dibabawon)
 Tausug Dabawefio Tagaydan
 Simunul Sambal Isamal
 Tagalog Iraya (Mangyan) Palawano
- g. *napuluq: Mamanua Aklanon Butuanon
 Capiznon Cebuano Bulalakaw
 Jaun-Jaun Hamtikon Masbatefio
 Datagnon Romblonon Sorsogonon
 Kamayu Waray
- h. *sapuluq: Dumagat (Casiguran) Ivatan Bagobo
 Buhid Gaddang Manobo (Tigwa)
 Magindanaon Tadyawan Itbayaten
 Subanon (Sindangan) Maranao Yogad
11. *yatut' (Demp.), R₂átús (Dyen) *hundred*
- a. *meRatus: Agta Iraya (Cagayan)
 Manobo (Kalamansig) Atta
 Maranao Magindanaon
 Bilaan (Koronodal) T'boli
 Talaud Itawes
 Malaweg Bilaan (Sarangani)
- b. *sanRatus: Kallahan (Keleyqiq) Aklanon Tausug
 Kalagan Hamtikon Palawano
 Datagnon Agutaynon Sorsogonon
 Tagbanwa (Kalamian)
- c. *sinRatus: Balangaw Kalinga Ifugao (Amganad)
- d. *Rasut: Bontoc Gaddang Ifugao (Batad)
 Isneg Kankanay Ifugao (Bayninan)
 Kalinga Sangir Kallahan (Keleyqiq)
 Sangil
- e. *sanaRasut: Itneg Ilocano
- f. *nanRatus: Manobo (Tigwa) Higaonon
- g. *esa ka Ratus: Dabawefio Hiligaynon Jaun-Jaun
 Kiniray'a Mandaya Romblonon
- h. *sa ka Ratus: Capiznon Cuyunen

1. *dahatus: Palun Mapun Sibutu Simunul
Siasi Yakan Kapul
12. *libu^c (Demp.), ribu (Dyen) *thousand*
- a. *meribu: Agta Bilaan (Koronodal) Iraya (Cagayan)
Atta Bilaan (Sarangani) Itawes
T'boli Malaweg
- b. *sinribu: Balangaw Ifugao (Bayninan)
Kalinga Ifugao (Amganad)
- c. *saṅribu: Dumagat (Casiguran) Pangasinan Datagnon
Bicol Alangan Sorsogonon
- d. *seṅibu: Manobo (Ilianen) Maranao Palawano
Manobo (Kalamansig) Magindanaon Tausug
Subanon (Sindangan) Subanon (Siocon) Tiruray
- e. *qumribuq: Ivatan Itbayaten
- f. *mararan: Kalagan Manobo (Sarangani)
Bagobo Mansaka
- g. *saribu: Gaddang Yogad
- h. *esa ka ribu: Gebuano Hiligaynon Butuanon
Capiznon Cuyunen Jaun-Jaun
Hanunoo Kamayu Kiniray'a
Romblonon
- i. *esa ḡ ribu: Bulalakaw Tagalog
- j. *daṅibu: Siasi Simunul Yakan

IV - SYNTACTIC²1. Ang + N_c (topic)³

- a. *ya + N_c (topic): Agta Itawes Iraya (Cagayan)
Malaweg Isamal Kalagan
Mamanua Sambal Tagakaolo
- b. *yaṅ + N_c (topic): Mandaya Mansaka Kalamian
- c. *nan + N_c (topic): Bontoc Ifugao (Amganad)
Kankanay Ifugao (Batad)
- d. *in + N_c (topic): Alangan Palawano Tausug
- e. *so + N_c (topic): Higaonon Inibaloi
Manobo (Ata) Maranao
Magindanaon Pangasinan
Subanon (Sindangan) Kallahan (Kayapa)
Kallahan (Keleyqiq)

- f. *sa + N_c (topic): Manobo (Kalamansig) Binukid
- g. *to + N_c (topic): Bagobo Batak
- h. *pag + N_c (topic): Iraya (Mangyan) Tadyawan
- i. *an + N_c (topic): Masbateño Sorsogonon Waray
- j. *in + N_c (topic): Ibanag Pampango
- k. *i + N_c (topic): Bilaan (Koronodal) Bilaan (Sarangani)
2. Si + N_p (topic)
- a. *i + N_p (topic): Dumagat (Casiguran) Tiruray Pampango
Sangir
3. Ng + N_c (possessor)
- a. *nan + N_c (possessor): Tagalog Dabaweño Cuyunen
Kamayu Mandaya Mansaka
- b. *na + N_c (possessor): Agta Isamal Mamanua
Atta Ibanag Tagakaolo
Gaddang
- c. *nan + N_c (possessor): Jaun-Jaun Bontoc Kankananay
Ifugao (Amganad)
- d. *it + N_c (possessor): Batak Aklanon Butuanon
Bulalakaw Datagnon Tagbanwa (Aborlan)
- e. *to + N_c (possessor): Manobo (Ata) Manobo (Dibabawon)
Manobo (Tigwa) Manobo (Ilianen)
Manobo (Western Bukidnon)
Manobo (Sarangani)
- f. *san + N_c (possessor): Capiznon Hiligaynon Kiniray'a
Masbateño Waray
- g. *ni + N_c (possessor): Dumagat (Casiguran) Kallahan (Keleyqiq)
Inibaloi Kallahan (Kayapa)
- h. *sa + N_c (possessor): Bol'anon Cebuano
- i. *ni + N_c (possessor): Pampango Romblonon
- j. *ya + N_c (possessor): Iraya (Cagayan) Malaweg
- k. *ta + N_c (possessor): Cagayancillo Tagbanwa (Kalamian)
- l. *nu + N_c (possessor): Maranao Magindanaon
Yogad Subanon (Sindangan)
Yami Ivatan

- m. *i + N_c (possessor): Bilaan (Sarangani) Bilaan (Koronodal)
- n. *del + N_c (possessor): Caviteño Zamboangeño
4. Ni + N_p (possessor)
- a. *si + N_p (possessor): Sibutu Siasi Simunul
Binukid Kankanay Palawano
Tausug
- b. *i + N_p (possessor): Manobo (Kalamansig) Itawes
Sangir Gaddang
5. Sa + N_c (location)
- a. *sa + N_c (location): Tagalog Dabaweño Magindanaon
Aklanon Tadyawan Tagaydan
Alangan Bicol Cebuano
Butuanon Isamal Bulalakaw
Kamayu Bol'anon Waray
Capiznon Hanunoo Masbateño
Hiligaynon Kiniray'a Sorsogonon
Hamtikon Maranao Tagakaolo
Tausug Datagnon Palawano
Romblonon Sambal Bantoanon
Iraya (Mangyan)
- b. *ta + N_c (location): Agta Bagobo Binukid
Atta Cagayancillo Iraya (Cagayan)
- c. *to + N_c (location): Manobo (Ata) Manobo (Dibabawon)
Manobo (Western Bukidnon) Manobo (Tigwa)
Manobo (Ilianen) Manobo (Sarangani)
Yogad
- d. *su + N_c (location): Sangir Talaud
- e. *tuŋ + N_c (location): Agutaynon Tagbanwa (Kalamian)
- f. *du + N_c (location): Yami Ivatan
- g. *kat + N_c (location): Tagbanwa (Aborlan) Batak
- h. *ad + N_c (location): Ifugao (Amganad) Bontoc
Pangasinan Balangaw
- i. *san + N_c (location): Mandaya Mansaka
- j. *si + N_c (location): Kapul Gaddang

6. Kay + Np (Indirect Object)

- a. *kay + Np (IO): Tagalog Aklanon Alangan
 Butuanon Bulalakaw Bol'anon
 Capiznon Cuyunen Hiligaynon
 Kiniray'a Hamtikon Datagnon
 Romblonon Palawano
- b. *tuŋ ni + Np (IO): Tagbanwa (Aborlan) Tagbanwa (Kalamian)
- c. *ki + Np (IO): Tagbanwa (Aborlan) Batak
 Bagobo Binukid
 Higaonon Bicol
 Manobo (Ata) Manobo (Western Bukidnon)
 Manobo (Ilianen) Manobo (Sarangani)
 Manobo (Dibabawon) Manobo (Tigwa)
 Maranao Bilaan (Sarangani)
- d. *kan + Np (IO): Itawes Hanunoo Sorsogonon
 Waray Mandaya Jaun-Jaun
 Tagakaolo Mamanua Masbateño
 Tausug
- e. *kani + Np (IO): Ilocano Magindanaon
- f. *kaŋ + Np (IO): Pampango Dabawēño Isamal
 Mansaka Butuanon Kamayu
 Bantoanon Cebuano Waray
- g. *ken + Np (IO): Bontoc Kankanay
- h. *kun + Np (IO): Caviteño Zamboangeño

7. Ang mga + N_c (Plural Topic)

- a. *aŋ maŋa + N_c (PT): Tagalog Kiniray'a Jaun-Jaun
 Bicol Hamtikon Kamayu
 Butuanon Datagnon Masbateño
 Cuyunen Sorsogonon Bulalakaw
 Waray Capiznon Cebuano
 Hiligaynon Bol'anon
- b. *ya maŋa + N_c (PT): Isamal Mamanua Kalagan
 Tagakaolo
- c. *yaŋ maŋa + N_c (PT): Tagbanwa (Kalamian) Dabawēño
- d. *su maŋa + N_c (PT): Higaonon Maranao
 Magindanaon Manobo (Sarangani)
- e. *in maŋa + N_c (PT): Palawano Alangan Tausug
 Tagaydan

- d. *A ay B (Mod): Bontoc Ifugao (Bayninan)
- e. *A pag B (Mod): Hanunoo Alangan Buhid
Tagaydan
- f. *A a B (Mod): Ilocano Ivatan Magindanaon
Yami Yogad Maranao
- g. *A ŋa B (Mod): Bagobo Kapul Butuanon
Cebuano Hiligaynon Ibanag
- h. *A no B (Mod): Manobo (Tigwa) Manobo (Dibabawon)
Manobo (Ilianen)
- i. *A ay B (Mod): Kankanay Bontoc Ifugao (Bayninan)
- j. *A ha B (Mod): Binukid Higaonon
- k. * an B (Mod): Ifugao (Amganad) Ifugao (Batad)
10. X ay Y, Y ϕ X (Predication)
- a. *X ay Y, Y X (Pred): Tagalog Agutaynon Aklanon
Alangan Bantoanon Bicol
Bulalakaw Cuyunen Kiniray'a
Hamtikon Datagnon Romblonon
Hanunoo
- b. *X e Y, YX (Pred): Ibanag Kallahan (Keleyqiq) Malaweg
- c. X Y, YX (Pred): Maranao Magindanaon

4.0 SUBGROUPING HYPOTHESIS

The Family Tree Diagram of 100 Philippine Languages (on folding page 165) presents the subgrouping hypothesis arrived at on the basis of the ESI's enumerated above: 4

An alternate hypothesis to the one given would be to link the 'Northern Philippine Group' (NP) to the 'Central Philippine Group' (CP) first, and then link the 'Southern Philippine Group' (SP) to the NP-CP. This would be supported by the following ESI's:

- a. NP-CP: II 5.g, 7.b, III 6.b, 6.c, 11.c, IV 3.c, 3.i;
- b. NP+CP-SP: I 3.ff, III 10.g, 10.f, IV 1.b, 1.g, 7.c, 7.g, 9.f.
- A second alternate hypothesis would be to link the NP to the SP first, and then link the CP to this NP + SP. This hypothesis would be supported by the following ESI's:
- a. NP-SP: I 3.hh, II 1.c, 1.g, 2.a, 2.e, 3.a, 4.a, 5.a, 6.a, 6.d, 8.a, III 1.g, 4.a, 5.a, 6.a, 7.b, 7.c, 11.a, 11.d, 12.a, IV 1.e, 2.a, 3.a, 3.b, 3.l, 4.b, 5.c, 5.j, 6.e;
- b. NP+SP-CP: II 5.g, 7.b, III 6.b, 6.c, 11.c, IV 3.c, 3.i.

The two alternate hypotheses given above, however, do not substantially alter the main lines of the hypothesis presented above, nor do they contradict the following interesting findings of this study: that

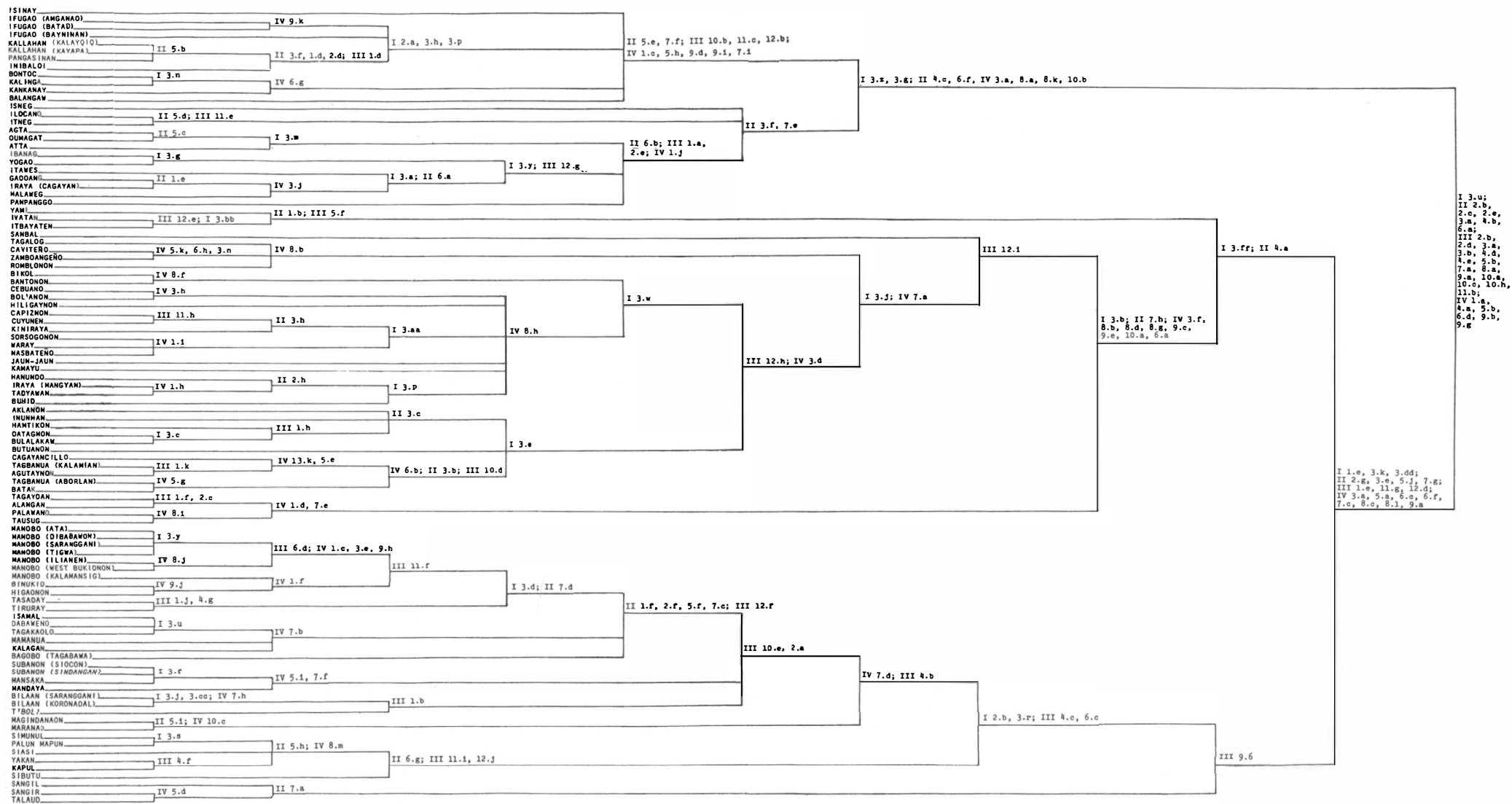
1. there is enough evidence for the genetic unity of the Philippine languages;
2. there are, apparently, three major subgroups within the Philippine subfamily, namely: the Northern Philippine, the Central Philippine, and the Southern Philippine groups;
3. the positions of most of these languages seem to correlate highly with their geographic location in the archipelago;
4. however, there are noteworthy exceptions to No. 3 above, namely; Kapul, Tausug, Ivatan, Itbayaten, and Yami;
5. the higher nodes in the family tree seem to have more ESI's to support them than the lower nodes, though there are some exceptions to this;
6. on the other hand, the ESI's on the lower nodes are shared by more languages in the subgroup than the ESI's on the higher nodes.

5.0 GLOTTOCHRONOLOGY

There remains the task of calculating the time depth of separation between the languages included in the study group. For this purpose, it is sufficient to present the percentages of cognates scored by each language with every other language.⁵ From these percentages, the formula $t = \frac{\log C}{r}$ can be used to obtain the time depth.

In preparing the charts below (on folding page 167), Reid's lists (1971) were used, which contain Swadesh's basic word list minus 'one member of frequently occurring doublets, such as husband/wife ... and relational words which have no unambiguous equivalents in Philippine languages, such as and, at, because, if and with. Words that generally have no indigenous equivalents such as freeze, snow, animal and correct. Altogether, 170 items were retained.' (viii). Several languages were not included in the charts (i.e. Yami, Zamboangeño, Caviteño, Jaun-Jaun, Bantoanon, Bol'anon, and Palawano). These remain to be done, except that Caviteño and Zamboangeño will not be included in the final tabulation, since work on both languages has shown that they turn out to be Indo-European languages on the basis of their basic vocabulary.

FAMILY TREE DIAGRAM OF 100 PHILIPPINE LANGUAGES



6.0 CONCLUSION

The study just concluded is one more step towards the complete classification of all the Philippine languages. Some 100 languages remain to be classified. Moreover, a comparison of the lexicons of these languages together with those of the 100 languages mentioned need to be done.

Only when this task is accomplished will reliable linguistic evidence for the cultural history and prehistorical contacts and migrations of the Filipino peoples be forthcoming.

N O T E S

1. Under the heading SIL and ALC are included a large number of individuals who have contributed data for this subgrouping project. Space limitations do not allow them to be mentioned individually. At any rate, their contributions are gratefully acknowledged in the files of both institutions.
2. The asterisked forms after the alphabetical subdivisions are not necessarily reconstructions. At this stage, they should be considered as merely convenient formulas for classifying the various ESI's.
3. In this section, the following symbols are used: N_c = common noun, N_p = personal noun (name), Re = reduplication, A' = words ending in consonants (except n , η or q), A'' = words ending in vowels, X = topic, Y = predicate. In subsections 9 and 10, there is reversing allowed, with only stylistic difference in meaning.
4. In the preparation of this hypothesis, I have profited from discussions with Patricia O. Afable of Yale University.
5. The languages in the two tables of cognate percentages above need to be further compared with each other, a task that awaits us.

BIBLIOGRAPHY

BRANDES, J.A.L.

- 1884 *Bijdrage tot de vergelijkende klankleer der westersche afdeeling van de Maleisch - Polynesische taal-familie.* Utrecht.

BRUGMANN, Karl

- 1884 'Zur Frage nach den Verwandtschaftsverhältnissen der indogermanischen Sprachen'. *Internationale Zeitschrift für Allgemeine Sprachwissenschaft.* 1.226-256.

CHRETIEN, Douglas C.

- 1951 'A classification of twenty-one Philippine languages'. *Philippine Journal of Science.* 91.485-506.

CONKLIN, Harold C.

- 1955 *Preliminary Linguistic survey of Mindanao.* Paper presented at Mindanao Conference, Chicago.
- 1952 *Outline Gazetteer of Native Philippine Ethnic and Linguistic Groups.* Mimeographed.

DEMPWOLFF, Otto

- 1934-38 *Vergleichende Lautlehre des austronesischen Wortschatzes.* I-III Band. ZES. 15, 17, 19. Berlin.

DYEN, ISIDORE A.

- 1965 'A Lexicostatistical Classification of the Austronesian Languages'. Indiana University Publications in Anthropology and Linguistics, Memoir 19 of the *International Journal of American Linguistics.* Baltimore. Waverly Press, Inc.

- FOX, Robert B., Willis E. SIBLEY and Fred EGGAN
 1954 'A preliminary glottochronology for northern Luzon'.
Asian Studies. 3.103-113.
- GRACE, George W.
 1964 'The linguistic evidence. Movement of the Malayo-
 Polynesians: 1500 B.C. to A.D. 500'. By Kwang-Chih Chang,
 George W. Grace, and William G. Solheim II. *Current
 Anthropology*. 5.359-406.
- HUMBOLDT, Wilhelm von
 1836-39 *Ueber die kawi-Sprache auf der Insel Java*. Berlin.
- KERN, Hendrik
 1887 'Klankverwisseling in de maleisch-polyneisische talen'.
BTLV. 5.333-343, 560-572.
- LLAMZON, Teodoro A.
 1969 *A Subgrouping of nine Philippine Languages*. The Hague.
 Martinus Nijhoff.
- REID, Lawrence A. ed
 1971 *Philippine Minor Languages: Word Lists and Phonologies*.
 Honolulu. University of Hawaii Press.
- SCHMIDT, Wilhelm P.
 1926 *Sprach familien and Sprachen Kreise der Erde*. Heidelberg.
- THALHEIMER, A.
 1908 *Beitrag zur Kenntnis der Pronomina personalia und
 possessiva der Sprachen mikronesiens*. Stuttgart.
- THOMAS, David and Alan HEALEY
 1962 'Some Philippine language subgroupings: A lexicostatistical
 study'. *Anthropological Linguistics*. 4.21-33.

RECONSIDERING THE NOTION OF *Focus* IN THE DESCRIPTION OF TAGALOG¹

Joseph F. Kess

Focus as a syntactic device has seen considerable use in the modern description of Tagalog and other Philippine languages. Such treatments have been relatively effective in their handling of certain aspects of Tagalog verbal structure, but at the same time have overlooked certain other important underlying considerations. Some of these are the question of focus affixes and their utility other than merely as a descriptive device of verbal morphology in Tagalog. Secondly, there is the question of just how meaningful the notion of focus is at all. Thirdly, there remains the question of just which features of the language lie submerged because of the limitations inherent in the frame of reference provided by the focus concept as applied thus far in the history of the linguistic description of Tagalog.

The fact is that a merely surface consideration of focus is not sufficient for a complete understanding of Tagalog verbal constructions, but represents only a partial approach to the problem. A bipartite approach must be used, the two levels of which together may provide a more complete picture of the language than has been the case. The direction of the argument is that the present interpretation of verbal constructions by focus accounts for only certain superficial features of the verb morphology. It provides convenient structural categories for verbal affixes, but cannot a priori predict what the semantic relationship of the affixed verb to the topic of the sentence will be. Nor, as a matter of fact, if such semantic relationships are reduced to a smaller and simpler number of possible case-like relationships between the affixed verb and the sentential topic, can one predict a priori what that case-like relationship will be from the affix exhibited by the verb.

A more satisfying alternative to this kind of treatment is to mark

verbs on a bipartite foundation of first verbal affix type and then notions of case function. In an accurate system of verb description, with correspondent verbal classification, it seems that the only way that this can be accomplished is with a system that cross-classifies its verbs both as to which verbal affixes they occur with (previously called focus), as well as which particular case relationships (to be designated as focus) these verbal affixes happen to mark with the sentence topic.

While verbal predicates in Tagalog do differ in the surface manifestations of structural arrangements which do occur, the sentence construction types cannot be said to be invariably signalled by the so-called focus affixes in the verbal construction. Moreover, such focus affixes cannot be described as invariably denoting the case-marking relationship between the sentence constituent in the topic position and the verbal predicate. Verb stems differ in respect to which verb stems occur with which verbal affixes. Secondly, such affixes may differ in respect to which case functions actually exist between the verbal predicate and the topic. Thirdly, verbal predicates so constructed may also differ in the other case relationships which they admit in the entire sentence structure.

If one is not to overlook such important considerations, the incompleteness of this approach can be compensated for by marking verb stems for the verbal affixes they may occur with, and in turn the resultant verbal predicates for the particular case-like function of the topic complement in the sentence. Here, if the term focus is to be retained for this latter feature, it is not very different from case grammar notions presented in recent arguments for the analysis of language, and such verbal predicate-topic relationships are easily translated into simple case relationships.

At this juncture it may be in order to quickly survey the earlier history of Tagalog description for some insights as to the origin of the particular descriptive philosophy so often employed in Tagalog. Tagalog, like a good many other Philippine languages, exhibits a set of unique structural arrangements in its paradigm of verbal construction types. This phenomenon has variously been termed *voice*, *case*, and *focus*, with the latter term becoming common in recent years. Descriptions employing the concept of focus emphasize its case-like function, claiming that the *dramatis personae* roles of the focused complement, or topic, are marked in the verb by certain focus-marking affixes, which are taken to indicate whether the topic plays the role of actor, object, beneficiary, instrument, or location.

Basically, the notion of focus and focus-types as they have been

used in Tagalog and related languages may be explained as follows. The major simple sentence type in Tagalog usually consists of at least a focused complement and a predicate. If the predicate is a verb, as it most often is, the focused complement is the topic of the predicate and is differentiated from non-focused complements in that it is introduced by the particle *ang* or contains a member of a pronominal or demonstrative substitute set associated with *ang*.² The predicate may also be a locative phrase or an adjective, in which case the topic is still introduced by *ang* or is an *ang* phrase substitute.

Though the focused complement is marked by the particle *ang* or one of its substitutes, its dramatis personae roles have been taken to be marked in the verb by certain affixes. These voice-marking affixes, which also mark tense, thus have been said to indicate whether the topic plays the role of actor, object, beneficiary or instrument, or location. Stems inflectable by such affixes are identified as verbs while other stems are nouns or descriptives. Verbal constructions, in turn, have been identified as focus constructions of one or another type by the various voice affixes in the verb. While the particular relationship of the topic to the verb has been said to exhibit overtones of a case-like nature, the relationship of the verb to non-topic complements has also been spoken of as a case relationship. In this sense, the particles which introduce the non-topic verbal complements have been occasionally called case-marking particles, while the particle which marks the topic is usually termed the topic-marking particle. Thus the case-like relationship of the topic to the verb, or the dramatis personae roles of the topic complement, have been taken to be explicitly marked in the verb, while those of the non-topic complement are marked by contrasting particles or contrasting pronominal sets.

This format of analysis has carried through ever since Blake and Bloomfield first proposed it for Tagalog and finds countless descriptive parallels in the discussion of many other Philippine languages. Taking but three examples of the many possible ones, one cannot help but note the similarity in description. For example, McKaughan, in an analysis of Maranao, outlines similar relationships which, he says, intersect. He remarks that "verbal affixes thus mark grammatical relations between verb and topic which intersect the relations marked by the particles used with other than topic substantives."³ These relationships for Maranao are actor, direct object, indirect object, and instrument.

Miller records a similar syntactic structure for Mamanwa. For Miller, "the term FOCUS as applied to Mamanwa refers to the significant

relationship which exists in a verbal clause between the action of its predicate and its actor, namely, Subject Focus; or between an action and its goal, namely, Object Focus; or between an action and the one on whose behalf the action is performed or the location of the action, namely, Referent Focus; or between an action and some other person or thing involved in the action, namely, Accessory Focus."⁴

Kerr's discussion of the verbal system of Cotabato Manobo lists "four distinctive types of relationship which the topic may contract with the verb, actor, object, instrument, or referent."⁵ Kerr notes that for Cotabato Manobo "the particular case-like relationship obtaining between the topic and the verb is indicated by the morphemic shape of the voice affix, not by any morphemic feature of the topic nominal expression itself."⁶

In summary, then, linguists have spoken of the distinctive nature of the Philippine verbal paradigm as being characterized by special voice-marking affixes. They have also called attention to the fact that, according to the focus type of the verb (as determined by the verbal affix), a particular sentence complement shall bear a special relationship to the verbal predicate. This complement is the focused nominal expression and has been termed the 'topic' of the sentence. It has also been said that the topic may contract at least four distinctive types of relationship with the verb, namely, actor, object or goal, instrument or accessory, and locative referent. These three - McKaughan, Miller, and Kerr - are only three examples of many descriptions which have made use of a similar framework, and one concludes that symmetry, compactness, and straightforward one to one relationships exist between verb affixes and case relationships in sentence structure. There are, of course, exceptions to this observation, as for example, the recent semantically-oriented treatment by Schachter and Otanes.⁷

It may be that Blake and Bloomfield's early studies set the precedent for the crystallization of verbal predicates in Tagalog into the four major focus types. The introduction of their descriptions, and further, Bloomfield's proffering of terminology for the four types, may have set a precedent for a good deal of grammatical thought in ensuing descriptions. In point of fact, the modern history of linguistic description for Tagalog verbs begins when Blake published some of his first articles in the *Journal of the American Oriental Society* at the turn of the century.⁸ Blake was followed and then paralleled by Bloomfield⁹ in Tagalog investigation, but one can easily imagine the authoritative influence exerted by the latter in certain quarters. Their particular orientations towards language, and specifically Tagalog, fused into a curious amalgam of mentalism and mechanism.

Blake's comments drew attention to what seemed to be semantic overtones to the focus or verbal construction types.

For Blake, it seemed that "in any given sentence the voice of the verb depends upon the relative importance of the various elements, the most important or emphatic idea being made the subject of the sentence. If this is the agent of the action expressed by the verb, the active voice is used; if it is any other element of the sentence, then one of the three passives is employed. In general, the in passive is used when the object of an action ... is made the subject; the i passive when the subject is the object of an action away from the agent ... or the instrument or case of the action; the an passive, when a place or anything regarded as a place stands as subject."¹⁰

Thus, it was first Blake who gave primacy to their semantic overtones, and a recent article by Hidalgo encapsulates what has been taken as a common assumption in the description of Tagalog ever since Blake published those first articles at the turn of the century. Like Blake's explanation, Hidalgo's conception of the notion of focus is such that "a constituent is brought into sharp perspective so that the attention of the listener is drawn closer to that constituent which is presumably in the speaker's mind. This element, which is in sharp perspective, or which is in focus, we call topic."¹¹

However, grammatical focus is not such that a sentence constituent is brought into some kind of sharp perspective so that the attention of the listener is drawn closer to that constituent, presumably foremost in the speaker's mind. A particular focus construction type does not mean that the particular focused topic is exclusively the focus or center of attention, although it may be related to the opposition of given versus new information, as hinted at by Buenaventura-Naylor.¹²

Here it might be beneficial to reconsider some of the possible notions fringing on the grammatical status of topic and focus. A particular focus construction type does not necessarily mean that the particular focused topic is exclusively the focus or center of attention. This must be obvious for at least several reasons. First, if this were the case, it would be impossible to make certain nominal phrases the center of attention for the simple reason that the particular verb in question does not admit (for reasons not too well defined) verbal forms which are the result of a particular focus affix type. Secondly, complements are rarely mentioned when verbal constructions are nominalized by position and case-marking particles indicating their function in sentences. If the notion of focus of attention is followed to its logical extremes, then a given sentence might have two, or perhaps more, foci of attention. Grammatically, this is obviously pos-

sible, but in any real sense, how must one then concentrate on several foci of attention. There is no longer any uniqueness attached to the item in focus.

Thirdly, if it is information content that focus is concerned with in a topic-comment relationship in a Tagalog sentence, it is the comment which provides information about the topic and makes for greater specificity. Conjecturing further, the attention given, if it is given in this fashion, would understandably gravitate toward the predicate, since this is where the greatest information is given content-wise.¹³ Lastly, the uses of focus and topic to underline center of attention and/or emphasis is in the last a purely cognitive performance paraphrase for what is essentially a grammatical phenomenon. Unless some empirical evidence is forthcoming, it is best to consider what has been termed focus simply a grammatical device. A parallel may be drawn here between English voice and its active and passive manifestations. There have been arguments for the semantic, stylistic, and emotional overtones of voice in English, but none of them entirely convincing. In fact, there would appear to be only frail empirical evidence for the hierarchy of priorities among the active, passive, negative, and passive negative. Recall here the rather strong debates, resulting confusion, and final disowning by some of Miller's project studying time lag as an indicator of priority of sentence types in the earlier version of generative transformational grammar.¹⁴

While it was Blake who may have contributed unwittingly to the center of attention characterization of focus, it may have been Bloomfield who helped to fix the number of verbal construction types by form alone. It is difficult to assess how the apparent readiness to equate specific affixes with specific focus types arose. Curiously enough, in an early article Blake had entertained the very notion that different case relations were shown in several ways by the four construction types in Tagalog. Having made mention of the four, marking them by their affix forms, Blake proceeded to mark some of the case-like relationships (using Latin-like cases) which might obtain between members of the sentence. Blake realized that the formal mechanics (the number of affixes - active, in, i, an - and the particles ang, nang, and sa) were indeed limited, but the number of real case-like relationships was not so limited nor were they exactly correspondent to the small number of four which formal properties later led him and others to provide fixed categories for.¹⁵

Extrapolating from Blake, suppose one does set about establishing and maintaining a set number of verbal construction types on the basis of the argument from formal properties. While the number of verbal

types may be held to a constant of four, there is nevertheless obvious overlap in the respective uses of the case-marking particles, *ang*, *nang*, and *sa*, and their respective pronominal (*ako*, *ko*, *akin*) sets. It is in this light not entirely unreasonable to expect some overlap between the various construction types themselves. Multiple overlap on the surface level is a phenomenon that Philippine languages exhibit in other areas, and even the idea of overlap in case representation is not a novel one. Blake and Bloomfield's early definitions were deliberately broad enough to accommodate some of the different case representations. Notably, some pedagogical texts like Larson and Aspillera point up the fact of incomplete paradigms and overlapping case representations. For example, Larson mentions overlap between the usage of *i-* as goal-, as instrument-, and as beneficiary-focus (this is apparently true, according to Larson, for the derived *mai-* affixed verbs as well).¹⁶ Aspillera's pedagogical treatment lists constructional possibilities for a sample of some 436 verbs, and the list is specific as to which are possible and which are not.¹⁷ Schachter also devotes part of an article to a discussion of several types of ambiguity in Tagalog, at least one of which is dependent upon the case-crossing functions of the *-an* construction as both a goal-focus and a locative-focus type.¹⁸

Incidentally, one notes that in early analyses, Blake and Bloomfield recognized several construction types, but at first termed these passive verbal constructions, with an implicit dichotomy between actives and passives and a simpler grouping thereof. For example, Bloomfield wrote of what he termed transient predicates ("transient words fall into four classes according to the four relations which a subject may bear to them when they are used as predicate")¹⁹ and christened these classes by the names active, direct passive, instrumental passive, and local passive.²⁰

This simpler dichotomy is somewhat maintained in Constantino's treatment of the syntax of a number of Philippine languages. Thus, for Constantino, "in an active sentence, the predicate verb is an active verb. An active verb is one which co-occurs with an actor subject ...; if the verb has an affix, it is an active suffix. A passive sentence is a verbal (definite) sentence in which the predicate verb is a passive verb. A passive verb co-occurs with a non-actor subject and has a passive actor complement in addition to the complements that occur with the active verb"²¹

A similar approach is seen in the more recent Schachter and Otnes reference grammar which notes that "while many of the transitive verbs that occur in basic sentences are actor-focus, an equal or greater number of such verbs select as topic something other than the performer

of the action. Any verb that does not focus upon the actor may be called a GOAL-FOCUS verb."²²

Linguists have not always agreed on the number of focus-types in Tagalog, and the argument as to how many focus types there really are has characterized more modern approaches. Some notion of the disagreement may be derived from even casually comparing Blake, Bloomfield, Wolfenden, Larson, and Bowen.²³ Perhaps one reason for the difference of opinion is a result of the attempt to make the framework more realistic without ever really breaking from it. For example, Bloomfield's four voice-modes are extended to five by Wolfenden. According to Wolfenden, Tagalog "voice-mode affixes specify subjective, objective, locative, implicative, and aptative relations" between the topic phrase and the predicating verb.²⁴ Unlike previous descriptions, Wolfenden's aptative voice "shows the topic to be able to undergo the action named by the verb stem", and mixes two previously identical focal relationships. Both actor and object complement may appear as topic of the same affix type.²⁵

Bowen, in an analysis for pedagogical purposes, outlines six focal categories: actor-focus, goal-focus, locative-focus, benefactive-focus, instrumental-focus, and causative-focus.²⁶ Topics of these constructions are taken to correspond to and convert with different complements in the actor-focus sentence, which is implied to be the reference point for analysis of other focal construction types.

Some comment may be made on an assumption which has been taken by some as being implied by the semantic center-of-attention notion in the interpretation of the nature of focus. This is the assumption that verbal constructions will have sentence-focus convertibility. In other words, if it is true that a given focus construction in an actor-focus sentence selects the actor of the sentence as the grammatical (and in this interpretation, the center of attention) topic, then it follows, or so it has been assumed by some, that other sentential complements can be placed in focus by the appropriate shift in topic- and case-marking particles as well as a shift in focus affix.

However, the fact is that free focus convertibility does not exist in this fashion for all verbal predicates for several reasons. First of all, verbs obviously differ in their potential to be inflected for various affixes which have been equated with focus types. Secondly, the relationship between an affix and a particular focus type is not invariant. Thirdly, in some cases where a verb is inflected for a given set of affixes, the semantic and/or case relationships between sentential components in a sentence of one focus type are not equivalent in their functions with sentential complements of another type.

Simply because a given focus type appears with a given verb is no guarantee that other focus-types may alternate with it in a manner which takes non-topic phrases of, say, an actor-focus (or any other sentence) and re-aligns them according to the desired focus type sentence. Namely, not all focus construction types equate on a one-to-one semantic basis, with all phrasal members accounted for as either focused or non-focused in normal sentence usage.

Some have on this basis suggested that the case-marking voices for which a verb is inflected may be employed to provide a framework for the classification of verbal types as well as verbal case functions. This point of view is implicit in Kerr's discussion of the case constellation of verbs (at least as far as classification according to verbal type goes), and explicitly stated in Kess' suggestion of the manner of presentation of syntactic features of Tagalog verbs. Kerr's suggestion was that the case-marking voices (a possibility of four) for which a verb is inflected may be grouped together and named the case constellation of the verb. This could have served as the basis of a verbal classificatory scheme, in which, according to Kerr, "the classification of a particular verb may be determined by setting up for it a limited transformation battery."²⁷

So also with Kess' treatment in which Tagalog verbs are examined for co-occurrence with a battery of primary affixes. Here verbs are found to differ in their capacity for focus affix potential, and as a result, are marked individually for such features. Resultant verbal constructions are also marked for co-occurrence with sentential complements in the various resulting sentential types.²⁸ A more interesting question is in fact contained therein. Is there any compelling reason why certain focus constructions appear more frequently than others. Secondly, is there any compelling reason why a single construction type or set of construction types (if the previous interpretation of focus is correct) appear for a given verbal root, and why others do not? There seems no point in completely denying the existence of connections between verb forms and underlying interpretations of semantic relationships which may indeed sometimes explain the intricacies of the Tagalog verbal scheme, but at the same time, there seems to be no simple and directly observable basis for their use and appearance other than pure statistical frequency appearance. In fact, recent studies in the acquisition of focus show that the process proceeds at a different pace for the several construction types.²⁹ One plausible reason for this being so is that such findings simply indicate the relative strength and exposure of children to the particular construction classes, either by virtue of the class as a whole or by

virtue of their exposure to common verbs which may show thus-and-so properties focus-wise. Both alternatives have an element of statistical frequency about them, and if a semantically-oriented answer is not forthcoming for the above questions, then one may be left with an answer of the same qualitative order.

Such classifications seem headed in the right theoretical direction, yet lack one important ingredient. Verbal types continue to be classified on the basis of formal properties alone, and semantic characterizations are given to such categories after their inception upon purely formal grounds. One might instead suggest that a similar type of classification be used, but one which couples formal properties and semantic equivalencies in terms of case relationships from the outset. The categories will not be as neat as before, but will ultimately provide more satisfying coverage. One must remember that information in terms of formal properties is not all that needs to be recorded. For such verbs which do take a given affix, the focus as defined in a semantic case-marking sense of what the role of the topic vis-a-vis the verb is not always invariant. To complete the informational set, one must then further ensure that information has been provided as to what focus type of relationship this particular affix is marking with this particular verb. As Buenaventura-Naylor has recently pointed out, "in the final analysis, the lexical content of the verb is the key factor; the verb stem determines which focus affixes may co-occur with it, the verb stem and its affix together determine the function and the semantic features of the topic as well as the number of obligatory participant roles of a focus construction."³⁰

The question is now how these groups shall be listed in the description. A method of verb stem classification along these lines is to indicate such an affix index for verbs and then the underlying case relationship of the topic to the verb which are shown by the index. This approach can make use of the suggestion of a complex symbol analysis of the syntactic features of formatives, since a method of description formulated to account for the fact that some aspect of linguistic structure are cross-classificationally rather than hierarchically arranged is exactly what is needed in this case. Such a system of representation would mean that verbal classification would be automatically built into lexicon representations of specific items. This would on the one hand provide for a built-in classification of verbal stems, and yet allow for any instantaneous classification of verbs on the basis of desired features, affix-wise, case-wise, or both.

In conclusion, description of verbal types in Tagalog may be greatly enhanced by dropping some previous connotations of focus and merely

considering the verbal affixes as surface properties. One must include another level, the semantic level of case relationships, and ask then, after having marked verbs for the appropriate affixes with which they may occur, which of the case-marking functions the resultant verbal construction in question is in fact marking with the topic. Here one could continue to entertain a notion of semantic focus, but it is only possible on this second level representing the underlying case relationship of the topic to the affixed verbal construction.

N O T E S

1. This is a somewhat revised and expanded version of a paper presented at the XVIII Annual Conference of the International Linguistic Association, Arequipa, Peru, March, 1973. This work has been assisted by a Canada Council Fellowship Award #W73 0201.
2. There are three marking particle sets in Tagalog, which may be labelled the ang, nang, and sa sets. The particle ang marks the topic; the particles nang and sa introduce syntactic complements other than that of topic and correspond to the case-marking particles of other Philippine languages.
3. H.P. McKaughan, 'Overt Relation Markers in Maranao', *Lg.* 38.47 (1962).
4. J. Miller, 'The Role of Verbal Stems in the Mamanwa Kernel Verbal', *Oceanic Linguistics* 3.88 (1964).
5. H.B. Kerr, 'The Case-Marking and Classifying Function of Cotabato Manobo Voice Affixes', *Oceanic Linguistics* 4.15-48 (1965).
6. *Ibid.*, 17.
7. P. Schachter and F.T. Otnes, *Tagalog Reference Grammar*, Berkeley: University of California Press, 1972.
8. F.R. Blake, 'Expression of Case by the Verb in Tagalog', *JAOS* 27.193-99 (1906); F.R. Blake, 'The Tagalog Verb', *JAOS* 36.396-414 (1916).

9. Bloomfield, *Tagalog Texts with Grammatical Analysis*, University of Illinois Studies in Language and Literature, Vol.III, No.3, 1917.
10. Blake, 'The Tagalog Verb', 411.
11. A.C. Hidalgo, 'Focus in Philippine Languages', *Philippine Journal of Linguistics* 1.27 (1970).
12. P. Buenaventura-Naylor, 'On Contextual Aspects of Topicalization', Paper presented at the First International Conference on Austronesian Languages, Jan., 1974, Honolulu, Hawaii.
13. Contrast the Japanese, where typically the verb and most information, comes at the end of sentences. A similar case has been pointed out for Chinese as well. See C.Y. Cheng, 'Toward a Theory of Subject Structure in Language with Application to Late Archaic Chinese', *JAOS* 91.6 (1971).
14. G.A. Miller, 'Some Psychological Studies of Grammar', *Amer. Psychol.* 17.748-62 (1962).
15. Blake, 'Expression of Case by the Verb in Tagalog', 183-85.
16. D.N. Larson, *Tagalog for Missionaries*, Manila, 1962.
17. P.S. Aspillera, *Lessons in Basic Tagalog*, Manila, 1956.
18. P. Schachter, 'Structural Ambiguity in Tagalog', *Language Learning* 11.135-45 (1961).
19. Bloomfield, op. cit., p. 153-154.
20. Ibid., p. 154.
21. E. Constantino, 'The Sentence Patterns of Twenty-Six Philippine Languages', *Lingua* 15.71-124 (1965).
22. Schachter and Otones, op. cit., p. 70.
23. F.R. Blake, *A Grammar of the Tagalog Language*, American Oriental Series, Vol. 1, New Haven, 1925; Bloomfield, op. cit.; E. Wolfenden, *A Restatement of Tagalog Grammar*, Manila, 1961; Larson, op. cit.;

- J.D. Bowen et al., *Beginning Tagalog*, Los Angeles: University of California Press, 1965.
24. Wolfenden, op. cit., p. 13.
25. Ibid., p. 15.
26. Bowen, op. cit.
27. Kerr, op. cit., 5.
28. J.F. Kess, *Syntactic Features of Tagalog Verbs*, Unpublished doctoral dissertation, University of Hawaii, 1967.
29. G.R. Tucker, 'Focus Acquisition by Filipino Children', *Philippine Journal of Psychology*, in press.
30. Buenaventura-Naylor, *Topic, Focus, and Emphasis in the Tagalog Verbal Clause*, Unpublished doctoral dissertation, University of Michigan, 1973.

THE INTERPRETATION OF POTENTIAL ACTION IN BIKOL VERBS

Malcolm Warren Mintz

1. Introduction
2. Affix Forms
3. Potential and Deliberate Actions
4. Nonvolitional and Volitional Actions
5. Conclusion

1. INTRODUCTION

Bikol is spoken by about three million speakers in the Philippines. The region, noted for its wide dialectal variation, comprises six provinces on the southern-most extension of the island of Luzon. The dialect discussed here is that of Naga City and is representative of the standard dialect of the region.

The verbs in Bikol, as in many other Philippine languages, take two general sets of affixes, one set which denotes a case relationship between the verb and the subject phrase of the sentence, and another set which adds a further semantic dimension to the meaning of the sentence (Blake 1925:38ff; Bloomfield 1917:402ff; McKaughan 1958:26ff). The first set may be referred to as 'case' affixes, and the second as 'semantic' (Mintz 1973). Semantic affixes can only occur in addition to or in combination with the case affixes, never alone. The semantic affix which is the subject of this paper is potential action.¹

Potential action affixes mark any action which has the potential of occurrence, whether that potential is volitional or nonvolitional. Nonvolitional actions are unplanned, mistakenly performed, or thwarted due to some physical short-coming on the part of the agent. If the result obtained is not that intended by the agent, or is not associated directly with the actions of an agent, the action may be said to be

nonvolitional. Volitional actions are more than deliberate. They require some special aptitude on the part of the agent. In Bikol, both volitional and nonvolitional potential actions are shown by the same semantic affix.

Potential action is an attribute of the speaker toward the agent and the action. If a speaker feels that an action is more, or less, than deliberate, he can use the potential action affix. The form of the affix alone indicates a deliberate-potential action distinction. Is the speaker correct in his assumption? If the speaker and the agent are the same, barring any attempt at irony or falsification, then we can assume that he is correct since the attributions are to his own actions. If the listener and the agent are the same, then the listener can validate or invalidate the use of potential action when he becomes the speaker. But if neither the speaker nor the listener is the agent, then the validity of the use of potential action must be based on shared speaker-listener experience and on an understanding of the agent and the nature of the action.

This mutual understanding is not only limited to situations where neither the speaker nor the listener is the agent of the action. The interpretation of volitional or nonvolitional action within the sphere of potential action must be also based on shared speaker-listener experience when the speaker himself is the agent of the action. If this were not the case, a speaker could not be sure his listener was interpreting the action correctly.

How volitional and nonvolitional actions are distinguished is the subject of this paper. The study opens with a presentation of the case affixes which, affixed singly to a verb base, indicate simple, deliberate actions. Following this is a brief outline of tense forms. This is to facilitate the recognition of verb forms used in example sentences. Next is a presentation of the potential action affixes in both their neutral, or infinitive, and inflected forms. The rest of the paper is divided into a comparison between potential and deliberate actions, and volitional and nonvolitional potential actions.

2. AFFIX FORMS

2.1. CASE AFFIXES

There are four case affix forms in Bikol: #mag#, #'i#, =han#, and =hon#. ² #Mag# indicates an agentive relationship between the verb and the subject phrase in the sentence. #'i# indicates that such a relationship may be either objective, instrumental, or benefactive; =han# that the potential relationships are either objective, dative, or

locative; and =hon# that the relationship is objective.³

#ku'a#	<i>get</i>	#mag#ku'a#	<i>to get</i>
#'uli'#	<i>return</i>	#'i#'uli'	<i>to return</i>
#'adal#	<i>study</i>	#'adal=an	<i>to study</i>
#basa#	<i>read</i>	#basa=hon#	<i>to read</i>

2.2. TENSES

Each of the verb bases may occur in an affixed neutral or infinitive form, as indicated above, or may be inflected for three tenses or aspects: future, past, and progressive.

Infinitive	Future	Past	Progressive
#mag#ku'a#	#ma:#ku'a#	#nag#ku'a#	#nag#kuku'a#
#'i#'uli'#	#'i#'u'uli'#	#'i#'=in=uli'#	#'i#'=in=u'uli'#
		#'i#pig#'uli'#	#'i#pig#'u'uli'#
#'adal=an#	#'a'adal=an#	#'=in=adal=an#	#'=in=a'adal=an#
		#pig#'adal=an#	#pig#'a'adal=an#
#basa=hon#	#babasa=hon#	#b=in=asa#	#b=in=abasa#
		#pig#basa#	#pig#babasa#

2.3. POTENTIAL ACTION AFFIXES

There are two potential action affixes, #maka# and #ma#. #Maka# is portmanteau, indicating, in addition to the semantics of potential action, an agentive case relationship between the verb and the sentence subject. #Ma# occurs with the three other case affixes.⁴

The following examples use the same verb bases presented above. The English translation is based on the volitional aspect of potential action.

#mag#ku'a#	<i>to get</i>	#maka#ku'a#	<i>to be able to get</i>
#'i#'uli'#	<i>to return</i>	#ma#'i#'uli'# ⁵	<i>to be able to return</i>
#'adal=an#	<i>to study</i>	#ma#'adal=an#	<i>to be able to study</i>
#basa=hon#	<i>to read</i>	#ma#basa#	<i>to be able to read</i>

Infinitive	Future	Past	Progressive
#maka#ku'a#	#makaka#ku'a#	#naka#ku'a#	#nakaka#ku'a#
#ma#'i#'uli'#	#ma#'i#'u'uli'#	#na#'i#'uli'#	#na#'i#'u'uli'#
#ma#'adal=an#	#ma#'a'adal=an#	#na#'adal=an#	#na#'a'adal=an#
#ma#basa#	#ma#babasa#	#na#basa#	#na#babasa#

3. POTENTIAL AND DELIBERATE ACTIONS

Potential and deliberate actions can formally be distinguished by the use of different sets of affixes. To show deliberate action, a case affix alone is used. To show potential action, a potential action affix occurs with the case affix. Any time a speaker wishes to show that more, or less, than normal effort is exerted in accomplishing a certain task, he uses the potential action affix. In the examples below, the speaker is also the agent. In the first sentence of each pair he attributes a simple, deliberate action to himself, and in the second, a potential action.

#NAG#KU'A# 'ako nin papel.
I TOOK some paper.

#NAKA#KU'A# 'ako nin papel.
I could GET some paper.

Da'i ko pa #'I#'=IN=U'ULI'# si libro.
I didn't RETURN the book yet.

Da'i ko pa #NA#'I#U'ULI'# si libro.
I couldn't RETURN the book yet.

The attribution of potential action may not be true. It may only be an assumption on the part of the speaker, especially if he is not the agent. In the next examples, the speaker is asking questions about the action of his listener.

#NA#BASA# mo na si "Gone With the Wind"?
Did you get the chance to READ "Gone With the Wind" yet?

#T=IN=APOS# mo na 'an assignment?
Did you FINISH the assignment yet?

When the listener, however, becomes the speaker, as is the case when he answers the question, he then validates, or invalidates the assumption of the speaker who asked the question. If he answers the first question, for example,

'Iyo, #NA#BASA# ko na.
Yes, I already got a chance to READ it.

then he is agreeing that it took more than simple, deliberate action to accomplish the task. If however, he answers,

'Iyo, #B=IN=ASA# ko na.
Yes, I already READ it.

he is indicating that the task was simply a deliberate one with no extra aptitude involved.

The same is true with the second question. An answer

'Iyo, #T=IN=APOS# ko na.
Yes, I FINISHED it.

indicates agreement with the questioner that the task was simply deliberate, while an answer

'Iyo, #NA#TAPOS# ko na.
Yes, I could FINISH it.

indicates that more than simple, deliberate effort was involved.

In the instance where neither the speaker nor the listener is the agent of the action, then the speaker must assume that the listener at least shares an understanding of the possible situation, and that the listener, at least for the present, is willing to accept that what the speaker says is true. For example, a speaker may say either of the following:

Mayo' si Jim na #'A'ADAL=AN#.
Jim has nothing to STUDY.

Mayo' si Jim na #MA#'A'ADAL=AN#.
There's nothing Jim can STUDY.

The use of distinct affixes indicates the difference between deliberate and potential action.

There is a further question, however. Why should a speaker assume his listener interprets potential action as volitional and not non-volitional? In other words, why should the speaker assume his speaker will understand Mayo' si Jim na #ma#'a'adal=an# as *There is nothing Jim can study* and not as *There is nothing Jim can mistakenly study*?

In this instance, and in the previous examples, volitional action is probably the most conventional interpretation shared by both the speaker and his listener. If the situation is not indicated as unusual, then why should the interpretation of the sentence be unusual?

What happens, however, when the situation is unusual and the speaker wants to convey the information that his use of the potential action affix is to be interpreted as nonvolitional and not volitional? The speaker has two recourses, and will probably use both of them. One is formal and one is contextual. To formally indicate that the potential action affix is to be interpreted as nonvolitional, length may be added to the final vowel of the affix. To indicate the situation is to be interpreted as unusual, the sentence may be expanded to further clarify the context. The following are examples.

#NAKA:#KU'A# 'ako nin papel, ta da'i ko 'aram na gagamiton pa.
I accidentally TOOK some paper because I didn't know that someone was still going to use it.

Maluya. #NA:#BASA# mo 'an "Gone With the Wind", pero "Uncle Tom's Cabin" 'an assignment ta.
What a mistake. You READ "Gone With the Wind", but our assignment is "Uncle Tom's Cabin".

Length added to the potential action affix in negative sentences also conveys nonvolitional action. Instead, however, of that action being interpreted as accidental, as was the case in the above sentences, such action receives an interpretation of thwarted volitional action.

Such thwarting usually arises from some physical shortcoming on the part of the agent.

Da'i ko #NA:#'I#'ULI'# si libro, ta hilang pa 'ako
I couldn't RETURN the book because I was still sick.

Da'i ko siya #NA:#HILING# ta ra'ot pa si salming ko.
I couldn't SEE her because my eyeglasses were still broken.

Da'i siya #NAKA:#DANGOG# ta kulog pa an talinga niya.
She couldn't HEAR because her ear still hurt.

4. NONVOLITIONAL AND VOLITIONAL ACTIONS

Just as there are contexts in which a speaker and listener tend to interpret potential action affixes as volitional, there are also contexts where such an interpretation is shared as nonvolitional. One of these contexts is that in which an agent does not appear in the sentence.

The agent, in sentences such as these, is not left out because it is understood, but because it is not important in the interpretation of the sentence. Nonvolitional action is conceived of in terms of result, not in terms of means, and due to such an interpretation an agent is often not necessary. This may differ from preferred expression in English. In English, for example, we might say 'I dropped the dish', whereas in Bikol, in the exact same situation, speakers say 'The dish fell'. In like manner, Bikol speakers say 'My watch got lost', 'The lamp got knocked over', etc., and not 'I lost my watch', 'He knocked over the lamp', etc. If more information is requested, it can be supplied in subsequent sentences.

The following are Bikol sentences in which the interpretation of the potential action affix is nonvolitional action.

#NA#RA'OT# 'an 'auto niya.
His car BROKE DOWN.

Tiba'ad #MA#BARI'# 'an silya kon dakolon magtukaw.
The chair might break if a lot of people sit on it.

#NA#HULOG# 'an plato.⁶
The plate FELL.

#NA#'ATI'=AN# 'an bado' ko.
My clothes got DIRTY.

These sentences cannot readily be interpreted as volitional action unless some attempt is made to clarify the context.

What if an agent were added to the sentences? Could these then receive an interpretation of volitional action? The answer to this depends both upon the intent of the speaker, and upon the shared speaker-listener understanding of the situation.

If a sentence such as

#NA#KAGAT# 'ako.
I got BITTEN.

received an agent and became

#NA#KAGAT# 'ako nin namok.
I got BITTEN by a mosquito.

the interpretation would still probably be nonvolitional. Deliberate action, which we will come to again shortly, may be attributed to the mosquito, but not, it seems, the extra degree of intent to become volitional action. This again brings us back to the question of assumption or attribution of potential action on the part of the speaker. Does a speaker believe a mosquito has the extra volition to bite someone? If he does believe that, does he share this assumption with his listener?

If the agent is inanimate, then the chance is so much greater that the interpretation will remain nonvolitional.

#NA#RA'OT 'an 'auto nin bagyo.
The storm DESTROYED the car.

What happens, however, if the agent is one to whom volitional action can be attributed? For example, does one interpret sentences such as the following as volitional or nonvolitional action?

#NA#RA'OT# ko 'an 'auto niya.
I accidentally RUINED his car.
#NA#'IPIT# niya 'an muro' ko sa puerta.
He CAUGHT my finger in the door.
#NA#WARA'# mo si libro niya?
Did you LOSE his book?

Do we use the above nonvolitional interpretations, or volitional interpretations such as the following?

I was able to RUIN his car.
He could CATCH my finger in the door.
Were you able to LOSE his book?

Probably the nonvolitional interpretation would be more common because of what the speaker and listener generally understand about the agent and the nature of the action. It is possible, however, that the speaker considers the agent malicious. He may then intend a volitional interpretation of the action. But, again, does the listener share his views? If not, then the speaker has the chance of being misunderstood. How can a speaker express his meaning unambiguously when there is a change of misinterpretation?

In the preceding section, volitional and nonvolitional actions were disambiguated by the addition of length to the potential action affix.

Length, however, can only be used to disambiguate actions which may be misinterpreted as volitional. If an action is interpreted as nonvolitional, length cannot induce a volitional interpretation. What the addition of length will do in these cases is further emphasize nonvolitional action. For example,

#NA#RA'OT# ko 'an 'auto niya.

will generally be interpreted as a nonvolitional action:

I accidentally RUINED his car.

If length is added to the potential action affix,

#NA:#RA'OT# ko 'an 'auto niya

then the sentence will even more emphatically be interpreted as nonvolitional action. There is no formal addition to the potential action affix so that a volitional action interpretation can unambiguously be made.

In cases such as the above, for the sake of clarity, a speaker would not use the potential action affix. He would probably use only a case affix for the expression of a simple, deliberate action. There is no doubt about the interpretation of the following sentences.

#PIG#RA'OT# ko 'an 'auto niya.

I purposely RUINED his car.

#'=IN=IPIT# niya 'an muro' ko sa puerta.

He purposely CAUGHT my finger in the door.

#PIG#WARA'# mo si libro niya?

Did you purposely LOSE his book?

There are also cases, again, those in which the result of the action is more important than the means, where volitional and nonvolitional action may remain ambiguous and not disturb communication. In such cases, the speaker does not care how his listener interprets the means. In the following sentence, for example, whether the agent was able to hear the President's speech, or just happened to hear it, does not seem to matter. What is important is whether he heard it or not.

#NA#DANGOG# mo 'an sinabi ni Marcos?

Did you HEAR what Marcos said?

In like manner, the following sentence may be interpreted either as 'I was able to see', or 'I happened to see Boyet at the market'.

#NA#HILING# ko si Boyet sa sa'od.

I SAW Boyet at the market.

As mentioned previously, disambiguation can be made in the direction of nonvolitional action by the addition of length. This would be done only if it were important to the speaker that he and his listener share the same interpretation of means. The two sentences below would be interpreted unambiguously as nonvolitional action.

#NA:#DANGOG# mo 'an sinabi ni Marcos?
Did you happen to HEAR what Marcos said?

#NA:#HILING# mo si Boyet sa sa'od?
Did you happen to SEE Boyet at the market?

If the above sentences were expressed as deliberate action, the interpretation, while unambiguous, would be slightly different.

#PIG#DANGOG# mo 'an sinabi ni Marcos sa radio?
Did you LISTEN to what Marcos said on the radio?

#H=IN=IHILING# ko si Boyet sa sa'od, pero da'i niya 'ako pig'i'intindi.
I was LOOKING at Boyet at the market, but he didn't pay any attention to me.

There are also examples where an action cannot be deliberate. Expression would be made with the potential action affix.

Pagduman mo sa Daraga, #NA#RISA# mo si dakulang simbahan sa ita'as nin bukid?
When you went to Daraga, did you NOTICE the big church on the hill?

Risa has no deliberate form. Expression of a simple, deliberate action would probably be made with the use of the base hiling *look at*.

5. CONCLUSION

Potential and deliberate action in Bikol can be formally distinguished by different verbal affixes. Within the sphere of potential action, however, there may be two interpretations, one volitional, and the other nonvolitional.

The attribution of potential action is made by the speaker. The interpretation of such action, however, is shared by both the speaker and listener. There are interpretations which both the speaker and listener clearly share as either volitional or nonvolitional action because of shared experience and a shared understanding of the agent and the nature of the action. There are other sentences which are ambiguous. A speaker may disambiguate such sentences formally or contextually.

Contextually he may indicate the unusual circumstance of the sentence, thereby clarifying the situation. Formally, he may disambiguate potential action in the direction of nonvolitional action by adding length to the potential action affix. If he wishes to disambiguate potential action in the direction of volitional action, he will have to forgo the use of the potential action affix, using simply a case affix. When the speaker does not care how his listener interprets the means of an action, as long as the result is clear, potential action may be left ambiguous. In all cases, formal and contextual specifications are made only when necessary and only to the degree necessary to clarify potentially ambiguous actions.

N O T E S

1. There are as many as sixteen semantic affixes which may be added to verb bases. Among these are affixes showing actions generalized over time or over a set of objects, social actions emphasizing the relationship between participants in the action, intensive actions, and repetitive actions occurring more than once in a set period of time. Semantic affixes also show actions that are incipient, those that are the consequence of previous actions, those that are directive or imperative, reciprocal, comitative, emphatically plural, mitigated in some respect, pending, and developed due to outside influences.
2. Verb bases and prefixes are bounded by #: #ku'a# *get*; #mag#. Suffixes are bounded initially by = and finally by #: =han#. Infixes, which are discussed in the section on tenses, are bounded by =: =in=. A glottal stop is indicated by an apostrophe, '.
3. Suffix initial h, as in =hon# and =han#, is deleted when suffixed to consonant final bases.
4. #Ma# and the case affix =hon# cannot occur at the same time on the same base. The case relationship indicated by =hon#, however, remains. A possible explanation is that =hon# is deleted after #ma# is prefixed. There is precedent for such deletion since =hon# is also deleted in the past and progressive verb inflection.
5. In Naga City, the preferred form is #'i#ka#'uli'#. The common replacement of #ma#'i# by #'i#ka# seems to be restricted to the Naga City area, reflecting, perhaps, a local development. The form presented as part of the above paradigm is the more common throughout the standard dialect region.

6. This is a verb base which generally takes the case affix #i# to indicate an objective case relationship between the verb and the sentence subject. This particular case affix contains a sense of 'alienable action', an action which results in the object ending up in a location different from where it was the start of the action. Interestingly, in the presence of the potential action affix when no agent is present in the sentence, such an affix is omitted.

BIBLIOGRAPHY

BLAKE, Frank R.

- 1925 'A grammar of the Tagalog language'. *American Oriental Series*, 1. New Haven, Connecticut. American Oriental Society.

BLOOMFIELD, Leonard

- 1917 'Tagalog texts with grammatical analysis'. *University of Illinois studies in language and literature*, 3:2-4. Urbana, Illinois. University of Illinois.

McKAUGHAN, Howard

- 1958 *The inflection and syntax of Maranao verbs*. Publications of the Institute of National Language. Manila. Bureau of Printing.

MINTZ, Malcolm W.

- 1973 *Case and semantic affixes of Bikol verbs*. Honolulu. University of Hawaii. Dissertation.

GAYO CONSONANT CORRESPONDENCES

H.L. Shorto

0. Introduction

1. *b

2. *d, *D, *Z

3. *R

4. *j

5. The Semivowel Rule

Appendix: Items from 214-word

Lexicostatistic List

0. INTRODUCTION

Gayo, an Austronesian language spoken in the interior of north-western Sumatra, shows a number of idiosyncratic consonant correspondences that cannot be accounted for by borrowing from neighbouring languages,¹ and must therefore result from shifts which occurred in Gayo itself. The peculiarity of the case is that they are found only in a minority of the lexical items for which Dempwolff (1938: vol.III) provided reconstructions; so far as the forms which fail to show them are identified as cognate loans, it is inherited and not borrowed vocabulary in Gayo that is statistically exceptional. Here we should consider that when borrowing takes place between languages as closely related genetically as are many contiguous ones in Indonesia, it is often a matter of replacing a form by another which differs perhaps only in one phoneme from it, in a more or less predictable way. The process bears more resemblance to that by which an English provincial accommodates his speech to metropolitan usage than to the loan process as usually studied. Widespread borrowing of this kind has been classically documented in Ngaju Dayak (Dyen, 1956). It is coming to appear

by no means abnormal in other Indonesian languages as they receive systematic examination;² and our analysis of the Gayo data will suggest among other things a new approach to those of Javanese, to which they bear an incidental resemblance.

Gayo has had little descriptive attention, which may account for a corresponding failure to attract the notice of comparatists.³ But its lexicon, fortunately, has been recorded at length by Hazeu (1907), from whose 1,148-page dictionary the forms cited here are taken.⁴

The shifts to be discussed affect the PAN consonants *b; *d, *D, *Z (the first two being merged in all, and all three in some, of the languages of the area); *R; and *j.⁵ They are in most environments

*b > \emptyset initially, w ~ \emptyset medially;
 *d etc. > r initially and medially;
 *R > \emptyset initially, γ ~ \emptyset medially, \emptyset finally;
 *j > \emptyset medially.

Words exhibiting these reflexes will be identified as items of the Gayo inherited vocabulary.

The lexemes which can be cited in evidence are not numerous, fewer than eighty in all. This has its own inconvenience; we shall be obliged at times to draw conclusions from a small number of examples, especially where environmental differences of treatment come into question. Some detailed formulations will therefore be provisional. In compensation, doublets showing the reflexes to be expected from borrowing are recorded in about a third of the cases, and then almost always the presumed loan is restricted semantically or collocationally by comparison with its shifted-reflex counterpart.

There are some items referable to PAN bases containing more than one of the relevant protophonemes in which only one of the reflexes shows the shift predicted. Such mixed cases cannot be explained from borrowing, at least without invoking a hypothetical contamination. On the assumption that they are regular I use them to deduce exception rules, in particular a 'semivowel rule' which inhibited the development of successive semivowels, or close non-obstruent segments.

I shall now set out the evidence for the shifts.

1. *b

Taking the reflexes in medial position first, *b is reflected as w in the environment *a—a. This has been accepted as the regular development in Malay (Dempwolff, 1937: vol.II, 20, § 70 (a); exceptions *ibid.*, 21, § 70 (e) 3). But in view of its coherence with the remaining reflexes I assume it to be regular in Gayo also and list all instances,

though borrowing from Malay is possible with some of them. Thus: (1) *babaq *mouth* > awah, ~ babah in phr. + Ach. babah; (2) *kabal *invulnerable* > kawal *guard, defence, guardian, defended area*; ⁶ (3) *kaban *companion, company* > kawan *herd, company* (cf. Mal. kawan); (4) *laban *opponent* > löwön ~ (old persons' speech in 1907) lawan (cf. Mal. lawan); (5) *taban *to hold fast, booty* > tawan *taken prisoner, abducted* (cf. Mal. tawan *to take prisoner*). Note also *tabaR > tawar, no. 75 below.

Following or preceding a back vowel, including o < *-aw, *-b- is generally lost. For this we have (6) *Dabuq *to fall* > tauh (with t- by dissimilatory devoicing, before the shift of *D-, of the first of two heterorganic voiced plosives; there is no contrary evidence in the inherited vocabulary; a similar development is seen in Ach.); (7) *z[ae]but *hairs* > dial. jaut *sugar-palm fibres*; (8) *k-ar-ebaw *buffalo* > koro (with vowel harmony as in nos. 32, 33, 49, 57, 74); (9) *tebus *sugarcane* > tu; (10) *luban *pit* > luang *hole, pit*; (11) *bubu(S) *fish-trap* > u ~ uu; (12) *subuk *to spy on* (Blust, 1971, no. 409) > suk.

It is reasonable to assume *b > (*)w as a first development in all these environments, followed by the loss of *w in contact with u, o. In this connexion two other items have to be noticed. In (13) *qabus *ash* ⁷ > wau, dial. au, *dry sand, sand mixed with ash from base of hearth* we have a case of metathesis with w- preserved under the monosyllable rule discussed in relation to no. 36 below. There is no direct evidence as to whether *b > w antedated the metathesis, but *awu > wau appears more probable than *abu > *bau. In (14) *quban *grey-haired* > wön = uwön ⁸ *grey hair*, however, it is likely that w- developed secondarily from *u- as in the reflexes of *buaq and *buSat, nos. 34-5 below.

In initial position *b is generally lost. No. 36 implies that the development took the same course *b > *w > ø as in the medial cases; the loss of *w parallels that of PAN *w (cf. no. 48, *waDa(S)). Thus: (15) *baqeRu(S) *new* ⁹ > ayu ~ rarely baru + Mal. b(ah)aru; (16) *batan *log ...* > atang *felled trunk, beam*, ~ batang *trunk, tree, plant* + Mal. batang *trunk*; (17) *batu[] *stone* > atu, dial. watu, ~ batu in phr. + Mal. batu; (18) *batuk *to cough* > atuk; (19) *buka *to open* > ukö, ~ occasionally bukö + Ach. or Mal. buka; (20) *bukuk (~ *bu-ŋ-kuk = UAN *buŋkuk) *crooked* > ukuk *to bend forward*, ~ bungkuk *warped, crooked* + Ach. bungko? < *bu-ŋ-kuk; (21) *bulan *moon* > ulön *moon, month*, ~ bulön *month* + Ach. bulöen or Mal. bulan; (22) *buluŋ *foliage* > ulung *leaf*; (23) *buluq *kind of bamboo* > uluh *bamboo*; (24) *bunuq *to kill* > unuh, ~ rarely bunuh + Mal. bunuh; (25) *buri(S) *to flow* > uri *descent of amniotic fluid*; (26) *buSuk *hair* ¹⁰ > uk; (27) *butuq *penis* > utuh; (28) *buzan *unmarried* > ujang *idem*, ~ bujang *marriageable [male]* + Mal. bujang *unmarried*. So too *babaq > awah, no. 1 above; *bubu(S) > u(u), no. 11; and note ayu:

Karo Batak bayu *to twill, make mats.*

Initial *be, and the secondary *bě- arising from vowel weakening in the antepenult, undergo one of two contextually determined developments. Each of these is paralleled in the treatment of *e following initial zero or laryngeal. The whole syllable is lost in three instances: (29) *belas *sympathy* > las *to regret*; (30) *betiis *calf of leg*¹¹ > tis; (31) *bi[tɪ]uka[] *intestines* > (*bětuka >) tukö *stomach*. The loss of *ě- following the disappearance of the consonant appears unitary with that in e.g. *[]e-m-pu *grandparents* > mpu. The contrary cases are associated with a type of vowel harmony found in the context of medial r or nasal, and also exemplified in nos. 8, 49, 57. Thus: (32) *beŋi(S) or *beŋuy¹² *night* > ingi; (33) *bin[e]siq *seed*¹³ > (*bėnih >) inih ~ eneh *idem* ~ elevated bėnih (beside röm) *paddy* + Mal. bėneh *seed*; and add *beRas > ɔrɔs, no. 74 below. Cf. with these e.g. *eZen > ɔrɔn (no. 54); *enem *six*¹⁴ > ɔnɔm, ~ nam in cpds. etc. + Ach. nam or Mal. ěnam.¹⁵

As in wön < *quban (no. 14), w- has probably developed secondarily from *u- in (34) *buaq *fruit* > wah = uah = uwah (phonologically one form, cf. n. 8); (35) *buSat *to lift*¹⁶ > wöt *to rise, get up, n(u)-wöt to lift up, pick up*. But it is w- < *b- that appears to be preserved in (36) *bajas *interior* > was ~ waas. For this reconstruction Dempwolff cited only a Tagalog word meaning 'north-west' and a Toba Batak one meaning 'dwelling', but when Karo Batak ba:s ~ bagas *interior* is added there is no reason to doubt the connexion. Since the expected Karo form is bagas, ba:s must be a loan from Gayo, showing that the loss of *-j- (below) preceded the shift of the initial. We may then formulate a rule, also operative in no. 13 above, that *w- (< *b-) was preserved in monosyllables except before a back vowel, where, as in medial position, it was lost (nos. 11, 26).¹⁷ The dialect variant watu (no. 17) provides further evidence of *b- > *w-.

There is no indication of a distinctive development of *b in final position; the inherited reflex is probably p as in rukup *upcurved [horns]* < *rukub *protection*.

2. *d, *D, *Z

In the inherited vocabulary *d, *D, and *Z generally merge as r initially and medially. The single exception, *Dabuq > tauh (no. 6 above), results from precedent dissimilatory devoicing.

Initially: (37) *diŋdiŋ *wall, partition* > rėring (with vowel weakening as in all inherited reflexes of 2(CVC) forms, e.g. tėnting *to sift by shaking* < *tiŋtiŋ); (38) *Dalem *inside, depth* > rölöm *deep, ~ dölöm inside, innermost* + Ach. or Mal. dalam; (39) *DataR *flat, level* > rata;

(40) *Deles to slide > rēlas collapsed; (41) *DuSa two¹⁸ > roa, ~ duö in cpds. etc. + Ach. or Mal. dua; (42) *[dD]aRaQ blood > rayoh; (43) *[dD]uRi[] thorn > rwi = ruwi = rui; (44) *Zaket ~ *Zeket to stōk¹⁹ > rakat ~ rakot hitting the mark, ~ dēkat ~ dēköt near + Mal. dēkat + Jav. dēkēt; (45) *Zalan path, road²⁰ > ralan gait, r-ēm-alan to go, ~ jölön way, manner + Ach. jalan idem... + Mal. jalan path, road, course to take....

Medially: (46) *hadep front²¹ > arap idem, ~ ödöp to face towards + Ach. adab idem; (47) *mudaq easy, cheap > murah cheap, open-handed, unstrenuous, ~ mudah easy + Ach. or Mal. mudah idem; (48) *waDa(S) to exist > ara present, existing; (49) *saDeR to prop against²² > sere ~ sere (with vowel harmony as in nos. 8 etc.; for -e < *-eR see below), ~ obsolescent (in 1907) söndör + Mal. sandar < *sa-n-DeR; (50) *quDI(S) rear²³ > p-ur-ön behind, after (: Karo Batak p-udi-:n henceforth); (51) *quDip to live, be alive > urip; (52) *tuDuq to drip > turuh letting rain in; (53) *ku[dD]en cooking-pot > kurön; (54) *eZen to squeeze out²⁴ > oron; (55) *quZan rain²⁵ > urön ~ elevated ujön + Ach. ujöan or Mal. ujan; (56) *tuZuq to point out²⁶ > turuh. Note further sara : Toba, Karo Batak sada one, taring : Karo tading left behind, left over.

The example of rēring < *diŋdiŋ (no. 37) against tēnting < *tiŋtiŋ suggests that the shift may have extended to medial position following a nasal, which was then dropped before the resultant r under pressure of the phonological pattern. This differs from the treatment of *b and *j, but at least one other apparent case can be cited: (57) *[]an[dD]uy to bathe > n-iri idem (with vowel harmony as in nos. 8 etc.), ~ mandi washing of corpse + Mal. m-andi. Cf. also n. 22.

A problem is posed by the word lo day. It was compared by Hazeu with Dayak andau, Jav. ěndon (~ andon), Tagalog arao, which would relate it to *qa(n)Zaw day, sun.²⁷ Different inherited developments of -nZ- and -nd-/-nD- appear prima facie unlikely, but no other obvious etymology is available. If we refer lo with Jav. ěndo-n to a variant (58) *qe-n-Zaw or *qe(n)Zaw of *qa(n)Zaw, it is possible to propose the generalization of a sandhi-form from the phrase mata n lo (< *n ro?) sun; but this solution is speculative.

*Z has not been reconstructed in final position. The reflexes of *-d and *-D, which do not apparently distinguish inherited words from borrowings, are respectively -t and -r as in laut sea, lake < *laud, pusör whorl < *puseD.

3. *R

As with *b, it will be helpful to begin with the reflexes in medial position, where *R is generally reflected as Gayo y. Thus (59) *paRaw

hoarse > payo; (60) *uRat *nerve, vein* > uyöt *muscle, sinew, nerve, fibre, root, creeper*,²⁸ ~ uröt *thread* + Ach. or Mal. urat; (61) *puRuq *quail* > puyuh (+ Ach., Mal. puyoh!); and *baqeRu(S) > ayu, no. 15 above; *[dD]aRaQ > rayoh, no. 42.

An earlier *-y- is to be inferred in (62) *kaRat *to gnaw, bite* > (*kayat >) ket *to bite*, ~ karat *rust; pressing, urgent* + Mal. karat *rust* and Ach. karat *pressing*, with the same development as is seen in e.g. *bayaD > ber (no. 77 below; b- by semivowel rule); and also in *buRaw > bio, no. 76.

Following or preceding a front vowel *-R- is lost, as *-b- is following or preceding a back one. Thus: (63) *uqaRi(S) or *quaRi(S) *sun, day*²⁹ > s-wöi = s-uöi *the day after tomorrow* (originally (*after*) *one day*), ~ in limited contexts ari (beside lo, no. 58) *day* + Mal. hari; (64) *ñiRu(S) *winnowing-tray*³⁰ > niu; and *[dD]uRi[] > rwi, no. 43 above. But the loss of *-R- in (65) *luRuq *to trickle, drizzle* > luh *tears*, ~ ruluh *fallen [fruit, leaves, hair]* + Ach. luroh *to fall* with regular metathesis, cannot be explained in this way - contrast no. 61. It is most simply accounted for by metathesis of l and *R, anterior to the shift, paralleling that of l and r:³¹ *luRuh > *Ruluh > *uluh > luh, though the loss of the first-syllable vowel remains problematic.

Reconstructions of initial *R are relatively infrequent; such evidence as there is of its treatment in Gayo points to a development *R- > (*y- > ?) ø. For this we have, besides the possible secondary case just quoted, only (66) *Rumaq *house, dwelling* > umah. Note also, however, ambu-ambu *fringe* beside rambu < *rambu(S), perhaps by hypercorrection at a stage when forms in ø < *R- were more widely competing with borrowed forms in r-.

Unlike the other protophonemes discussed here, *R has characteristic Gayo reflexes in all three positions; and with *-R we are on firmer ground. A development *-R > *-y underlies the changes *-aR, *-eR > (*-ay >) -e ~ -ε and *-iR > -i, exemplified in (67) *lapaR *hungry* > lape *famine*; (68) *deqeR *to hear*: pēnge ~ pēnge ~ t(ēn)ēnge ~ nēnge apparently < *pērēnge (or *pēnēnge?) < *pVn-deqeR, etc.;³² (69) *ikuR *tail* > (*ukiR by metathesis³³ >) uki; and in *saDeR > sere ~ sere, no. 49 above.

*DataR > rata, no. 39, is an exception. Ach. has rata *flat, level*, apparently + Gayo, beside a less frequent regular data; perhaps this is a case of double borrowing, with Gayo -a, for *-e, normalized on Ach.

4. *j

No reconstructions of *j in initial position have been proposed. Finally, and medially following a nasal, Gayo shows the velar reflexes typical of Batak languages: -k and -ng(g)-, as in *pusək navel* < *pusej, *ngi* ~ *nggi younger sibling* < *a-n-ji(S). Medially in the absence of nasal augment, however, *j is lost in the inherited vocabulary whatever the vocalism. Thus: (70) *qapeju(S) *gall* > *pau*; (71) *pija(S) *how much?* > *piö-n*; (72) *ijun *nose* > *iung*; and *bajas > *wa(a)s*, no. 36 above.³⁴

5. THE SEMIVOWEL RULE

In a number of cases involving reconstructions where *b is followed later by *R the Gayo form shows one of the expected reflexes, but not the other. The shift of *b but not that of *R is found in (73) *baRa[] *live coals*³⁵ > (*ara-ara > *ərara >) *rara* (+ Karo Batak *rara to glow red*; for the reduplication cf. Jav. *wa-wa*), ~ *börö* in phr. + Mal. *bara*; (74) *beRaS *husked rice* > *orəs*; and, with medial *b and final *R, in (75) *taBaR *antidote* > *tawar counteracting, antidote*. The converse is seen in (76) *buRaw *to drive away*³⁶ > (*buyo >) *bio to drive [cattle], drive away, drive out*. The only item in which both *b and *R appear shifted is *baqeRu(S) > *ayu*, no. 16, which as an original trisyllable with an intervening phoneme may be subject to a special rule.

It is possible to account for these instances by postulating a 'semivowel rule' that inhibited the development of sequences *wVy. But in view of the divergent treatment of no. 76 this depends on the assumption, not in itself implausible, that the initial stages of both shifts were broadly contemporaneous and that their detailed chronology varied according to the nature of the neighbouring vowels.

The hypothesis can be tested by reference to reconstructions in which *b is followed by *y, where we should expect *b > *w to be inhibited wherever inherited *y had remained unshifted. This is borne out by (77) *bayaD *to pay* > *ber*; if *ber* is a loan, it must be an early one, since *-aya- > -ε- is characteristic of Gayo (cf. no. 62 above, and contrast *bayang shadow* + Ach. or Mal. *bayang idem* < *bayaŋ *to sway*). Though *ayu* : Karo Batak *bayu* (p. 201) seems to contradict this, it may be on a par with *ayu* < *baqeRu(S). Correspondingly, with medial *b (78) *labay *yarn* > *labe to skein yarn* may well belong to the inherited vocabulary.

I have found no evidence bearing on the treatment of the converse sequence, PAN *RVb-, in inherited vocabulary.³⁷ But *yVy as well as *wVy may have been inhibited if (79) *layaR *sail* > *reI* (with metathesis as seemingly in no. 65) is an inherited form. As a nautical term per-

haps it is unlikely to be. Certainly *wVw cannot have been ruled out in view of *babaq > awah, no. 1.

However, we can take the semivowel rule further. No instances of the shift of *b before *i have been found apart from the trisyllabic *bi[tT]uka[] > tukö, *bin[e]Siq > 'inih etc. (nos. 31, 33), where antepenult vowel weakening may well have come first.³⁸ If the rule also inhibited the sequence *w + close front vowel, the normal development will have been as in bintang *star* (<?) *bintaŋ, lebih *more* (<?) *lebiq; the want of items showing the shift is explained, but inherited forms will not usually be distinguishable from borrowed ones.

A similar inhibition of the sequence close back vowel + *y could have affected unstressed syllables only in view of nos. 60-1, but would account for the lack of instances of the shift of final *R following *u. PAN *-uy yields -i (as in no. 57, *[]an[dD]uy > n-iri); this development may have come earlier, but it may reflect the same pressure. Thus again e.g. kumur *to gargle* (<?) *kumuR may be an inherited form.

The forms which show these reflexes are a minority of all those that can be related to Dempwolff's reconstructions containing the corresponding protophonemes. I omit from the tabulation below items the occurrence of which is stylistically or collocationally restricted, as well as those in which distinct reflexes may be precluded by the semivowel rule. I have found the following numbers of items with the respective reflexes:

	inherited	borrowed
*b-	27 (nos. 1, 11, 15-36, 73-4, 76)	63
*-b-	15 (nos. 1-14, 75)	26
*d-/*D-/*Z-	10 (nos. 6, 37-45)	25
*-d- etc.	11-12 (nos. 46-56, ?58)	24
*-nd- etc.	2-3 (nos. 37, 57, ?58)	18
*R-	1 (no. 66)	6
*-R-	13 (nos. 15, 42-3, 59-65, 73-4, 76)	23
*-R	5 (nos. 49, 67-9, 75)	8
*-j-	4 (nos. 36, 70-2)	4

Overall, correcting for items which figure twice in the count, inherited vocabulary amounts to about 30 per cent of the whole sample.

To list all items marked as borrowed by these criteria would inordinately lengthen this article. In the Appendix I note all items in either category which figure in a 214-word lexicostatistic list. In this core vocabulary the proportion of inherited items, similarly calculated,³⁹ rises to 56 per cent, reinforcing the presumption that the method I have followed to identify them is valid; but the presence of

so many cognate borrowings even here is remarkable. The incidence of cognate borrowing in Achinese is of the order of 50 per cent for the Dempwolffian vocabulary; Gayo may well prove to be an extreme case of linguistic acculturation.

Lessons may be drawn even from extreme cases. If we tabulate the Gayo reflexes without discriminating between inherited and borrowed forms, in order of frequency, we have

*b:	b, \emptyset (\sim) w (finally p);
*d, *D:	d, r, (l) (finally t < *d, r < *D);
*Z:	r, j, d;
*R:	r, \emptyset (\sim) y;
*j:	\emptyset , d, r (finally k).

We may compare this with a corresponding tabulation of the reflexes in Javanese as usually stated:

*b:	w, b;
*d, *D:	d, d, r, finally also t;
*Z:	d, r;
*R:	\emptyset , r;
*j:	r,

and proceed to establish the Jav. correspondents of the Gayo forms treated in this article.

From *b-, *-b- Jav. has w in 19 cases, and doublets in w, b in seven more, against b in three cases only:

(w) nos. 2, kawal *resistance, defended*; 4, lawan; 5, tawan *booty, prisoners*; 6, dawuh; 10, luwang; 11, wuwu; 13, awu *ash*; 14, uwan *grey hair*; 15, wau *just now*; 16, watang *stave, felled trunk*; 18, watuk; 21, wulan; 23, wuluh *bamboo*; 26, wok *beard*; 29, wēlas *sympathy*; 30, wētis *thigh, calf*; 33, winèh, obsolescent winih; 34, woh; 75, tawar *made harmless*; (w \sim b) nos. 17, watu \sim batu in phr.; 20, wungkuk *crooked* \sim bungkuk (*with*) *back bent*; 28, wujang *unmarried* \sim bujang *marriageable*; 32, wēngi \sim bēngi; 35, a-bot *heavy, weight*, wo-wot-an \sim bo-bot *weight*; 73, wa-wa \sim barah; 74, wos \sim bēras; (b) nos. 8, kēbo < *kebaw; 9, tēbu; 19, buka *breaking of fast*.

From *D-, *-d-, *-D- Jav. has r in nine cases, plus one more with a doublet in d, against one each of d and d:

(r) nos. 39, rata; 41, lo-ro; 42, rah; 43, ri; 46, arēp *to want to, intend to*; 47, murah *cheap*; 48, ora *not to be*; 50, w-uri *behind, after* \sim b-uri *rear*; 52, turuh *to let rain in*; (r \sim d) no. 51, urip \sim udip; (d) no. 6, dawuh; (d) no. 38, dalēm *home, inner room*.

From *Z-, *-Z- Jav. has d in five cases, against one of doublets in d, r:

(d) nos. 45, *dalan path, road*; 54, *ěděn*; 55, *udan*; 56, *tuduh*; 58, *ěndo-n*; (d ~ r) no. 44, *děkět near ~ rakět intimately connected*.

From *R-, *-R-, *-R Jav. has \emptyset in eleven cases, plus two more with doublets in r, against one of r alone:

(\emptyset) nos. 15, *wau just now*; 39, *rata*; 42, *rah*; 43, *ri*; 60, *ot-ot muscle, sinew, nerve, vein*; 63, *we-we [rain] while sun is shining*; (64, OJav. *nyiyu*;) 65, *luh tears*; 66, *omah*; 67, *lapa hunger*; 68, *rungu*; (\emptyset ~ r) nos. 73, *wa-wa ~ barah*; 74, *wos ~ bėras*; (r) no. 75, *tawar made harmless*.

From *-j- Jav. has regularly r. There are three cases: nos. 70, *ampėru ~ r-ěmpėlu*; 71, *pira*; 72, *irung*.

Thus - omitting *j, where no problem arises - in the majority of cases Jav. shows the following correspondences to the Gayo inherited reflexes: *b > w, Gayo \emptyset ~ w; *d, *D > Jav., Gayo r; *Z > d, Gayo r; *R > \emptyset , Gayo \emptyset ~ γ . The pattern of shifting in the two languages is similar, but not identical: Jav. merges (in part) *r, *d, *D, *j, Gayo *r, *d, *D, *Z; the Gayo treatment of *j is wholly idiosyncratic. No one, I take it, will attribute the forms examined above to borrowings from Javanese that are not found in any of the languages that separate Gayo from the outside world, still less argue that Gayo is a Javanese dialect brought to its present location by unimagined historical events. Some may nevertheless be tempted to construct new or revised proto-phonemes on the basis of the Gayo-Jav. agreements; but this I believe to be a pseudo-solution. Just as there are some correspondences of e.g. Jav. b to Gayo \emptyset ~ w, so there are others, more frequent, of e.g. Jav. w to Gayo b (thus Jav. *wasuh ~ wasoh to wash*; Gayo *basuh*, Appendix). This is what one would expect if Gayo, Jav. b both resulted from borrowing, but the incidence was higher in Gayo. The coincidences of Gayo w/ \emptyset and Jav. w are to be attributed to the higher retention probability of certain lexemes, a topic to which I shall return; and it seems that Gayo with its exceptionally high borrowing-rate provides a net to catch some of the most retainable.

Many attempts have been made to explain the multiple reflexes found in Jav.; beginning with Dempwolff's *Tendenz zur Lautverschiebung*, a thin disguise for 'sporadic change' which we should hasten to discard. To account for w \neq b Haudricourt at one time entertained the idea of reconstructing a labiovelar order of consonants, but later abandoned it in favour of the filling of a case vide by interdialectal borrowing (Haudricourt, 1964, 118; cf. idem, 1951, 144-5). The apical stops have begotten a whole literature, which I shall not attempt to summarize here; for a critical discussion see Dahl, 1973, ch. 14. Dahl agrees with Haudricourt (1964, 110-111) that for Jav. - so a fortiori for

western Indonesian languages in general - the distinction between *d and *D in non-final positions is likely to be a ghostly one. But his own solution of the problem, which entails reconstructing three phonemes *d₁, *d₂, *d₃ on the basis of Formosan reflexes, is no more help than others in disentangling the Gayo data. All three of Dahl's phonemes are represented in the material we have cited, at least in initial position: *d₁ (no. 43), *d₂ (39, 40; in medial position 51), *d₃ (41). What is striking in relation to our findings is that previous authors are agreed in treating Jav. r, the most frequent reflex in the vocabulary examined here,⁴⁰ as the secondary development; see most recently Dahl, § 14.19. Finally, Jav. and other reflexes of *R were investigated by Dyen (1953b), who codified four sets of correspondences as *R₁-*R₄. His proposals are criticized by Dahl, who goes so far as to assert that 'the only inherited reflex of *R is ø in Jav. ...' (§ 17.11). Dyen's *R₄ is not represented in our material,⁴¹ but all the others are: *R₁ (nos. 15, 42), *R₂ (60, 66), *R₃ (76).

The retention of inherited forms in a situation which favours cognate borrowing is not solely controlled by their high frequency or core function. Their chance of retention will be multiplied when no cognates exist in the languages from which loans are being drawn, and equally when semantic shift removes them from lexical equivalence with cognates (so 'descent of amniotic fluid', no. 25). Dahl has nevertheless noticed the high frequency of some of the Jav. items with r < *d/*D. Jav. r < *j, d < *Z, ø < *R all are, or may without great objection be considered as, regular inherited reflexes. What an examination of the Gayo facts suggests is the possibility that w < *b, r < *d/*D have the same status. The evidence now coming to light of the extent to which cognate loans may permeate Indonesian languages at least commends a re-examination of the Jav. data, in which due attention might be paid to the character of the items in which the various reflexes appear, and the co-occurrence of the reflexes which are in question. Here let us notice that whereas discrepancies in many languages are naturally discussed in terms of contact between languages, corresponding discussions of Jav. have more often fastened on contact between dialects, including those peculiar to classes or age-groups. But - as is argued in Dahl's examination of *R - Madurese, Sundanese and indeed Malay have all played a role in the historical ecology of Jav. It is true that r corresponding to PAN *d/*D cannot be accounted for by borrowing from any of these three languages; but, if r is accepted as the regular reflex, d and d can be!

APPENDIX

ITEMS FROM 214-WORD LEXICOSTATISTIC LIST

Note the difference of aim in historical and lexicostatistical comparison; items like 'sand' are classed as inherited but, having undergone semantic change, would not be reckoned as lexicostatistical equivalents of their cognates. Loans from non-Austronesian languages are excluded.

Figures in parentheses are the numbers under which items are cited in the text of the article.

*b.

Inherited: *night* ingi (32), *sand* wau (13), *stone* atu (17), *fire* rara (73), *leaf* ulung (22), *seed* inih (33), *berry* [= *fruit*] wah (34), *mouth* awah (1), *intestines* tětukö, *belly* tukö (31), *hair* uk (26), *new* ayu (15), *to kill* unuh (24). 13 items.

Borrowed: *to wash* basuh ← Mal. basoh < *basuq, *to split* blah ← Ach. blah or Mal. bēlah < *belaq, *tree* batang kayu (see 16), *flower* bungö ← Mal. bunga or Ach. bungong < *buṅa, *to smell* bau ← Mal. bau or Ach. beə < *baSu, *rotten* buruk ← Karo Batak buruk < *buRuk, *right, true* bēnar ← Mal. benar or Ach. böna < *bener, *wet* basah ← Ach. or Mal. basah < *bas[ae]q (cf. Dahl, 1973, § 8.2), *work* buöt ← Ach. buət < *buSat. 9 items.

Ambiguous: *star* bintang (+ Ach. or Mal. bintang?) < *bintaŋ, *to count* bilang (+ Mal. bilang or Ach. bilöeng?) < *bilaŋ. But *all* beh ← Ach. abeh *over, finished* < *Sabis *all*⁴² is marked as a loan by its final. 2 + 1 items.

*d etc.

Inherited: *to walk* rēmalan (45), *day* lo (58), *two* roa (41), *sun* mata n lo (58), *rain* urön (55), *blood* rayoh (42), *to live* urip (51); and putatively *to hear* pēnge (68). 8 items.

Borrowed: *near* dĕkat (see 44), *twenty* duö puluh (see 41), *tongue* delah + Ach. dilah < *dilaq.⁴³ 3 items.

*R.

Inherited: *fire* rara (73), *root* uyöt (60), *blood* rayoh (42), *tail* uki (69), *to hear* pĕnge (68), *new* ayu (15). 6 items.

Borrowed: *left* kiri + Mal. k-iri < *wiRi, *to flow* jaril (with regular metathesis) + Mal. j-alir < *aliR, *to cut* kĕrat + Mal. kĕrat < *keRet, *hundred* ratus + Mal. ratus < *Ratus, *dry* kring + Mal. kĕring < *keRiŋ, *rotten* buruk + Karo Batak buruk < *buRuk. 6 items.

*j.

Inherited: *nose* iung (72). 1 item.

N O T E S

1. The Gayo language boundary is for more than three-quarters of its length with Achinese, to seaward; in the east with the Malay-speaking area centred on Medan, and in the south-east with Karo Batak. Malay loans, which are numerous, are likely to have entered Gayo via Achinese as well as directly.

2. Cf. for Achinese Shorto, 1975, 100-1.

3. Ferrand (1924, 419) quoted item 17 below, in a general context.

4. The following changes in transcription have been made: ch, j, ny, y for Hazeu's tj, dj, nj, j; e, ε, ɔ, o, u for his é, è, ò, ò, oe.

5. Dyen's 'Proto-Malayo-Polynesian' conventions of transcription are followed here; departures from his published reconstructions are indicated in footnotes.

Abbreviations: Ach. = Achinese, Jav. = Javanese, Mal. = Malay, OJav. = Old Javanese, PAN = Proto-Austronesian, UAN = Uraustronesisch, as reconstructed by Dempwolff; cpds. = compounds, dial. = dialectal, (in) phr. = in specific phrases only; C = consonant, V = vowel.

6. But in this sense, with Ach. kaway, Jav. kawal, perhaps a loan from Tamil.

7. Dyen, 1953a, § 112, constructed *abuS. Dahl, § 12.2, has *qabu, and *q- is confirmed by the mainland forms: Shorto, 1975, 90, n. 25.

8. It appears from Hazeu, ix-x, that these spellings represent one phonological form /wön/.

9. Uniting UAN *bayu', *b-ah-ayu', *bəyu', *b-ah-əyu': Dyen, 1953a, § 51.
10. Dyen, 1965, revised his earlier *buhuk to *buS₁₄ek. However, Blust (1969, 91 n.) observes that 'the Bornean evidence would be better accounted for by a reconstruction PAN *buS₁₄uk', and this is equally true of the forms in Ach. and the mainland Austronesian languages.
11. Uniting UAN *bətit' *lower leg*, *bitit' *calf*: Dyen, 1953a, § 57.
12. i.e. UAN *bəŋ[i'].
13. Uniting UAN *binih, *bənih, on the assumption that Karo Batak bñih, Toba boni + Mal. bñih. Some such reconstruction is in any case needed to account for Tagalog binhiq. I take the first i/e of the Gayo reflex to arise secondarily through vowel harmony in view of the most probable formulation of the semivowel rule, below.
14. So Dyen, 1953a, § 133. In 1965 Dyen constructed *'enem[]; Dahl, §§ 13.5, 15.2, constructs *uən₁əm.
15. We cannot explain under this rubric *Sepat *four* (*x₂epat[], Dyen, 1965; *Səpat₁, Dahl, § 14.11) > əpat, ~ mpat in cpds. etc. + Mal. ěmpat < *Se-m-pat. Here ə- probably results from analogy.
16. Two lexemes *to make, do* and *to lift* are to be extracted from *buSat = UAN *bu'at; the forms are distinct in Ach. (Shorto, 1975, 93; 94 and n. 32). *To make, do* is represented by a loan in Gayo; see Appendix, *work*.
17. Note also wöih, dial. öih, öis *water* : literary Jav. warih (< *[bw]ajiq?). (aa represents a long vowel, öi and au short diphthongs: Hazeu, x, xii.)
18. Dyen, 1965, constructed *DewS₃a. I follow Dahl (§ 14.4) in discarding his reconstruction of *ew, and in restoring Dempwolff's *u in this word.
19. *Zeket uniting UAN *dəkət, *d'əkət: Dahl, § 16.3. I add the variant from Gayo and Jav. rakēt *intimately connected*. This derivation appears preferable to Blust's (1971, no. 355) variant *reke(Ct), from Mal., Maranao, which may be susceptible of a loan interpretation.

20. Uniting UAN **dalan*, **d'alan*: Dyen, 1951.
21. Dyen's *[q^h]ade^p (1953a, § 124) is disambiguated by Cham *ana?* in *front (of)*: Shorto, 1975, 91, n. 25.
22. Conceivably **sa-n-DeR*, cf. below; but Ach. has the form without nasal augment.
23. UAN **hudi'*. **q-* is established by Cham *hatèy*: Shorto, loc. cit.
24. Dyen, 1951, revising UAN **hadən*; see further Dahl, § 16.3.
25. Uniting UAN **'udan*, **hud'an*: Dyen, 1951.
26. Uniting UAN **tuduh* to *point out*, less Mal. and Tagalog forms, and *[t]und'uk to *show*: *ibid.*
27. *Ibid.*, uniting UAN **'an(d)av* and **ha(ŋ)g'av*. Dahl, § 16.2-3, prefers **qa(N)g'au*, unhelpful here.
28. Reconstructed as **'uR₂aC[]* in Dyen, 1965, as **uyat₂* by Dahl, § 14.14. *Creep*er by paronymic attraction of **waRe[dj]* (Blust, 1971, no. 438, partly disambiguated by Ach. *uret*).
29. See Shorto, 1975, 91, n. 25. Dyen, 1962, constructed **WaRi(S)*.
30. UAN **ni[l]u'*, adding OJav. *nyiyu* cited by Hazeu.
31. An early date for this metathesis is not incompatible with its occurrence in such later loans as *ruluh*, which may be ascribed to pressure of the phonological pattern.
32. Unless **-nd-* etc. > *-r-* (nos. 37, 57) is excluded between unstressed vowels, the intermediate forms should be **pěřenge* etc.; if so, *tě-něnge* secondarily from *něnge*, the regular nasal form of *těnge*. Late trisyllabism must account for the absence of vowel harmony; contrast no. 33, where contraction will have taken place earlier.
33. Also in Hova *uhi*.
34. Note also *rangang soot*, to be connected with **qajen charcoal* (cf. semantically Ach. *adang*; and for **q-* Shorto, 1975, 91, n. 25); it may

reflect *q-ar-a-n-jəŋ or a reduplicated *q-ar-ajəŋ qajəŋ. The doublet arang *charcoal* is prima facie ← Mal. arang ← Jav. arəŋg, but could perfectly well be < *q-ar-ajəŋ if the loss of *-j- preceded antepenult weakening (and then → Mal.?).

35. Dyen, 1965, constructed *baR₁aH[].

36. Dyen, 1953b, constructed *buR₃ew from UAN *buyav and *buļu' to hunt.

37. I cannot at present account for *teRab (rather *tuRaeb?) to *belch* > tɔrɔp, which appears from its vocalism to belong to the inherited vocabulary. Dempwolff's reconstruction needs revision in view of Cebuano Bisayan dug-ab, tug-ab (and Ach. gōrō?ɔb ~ gōmō?ɔb!); Tagalog tigab to *yawn* belongs elsewhere.

38. The later weakening assumed in no. 73 may then be ascribed to pressure of the phonological pattern.

39. i.e. counting one for two items referred to the same base, and excluding those where the reflexes are not distinct.

40. Also in five items where Gayo has cognate loans: la-ra *virgin* < *DaRa(S), ron, ron-don *leaf* < *DaSun, ratu *ruler* < *[dD]atu, urang *crustacean* < *quDaŋ, pərih ~ pərih *smarting* < *pe[dD]iq.

41. But is, along with *R₁-*R₃, in cognate loans.

42. Correcting UAN *[']abih: Jav. wis, Mal. habis, Cham apì:h, Tagalog abas, Hova avi, etc.

43. Not, as Dahl, § 16.4, *Zilaq in view of Ach. and Cham tàla:h.

BIBLIOGRAPHY

BLUST, R.A.

- 1969 'Some new Proto-Austronesian trisyllables'. *Oceanic Linguistics*, 8/2:85-104.
- 1971 'Proto-Austronesian addenda'. *Oceanic Linguistics*, 9/2:104-62.

DAHL, O.C.

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

DEMPWOLFF, O.

- 1934-8 *Vergleichende Lautlehre des austronesischen Wortschatzes. Zeitschrift für Eingeborenen-Sprachen*, Beihefte 15, 17, 19. Berlin. Dietrich Reimer. [Reprinted Nendeln. Kraus Reprint, 1969.]

DYEN, I.

- 1951 'Proto-Malayo-Polynesian *Z'. *Language*, 27/4:534-40.
- 1953a *The Proto-Malayo-Polynesian laryngeals*. Baltimore. Linguistic Society of America.
- 1953b 'Dempwolff's *R'. *Language*, 29/3:359-66.
- 1956 'The Ngaju-Dayak 'old speech stratum''. *Language*, 32/1:83-7.
- 1962 'Some new Proto-Malayopolynesian initial phonemes'. *Journal of the American Oriental Society*, 82/2:214-15.

DYEN, I.

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

FERRAND, G.

- 1924 'Langues malayo-polynésiennes'. In: A. Meillet and M.
Cohen eds., *Les langues du monde*. Paris. Champion.

HAUDRICOURT, A.G.

- 1951 'Variations parallèles en mélanésien'. *Bulletin de la
Société de Linguistique de Paris*, 47/1:140-53.
- 1964 'Problèmes de comparatisme austronésien: la phonologie
diachronique des corrélations et la reconstruction du
système consonantique'. *Bulletin de la Société de Lin-
guistique de Paris*, 59/1:105-18.

HAZEU, G.A.J.

- 1907 *Gajōsch-nederlandsch woordenboek met nederlandsch-gajōsch
register*. Batavia. Landsdrukkerij.

SHORTO, H.L.

- 1975 'Achinese and Mainland Austronesian'. *Bulletin of the
School of Oriental and African Studies*, 38/1:81-102.

THE FUNCTIONS OF INDONESIAN IN CENTRAL JAVA¹

John U. Wolff

1. Introduction
2. Form of Indonesian used by Javanese speakers to other Javanese speakers
3. Factors which lead to a choice of Indonesian as opposed to Javanese
4. Indonesian in the Peranakan community

1. INTRODUCTION

The population of Central Java (the area around the cities of Yogyakarta and Surakarta on the island of Java, Indonesia) is to a large extent bilingual. First, there is the language native to the region, Javanese, which is the mother tongue of the vast majority of the native-born population, and second there is Indonesian, the national language of Indonesia, which is very widely known and used in daily life, especially by the generations which have grown up since World War II. In this paper we shall discuss the function which Indonesian occupies in the Javanese speech community of Central Java.² What we describe for Central Java also holds true in its broad outlines for other areas where Javanese is spoken, but there are differences which arise from difference in attitudes toward Javanese in Central Java as opposed to Javanese elsewhere.

To start out with, we must distinguish two subcommunities of the Javanese speech community whose speech and behavior differ markedly: the Peranakan, Javanese speakers of Chinese ancestry, and the Pribumi, the rest (for the most part, people of native Javanese ancestry). Although these communities speak the same language and have absolutely no difficulty understanding each other's speech, the function of Indonesian and attitudes toward it in the two communities are so dif-

ferent that we must deal with them separately.

Let us start with the Pribumi speech community (or rather the sizeable portion of the Pribumi community which has a good command of Indonesian and uses it frequently in daily life). In Central Java, as throughout Indonesia, Indonesian is the official language and is used for the kind of functions which official languages commonly occupy: in schools above the first few grades, for any activities involving the government (courts, military, public health, agricultural extension, etc.), for national mass media, advertising, and the like, and other such activities which are carried out on a national scale. Indonesian is also a language of wider communication. It is the language of business above the bazaar level and is the language of communication with non-Javanese. Indonesian, thus, also has the function of creating distance and clarifies the outsider's status as an outsider. The corollary of this is that Javanese (as is true also for other regional languages) is the code for in-group identification, the recognition of a person as 'one of us'; and Javanese have the tendency to use Indonesian with all non-Javanese, even those who know Javanese.

The activities for which Indonesian is typically used are prestigious activities, and in using Indonesian people get the kind of prestige one would get from engaging in these activities. Thus, Indonesian functions as a means for conferring prestige of a certain sort. For example, Indonesian is associated with education. A well-educated person is fluent in Indonesian, and therefore fluency in Indonesian gives one the prestige of being well-educated. Indonesian is also the language of Jakarta, the capital city, and as such is associated with people who have gone places, especially with the elite (most of whom live or have lived in the capital or at least frequently go there). It is the language which non-Javanese use, and since these people in Central Java are often perceived to be economically better off than the ordinary Javanese, their language has an aura of prestige. On account of these factors, Indonesian has become an important code among Javanese: it is used not only to communicate with non-Javanese or to talk about subjects for which Javanese is normally not used, but it is also a device for asserting the status or right to prestige which is ascribed to speakers of Indonesian.

Finally, Indonesian also has the function of avoiding Javanese, where the use of Javanese involves a speech-level choice which would create a feeling of awkwardness. (We shall discuss this function in Section 3.1., below.)

1.1. SPEECH LEVELS

In order to understand the functions which Indonesian occupies in Central Java we must give a short description of Javanese speech levels.³ As is well known, Javanese has speech levels, a series of alternative vocabularies, the choice of which depends upon the relative statuses of the speaker and the interlocutor and their degree of intimacy. A person of low status gives a high level speech to his superior. A high-status person gives low level speech to his inferior. Intimates give each other low level. In other words, the choice of a lower or higher speech level is governed by considerations very similar in type to the ones which govern the choice of tu or vous in French, du or Sie in German, ty or vy in Russian, and so forth. There are differences: whereas the tu-vous alternation involves only these pronouns and verbal agreement in French, the Javanese levels involve close to a thousand vocabulary items. Further, in the European languages there is only a two-way choice: one speaks either on a vous level or on a tu level; in Javanese, level choice is a cline: one may speak on a purely low level or on a purely high level or on any of an infinite number of levels in between, depending on how many and for which particular meanings one chooses the high as opposed to the low alternative form.⁴ In Javanese the picture is further complicated by the occurrence of honorific vocabulary which gives honor to the person spoken or referred to and which is employed or not employed irrespective of the speech level. The presence or absence of honorific vocabulary is also referred to by the term 'speech level'.

2. FORM OF INDONESIAN USED BY JAVANESE SPEAKERS TO OTHER JAVANESE SPEAKERS

With this brief description of what speech levels are, we are now in a position to understand the form which Indonesian takes when used in everyday conversations among Javanese. First, except for formal speech on formal occasions or discussions of an official nature, there is practically no conversation purely in Indonesian (or, for that matter, in many circles, purely in Javanese). Instead, we find a constant switch from Indonesian to Javanese and back. It is possible to do this because the syntactic structures of the two languages are very close. Now this switch is by no means random. The choice of Indonesian forms is governed by factors or motives which we shall examine in the succeeding section. Further, there are certain forms in Javanese which very clearly indicate speech level (functors, demonstratives and pronouns and certain other words of high frequency in conversation), and

these forms have a strong tendency to be put in Javanese so that the speech level is clear. (When the purpose of a shift to Indonesian is to obscure the speech level, as we describe in Section 3.1., below, these forms are put in Indonesian). The following citation exemplifies the shift to Indonesian where Javanese forms are interspersed to preserve the speech level. The speaker has shifted to Indonesian because of the subject (school), but the Indonesian is broken by forms which clearly indicate a high speech level (*meniko this*, *éngkang marker*, etc). In this citation and all citations in this paper forms that are Indonesian and not Javanese are capitalised.⁵

1. Lajeng kulo raq énggéh meniko kéngéng pón wastani dados PANI-TIYA MENERIMA MURÉT. Wonten maléh anu meniko, MASALAH laré éngkang NILÉNYA KURANG BEGITU BAÉQ, ATÓ ANAQ YHANG DHATANG meniko BISA DITERIMA.

further I you-know also umm could passive-marker be-called be THE-COMMITTEE TO-ACCEPT PUPILS. There-is also umm this PROBLEM child who HIS-GRADES NOT-SO VERY GOOD OR CHILD THAT COMES that-one CAN BE-ACCEPTED

'Further, I could also be called the admissions committee. Further there is, umm, the problem of the child whose grades aren't good enough or whether the child that comes can be admitted.'

In a similar way, Javanese forms of lower level are inserted in between Indonesian forms in utterances which are at a lower level.

3. FACTORS WHICH LEAD TO A CHOICE OF INDONESIAN AS OPPOSED TO JAVANESE

In Java Indonesian functions much as the High forms of diglossia function in the four speech communities which Ferguson describes and Javanese functions much like the Low forms. We find that Indonesian is used for most of the functions which Ferguson outlines for the High form in diglossia: personal letters, political matters, university lectures, news broadcasts, newspaper editorials or news stories; and Javanese is used for the functions which he lists as typically performed by Low: instructions to servants, waiters, workmen, clerks, conversation with family friends, colleagues, radio 'soap opera', caption on political cartoon (in publications whose readership is largely Javanese). The only exception is that poetry and sermons are in Javanese (if aimed at a Javanese audience). Whether or not the term diglossia should properly be applied to the Javanese speech community is a question we cannot consider here. There are enough differences in language attitudes and forms which the alternative codes take between the Javanese speech community and Ferguson's illustrations

that the question needs further consideration. Suffice it to say that some functions of Indonesian are much like the function of High speech in diglossia and we will refer to these functions by the term 'high speech in diglossia'. Further, just as in diglossia there is no rigid separation of High and Low (there seems to be a switch back and forth and various intermediate forms), so in Central Java we find switching from Javanese to Indonesian and back again.⁶ Citation one, Section 2, above, exemplifies this formal feature and also the choice of Indonesian elicited by the subject matter.

It is important to note that not only the subject matter leads to the choice of Indonesian but also the setting: conversations conducted in a setting (either in the location or with the mood of a certain setting) where Indonesian is the normal code tend to be in Indonesian (or, more accurately, in an Indonesian - Javanese mixture). For example, formal meetings among people who know each other from work or school are commonly carried out with Indonesian. The following citation is from the meeting of a group of students who are planning an outing. We find a mixture of Indonesian and low-level Javanese. (The normal code among students for social intercourse is low-level Javanese.) Utterances directed to someone are in Javanese, whereas those of an official nature discussing the subject of the meeting are in Indonesian.

2. A. Njók kuwi BIYAYANÉ piro?
and that ITS-COST how-much
'What is THE COST?'
- B. Telóng atós sèket, BERMALAM DIBUATKAN KÉMAH
three hundred fifty OVERNIGHT BE-USED-FOR CAMPING
'Three hundred and fifty, FOR THE CAMPING FEES OVERNIGHT.'
- A. Nèq utang ólèh, óra?
if owe can not
'Can we pay later, or not?'
- B. MAKAN DUWA KALI.
eat two times
'You get two meals.'
- C. Ngandel óra?
believe not
'Do you believe that?'
- D. (to C) Kowé dikongkon meneng whaé.
you ordered be-quiet just
'Hey, it's not your turn to talk!'

Pretending that one is in a certain situation is sufficient grounds for switching to Indonesian. In the following citation the switch to Indonesian is a way of jesting: a young man jokingly upbraids his cousin, a doctor, for not giving a clear explanation of some medical problem. The doctor was intentionally obscure, and the cousin by choosing Indonesian creates the scene of a courtroom:

3. Lha INI TERDHAQWA anu, MEMBERIKAN KETERANGAN YHANG MENYULÉTKAN *and THIS IS CHARGED UMM, GIVE EXPLANATION THAT MAKE-DIFFICULT 'THIS MAN IS CHARGED WITH GIVING EXPLANATIONS WHICH LEAD TO EVEN MORE CONFUSION.'*

This use of Indonesian for official matters thus allows Indonesian to have the function of creating a scene or atmosphere of seriousness. Thus, an utterance can be given importance by choice of Indonesian. The following citation illustrates this use of Indonesian.

A servant whose master had promised him a government job asks permission to quit. The master threatens to withdraw the application for a government appointment if he does not stay on. The threat is with Indonesian. The Javanese itself is on a mid level, not the highest and not the lowest, and very formal in style:

4. Dadi kulo iséq iso MEMPENGARUI ajengo sampan mpón dugi TAHAP semanten niko. SAYA BILANG Dhéq Ratno nyemlangi DIPUTOSKAN SAJA, kulo saget.

so I still can INFLUENCE although you already arrived stage that-far that. I SAY title name are-in-danger-of BE-CUT-OFF just I can 'So, I could still INFLUENCE (the decision about your job), even though it has already reached this STAGE. I AM TELLING YOU, Ratno, you are running the risk of HAVING YOUR APPLICATION WITHDRAWN. I could just do that.'

The use of Indonesian for education leads to its use as a device to show the world that one is not ignorant. In fact Indonesian is frequently resorted to as a self-defense mechanism. The following citation shows a combination of these factors. The speaker (a clerk for our project) reports to an outsider about a co-worker (a student) of whom he is jealous. He uses Indonesian to show that he himself is just as educated as the student and also to underline the officiality and importance of his deductions. The speech level is high, but the speaker becomes so upset as he proceeds that he loses control over the speech levels as well as the syntactic construction.

5. Méng kulo-meniko kuwatos kulo-menio, ong larélaré mriki mahaséswa mriki nio ... és yhéh anu, sami pinterpinter nai mrayu

NANTI KALOQ ORANG ... KELIHATAN ORANG-ASÉNG rós dhidhekatí. Lajeng, yho, NANTI teros MUDHAH DHIANGGAP SEBAGÉ ... DHIJADHIKAN ASISTÈNNYA.

only I am-concerned I, because boys here students here this ...
 OK yes umm, plural smart at flattery LATER WHEN PERSON THEY-SEE FOREIGNER then APPROACH HIM. Then, umm, LATER then EASY BE-CONSIDERED AS ... HE-MADE HIS-ASSISTANT.

'Only I am concerned because these young people, these students here ... um, you know, they are good at flattery, and SO IF SOMEONE, ... IF THEY SEE A FOREIGNER, THEY JUST GO RIGHT UP TO HIM. So, umm, THEN THE FOREIGNER PROBABLY THINKS THAT THEY ARE umm ... SO HE HIRES THEM AS ASSISTANTS.

3.1. INDONESIAN AS A DEVICE FOR AVOIDING SPEECH LEVEL CHOICE

Indonesian is frequently used as a device for obviating references to status or intimacy which Javanese makes clear. As we mentioned above, in such situations Javanese forms which clearly indicate speech level are avoided at all costs. Situations which lead to the choice of Indonesian as a neutral speech level:

- 1) where the relationships between the participants in the conversation are such that two different speech-level choices which conflict with one another are called for - i.e., where two factors which lead to a certain speech level choice are in conflict;
- 2) where there has been a change in the course of time in the relative status or relationship between the participants or where modern life conflicts with older usage;
- 3) where the Javanese calls for a choice between an honorific or its absence but where the status of the person spoken or referred to is too high for the absence of an honorific but too low for the employment of an honorific;
- 4) where someone has used an inappropriate speech level. It is important to note that the switch to Indonesian is not freely available. Because Indonesian functions as an official language or like High speech of diglossia, it creates a feeling of distance, over-emphasis or pretentiousness which must be balanced against the difficulties posed by speech-level choice. Usually, we find that speakers cannot decide and end up switching back and forth from pure Javanese in the wrong speech level to Indonesian (or part Indonesian) and back again to Javanese.

A typical example of the choice of Indonesian as a device for obviating conflicts is the situation in which two persons of widely differing age work or study together in the same institution and have

exactly the same rank. People who work together are on a pseudo-intimate level, very much like co-workers in America who are on a first-name basis: they are not intimate in reality, but custom requires them to speak as if they were. However, it creates conflicts for a person to use low-level speech to someone old enough to be his father or mother, especially if he is, in fact, not really intimate with the addressee, but is just in this pseudo-intimate relationship. In such cases we find constant shift.

Similar behavior is evoked by a change in status. For example, a village school teacher meets an old pupil of his who has meanwhile gone on to get a PhD. The former relation was low-level speech on the part of the teacher and high on the part of the student. They meet again. The teacher should not speak low to a PhD. On the other hand, older persons who have known someone since childhood and were at one time the child's superior should continue addressing the person with low-level speech. Thus when the teacher and the student meet, the student continues speaking high-level speech, but the teacher switches from Javanese low to Indonesian mixture to Javanese high to Indonesian and so forth. The Indonesian mixture is the sort that obscures the speech level.

A momentary shift to Indonesian also may serve the function of avoiding the choice or absence of an honorific where the decision is a difficult one. In the following citation the discussion is about where an American will live. The portion of the utterance which means 'to live somewhere' is put into Indonesian to avoid the Javanese form which clearly ascribes status. *Manggèn live* is not high enough for an American and *lenggah reside* is too high for this particular American, who is just a young student.

6. Lha, lajeng saqmeniko kepéngén nglajengaken Boso Éndónésianipón, patang wulan maléh ngaten. Mawi tèhnik meniko, anu, HIDHÓP JADHI SATU RUMAH TANGGA.

'Anyway, now he wants to improve his Indonesian for four more months. Using that technique, umm, *LIVING IN THE SAME HOUSEHOLD (with Indonesians)*.'

A closely related function to these avoidance usages is the use of Indonesian in response to an interlocutor who one feels has given the wrong speech level. For example, a college student addresses a vendor older than himself with low-level speech. The use of low-level speech to an older stranger is clear evidence of a wide social gap between the speaker (high status) and the interlocutor (low status). If the vendor were to answer at a high speech level it would be a clear acknowledge-

ment that his status is very much inferior to that of the student. On the other hand, if the vendor were to respond with low-level speech, he would be pretending to be in the same social class as the student (i.e., a student himself, clearly not the case). In fact he tries speaking low-level Javanese but is uncomfortable and switches to Indonesian. But since Indonesian gives an aura of pretentiousness, the vendor goes back to Javanese, and so forth:

7. Student: Nèq kowé dhéwé sóq, opo slama rong taón ki yho meneng whaé?

'How about you yourself, will you also abstain for two years?'

Vendor: Ha iyo nó umóm poqé dhaérah nggonaku kuwi, KECUALI ADHA YANG NDHAQ SADHAR, YHA, ITU.

'Yes, that is common in my AREA, in my place. UNLESS THEY DON'T KNOW ABOUT IT (family planning), THAT'S WHAT THEY DO.'

Also related to these functions of Indonesian is the function of Indonesian to cover up incompetence in Javanese. Javanese is a language with strong traditions of correctness, both in dialect and in proper speech-level usage. (In this way Javanese differs strikingly from the Low speech of the diglossic communities Ferguson describes.) Speakers who use a substandard dialect or do not follow the rules of speech-level usage which are considered correct (as is frequently the case) may well take refuge in Indonesian. A case of this type of motivation is the speech of a teacher from a poor peasant family who switches periodically into Indonesian in speaking to his fellow teachers (and others of respectable status). He uses enough Javanese admixture to keep the feeling of speaking Javanese (avoid pretentiousness) but consistently avoids Javanese forms which involve difficulties with honorifics (or their absence). In these cases he uses the Indonesian analogue or leaves the word unspoken. For example, in the following citation he puts into Indonesian the word for 'give' and leaves unspoken the word for 'bought' (since these forms involve a three-way choice depending on the relative statuses of the giver or buyer and receiver):

8. Aspileks. DHIBERI DHULU resep Dhoqter Dewi, malah rong taq ...
'Aspilex. Dr. Dewi GAVE HIM a prescription BEFORE, but I haven't ... (bought it for him yet).'

This discussion by no means exhausts the functions of Indonesian, the motivations for the choice of Indonesian. Indonesian has other

important rhetorical functions: it may be used for softening and euphemism; it may be used as a device for addressing several interlocutors; it may be a device for keeping apart different threads of a narrative. All of these uses spring from Indonesian's function analogous to high speech in diglossia. Interests of brevity prevent us from further discussion of these functions.

4. INDONESIAN IN THE PERANAKAN COMMUNITY

For the Peranakan subcommunity, much as for the Pribumi community, Indonesian functions as the high speech of diglossia, at least for the younger generations educated after the War. For older generations, who rarely received education in Indonesian, these functions of Indonesian are greatly reduced, even though practically all Javanese-speaking Peranakans also can speak Indonesian.⁷ What is more interesting and even startling are the differences between the two communities in the function of Indonesian.

In the first place, Peranakan speech does not employ Javanese speech levels or honorifics (except perhaps on the part of a few members who have learned them much as one might learn a foreign language). Thus for Peranakans Indonesian does not have the functions of speech-level avoidance that we discussed above in Section 3.1. More interesting yet is the use of Indonesian forms as speech level forms. The admixture of Indonesian forms into the Javanese speech makes for a higher level of speech, i.e. speech with Indonesian admixture increases the distance between the speaker and the interlocutor and ascribes status to the interlocutor. A person of high status received utterances on a high level (Javanese with Indonesian admixture) and a person of low status receives utterances on a low level (Javanese with little or no Indonesian admixture). Intimates exchange low level speech. As in the case of Pribumi Javanese speech levels, the Peranakan speech levels are a cline: a speaker may choose a low level (speech with no Indonesian) or high (speech with few or no Javanese forms) or, as is most frequently the case, speak on one of an infinite number of levels depending on how many of the forms are Indonesian and which ones they are. The following citation illustrates this function of Indonesian as a speech level. The utterance is on the high level (with a minimum of forms left in Javanese). However, it contains a quote of what the speaker thought to himself, which is on low level. (Quotations in both Peranakan and Pribumi speech are made on the level of the original utterance, and of course thoughts are on the lowest, most intimate level.) In the examples in this section forms which are Indonesian

only are Capitalised and forms which exist in both Javanese and Indonesian are underlined. Purely Javanese forms are unmarked:

9. Tór LAGI dipikéripikér, nèq DIBAWAQ SINI MALEM, NANTI, APA, "dadiné ra nono nggoné." NDAQ ADA TEMPATé to.

Further THEN I-thought-it-over, if I-BRING-IT HERE AT-NIGHT THEN, UMM, "so no there-is place-for-it" NO THERE-IS PLACE-for-it you know?

'Then I THOUGHT if I BRING IT HERE AT NIGHT, UMM, (I said to myself), "Then there won't be room to park it." You know, THERE WON'T BE ANY PLACE TO LEAVE it.'

4.1. FORMS WHICH INDONESIAN TAKES IN THE PERANAKAN COMMUNITY AS COMPARED WITH INDONESIAN IN THE PRIBUMI COMMUNITY

The use of Indonesian as a speech-level indicator in Peranakan speech leads to what seems at first blush to be a random mixture of Javanese and Indonesian, especially at a speech level intermediate between the highest and the lowest level. Further, the choice of Indonesian has absolutely no connection with the phrase structure, again, as we shall see, a product of its function as a speech level. The following citation shows speech on a mid (neither high nor low) speech level. Brackets separate the immediate constituents in this example:

10. [Mamaé ITU] [SUDAH rondo], [dadiné ngerjaqno japét]. [ITU [séng njladrèni]], [YHA] [SUDAH diwarai]. [ADA séng mboqmbog] [KAN] [SUDAH biasa], [Èl], [wong [lé bikén] [SUDAH suwi]], [YHA].

Her-mother THAT ALREADY widow so makes cakes. THAT the one-who makes-dough, PARTICLE, ALREADY be-told THERE-ARE the-one-who old-women PARTICLE ALREADY experienced name because THE-ACTION-OF MAKE ALREADY long-time particle

'Her mother is a widow, so she has been making these cakes. The ones who make the batter already know how to do it themselves because they have been shown how. There are some old women who are experienced, El, because they have been doing it for really a long time.'

If we compare this citation with our first citation in Section 2, we can see that when Indonesian functions as the high language of diglossia, the switch from Javanese to Indonesian follows very closely the phrase structure of the sentence. The only exceptions are the forms interspersed to keep the speech level clear:

11. Wonten maléh anu meniko, MASALAH [laré [éngkang] [NILÉNYA KURANG BEGITU BAÉQ]], [ató] [[[ANAQ- YHANG DHATANG] meniko] [BISA DHITERIMA]].

there-is also umm this PROBLEM child WHO HIS GRADES NOT-SO VERY GOOD OR CHILD THAT COMES that-one CAN BE-ACCEPTED

'Further, there is, umm, the problem of the child whose grades aren't good enough or whether the child that comes can be admitted.'

These examples also illustrate the difference in the semantic character of the forms which are put into Indonesian as a result of these different functions. When Indonesian functions like a High form of diglossia, it is important to make the sentence readily identifiable as Indonesian. Accordingly, whole phrases are put into Indonesian and markers are largely Indonesian. This tendency is offset by the need to make the speech level clear, in which case these markers may be left in Javanese. Thus we see Indonesian markers in citation one: *-nya his*, *ató or*, *yhang* 'grammatical particle', *bisa can*, *dhi-* 'passive prefix'. (We do find some high-level Javanese markers.) In Peranakan speech the markers are not strong indicators of speech level and are usually kept in Javanese. Thus, in citation ten we have all Javanese markers: *-né the* (= Indon *-nya*), *-no* 'transitive verb suffix', *séng* (= Indonesian *yang*), *lé* 'nominalizing particle'. On the other hand certain forms of high frequency and high communicative importance tend to be put into Indonesian as long as the level is above the lowest. In citation ten we have the following forms in Indonesian: *itu that*, *sudah* 'aspect marker', *yha* 'particle asking if the interlocutor is following', *ada there is*, *kan don't you know*, and only one contentive *bikén make*. These are just exactly the forms which would be left in Javanese in Pribumi speech, as they most clearly indicate speech level. Thus, the function of Indonesian as a speech level in Peranakan speech gives rise to an Indonesian - Javanese admixture very different in type from Indonesian in its function analogous to the high speech of diglossia.

N O T E S

1. The results reported in this paper stem from a research project undertaken jointly by me and Dr. Soepomo Poedjosoedarmo of Sanata Dharma Teacher's College, Yogyakarta. The aim of our project was to determine the various communicative codes which exist in the Javanese speech community and how they function. The basis of our research is tapes of conversations involving people from all walks of life and on a large variety of subjects recorded as the conversations happened to take place, usually unbeknownst to the participants. Our total recordings amount to more than one-hundred and fifty hours, and the quotations here come from these recordings. We hope to publish our full report in monograph form shortly. Our research was financed by the Ford Foundation, to whom we express gratitude. I also express gratitude to Dr. Soepomo. The conclusion drawn here are my own and I am solely responsible for errors, but without his joint effort in the tedious job of collecting the data and without the endless hours of discussion involved in the interpretation the materials and the final write-up I would have had no basis for preparing this paper.

2. We stick to the sociolinguistic concepts which are by now well known and use terms current in sociolinguistic literature. The basic concept is that of a sociolinguistic variable (Gumperz and Hymes: 18-20). Sociolinguistic variables are alternate forms with the same denotation (forms that are referentially equivalent) whose selection carries social significance for some speaker (i.e., the choice of which is motivated by factors of social context such as scene, setting, key, subject matter, speaker, interlocutors, other parties present or involved, et al. - Cf G and H: 35-71). Sociolinguistic variables tend to occur in groupings (co-occurrent clusters - G and H: 21). Groupings of stylistic variants which tend to cooccur we call a code. The sociolinguistic meaning which a code has its function. The function can be

described in terms of the factors which motivate code selection (the factors of scene, setting, key, etc. mentioned above) or it can be described in terms of reactions by members of the speech community to code selection. We follow Gumperz also in our use of the terms speech and speech community (G and H: 53, 54): speech is the sum total of forms which an individual uses at a single occasion or on many occasions (or the surrogate thereof - writing, etc.). A speech community is a group which shares rules for the conduct and interpretation of speech and rules for the interpretation of at least one code.

3. A well-known and good introduction to Javanese speech-level usage is given by Geertz 1960. Our forth-coming monograph (Soepomo and Wolff) will treat these matters in great detail.

4. This statement is an oversimplification. As Friederich points out, there is in Russian a wide range of level distinctions, which in many ways seems to be analogous to the numerous distinctions available in Javanese. The other languages of Europe have (or at least formerly had) such an apparatus.

5. We transcribe all utterances phonemically but omit intonation markings. Since Javanese speakers use the same phonemic system for both Indonesian and Javanese, the same transcription will do for all forms we cite. The following chart gives the Pribumi phonemic system. The Peranakan system differs from the Pribumi mainly in that the apico-alveolar series is merged with the apico-dental series, and we transcribe Peranakan utterances using only the apico-dental symbols.

	labial	apico-dental	apico-alveolar	palatal	velar	glottal
fortis stop	p	t	t̚	c	k	ʔ
lenis stop (pharyngealized)	b	d	d̚	j	g	-
prenasalized voiced stop	mb	nd	nd̚	nj	ŋg	-
nasals	m	n	-	ɲ	ŋ	-
fortis continuents	w	l, r	-	y	-	-
lenis continuents (pharyngealized)	wh	lh, nh, rh	-	yh	-	h

Vowels:

	front	central	back
high	i		u
mid	e		ó
low	ɛ	ə	o
very low		a	

Note on transcription: /t/ is transcribed as th, /d/ as dh, and /nd/ as ndh. /ŋ/ is transcribed as ng; /ŋg/ as ngg; /ʔ/ as q; /e/ as é; /ɛ/ as è; /ə/ as e.

6. Ferguson's original article did not make mention of this important feature of code choice in diglossic communities, but subsequent discussions make it clear that much of the speech in diglossic communities consists of a mixture of high and low. Cf. the discussion to Ferguson, 1962.

7. We have conducted no language ability tests, but we are confident of the accuracy of this statement. The main reason is that Indonesian functions as a speech level in the Peranakan community (as we shall describe below), and Peranakans tend to learn Indonesian forms very early in their speech development. We have recordings of adults using Indonesian as well as Javanese forms to infants, and recordings of children below the age of three using Indonesian (as well as Javanese) forms. I have personally never met a Peranakan who did not speak fluent Indonesian (though not necessarily what the speech community considers 'good' or 'correct' Indonesian). This includes a moron, who often made errors in code choice because of faulty social judgment, but never from lack of control of the Indonesian as opposed to the Javanese codes. Totoks (native speakers of Chinese) insofar as they enter the Central Javanese speech community at all, also seem to know Indonesian at least as well as they know Javanese, and often far better. Peranakans in Central Java most frequently think of themselves as native speakers of Indonesian and not Javanese speakers. In a survey conducted by Willmott in the fifties as to language of the home, approximately four times as many Peranakans gave Indonesian as gave Javanese (Willmott: 112). Willmott's figures do not accurately reflect the extent to which Javanese is used, and it is clear why: Peranakan speech consists of a mixture of Javanese and Indonesian, and since Indonesian forms are the better (the higher level, as we shall see), it is not unexpected that Peranakan respondents should describe their home speech as Indonesian. That the language of the Peranakans in Central Java is

Javanese, not Indonesian, is without question, even if Peranakan utterances are often replete with Indonesian forms. We cannot go into the reasons here. It is enough to say that surely all Peranakans in Semarang, except for those who originate outside of Java, are Javanese speakers. In any case Willmott's figures back up nicely our impression that Indonesian is almost universally known and used in the Central Javanese Peranakan speech community.

BIBLIOGRAPHY

FERGUSON, C.

1959 'Diglossia'. *Word* 15:325-40.

1962 'Problems of Teaching Languages with Diglossia'. In *Georgetown University Monograph Series on Languages and Linguistics: Monograph No. 15, Report of the Thirteenth Annual Round Table Meeting 165-77*. Washington, D.C. Georgetown University Press.

FRIEDERICH, Paul

1966 'Structural implications of Russian pronominal usage'. In William Bright (ed.) *Sociolinguistics* 214-53. The Hague. Mouton.

GEERTZ, Clifford

1960 *The Religion of Java* 248-260. Glencoe, Illinois. The Free Press.

GUMPERZ, John and Dell HYMES

1972 *Directions in Sociolinguistics*. New York. Holt, Rinehart and Winston.

WILLMOTT, Donald E.

1960 *The Chinese of Semarang*. Ithaca, New York. Cornell University Press.

FORMOSAN REFLEXES OF PAN NASAL/ORALS¹

Paul K. Benedict

The P[roto]A[ustro]N[esian] nasal/orals, consisting of nasal + homorganic obstruent or spirant, correspond directly to similar elements in P[roto]M[iao]Y[ao] and have regular reflexes in P[roto]T[hai] (Benedict 1975: Introduction to Glossary). It is a puzzling fact, however, that these elements are poorly represented in all three groups of Formosan languages (East Formosan = Paiwanic, Atayalic, Tsouic). At the 'other end' of the AN domain, the P[roto]E[astern]O[ceanic] reflexes have presented a problem, but the reflexes worked out by Biggs (1965) appear to fit fairly well in the over-all A[ustro]T[hai] pattern:

TABLE 1

PEO. CONSONANTS AND THEIR PAN. AND PPN. CORRESPONDENCES

PAN	p	b	mp	mb	t	nt	d	D	nd	nD	l	r	s	z	c	j	Z	ns	nz	nc	nj	nZ
PEO	p		mp		t	nt	d		nd		l	r		s						ns		
PPN	f		p		t		r		l					s						h		
PAN	k	g	ŋk	ŋg	m	n	ñ	ŋ	w	q	R	h	y									
PEO	k		nk		m	n		ŋ	w	?	R	ø	y									
PPN		k			m	n		ŋ	w	?	ø	ø	(ø)									

TABLE 2
ORAL AND NASAL GRADE REFLEXES OF PEO. PHONEMES

PEO.	*p	*mp	*t	*nt	*d	*nd	*s	*ns	*k	*ŋk
Fijian	v	b	t	d	r	dr	s	c	k	q
Samoan	f	p	t	t	l	l	s	∅	?	k
Tongan	f	p	t	t	∅	l	h	h	k	k
Maori	wh/h	p	t	t	r	r	h	∅	k	k
Saʔa	h	p/q	∅	d	r	d	t/s	d	?	k
Nggela	p/v	mb	t	nd	r	nd	s	h	k/g	ngg
Mota	v/w	p/q	t	t	r	r	s	s	k/g	q

Biggs, who distinguishes between oral 'grade' and nasal 'grade', points out that there are many doublets in the daughter languages (e.g. Fi. kari *scrape*, qari *scratch*) and that the correlation among cognates is imperfect. Many scholars, including Biggs (op cit.) and Haudricourt (1962), have speculated as to a possible underlying morphemic process, with the latter writer (Haudricourt 1965) specifically pointing to the rarity of this 'nasal grade' in Formosa as an indication of a relatively late origin for the feature. This would indeed represent a cogent argument for a late origin, given the generally archaic nature of Formosan phonology, were it not for the correspondences with nasal/orals or specific reflexes in the mainland AT languages; cf. the following table (from Benedict 1975):

TABLE 3
AUSTRO-THAI CONSONANTS (NASAL/ORAL AND NASAL)

AT	Indo-nesian	Form: East	Kadai [>Thai]	Miao-Yao
mp	mp/p	b	b	mp(h)
mb	mb/b	mbʷm	mbʷm	mb
m	m	m	m	m
nt	nt/t	ntʷd	d	nt(h)
nd	nd/d	n	n	nd
n	n	n	n	n
nts	nt'/t'	dz	dzʷz	nts(h)
ndz	nd'/d'	-	n	ndz
ns	nh	ns	n/z	s
nz	nd/d	n	n	ń

Table 3 (cont.)

AT	Indo-nesian	Form: East	Kadai [>Thai]	Miao-Yao
nc	ńk'/k'	-	j	-
nj	ńg'/g'	-	n	-
[ńś]	ń	-	z	ń~n
[ńź]	nz	-	ńj [>ń]	-
ńy	ń	-	ńy [>ń]	-
ń	ń	ń	ń	ń
ŋk	ŋk/k	g	g	ŋk(h)~nts(h)
ŋg	ŋg/g	-	ŋ/ϕ	ŋg~ndz
ŋ	ŋ	ŋ	ŋ/ϕ	ń
Nq	(ŋ)k/k	ŋ~k	G [>ɣ]	Nq(h)
NG	ŋ(g)/g	-	N [>n]	NG~ŋg
N	-	-	N	ŋ

Further details are spelled out in an additional table from the same source (Benedict 1975):

TABLE 4

AUSTRO-THAI REFLEXES FOR ORALS AND NASAL-ORALS

Austro-Thai	*p	*mp	*b	*mb
Hova (IN)	f	(m)p	v	mb
Proto-Oceanic	*p	*mp	*p	*mp
Samoan (PN)	f	p	f	p
Dobu (SEP)	ϕ	b	ϕ	b
Tsou (Form.)	p	b	f	b
Pazeh (Form.)	p	b	b	m
Thai	*p	*b	*b	*m
Lao	p (h.t.)	p (l.t.)	p (l.t.)	m
Miao: PE	*p (h.t.)	*p (h.t.)	*p (l.t.)	*m
Yao: PY	*p (h.t.)	*b (h.t.)	*p (l.t.)	*b (l.t.)

As indicated in the above table, which makes use of the labial series by way of illustration, the P[roto]Y[ao] reflexes for P[roto]AT *mp and *mb parallel those of Tsou (Formosa) but with a tonal distinction, while the PT reflexes throughout are precisely those of Pazeh (Formosa). The material bearing on these reflexes in Tsou, Pazeh and other Formosan languages is in general very scanty, especially for the Atayalic and Tsouic groups. Atayal itself lacks nasal/orals and the few examples

(labials only) found in the related Sedik appear to be largely secondary; cf. Sed. *simburagan spear* (cf. Ata: Ci'uli dial. *sinbaʒanan*); *kəmpah work* (cf. *komopach*, id., cited by Bullock 1874); *?mpusal 20* (cf. Ata: Ci'uli *mapusal* < *ma/pusal); *səmpu count*, from *s/m/[i]pu[ɣ] (cf. Pai. *səmpu*, Bun. *maʒipul*). There is, however, one bit of evidence for Atayalic *b as the reflex of PAN *mp: cf. Ata: Ci'uli *bokwi the back*, *bukwi back/behind*; Sed. *bukwi the back*; *back/behind*, from *buku[r] via *bukui; IN *puŋku[r] *hind-part* (Ja. *behind*); PEO *mpuku: Fi. *buku pointed hind-end, tail* < PAN *(m)pu(ŋ)kur, a comparison greatly strengthened by the Pazez cognate (below). Atayal also has the doublet *baŋ-lyeq ~ maŋ-lyeq hibiscus*, from *(m)ba[ɣ]u-; IN *bayu = *(m)bayu (Hova *varu = baru*), id. Initial *(N)q- is to be reconstructed for Atayalic for the following root: Ata. *qumah to work the field*, *qəmayah field (dry)*, from *q u m/al/ah (Ata. -y- < *-l-); Sed. *kəmpah* < *kumupah *work* (above), *kəmpahan field (dry)*, from *kumu[h]- < **Nqumu[h]- < *Nquma[h]- (vocalis assim. as in Rukai, below) < Atayalic *(N)qum[ah]; East Formosan *qumah and (Sai.) *qum/qumah *field (dry)*; also (Pax.) *quma/mah and (Ruk.) *qumu/umah [*< *qumah/qumah by assim.*], id.; also (Ami Sir.) *m/qumah *work*; IN *?uma *cultivated field*; PEO *?uma *garden; to work/plant/clear ground*; PPN *kumala *sweet potato (= the cultivated plant)*; contrast same infix form in Ata. = *the cultivated land*, from *kum/al/a < *Nqum/al/a[h]; PT *man *potato/sweet potato/yam* [generic term], from *[qu]mal[ah]; PLi *mwal *sweet potato*, from *umal < *[q]umal[ah] < *[q]um/al[/ah]. Finally, Ata. r < *d ~ *dz apparently reflects an earlier *nt in *pareq < *padiq intestinal worm*, from *pantiaq < *pantiaq (regular Formosan shift); PT *tiaq*, id., from *[]tyak < PAT *[pa](n)tiaq; also an earlier *nts in *ramo? < *dzamu? blood*, from *ntsamu? (see below), while Ata. ŋ reflects an earlier *ŋg in *ŋəbuŋ wasp*, from *ŋubūŋ (stressed form); P[roto]Li (Kadai) *voŋŋ < *bwoŋŋ < []obŋŋ *classifier for bee*; P[roto]M[iao] *ŋg[ai] *wasp/bee*, from PMY *ŋg[ou] < *ŋgow < *ŋgob[ŋ] < PAT *ŋgobŋ.

The Tsouic evidence for reflexes of PAN nasal/orals is even more limited, if anything, with almost no reliable comparative material bearing on the problem. Tsou and Kananabu occasionally show clusters (probably secondary) of this type; cf. Tsou *nsou breath*, *eansoua breath*; Tsou *mpusku 20* from a prefixed *ma/ form (cf. Sedik, above; also Kan. *mapusau*, Saa. *mapua* < *ma/pusal); Kan. *kinti thorn*; Kan. *ŋkou monkey* (cf. Tsou *ŋu?xou ~ ŋu?hou ~ ŋhou*); Kan. *tsuŋkutsu bridge* (cf. Saa. *tokoso*). As indicated in the above table, Tsou regularly has initial f- for PAN (and Tsouic) *b; Tsou initial b- appears to be the reflex of PAN *mb- as well as *mp-; cf. Tsou *ba?i grandmother* (also in East Formosan: Kuv. *ba:i*, id.); IN *bayi = (m)ba[?]i (Tg. *ba:?i mother/*

grandmother, Ja. bayi *infant*); Tsou boki *penis*, from *buki < *mpuki < *mpuNqi (see below); IN *puki *vulva* (these two body parts are often found in association). Another bit of evidence indicates that the Kan. reflex of PAN *mb- is m-; cf. Kan. mumu *breast*; Ata. bubu?, id. (perhaps also Sed. bubu *mother*, if not related to *grandfather* root, below), from *(m)bu(m)bu[?]; cf. also Saa. maŋusipi *alive/live* < *ma/Nqutsip; IN *ʔuɖip, id. < AT *(N)quzip. This scanty evidence for the labial series suggests that *mb- and perhaps *mp- are to be reconstructed at the Proto-Tsouic level, but the necessary comparative material is not at hand.

In contrast to Atayalic and Tsouic, the East Formosan languages present a sizeable body of evidence for specific reflexes of PAN nasal/orals. The evidence is best for the labial series, very scant for the velar and post-velar series, as seen in the following:

PAN *mp: the regular reflex appears to be *b; Sai. -mp- is probably secondary in rampuz *ten* (Og.-As.; cf. lanpəz [Tsuchida]) and perhaps also in ʔampowa *why*; Puy: Rikavong dial. has təmpok < *təmpuk *hit (with fist)* possibly related to PAN *(m)puk(m)puk, as represented by Sir. bouchbouck (= bukbuk), id.; IN *pukpuk *beat with a tool*. The *b reflex appears in Paz. bukun *the back; back/behind*, from PAN *(m)pu(ŋ)kur (cf. the Atayal cognate, above); also Pai. vuvu, Ami fufu, Fav. boeboe (= bubu) *grandparents* (perhaps also Sed. bubu *mother*, above), from *bubu < *mpumpu; IN *pu *sir* [term of respect for older males], a doublet of *ə(m)pu *grandfather/grandson* (= *grandparent/grandchild*); PO *mpu: Fi. bu *father's parents* (Bau), *mother's mother* (Nausori), also *mpumpu: Motu bubu *term of address to elders* [the normal term for 'grandparent' in Malekula and other parts of the New Hebrides]; PT *phu ~ *bu *male (human, animals, birds)*, from *(m)pu; P[roto]K[am]S[ui] (Kadai) *bu *father*, from *mpu; also Puy: Hinan dial. tabi *mortar* (= *the pounder*), Ruk. /bi:bi *wipe* (partially reduplicated form); IN *ta(m)pi *remove dust and chaff, winnow* (NgD. *pound rice*); PEO *ta(m)pi: Fu. ta/tafi *sweep* (Demp. *cleanse*), tapi *wash* (Demp.), Sm. tafi *sweep*, ma/tafi *wiped away* (Demp.), tapi *wipe off* (Demp.); PT *wi fan, from *(bi)bi < *[ta]mpi/mpi (cf. Rukai); Diol pi (l.t.) (*comp.*) fan, from *bi < *mpi; PKS *bi: Mak pəi (l.t.) fan, pəi pəi *winnow (rice)*, from [ta]mpi and *[ta]mpi/mpi; P[roto]Y[ao] *pei *to skim off [rice water]* (Highland Yao), from *pi < *[ta]pi; also Ami maʔfar [< *ma/q[ə]bar], Fav. mabart [app. = *maʔbar] *fly* (Thao has marfað < *ma/r/ba[R]); IN *hampa[r] = *[ʔ]ampa[ɣ] *spread out, stretch out* (Ml. hampar *spread out (as mats, carpets)*, hampar/an *carpet*; Ja. hampar (*spread the wings*) = *swoop (birds)*; Chamorro gwafag *mat* [Dahl cit.]); PEO *ʔempa (from ə vowel doublet; cf. Ami): To. ʔe/ʔepa *spread out*, Sm. epa *mat*, from

PAN *q[a]mpa[R]. In addition to the above, the *b reflex also appears in East Formosan (Ami Pai. Ruk. Paz.) *(ta)balana(n) *arm/armpit/shoulder/wing*, from *(ta)mp/al/ana(/n), also *balanalan arm/armpit (Pai. only), from *mp/al/ana/l/ana/n (cf. IN *lənən *forearm*, app. from the reduplicated root through stress reduction [*a>ə] in the final segments); N. Thai *vian *hand*, from *pwian < *[t]paŋ; Laqua (Kadai) paŋ *arm*, Lati (Kadai) ta pə ~ peŋ *shoulder*, from *(ta)p[a]ŋ; P[roto]W[estern]M[iao] *mpaŋ *hand/arm* from PMY *mp[aa]ŋ, from *[t]ampaŋ < PAT *ta(m)paŋ[a].

PAN *mb: appears in two Saisiat entries, both with significant cognates in IN; cf. Sai. tombok (Tsuchida), tomobok (Og.-As.) (with epenthetic vowel) < *tumbuk *kill* (= *run a person through*); IN *tə(m)buk *perforate*, from *ta(m)buk (stressed form), also *tumbuk *thrust through* (Hova *perforated*) (assim. form, as in Saisiat); PEO *tompu: Fi. tobu *hole in river bed*, from *tambu[k]; PT *ʔbuak *tube/pipe (water)/quiver/container for chopsticks* (Tho-Nung), also *hmook *quiver/tube/gun* (SW), from *hmuak; both from an earlier *ʔ(m)buak < *[t]a(m)buk < PAT *[t][a](m)buk also Sai. rimbutul *grass*; cf. IN *rumpu, id.; Ong-Be (Kadai) bət ~ bət < *pot, id. < PAT *[ro](m)pot. This element also appears in Thao qumbu *snake* (app. isolated form), but it is best represented in Bunun; cf. Bun. tambo < *tambu *wet (rice) field* (app. isolated form), also the following two forms, which have IN cognates showing *m rather than the anticipated *mb: Bun. ſumbaŋ < *tumbaŋ *breathe*; IN *t'umaŋət *spirit* (Geist), from *tsumaŋ/t (unstressed suffix); PT *[h]ma[a]ŋ *imaginary evil spirit* (Ahom); N. Thai *mwaŋ *genie*, from *[ts]um[b]aŋ; PKS *hmaŋ *spirit/demon/ghost* (Mak), from *s[]maŋ < *ts[]m[b]aŋ; also Bun. tomboś ~ tumbəś (assim.) *body louse*, Thao tumbus, id. (app. loan from Bunun), Sai. somäh < *tumas, id., Ami tumus ~ tomus ~ toməs *body louse/flea*, Kuv. tuməs *body louse/gnat* (cf. Bunun/Thao assim.); IN *tuma *louse, clothes louse*; PT *hmat *body louse/plant louse/flea*, also *bat *kind of dog louse* (Ahom), from an original *ʔ(m)bat; PKS *hmat *flea*; P[roto]Li *m[ua]t, id., from *[]um[b]at < PAT *pru(m)baś.² In the latter root the PT doublet furnishes support for the reconstruction of PAT medial *-mb- rather than *-m-; a third root, however, with Bunun medial -m-, must be reconstructed with *(m)b- on the basis of the East Formosan forms: Pai. quvis ~ qovis, Puy. ʔovi ~ obi, Ruk. obisi ~ ubisi ~ ubusi [assim.], Thao qu:mis, Ami kuməs ~ koməs, Bun. komis *pubic hair* (Pai. also *axillary hair*) < *(N)qu(m)bis (see below for initial); also (second. voicing < reduplication) Sai. romis < *[G]umis *beard*; Atayal has kumis *body hair/public hair/feather/down/beard*; Tsou has fusifusi < *[qu]bus/[qu]bus (cf. Rukai assim.), also the apparent doublet: muʔmuu ~ mʔumʔu *body hair/beard*, from *qumu[s]/qumu[s] (cf. Kan. mu:mus *beard*); IN *kumit'

[< *kumits] *beard*, also (second, voicing; cf. Saisiat) gumi [< *gumis], id.; PT *hmooy *pubic/axillary hair* (Shan also *beard*), from *qomiy < *qomis; Dioi mi (h.t.), from *hmi < *q[ə]mis (with unstressed vowel); also PT *m[o]m *beard* (Tho-Nung; cf. Tsou/Kan.); Dioi məm ~ mum *body hair/beard*; Pli *mii̯m *beard/whiskers* < PAT *qo(m)bi(t)s. In still another root medial *(m)b- must be reconstructed at the PAT level on the basis of the PMY cognate despite the lack of mb as a reflex even in Bunun (the IN cognate is lacking): East Formosan *tūmay; Tsouic *tsumay; Atyalic *k[r]umay (Sed. kumay ~ sumay) *bear*, n.; PT *hmi, id., from *qmay (*-ay > *-i after *q), also the doublet *hmiay *bear, large sp.* (Lao), from *hmyay < *qəmay (with unstressed ə vocalism); N. Thai *mui (h.t.), from *hmuy < *qmuay (influence from *q, as in PT) < *qumay < *[kr]um[b]ay; PKS *?muy ~ *myay, from *qumay ~ *qəmay (cf. the PT doublet); Li (White Sand) moi; PMY *krop < *krup < *krub[ay] < PAT *kru(m)bay. Three East Formosan roots must be reconstructed with *mb or *(m)b on the basis of comparisons with IN and/or Atyalic: Paz. muta:mak *cut (person)*, from *m/tambak; IN təbak *chopping knife* (Tg.), *to clear forest (= chop down trees)* (TB.), from *tabak (with stressed vowel); PT *vak *cut/chop/mince*, from *[]bak < PAT *[t][a](m)bak; also Ami kahmmau ~ kahmau ~ dahmau ~ ahmau *light (weight)*, from */h[ə]mbaw; Ata. lahbao ~ ləhbao < */h[ə]baw, id.; PT *?baw ~ *[h]maw (Ahom), id. < PAT *h[ə](m)baw; also Sir. mani < *mali[s], Kuv. balis, Thao baliθ *iron*, from *(m)balis (with 'irregular' final, possibly through assimilation to the front vowel, and a likely loan from Atyalic); Ata. baliq < *ba[x]liaq *iron/metal*; IN *bat'i = *(m)bat'i, id. (Hova basi *musket*, although Dahl considers this word a loan), from *(m)bakli(aq); also (doublet with unstressed vowel) *bət'i (Hova vi *iron*); PEO *pesi: Fi. vesi *name of a spear*; PT *hle̯k < *hliak, id. (second. vowel shortening before original *-q); N. Thai *mwa, id., from *mba[h]liak; PKS *qhlet < *qhliak, id. (cf. Thai); Lakkia (Kadai) khyək, id., from *m[]khyak < *mb[a]khliak; PMY *hliā?, id. < PAT *(m)baxliaq; cf. also the Formosan forms for *banana*: *bulibul (Paz. Sed.), *b[u]li- (Kuv.), *bulbul (Pai. Puy. Ruk. Bun.), *ta/bulbul (Kan. Saa.), *bulbil (Sir.), *bilbil (Thao) and *bilpil (fav.), from *buli/buli (with varying degrees of vocalic assim.), perhaps from an earlier *mpuli/(m)puli (cf. the Fav. form), bringing these Formosan forms in line with IN *pun[t]i [<*puniti < *pulipili < *puli/puli], id., and PT *pli *banana flower/bud* < PAT *p[u]li/p[u]li. Finally, an initial *mb- with epenthetic vowel (cf. Sai. doublet for *kill*, above) is apparently represented by W. Ruk. movoroko *monkey*, from *mbu[ɣ]uk < *mbə[ɣ]uk (assim.); IN *bə[r]uk = *bəyuk, id. (Demp: Ml. bəru? NgD. beruk; add Old Ja. wruk, Busang vərək

and [metathesis] Kadazan gobuk, Timugon gabuk); Ong-Be (Kadai) ma-lu, id. (not analyzed); PLI *nuk, id., from *mruk < *mbruk; Laqua (Kadai) tək, id., from *prək < *br[u]k < PAT *(m)b[ə]yuk. The occasional appearance of *m rather than *mb as a (reconstructed) reflex in IN or Formosa (*spirit, louse, body hair, bear*, above) remains a problem, but both IN and PEO (PPN) appear to have *m as a reflex for PAN *mb and even *mp (apparently via *mb) in rare instances; cf. IN *baliw *change* (*Veränderung*): Tg. baliw *deranged* (= *altered mind*), maliw *alteration* (*Verwandlung*), a doublet from *mbaliw; IN *pu(n)dul = *(m)pu(n)dul *cut off* (Hova mundru), also *putul (app. for *putul) *break off/cut off*; PPN mutu *cut off/ended*, from *mbutu[1] < *mputu[1]; PT *toon *castrate* (Shan also *break off the head of a plant*), from *[]oton < PAT *[(m)p]oto!; IN *pat'aŋ *pair*; PPN *maasaŋa *twin*, from *mba(a)saŋa < *mpasaŋ/a; PT *sooŋ *two* (perhaps via back-loan from Chinese) < PAT *[(m)pa]ts[a]ŋ.

PAN *nt: Bunun and Saisiat maintain this element in one well-represented root: Bun. bintoqan, Sai. bintö?än, Pai. vituqan, Puy. vitu?ən *star* (cf. also Paz. bintun < *bintul; Bullock [1874] cites bintul; app. an infix form); IN *bi[t]u?ən. The nasal feature appears in IN as well as Formosa in the following: Kuv. bu:tis *calf*, Thao buntuθ < *buntis (assim.) *lower leg*; IN *buntit' *shin* (NgD. buntis), also (with unstressed vowel) *bə(n)tit', id. (Ja. wəntis, Ml. bētis) and (assim. form; contrast Thao assim.) *bitit' *calf* (TB. bitis, Hova vitsi) < PAN *bu(n)ti(t)s; Thao also has bantaθ *leg*, an apparent loan from Bun. bantaś *leg/foot*, possibly cognate with the above root. Thao has additional forms with medial *-nt-; cf. Thao tantu:qas *older brother* < PAN *tuqas *old/elder*; Thao muntu:muq *lie/recline*, contrasting with Thao mu:taq *vomit* (Bun. motah, Pai. mut'aq, Puy. mutaq, Ami ma?otaq < *m/utaq); cf. also Bun. śibuntus *pull*; Bun. parantahon *tie*; Sai. komontotol < *k/muntutul *push* (see below). The anticipated *nt > *d shift appears to be rare; cf. Paz. dulut *tail*, from *ntulut < *(bu)nt/1/ut; IN buntut, id.; PT *sut *end*, from *t/r/ut < *[bu]t/1/ut; PKS *[]zot (h.t.) *tail*, from *[]nsot < *[bu]nt/1/ut; Ong-Be (Kadai) tu?, id., from *sut; PLI *sut, id.; also Sai. komontotol *push* (above), Paz. mudu:dun, id., apparently from *muntu:dun < *muntu:tul (assim.) but note Yami padundunun, id., possibly of similar origin (< *pa/ntuntun/n < *pa/ntuntul/n).

PAN *nd: exceedingly rare in Formosa: cf. Thao mundada:n *walk* (cf. Bun. mudada'an); Bun. ?indin *this*, probably of secondary origin. Three roots, all with the *d ~ *n type of alternation, require reconstruction with *(n)d: W. Ruk: Maga dəga, Tona nəga, Mantauran nəka (also

Thomson [1873] *denga*) *one*, from *(n)də(ŋ)ga; PT *(h)nŋ, id., from *()ndŋ < *()ndyaŋ < PAT *() (n)da(ŋ)g[a]; also Pai. *budas* ~ *vudas*, Sai. *bunaz* ~ *bonaz*, Thao *bu:nał*, Paz. *bunat*, Fav. *bonnad*, Bun. *da'as* < *bu(n)daj *sand*; PT *draay, id.; PLi *phəw, id., from *bəw[draj]; Lati *ñá*, id., from *nd[r]ǎ[j]; < PAT *bəw(n)draj; Sir. /lamag/ *burning* (Dutch *brandt*), Fav. *ramal burnt field*, Ami *namar fire* < *(n)damay = *(n)damay *resin/light/torch*; PPN *marama *light*, *malama *moon/month* (o. *shine*), both from *ma/(n)dama[ɣ]; PT *hmay *burn* < PAT *[(n)da]maR.

PAN *nt: represented by d in Kuvalan and by t in Thao, which regularly has θ for East Formosan *t (= C [Dyen]), from an earlier cluster (CL): θaw *man*, θa:qi *excrement*, θu:may *bear*, ma:θa *eye*, ma:θay *die*, ku:θu *head louse* (> s before m: smaqs *sew*); contrast Thao *ɬari:ŋa ear*, from *ntaliŋa; IN *taliŋa; PEO *(n)taliŋa: Fi. *daliŋa*, Sa. 'iŋiŋe; PKS *qha < *qh[r]a[liŋa] < PAT *(N)qraŋ[a]; also Thao *ɬa:ri taro*, Kuv. *dari sweet potato* (Taintor 1874), both from *ntali (Ami *tali*, Bun. *taye*, Buk. *ta'i taro*; also Atayalic: Sed. *sari*, id.); cf. IN *talet', id., probably from *talit' < *talits (with unstressed *i > *ə) but the final is irregular in any event; PEO *ntalo(s), id.; PMY *ndoi *yam* (also *edible tuber, potato and sweet potato*), from *ndawi < *ndali (*l > w is regular MY shift), with final corresponding to the Formosan form, from PAT *(m)p[r,l]ali ~ *mb[r,l]ali.

PAN *nts: represented by East Formosan *dz (and Ata. r < *dz), on the basis of the following PMY correspondence: Pai. *d'amuq* ~ *d'amoq*, Puy. *damok*, Sai. *ramo*, Paz. *damu* < *dzamu[?] *blood*; Ata. *ramo?* < *dzamu?, id.; PMY *ńcyaam, id., from *[]ntsaam (palatalized) < PAT *()ntsa[a]m[u?].

PAN *ns: appears in two Bunun forms, one with an excellent IN (and PT, PMY) correspondence; cf. Bun. ?insunun *push* (app. isolated form); Bun. *binsah* < *binsaq, Sai. *biŋsi* < *bi[n]sa[q] (assim. backwards) *seed*; IN *bini? = *binhi? (Tg. *binhi?*), id., from *binha? (assim. as in Saisiat), also *bani? (doublet with unstressed vowel); PT *van ~ *ban, id., from *b(w)an < *(ba/)ban[saq] < *bin[saq] (partial redupl. and assim. forwards; contrast Sai. and IN); PY *sa? *sesame* (the seeds *par excellence*), from *[bin]saq < *PAT *()binsaq.

PAN *nz: initial *(n)z- is to be reconstructed for East Formosan for the following root: Pai. *zalum*, Sai. *ralom*, Paz. *dalum*, Thao *θa:ðum*, Bun. *danom* (n < *l), Kuv. *ranum* (n < *l), Sir. *salom water*, from *zalum, but Ami *nanom* < *nalom (assim.) < *nzalom, Puy. *źanum* < *ńźalum (assim.); IN *[dd]anum = *[z]anum, id.; PPN *lanu *liquid/fresh water*, from *ndanu[m]; PT *nam ~ *naam *water*, from *nza(l)am (regular intervocalic loss of *l) < *nza(l)om (vocalic assim.); Dloi *ram*, from *nram <

*nz(a)lam; Lakkia (Kadai) num (without assim.); PL1 *nom (also without assim.) < PAT *() (n)zalom.

PAN *ŋk: rare in occurrence, the few available examples found either in isolates or in secondary developments; cf. Bun. taŋkinuð *back/behind*; Sai. riŋkəlan *thigh*; Sai. miŋkoriŋan *woman* (cf. korkoriŋ *child*); Puy. taŋkar(/kar) *dry* [field]; Puy. ?iŋku < *in-ku I (cf. ?inu < *in-su *thou*, ?inmu *you*). There is some evidence for the anticipated *ŋk > g shift; cf. Sir. vugot *bind*, from *bugu[ts] < *buŋku[ts]; IN *bəkət' = *bəŋkət' *bundle* (Tg. bigkis *bundle*; *bound, to tie*, Ml. bėrkas, TB. borhos *bundle*), from *bəŋkut' (assim. backwards), also *buŋkut' *bundle* (assim. forwards); PEO *po(ŋ)ko: Sa. ho?o *to bind magically*, i/hə?o ~ hoko *bundle*; PT *koot *embrace/enlace*, from *okot < *[b]okots; Ong-Be (Kadai)kət (l.t.) *tie up/enlace/knot*, from *gət < *oŋkot < *[b]oŋkots < PAT *[b]ə[ɣ][ə]-(ŋ)kots; also Pai: Kulalo t'akit (Dahl cit.), Makazayazaya (Ferrell) t'agit *knife/sword*, from *ta(ŋ)kits; cf. IN *ta(ŋ)kit' *ward off* (NgD. taŋkis *parry*, takis *push away*); PT *kiit ~ *kiat (Siamese/Lao) *hinder/prevent*, from *[]ki(y)at; also *geet (Shan) *head off/thwart/hinder/obstruct*, from *ŋkeet []*ŋkiat < *[ta]ŋkiats < PAT *[ta](ŋ)ki(y)ats.

*ŋg: see above (*wasp/bee* under Atayal, also *one*); no certain examples of the anticipated *ŋg > *ŋ shift for East Formosan have been uncovered.

*Nq: extremely rare in East Formosan; cf. Bun. madaŋqas ~ madaŋxas ~ madaŋhas *red* (app. isolated form). The regular shift appears to be to k, rarely to ŋ (cf. Saa. *live*, above); cf. *(N)qu(m)bis *body hair/beard* (above); also Ami: Tauran poki *vulva*, from *puNqi, as shown by the doublet: Bun. po?o < *puqu < *puqi (assim.), id. (Tsou boki *penis* < *mpuNqi, above); IN *puki *vulva*; PT *hi, id., from *hNi < *hNGi, from *hNqi (second. nasalization) < *[pu]Nqi (second. aspiration); KS *hŋi: Mak hŋei (h.t.), from *hNi (as in PT); PM *b[ou]? ~ *bi? (White Miao), from *buq[i] ~ *biq[i] (second. voicing and vocalic assim.; contrast the Bunun assim.); also Pai. tsaqi, Puy. ta?i, Ami ta?e, Sai. sə?i, Thao θa:qi, Fav. tshe < *taqi *excrement* (Tsuic *taqi), but Ruk. tsaki ~ tsake, Bun. take, Paz. saik [< *saki], from *taNqi; also Sir. taiŋ [< *taŋi], from *taŋgi < *taNqi (second. nasalization; cf. Saa. *alive/live*, also PT *vulva*, above); IN *ta?i = *(n)ta?i *excrement*, but Kalagan takki, Singhi toki (cited by Blust 1973), from *taN?i = *taNqi, with *Nq > k as in East Formosan; PEO *(n)ta?e, id.: F1. de, Sa. ae; PT *khi *excrement/defecate* (irreg. for *xi < *qi); N. Thai *yai *excrement*, from *Gay < *Nqay < *[]aŋqi; PKS *()qe < *()qay, id.; Lakkia (Kadai) kwei (l.t.), id., from *[G]ei < *[Nq]ei < *[]aŋqi; Ong-Be (Kadai) kai (l.t.) id., from *[G]ay (as in N. Thai); PL1 *hay,

Table 5 (cont.)

Paz.	b	m	{ ^{nt} _d	n	-	d	-	-	-	k
Pa1.	b/v	-	-	-	-	d'	-	-	g	k
Am1	f	m	-	n	-	-	-	n	-	k
Kuv.	-	m	-	-	d	-	-	-	-	-
Sir.	b/v	m	-	-	-	-	-	-	g	{ ₀ ^k
Fav.	b	-	-	n	-	-	-	-	-	-

NOTES

1. Sources as in Benedict 1975, principally Ogawa and Asai (1935) (in Japanese), with normalized phonetic orthography, and Ferrell (1969), which includes the citations from early sources; IN forms cited after Dempwolff (1930), with modifications by Dyen (notably ? for medial and final h) and the occasional indications of nasal/orals, especially *(m)b for *b (almost always on the basis of b rather than v as a reflex in Javanese or Hova); PEO reconstructions along the lines suggested by Biggs (1965); PT and PMY reconstructions as in Benedict 1975. Abbreviations: AN Austronesian; AT Austro-Thai; Ata. Atayal; Bun. Bunun; Demp. Dempwolff; E. Eastern [Rukai]; Fav. Favorlang; Fi. Fiji; Fu. Futuna; h.t. high tone; IN Indonesian; Ja. Javanese; Kan. Kanakanabu; Kuv. Kualan; l.t. low tone; Ml. Malay; N. Northern [Thai]; NgD. Ngaju-Dayak; Og.-As. Ogawa and Asai (1935); Pai. Paiwan; PAN Proto-Austronesian; PAT Proto-Austro-Thai; Paz. Pazeh; PE Proto-Eastern [Miao]; PEO Proto-Eastern Oceanic; PKS Proto-Kam-Sui; PLi Proto-Li; PM Proto-Miao; PMY Proto-Miao-Yao; PN Polynesian; PPN Proto-Polynesian; PT Proto-Thai; Puy. Puyama; PWM Proto-Western Miao; PY Proto-Yao; Ruk. Rukai; Saa. Saaroa; Sai. Saisiat; Sed. Sedik; SEP Southeast Papua; Sir. Siraya; s.t. same tone; SW Southwest [Thai]; Tg. Tagalog; To. Tongan; W. Western [Rukai].

2. A similar but distinct root is represented by Bun. *tumbi* ~ *tumbe flea/bedbug*, from **tumbi*[q] (irreg. loss of final); cf. Puy. *ta^htumuq*, Pai. *tsatsumuq bedbug*, from **ta^htumbiq* (with vocalic assim.); also Ata: *Squ^hliq sumiq*, *Ci^huli lumi?*, Sed. *tsumiq body louse*, from **CLumbiq* (CL= consonant cluster) < *(CLa(CLumbiq; no MY or Kadai cognates of this root have yet been uncovered.

BIBLIOGRAPHY

BENEDICT, Paul K.

- 1975 *Austro-Thai: Language and Culture, with a Glossary of Roots*. New Haven. HRAF Press.

BIGGS, Bruce

- 1965 'Direct and Indirect Inheritance in Rotuman'. *Lingua* 14: 383-415.

BLUST, Robert A.

- 1973 'Additions to 'Proto-Austronesian addenda' and 'Proto-Oceanic addenda with cognates in non-Oceanic Austronesian languages' - II. *Working Papers in Linguistics* 5/3:33-61.

BULLOCK, T.L.

- 1874 'Formosan Dialects and their Connection with the Malay'. *China Review* 3:38-46 (cit. from Ferrell, 1969).

DAHL, Otto C.

- 1973 *Proto-Austronesian, Scandinavian Institute of Asia Studies Monograph Series, No.15*.

DEMPWOLFF, Otto

- 1930 *Vergleichende Lautlehre des austronesischen Wortschatzes, Bd.3: Austronesisches Wörterverzeichnis, Beihefte zur Zeitschrift für Eingeborenen Sprachen Bd.19*.

ELKINS, Richard E.

- 1974 *A Proto-Manobo Word List*. Paper presented at the First International Conference on Comparative Austronesian Linguistics, Honolulu.

FERRELL, R.

- 1969 *Taiwan Aboriginal Groups: Problems in Cultural and Linguistic Classification*. Institute of Ethnology. Academia Sinica. Monograph 17 (includes Comparative Vocabulary).

HAUDRICOURT, André G.

- 1962 'Comment' on Capell: 'Oceanic Linguistics Today'. *Current Anthropology* 3/4:410.
- 1965 'Problems of Austronesian Comparative Philology'. *Lingua* 14:315-29.

OGAWA, A. and Erin ASAI

- 1935 *Myths and Traditions of the Native Tribes of Formosa*. Taihoku, Taipei. Taihoku Imperial University. In Japanese.

TAINTOR, E.C.

- 1874 'The Aborigines of Northern Formosa'. *Journal of the North China Branch of the Royal Asiatic Society* 10:53-88 (cit. from Ferrell 1969).

THOMSON, J.

- 1873 'Notes of a Journey in Southern Formosa'. *Journal of the Royal Geographical Society* 43:97-107 (cit. from Ferrell 1969).

TSUCHIDA, Shigeru

- 1968 *Kavalan Word List and Saisiyat Vocabulary* (cit. from Ferrell 1969).

SOUND SYMBOLISM AND KHASI ADVERBS

Lili Rabel-Heymann

Khasi, a language of the Austro-Asiatic language family and spoken by about 300,000 speakers in the State of Meghalaya (formerly Western Assam), displays a fascinating array of verb qualifying adverbs. They fall into two classes, general adverbs and proper adverbs (Rabel 1961, p.63). General adverbs, which are few in number, occur with many different verbs: *bha good, well, no' away, 'e' very*, as in *snew bha to feel well, le' no' to go away, klet no' to forget completely, lnno' 'e' to be very much surprised*. Proper adverbs act like satellites to a limited number of verbs only. These are the adverbs under discussion in this paper. Some verbs have a large number of satellite adverbs, others have only one or two. The number of adverbs seems to depend on the frequency of the verb itself. *kren to speak* has 77 adverbs, *yaaj to go* 66, and *le' to do* has 150. The adverbs of *le'* can be subdivided into those indicating the effects of mental states and feelings on the bodily features (38), dress and movements (61), cleanliness, thrift and extravagance (14), disposition of mind and character (37) (Bareh p.64 ff).

Since it would lead too far to demonstrate 150 examples I will use *'iw to smell* with its 15 satellites as an illustration.

- 'iw bi' to have an aromatic sweet smell*
- 'iw hek to be smelly*
- 'iw jlep to smell of rotten fish*
- 'iw jli to smell nauseating*
- 'iw krteŋ to have a smell of a thing bearing that name*
- 'iw khoŋ to have an unpleasant smell*
- 'iw lhop to smell close and stuffy*
- 'iw li to smell of fresh fish*
- 'iw lnəaw to have the smell of a long unoccupied house*

?iw pdu to have a slightly bad smell

?iw riem to have a very bad smell

?iw sit to smell of urine

?iw sien to have a pungent smell

?iw sma to have a rotten smell

?iw tuŋ to stink

Could it be possible that the vowel i in the only two pleasant smells has any significance? or that the initial cluster j|- in jlep and jli denotes any similarity in meaning?

In preparing cards for my Khasi dictionary I noticed that adverbs - and even some verbs - of related meanings began with the same consonant clusters. The idea came to mind that certain sounds actually symbolized certain meanings in adverbs, and I began to assemble and classify initial consonant clusters according to large and not too narrowly defined semantic territories. There seem to be more Khasi adverbs with initial j and l than with any other sound; I therefore tabulated all adverbs with j- and l-, adding adverbs with k- for a more complete sampling.

I am herewith presenting my material which shall serve to illustrate a theory and not to prove a fact. I will indicate counter-evidence in appropriate places and will caution against too much confidence in my theory.

Sound symbolism works in more than one way: vowels in the adverb's major syllable (Rabel 1961 p.20) reflect the size of the person or object described by the adverb; these vowel values are quite predictable and can be compared to associating the i-vowel in the Germanic languages with smallness (teeny-weeny, itsy-bitsy, nitty-gritty) or Italian adjectival suffixes in -ino, -one, -accio and -uccio.

In the limited corpus chosen for this paper certain initial consonants or consonant clusters may relate to shape, position, protrusion, movement, noise, speed, and unpleasantness.

SHAPE. Adverbs whose first syllable begins with liquid + nasal are especially prominent in their designation of a person's or an object's shape.

lmoñ cut short (of small things)

lmuñ cut short (of big things)

lman-lman fleshy (of a young baby)

lmum-lmum fleshy and plump (of an adult)

lme? young looking in spite of old age

From these five examples one could isolate a morpheme lm- and assign a meaning 'short and fleshy' to it. Further examples permit a more detailed morpheme lmmVC¹:

Immaŋ *without ornament, cut off*

Immen *hands without fingers, feet without toes, cut off.*

The foregoing examples are contradicted by Immu' *branch with green leaves* which seems to imply *with projections intact* so that I must revise Imm- to mean *with respect to projections or extensions*.

From Imphooñ *smooth and small* and Imphuoñ *smooth and big* one can isolate Imph- with a semantic designation *smooth*. The morpheme Iŋ^oVC is exemplified by Iŋ^oan *short and stout*, Iŋ^oren *with hands on hips like a fat man*, and Iŋ^our *like a big circle*. These three examples together with the noun ka Iŋ^ouñ *circumference* would allow me to assign a meaning 'of roundish proportions' to the morpheme. Unfortunately, I have an item that does not fit semantically, Iŋ^ou' *to be full of fire and smoke* and perhaps I have to withdraw Iŋ^oVC.

I^oVC *round and short* derives from I^oañ *squat* (persons), I^oar *short and round*, and I^oooñ *beautifully round*. IkVC *plump and squat* derives from lker-lker *pulpy, flabby*, lkur-lkur *nice and plump*, lkay *plump*, and lkut *short and square* like a log. (lko *to fall apart* would not fit this morpheme formula and therefore presents no counter-evidence.) IkhVC *like a lonely left-over piece* derives from Ikhəŋ *lonely and helpless*, Ikhəy *to be tender* (of meat), Ikhəj *like a lonely little piece left over*, Ikhuj *like a good solid piece left over*, and ki Ikhit *small and useless things*.

POSITION. An object, and sometimes a person or an animal, may be lying, standing, hanging or trailing:

yeŋ kryep or kryap *standing like a diseased fowl*²

yeŋ krŋəŋ *stand like a truant boy*

šəŋ krŋəŋ *stand like a modest little boy*

šəŋ krnaay *stand in the same place*

krpaŋ *aside, separately*

bha briw khrwan khrwe' *be tall and handsome in stature.*

We could isolate kry-, krŋ-, krn-, krp-, and khrw- as morphemes 'in a standing position' because krw- and krs- do not refer to standing, so that we cannot call the minor syllable kr- by itself a morpheme.

kntep *cowardly, i.e. with drooping tail like a frightened fox or dog*. kntar, kntup and knteer all refer to *a way of standing cowardly like a frightened fox or dog with a drooping tail*, so that kntVC can be called *standing cowardly*.

knthew *with a long tail, in ragged clothes*, knthay *beautifully dressed with the boh-khaila flowing*, and knthem *flourishing an overlapping dhoti* can be grouped under knthVC *standing or walking with something trailing*.³

knji? *raised on tiptoe*, knjin *on tiptoe*, knjoon *high up on the top*, and knjoñ and knjuoñ *aloft* clearly contain a morpheme knjVC with the meaning *raised up*. This morpheme could perhaps be incorporated into knnVt *like a bird's uplifted tail* as illustrated by knnit *lifted up like a bird's tail*, knnet, knnut *with a long and pointed tail of a big bird*.

In all the foregoing examples the minor syllable kn- alone cannot be isolated because of its high frequency as the first syllable in countless nouns and verbs with unrelated meanings.

There are three possible morphemes for *in a hanging or dangling position*, krdVC, ldVC, and lJVη: deη *to wear* is used with krdat *dangling*, krdot *hanging like a small ball*, and krduť *like a swollen earlobe*, krdaap *as if wearing something hanging*, krdeep *as if wearing an old coat*. ηat *to fall into, to be caught in a snare* is used with ldat *hanging upside down*, ldet *dangling like a dead bird*, and ldit *dangling like a small bird*. sdien *to hang* is used with lday *dangling like a parrot*, ldoy *hanging like a small ball*, and lduy *hanging like a heavy ball*. sdren is also used with ljin, ljeη, ljaη, ljoη, and ljuη all meaning *hanging down without touching the ground*.

knrVC could be glossed with *horns*: yeη knram or knrem means *to stand with horns or moustache ends pointing in opposite directions* and yeη knreη *stand with long horns* (like a stag or bull).

lbVC apparently means *with flesh exposed* from ka lboη *the thigh*, lbiη *fleshy and big*, lbaη *with trousers rolled up*, lboj and lbuj *short and naked*, and lbew *naked*.

lηkCVC means *to lie flat on the ground* deriving from lηkraη and lηkreη *lying on the back like a drunkard*, thia? lkdaη *lying helpless on the back*, lηknap, kηkniap, lηknep *lying thin and flat like a coin*, and ka lηknep *the surface area*.

PROTRUSION. kntir *with a small tail*, knton *like a small hill* and kntoy *like a small protruberance* can be united under kntVC *protruding*.

SPEED. I found two examples where the final sounds could be isolated as possible morphemes CCek and CCoη both with the meaning *suddenly*: baam klek *to eat suddenly*, phrnaay krek *to sparkle suddenly*, yeη khek *to stop abruptly*, and jlek *at once*. Further, mi? phnoη or mi? jmoη *to appear suddenly*. Since such morpheme structures do not fit into the overall language pattern I will dismiss these examples as accidental.

NOISE. Only two examples can be united under k[h]rVC⁴ *with a crash*: kllon khram *to fall with a crash* and kram-kram, krum, krim *with a crash*.

lnthVC has the meaning *destroy by throwing noisily* derived from lnthem *throw with stones*, lntheer *throw with stones continuously*, and lnthaaw *smash to pieces*.

UNPLEASANTNESS. All words beginning with ja- and not continuing with l²-, n-, or r- designate some kind of unpleasantness, most often carelessness or dirtiness: japhrut-japhret *in a careless way*, jawut-jawet *work carelessly*, ja-krak *in a careless manner*, kren ja-kter *slur words like a drunkard*, ja-knep *to be sloppy*, ja-khlia *to be dirty*, ja-lbaañ⁵ *with a dirty appearance*, ja-tbe *with something dirty sticking on the lips*, ja-pnek *sticky*. Although I would like to establish a morpheme for *careless* and *dirty* I believe I cannot formulate it precisely enough to be convincing.

In the case of lη-, jrh-, and jrt- these doubts are not necessary: jrtVC sitting *rude and lazy* is derived from jrtien *to stretch the legs while sitting* [impolite behavior], and jrten *to be lazy and indolent*. thia² lηgay *to lie without a care*, šon lηηe² *to sit with head thrown backwards*, le² lηgia² *with head thrown backwards*, and šon lηnej *to sit firmly with the whole weight of the body* all express the same unacceptable, bad behavior. jrhVC means *nose or throat trouble* or *having the symptoms of a head cold* and is derived from jrhi² *to sneeze*, jrhi² *to choke, gag*, and jrho² *to cough*.

My final example lhVC means *a hollow with bad smell* derived from lhi² *to fester more and more*, lho² *to smell bad*, lhen *stagnant*, lhin-lhin *overflowing*, lhon *stagnant*, and ka lhu² *valley, hollow*. snew lhop *stuffy stifling*, and šit lhep *to be oppressively hot* belong to the same morpheme.

I am on much firmer ground in associating vowel symbols with semantic content. The glosses in two Khasi dictionaries (Nissor Singh 1906, Kharkongngor) and explanations by my informants bear out the following vowel associations in adverbs which are about 90 percent reliable:

- i, ie: *small, light, dainty, cute;*
- ia: *young, tender, thin, flat, pretty, modest, quick, attractive;*
- o, oo, uo: *small, short, smooth, slender, delicate, feeble,*
pertaining to babies;
- a, aa: *medium-sized, small and flat, pertaining to stout persons;*
- e, ee: *big, strong, tall, pertaining to grown-ups;*
- u: *big, large, fat, ungainly, plump, fleshy, heavy, pertaining*
to old persons.

Many adverbs occur in partial or complete ablaut series ranging from two to five related forms. The following vowel alternations occur: i/e, a/e, e/u, o/u, a/u, i/a/e, i/a/o, i/e/u. Sometimes the entire series of i/ia/o/a/e/u occurs. Although all of my examples can be found in the Khasi dictionaries, it is my feeling that proficient Khasi speakers can make up new forms on the spur of the moment using the

vocalic associations mentioned above. Examples:

- i/e saat bha jlin *to comb neatly*
 jron jlen *to be handsomely tall*
 dem lpen *to lie like a snake*
 thia[?] lphen *to lie like a fat and hefty person*⁶
 thia[?] lphin *to lie like a small and thin person*
- a/e jem btaj-btaj or btej-btej *to be sticky and muddy*
 yaaj šen-šen *to walk staggeringly*
 yaaj šan-šan *to walk like a drunkard*
- o/u spon knthron *to wear something crest-like* (e.g. a turban)
 (by a small person)
 spon knthron *to wear s. like a turban* (by an older person)
 yaaj dot-dot *to walk shakily like an old man* (small)
 yaaj dut-dut *to walk shakily like a bigger person*
 sŋaaj tkor-tkor *to be fat and tender* (a small animal)
 sŋaaj tkur-tkur *to be fat and tender* (bigger animal)
- ia/e riam jriaŋ or jreŋ *to dress nicely*
- ia/a/e lŋkniap, lŋknap, lŋknep *flat on the ground* (for objects thin
 and small, thin and medium, thin and big)
- i/a/e jrtreŋ ltir, ltar, lter *to stretch one's legs at full length*
 rit jkiŋ *to be small, but straight and slender*
 jron jkaŋ or jkeŋ *to be tall and slender*
- i/e/u dem knnit *to sit with the tail lifted up* (small bird)
 jron knnet or knnut *with the tail lifted up* (bigger bird)
- oo/uo bha briw lmphooŋ or lmphuoŋ *to be beautifully smooth*
- a/o/e jem krwap, krwop, or krwep *to be easily bent, pliable*
- a/ia/o/e/u dem *to lie, sit* is used with
 lŋsnap *the way a moth sits on the window pane*
 lŋsniap *the way a beautiful butterfly sits*
 lŋsnop *the way a baby chick sits*
 lŋsnep *the way a frog sits*
 lŋsnup *the way a turtle sits*

Perhaps one third of all adverbs are reduplicated, most often resulting in a simple repetition of the adverb itself: jem jer-*jer to be soft like jelly*, jem tiŋ-tiŋ *to be pliable like a string*, jem smop-smop *to be soft and slimy*, dap klaŋ-klaŋ *how a place full of water glitters*, jaaw jop-jop *to fall in drops*, šoŋ kjoŋ-kjoŋ *to sit high up in a tree* (like birds), la[?] ley-ley *to be able to without fail*, wan kum-kum *to be actively engaged in*, luŋ jiap-jiap *to be very young and tender*, jhi[?] jaw-jaw *to be in wet clothes*. Most of the reduplicated forms seem to have reiterative force.

A few of the reduplicated adverbs display two different vowels, thus combining e.g. the vowel for small o with that for huge u or that of huge with that of ungainly e. It is possible that such an adverb has a generalizing function: khmi[?] luŋ-luŋ leŋ-leŋ *to search hurriedly*, le[?] japhrut-japhret *to act in a careless way*, Trey jawut-jawet *to work carelessly*, baam jrup-jrap *to eat at the same time*, yatrey jhup-jhap *to work simultaneously*, ja-tmut ja-tmat *to be covered with dirt and grime*, jirwit-jirwat *act in a round-about way, wasting time*.

i/a vowel alternations usually occur in a doubly reduplicated adverb: knja[?] jik-jik jak-jak *to be absolutely silent* or knja[?] jlim-jlim jlam-jlam with the same meaning, thŋan liw-liw laaw-laaw *to be hungry or craving for something*. The concept of size, i.e. small vs. medium sized, cannot possibly play a role here, rather I believe that the adverb has a superlative force.

Such reduplicated adverbs are sometimes used as verbs or as nouns: loy-loy luy-luy *to be tender-hearted or innocent*, li-li pem-pem *to be destitute*, ki tum ki tam *articles of furniture*, ki tiar ki tar *things lying around*, lŋkrum lŋkram with the same meaning, ka lat-lat *a current*. I assume the function of double reduplication here is distributive.

The Khasi language has adverbs for every situation in life, for every smell, noise, size, speed, mood, and feeling. It takes training and love for the language to become proficient in the use of the proper adverb for every occasion. Since many young Khasis are being educated entirely in English-speaking colleges it is unfortunately possible that such versatility of language is becoming a lost art for the young generation.

SUMMARY

ja---	<i>unpleasant, dirty</i>
jr ^h --	<i>with symptoms of a headcold</i>
jrt--	<i>sitting rude and lazily</i>
knj--	<i>raised up</i>
knn-t	<i>like a bird's uplifted tail</i>
knr--	<i>with horns</i>
knt--	<i>protruding; standing cowardly</i>
knth--	<i>standing or walking with something trailing</i>
krd--	<i>in a hanging or dangling position</i>
krn--	<i>in a standing position</i>
kr ^o --	
krp--	
kry--	
krw--	
k[h]r--	<i>with a crash</i>
l ^o --	<i>round and short</i>
lb--	<i>with flesh exposed</i>
ld--	<i>in a hanging or dangling position</i>
lj--	
lh--	<i>a hollow with bad smell</i>
lk--	<i>plump and squat</i>
lkh--	<i>like a lonely left-over piece</i>
lm--	<i>short and fleshy</i>
lmm--	<i>with respect to projections or extensions</i>
lmp--	<i>smooth</i>
lnth--	<i>destroy by throwing noisily</i>
l ^o --	<i>sitting impolitely</i>
l ^o '--	<i>of roundish proportions</i>
l ^o k--	<i>lying flat on the ground</i>

N O T E S

1. V= vowel, C= consonant
2. The first word in these examples is a verb; verbs are listed whenever possible and omitted in case of doubt.
3. knt- and knth- should probably be considered related morphemes, both semantically and phonologically; Eugenie J.A. Henderson apparently agrees to some degree with Pater Schmidt's theories on this topic - see her forthcoming papers in *Oceanic Linguistics* (Proceedings of the First International Conference on Austro-Asiatic Linguistics, January, 1973, Honolulu.)
4. krVC and khrVC are to be considered related - see note 3.
5. The hyphen indicates an internal open juncture to show that the words are not to be read *jal-baañ, *jat-be, *jap-nek.
6. lpen and lphen probably represent related morphemes - see note 3.

BIBLIOGRAPHY

BAREH, U Mondon

- 1929 *Khasi-English Course and Grammar for Schools and Colleges.*
Shillong. Ri Khasi Press.

KHARKONGNGOR, U Iarington

- 1968 *Ka, Dienshonhi, a Khasi-Khasi Dictionary.* Shillong. Ri
Khasi Press.

RABEL, Lili

- 1961 *Khasi, a Language of Assam.* Baton Rouge. Louisiana State
University Press.
- 1968 'Redundants in Khasi'. *Studies in Indian Linguistics.*
Annamalainagar. Deccan College and Poona University.

SINGH, U Nissor

- 1906 *Khasi-English Dictionary.* Shillong. Eastern Bengal and
Assam Secretaria Press.