Malayic varieties of Kelantan and Terengganu: description and linguistic history

Wu, J.

CHAPTER 7

Phonological history

7.1 Introduction

This chapter examines the phonological histories of NEPMs. A top-down approach is adopted: by comparing pre-existing reconstructions of the ancestral language with their reflexes in the present-day daughter languages, sound changes that have taken place over time are established.

The hypothetical ancestral language of all contemporary Malayic varieties is Proto Malayic (henceforth PM), which has been reconstructed using the historical comparative method by Adelaar (1992) [1985]. As the internal subgrouping of Malayic is much debated (see §1.2), PM is considered the most recent common ancestral language from which NEPMs have developed, and it serves as the point of reference for establishing phonological changes.

While Adelaar's PM reconstructions are widely accepted, some of them have been subject of controversy. Meanwhile, a wealth of additional data has become available over the past few decades, which should be used to evaluate the original reconstructions. This chapter therefore also includes a critical examination and updates on the reconstructions of some PM phonemes and phonotactics, particularly by drawing on relevant NEPM data.

The representation and transcription of data in this chapter adhere to the following conventions. Unless otherwise noted, PM reconstructions are taken from Adelaar (1992), and higher-order reconstructions such as Proto Malayo-Polynesian (PMP) and Proto Austronesian (PAn) are cited from the online Austronesian Comparative Dictionary (ACD) (Blust & Trussel 2023). These reconstructions are marked by an asterisk "*". Reflexes in NEPMs are from my own databases, transcribed in phonemic forms and presented in italic. ITM data are represented by the Dusun subvariety, but other subvarieties are also discussed when necessary. Inferred forms at intermediate stages from a reconstructed ancestral language to present-day varieties are marked by a plus sign "⁺". A cross sign "ˣ" indicates an unreconstructable or unattested form. In addition to reflexes in NEPMs, SM cognates are given for comparisons. In cases when no PM reconstructions are available, SM cognates are taken as close approximations to probable PM reconstructions. Glosses in tables are for the reflexes in present-day languages. In cases of semantic shifts, glosses for PM reconstructions or some of the cognates are offered in notes. A vertical bar "|" indicates a historical morpheme boundary which has been fossilised. In contrast, a hyphen "-" marks a morpheme boundary in the active morphology. A triple-dot "..." means no data is available, and a dash "–" means no reflex or no cognate.

The remainder of the chapter starts with the introduction of PM phonemes and phonotactics in §7.2. Reflexes of PM consonants and vowels in disyllables are presented in §7.3 and §7.4 respectively. Specific types of changes involving syllable reduction (trisyllables to disyllables and some disyllables to monosyllables) are discussed in §7.5. The relative chronology of sound changes is established in §7.6. §7.7 summarises this chapter.

7.2 Some notes on the reconstruction of PM

As mentioned in §1.2, the reconstruction of PM in Adelaar (1992) was primarily based on six Malayic varieties: SM, Minangkabau (central-west Sumatra), Banjar Hulu (southeast Borneo), Seraway (southwest Sumatra), Iban (northwest Borneo) and Jakarta Malay (Java). SM had long been used as material for higher-order reconstructions in the Austronesian family and a yardstick for comparisons with Malay-like languages, while the other five varieties were selected because they "show important phonological retentions from PAn/PMP" (Adelaar 1992: 3).

Table 7.1 and Table 7.2 display the phoneme inventory of PM. The consonant inventory included nine stops, four nasals, two fricatives, two liquids and two glides.[46] The liquid *r was phonetically a velar (or possibly uvular) fricative. Doubts have been expressed about the reconstruction of *ʔ, which will be discussed in more detail in §7.3.1.

Table 7.1: Consonant inventory of PM (Adelaar 1992: 102)

		Labial	Dental	Alveolar	Palatal	Velar	Glottal
Stops	voiceless	*p	*t		*c	*k	*ʔ
	voiced	*b		*d	*j	*g	
Nasals		*m		*n	*ɲ	*ŋ	
Fricatives				*s			*h
Liquids				*l		*r	
Glides		*w			*y		

(j = IPA /ɟ/, y = IPA /j/, r = IPA /ɣ/)

In the vowel system, four monophthongs *a, *ə, *i and *u were reconstructed along with two word-final "diphthongs" *aw and *ay, which were in fact sequences of *a + glide *w or *y (see §7.3.5).

Table 7.2: Vowel inventory of PM (Adelaar 1992: 102)

	Front	Central	Back
High	*i		*u
Mid		*ə	
Low		*a	

(diphthongs: *-aw, *-ay)

At the word level, PM lexemes were typically disyllabic. In Adelaar's reconstruction, all four vowels could occur in either penultimate or final syllables, but the presence of *ə in final syllables has been contested (see §7.4.3).

[46] Anderbeck (in print) considers *c and *j as palatal affricates. Regardless of whether *c and *j are stops or affricates, it is safe to say that their phonological behaviour is comparable to other sets of stops. Therefore, I maintain Adelaar's reconstruction here.

Word-medially, heterosyllabic consonant sequences could consist of a nasal + a homorganic stop or a velar nasal + *s. Some trisyllabic lexemes were also reconstructed in PM. Adelaar used a capital *A to indicate an uncertain reconstruction of some antepenultimate vowels, which could be either *a or *ə.

7.3 Reflexes of PM consonants

Initial and intervocalic consonants in PM are relatively well-preserved in NEPMs, whereas PM final consonants have undergone a number of changes with an overall tendency of merging, debuccalisation and reduction. PM word-medial consonant sequences have also been reduced. Table 7.3 to Table 7.5 present overviews of regular reflexes of PM consonants and consonant sequences. A few unconditioned phonemic splits are marked by a slash "/". For instance, the loss and retention of PM *-ʔ (if PM *-ʔ is reconstructable) are not conditioned by any clear environments. The reflexes of *-ʔ are thus given as *Ø/ʔ*, with *Ø* preceding *ʔ* indicating that *Ø* is the more common outcome.

The following sound changes are summarised for the development of consonants from PM to NEPMs:

1) Merger of all final stops (*p, *t, *k and sporadic retention of *ʔ) to a glottal stop;
2) Sporadic loss of final *ʔ;
3) Merger of all final nasals to a velar nasal *ŋ*, which was subsequently lost in KM following *a and *ə;
4) Loss of initial *h and intervocalic *h between non-identical vowels;
5) Merger of final *s and *h to *-h*;
6) Loss of initial *r preceding *i and *u in Dusun;
7) Loss of final liquids;
8) Loss of final glides;
9) Reduction of nasal + voiceless obstruent sequences to the obstruent components.

Reflexes of PM consonant phonemes and sequences, as well as the sound changes summarised above, are discussed in more detail and exemplified in the following sections. Some irregular changes will also be noted.

Table 7.3: Overview of reflexes of PM initial and intervocalic consonants

PM	Env.	KM	CTM	Dusun	SM
*p		*p*	*p*	*p*	*p*
*t		*t*	*t*	*t*	*t*
*c		*c*	*c*	*c*	*c*
*k		*k*	*k*	*k*	*k*
*b	*a_*a	*w*	*w*	*w*	*w*
	elsewhere	*b*	*b*	*b*	*b*
*d		*d*	*d*	*d*	*d*
*j		*j*	*j*	*j*	*j*
*g		*g*	*g*	*g*	*g*
*m		*m*	*m*	*m*	*m*
*n		*n*	*n*	*n*	*n*
*ɲ		*ɲ*	*ɲ*	*ɲ*	*ɲ*
*ŋ		*ŋ*	*ŋ*	*ŋ*	*ŋ*
*s		*s*	*s*	*s*	*s*
*h	#_	Ø	Ø	Ø	*h*/Ø
	$*V_x\_*V_x$	*h*	*h*	*h*	*h*
	$*V_x\_*V_y$	Ø	Ø	Ø	*h*
*l		*l*	*l*	*l*	*l*
*r	#_*i, *u	*ɣ*	*ɣ*	Ø	*r*
	#_*a, *ə	*ɣ*	*ɣ*	*ɣ*	*r*
	*V_*V	*ɣ*	*ɣ*	*ɣ*	*r*
*w		*w*	*w*	*w*	*w*
*y		*y*	*y*	*y*	*y*

Table 7.4: Overview of reflexes of PM final consonants

PM	Env.	KM	CTM	Dusun	SM
*p		*ʔ*	*ʔ*	*ʔ*	*p*
*t		*ʔ*	*ʔ*	*ʔ*	*t*
*k		*ʔ*	*ʔ*	*ʔ*	*k*
*ʔ		*Ø/ʔ*	*Ø/ʔ*	*Ø/ʔ*	*Ø/k*
*m	*a, *ə_#	*Ø*	*ŋ*	*ŋ*	*m*
	*i, *u_#	*ŋ*	*ŋ*	*ŋ*	*m*
*n	*a, *ə_#	*Ø*	*ŋ*	*ŋ*	*n*
	*i, *u_#	*ŋ*	*ŋ*	*ŋ*	*n*
*ŋ	*a, *ə_#	*Ø*	*ŋ*	*ŋ*	*ŋ*
	*i, *u_#	*ŋ*	*ŋ*	*ŋ*	*ŋ*
*s		*h*	*h*	*h*	*s*
*h		*h*	*h*	*h*	*h*
*l		*Ø*	*Ø*	*Ø*	*l*
*r		*Ø*	*Ø*	*Ø*	*r*
*w		*Ø*	*Ø*	*Ø*	*w*
*y		*Ø*	*Ø*	*Ø*	*y*

Table 7.5: Overview of reflexes of PM consonant sequences

PM	KM	CTM	Dusun	SM
*-mp-	*-p-*	*-p-*	*-p-*	*-mp-*
*-nt-	*-t-*	*-t-*	*-t-*	*-nt-*
*-ɲc-	*-c-*	*-c-*	*-c-*	*-ɲc-*
*-ŋk-	*-k-*	*-k-*	*-k-*	*-ŋk-*
*-ŋs-	*-s-*	*-s-*	*-s-*	*-ŋs-*
*-mb-	*-mb-*	*-mb-*	*-mb-*	*-mb-*
*-nd-	*-nd-*	*-nd-*	*-nd-*	*-nd-*
*-ɲj-	*-ɲj-*	*-ɲj-*	*-ɲj-*	*-ɲj-*
*-ŋg-	*-ŋg-*	*-ŋg-*	*-ŋg-*	*-ŋg-*

7.3.1 PM stops

All PM stops except for *ʔ occurred initially and intervocalically. In these two positions, stops are generally retained without changes except that intervocalic *b between two *a underwent lenition and became *-w-* (which also happened in SM). Examples are given in Table 7.6.

Table 7.6: Reflexes of PM initial and intervocalic stops

PM	KM	CTM	Dusun	SM	Gloss
Initial stops					
***p**asir	***p**ase*	***p**ase*	***p**asi*	***p**asir*	'sand'
***t**aŋan	***t**aŋɛ*	***t**aŋaŋ*	***t**aŋaŋ*	***t**aŋan*	'hand'
***c**aciŋ	***c**aciŋ*	***c**aciŋ*	***c**acɛiŋ*	***c**aciŋ*	'worm'
***k**ulit	***k**uleʔ*	***k**uleʔ*	***k**uliʔ*	***k**ulit*	'skin'
***b**akar	***b**aka*	***b**akɔ*	***b**akɔ*	***b**akar*	'to burn'
***d**agiŋ	***d**agiŋ*	***d**agiŋ*	***d**agɛiŋ*	***d**agiŋ*	'meat'
***j**auh	***j**aoh*	***j**aoh*	***j**auh*	***j**auh*	'far'
***g**igi	***g**igi*	***g**igi*	***g**igɛi*	***g**igi*	'tooth'
Intervocalic stops					
*a**p**i	*a**p**i*	*a**p**i*	*a**p**ɛi*	*a**p**i*	'fire'
*pu**t**ih	*pu**t**eh*	*pu**t**eh*	*pu**t**ɛih*	*pu**t**ih*	'white'
*pə**c**ah	*pə**c**ɔh*	*pə**c**ɔh*	*pə**c**ɔh*	*pə**c**ah*	'to break'
*sa**k**it	*sa**k**eʔ*	*sa**k**eʔ*	*sa**k**iʔ*	*sa**k**it*	'sick'
*də**b**u	*də**b**u*	*də**b**u*	*də**b**əʊ*	*də**b**u*	'dust'
*ba**b**ah	*ba**w**ɔh*	*ba**w**ɔh*	*ba**w**ɔh*	*ba**w**ah*	'below'
*hi**d**up	*i**d**oʔ*	*i**d**oʔ*	*i**d**uʔ*	*hi**d**up*	'to live'
*ta**j**əm	*ta**j**ɛ*	*ta**j**aŋ*	*ta**j**aŋ*	*ta**j**am*	'sharp'
*ti**g**a	*ti**g**ɔ*	*ti**g**ə*	*ti**g**ɛ*	*ti**g**a*	'three'

In one instance, *k- is unexpectedly reflected as *g-* in KM and Dusun: PM *kutu 'head louse' > KM *gutu*, CTM *kutu*, Dusun *gutəʊ*. The sporadic change of *k- > *g-* (or the other way around) is not uncommon throughout the history of Malayic and Austronesian languages (Adelaar 1992: 61–62; Blust 1996). An initial *g-* in this particular word can also be found in many languages within and outside of Malayic. Within Malayic, Salako and various varieties of Kendayan, Keninjal, Sarawak Malay, Sambas Malay, and Sosok

Malay all have *gutu* (Collins 1987; Adelaar 2005b; Anderbeck & Cooper 2017; Smith 2017). Outside Malayic, Punan (central-north Borneo) and many Land-Dayak languages have *gutu*(*h,ʔ*), and a few Oceanic languages such as Kokota and Roviana (both Northwest Solomonic, Solomon Islands) also have *gutu* (Blust & Trussel 2023), all reflecting PMP *kutu 'head louse'.

PM stops occurring in final position were *p, *t, *k and *ʔ. Voiced stops and palatals were not allowed in this position. *-p, *-t and *-k have merged to *-ʔ* in NEPMs, as demonstrated by the examples in Table 7.7.

Table 7.7: Reflexes of PM *-p, *-t and *-k

PM	KM	CTM	Dusun	SM	Gloss
*saya**p**	*saya**ʔ***	*saya**ʔ***	*saya**ʔ***	*saya**p***	'wing'
*hidu**p**	*ido**ʔ***	*ido**ʔ***	*idu**ʔ***	*hidu**p***	'to live'
*laŋi**t**	*laŋi**ʔ***	*laŋi**ʔ***	*laŋi**ʔ***	*laŋi**t***	'sky'
*mulu**t**	*mulo**ʔ***	*mulo**ʔ***	*mulu**ʔ***	*mulu**t***	'mouth'
*tasi**k**	*tase**ʔ***	*tase**ʔ***	*tasɛi**ʔ***	*tasi**k***	'lake'
*dudu**k**	*dudo**ʔ***	*dudo**ʔ***	*dudəʊ**ʔ***	*dudu**k***	'to sit'

The development of PM *-ʔ merits a separate discussion. While most *-ʔ* in NEPMs are regular reflexes of *-p, *-t and *-k, some *-ʔ* cannot be traced back to such origins. Yet still, whether they are reflexes of *-ʔ is not completely clear, as it is controversial whether *-ʔ should be reconstructed at all. Since the status of *-ʔ has direct consequences on the analysis of these *-ʔ* in NEPMs, the reconstruction of *-ʔ will be reexamined, and the relevance of NEPM data will be explored. The short conclusion is that while a final verdict on the issue is still lacking, it is reasonable to assume that NEPMs likely developed *-ʔ* from secondary origins rather than inherited *-ʔ.

Suppose PM *-ʔ is taken at face value – it only occurred finally, and it was generally lost in NEPMs except in a few cases:

Table 7.8: Reflexes of PM *-ʔ

PM	KM	CTM	Dusun	SM	Gloss
PM *-ʔ > Ø					
*bukaʔ	*bukɔ*	*bukə*	*bukɛ*	*buka*	'to open'
*naŋkaʔ	*nakɔ*	*nakə*	*nakɛ*	*naŋka*	'jackfruit'
*lagiʔ	*lagi*	*lagi*	*lagɛi*	*lagi*	'again'
*m|andiʔ	*mandi*	*mandi*	*mandɛi*	*mandi*	'to bathe'
*daguʔ	*dagu*	*dagu*	*dagəʊ*	*dagu*	'chin'
*akuʔ	*ŋŋ|aku*	*ŋŋ|aku*	*ŋŋ|akəʊ*	*məŋ-aku*	'to confess'
PM *-ʔ > *-ʔ*					
*nasiʔ	*nasiʔ*	*nasiʔ*	*nasiʔ*	*nasi*	'cooked rice'
*tahiʔ	*taiʔ*	*taiʔ*	*taiʔ*	*tahi*	'excrement'

The primary evidence for reconstructing PM *-ʔ comes from Iban *-ʔ*, to which most other Malayic varieties have *Ø* as a correspondence.[47] Compare the following cognate sets:

1) Iban *naŋkaʔ* 'jackfruit', SM and Banjar Hulu *naŋka*, Seraway *naŋko*, Jakarta Malay *naŋkè*;
2) Iban *asiʔ* 'cooked rice', SM, Banjar Hulu, Minangkabau, Seraway and Jakarta Malay *nasi*.[48]

These correspondences yielded the reconstructions PM *naŋkaʔ and *nasiʔ. However, whether *-ʔ should be reconstructed depends on the interpretation of Iban *-ʔ*, whether it is considered a retention or an innovation. This also raises further questions: if Iban *-ʔ* is a retention, which higher-order proto phoneme(s) does it reflect? On the other hand, if it is an innovation, how can its origin be explained? There are two opposing views in the literature, both of which will be briefly summarised.

The first interpretation suggests that Iban *-ʔ* is a retention from PAn/PMP, and *-ʔ should be reconstructed in PM. Based on a number of instances where Iban *-ʔ* corresponds to a final consonant in some central

[47] There are some exceptions where Iban *-ʔ* corresponds to SM *-k*, which are often kinship terms. They will be set aside for now and discussed later in this section.

[48] The correspondence of Iban *Ø-* and others *n-* is unexplained.

Philippine and Formosan languages, Zorc (1982, 1996) argues that Iban *-ʔ* is a PAn/PMP retention, which originated from the merger of PAn *-S, *-H and *-ʔ:[49]

1) PAn *-S, *-H > PMP *-h > Iban *-ʔ*
2) PAn *-ʔ > PMP *-ʔ > Iban *-ʔ*

Zorc's arguments and Iban data were critically evaluated in Adelaar (1992: 63–67), who agreed that Iban *-ʔ* does have correspondences in Philippine and Formosan languages to a certain extent, and at least PAn *-ʔ > PMP *-ʔ > Iban *-ʔ* seemed well-supported. He tentatively accepted Zorc's theory, and reconstructed PM *-ʔ if Iban has *-ʔ*, PM *Ø if Iban has *Ø*, and PM *(-ʔ) if Iban provides no evidence or lacks a reflex.

There has been, however, criticism towards Zorc's analysis, as neither sound change proposed above holds scrutiny. First, PAn *-S, *-H > PMP *-h > Iban *-ʔ* cannot be sustained by Zorc's own material, given the many counter-examples where Iban fails to reflect this sound change. For instance, PAn *təbuS 'sugarcane' and *sikuH 'elbow' were reconstructed based on Philippine and Formosan evidence, but their reflexes in Iban are *təbu* and *siku*, which have final *Ø* instead of expected *-ʔ*. Adelaar (1992: 63, f.n. 102) worked out the following statistics based on Zorc's data: in only three out of nine instances does Iban *-ʔ* reflect PAn *-S, and in eleven out of seventeen instances, Iban *-ʔ* reflects PAn *-H. Second, while there are instances demonstrating the correspondence of Iban *-ʔ* : Philippine *-ʔ* : Formosan *-ʔ*, there is an equal number or more instances where such a correspondence is not present. Wolff (2009: 122–124) examined the correlation between *-ʔ* in Iban and Bunun (Formosan): out of 23 cognate sets, fifteen have *-ʔ* in either or both languages. In only six out of the fifteen sets Iban *-ʔ* agrees with Bunun

[49] The sound correspondences providing evidence for PAn reconstructions are as follows (limited to final position, after Zorc 1982):

a) Iban *-ʔ* : Philippine *-h* : Formosan *-s* or *-ʃ*, pointing to PAn *-S;
b) Iban *-ʔ* : Philippine *-h* : Formosan *-h*, pointing to PAn *-H;
c) Iban *-ʔ* : Philippine *-ʔ* : Formosan *-ʔ*, pointing to PAn *-ʔ.

Another set of relevant sound correspondence is Iban *-h* : Philippine *-ʔ* : Formosan *-q* or *-ʔ*, pointing to PAn *-q (possibly a uvular stop). The last set of sound correspondence is fairly regular and not at issue here, but note that Philippine *-ʔ* can be reflexes of either PAn *-ʔ or *-q when Formosan evidence is lacking, hence Iban *-ʔ* is crucial in supporting the reconstruction of *-ʔ instead of *-q (at least at the PMP level).

-*ʔ*, which indicates a rather poor correlation. In a similar vein, Blust (2009: 563–568) tested the correlation between Iban -*V*(*ʔ*) : Tagalog -*V*(*ʔ*), and concluded that Iban -*Vʔ* : Tagalog-*Vʔ* correlates less than chance frequency, and Iban -*Vʔ* more often correlates with Tagalog -*V*. These mismatches weaken the claim that -*ʔ* is a retention and an indicator of genetic affinity; instead, Iban -*ʔ* is more likely an accretion in words with original final vowels, i.e., an innovation. Blust (2009: 562) also drew attention to the observation that some -*ʔ* in Iban originated from some other known sources:

1) Iban -*ʔ* < PMP *-R, as in *ai**ʔ*** 'water' < *wahi**R**;
2) Iban -*ʔ* < PMP *-q, as in *lua**ʔ*** 'spit out' < *lua**q**;
3) Iban -*Vʔ* < PMP diphthongs, as in *kayu**ʔ*** 'tree' < *kah**iw**.

The occurrence of Iban -*ʔ* in these cases can only be attributed to secondary developments, which further supports the idea that Iban -*ʔ* with unknown sources is likely secondary rather than inherited.

Assuming that Iban -*ʔ* is an innovation, several hypotheses have been put forth to account for its introduction. It might have been borrowed from an unknown source language which had -*ʔ* as a regular reflex of PMP *-q (Blust 2009: 562) or perhaps had a non-phonemic [-ʔ] added to all words with final vowels (Wolff 2009: 124). Another possibility is that it represents a fossilised grammatical marker, which only survived in a small number of forms, as illustrated by Iban noun-verb alternations like *asuʔ* 'dog' : *ŋ-asu* 'to hunt (using dogs)', *dua* 'two' : *bə-duaʔ* 'to divide', *tusu* 'breast' : *tusuʔ* 'to suck' and *aku* '1SG' : *akuʔ* 'to confess' (Nothofer 1996: 42; Zorc 1996: 47; Blust 2009: 566).[50] Unfortunately, none of the explanations is entirely satisfactory and convincing, and it remains obscure why -*ʔ* was only added to some vowel-final words but not others.

The discussions above have focused on the correspondences of Iban -*ʔ* in languages outside the Malayic group. Before drawing any definitive conclusions, there is in fact a third hypothesis to consider regarding the origin of Iban -*ʔ*: it could be an innovation with reference to PAn/PMP, but the innovation might have taken place in PM or pre-PM. As more data come into sight, it becomes clear that a phonemic -*ʔ* is also present in other Malayic varieties in western Borneo and southeastern Sumatra, some of which show a fairly

[50] Three out of the four pairs have the nominal forms ending in vowels and the verbal forms ending in -*ʔ*, but *asuʔ* 'dog' : *ŋ-asu* 'to hunt (using dogs)' shows a reverse pattern.

high degree of agreement with Iban *-ʔ*. Nothofer (1996: 39) demonstrates the correspondence of *-ʔ* in Iban, Sarawak Malay, Salako, Ketapang and Bangka Malay, postulating that *-ʔ* must have developed in an immediate common ancestor of these varieties, referred to as Proto Western Bornean (or Northwest Bornean in Nothofer 1997). It is essential to note that Nothofer's subgrouping hypothesis was formulated based on the assumption that PM had no final glottal stop, and these Bornean and Sumatran Malayic varieties innovated *-ʔ* post-PM. The correlation of *-ʔ* in these varieties does suggest a shared innovation; however, without a determined internal subgrouping based on other evidence, it is also possible that the ancestral language that innovated *-ʔ was actually PM rather than Proto Western Bornean. It could be that PM innovated *-ʔ, which was inherited in Iban and some other varieties, but lost elsewhere to varying degrees.[51] In either scenario, one proto language underwent the innovation of sporadic accretion of *-ʔ*. In fact, if regular correspondences with Iban/western Bornean *-ʔ* can be found outside Borneo and southeast Sumatra, the reconstruction of *-ʔ could gain a more secure foundation. *-ʔ* in Iban and other varieties would then be retentions instead of innovations, therefore offering no value in subgrouping.

In a nutshell, there is weak extra-Malayic evidence supporting PM *-ʔ, but internal evidence from Malayic provides a somewhat ambiguous picture. Some Malayic varieties have developed *-ʔ* in original vowel-final words, but whether this innovation can be traced back to PM is uncertain.

Returning to data from NEPMs, the foregoing discussion holds relevance for two reasons. First, as mentioned earlier, some *-ʔ* in NEPMs cannot be traced back to PM *-p, *-t or *-k, and they seem to reflect *-ʔ at first glance. Some instances of *-ʔ* actually correspond to Iban *-ʔ*, even though NEPMs are not geographically adjacent to Iban-speaking area. While this observation does provide some initial support for the reconstruction of *-ʔ, a closer examination reveals that the number of such correspondences is too limited to carry significant weight. Second, there's at least one example where KM exhibits a noun-verb alternation signalled by *Ø* : *-ʔ*, similar to what has been reported in Iban. This suggests the possibility that *-ʔ* might also be an innovation in KM. These two observations are elaborated below.

[51] A similar view is implicitly expressed in Adelaar (2004b: 20), in light of the fact that western Borneo is supposedly the Malayic homeland itself. Thurgood (1998: 308) suggests that phonemic *-ʔ* in present-day Malayic varieties could result from contact with Mon-Khmer languages, but also implied that the contact could have taken place pre-PM.

A total of 85 PM reconstructions with *-ʔ are found in Adelaar (1992), based on Iban -*ʔ*. Taking these 85 Iban words as a baseline for comparisons, I established 40 cognate sets between Iban, NEPMs and SM. For the remaining 45 Iban words, either no cognate was found or no data was available. The correspondences of final segments were compared and are presented in Table 7.9. In all instances, NEPMs agree with each other in having either -*ʔ* or *Ø*. Their correspondences with Iban and SM are grouped into three classes: in seven cognate sets, NEPMs agree with Iban in having -*ʔ*, to which SM has -*k* as a correspondence. In two sets, NEPMs and Iban have -*ʔ*, but SM has *Ø*. In the other 31 sets, NEPMs agree with SM in having *Ø*, whereas Iban has -*ʔ*.

Table 7.9: Correspondences of KM, CTM, Dusun and SM finals with Iban -*ʔ*

Iban	KM	CTM	Dusun	SM	Gloss
Iban -*ʔ* : KM -*ʔ* : CTM -*ʔ* : Dusun -*ʔ* : SM -*k*					
*adi**ʔ***	*adi**ʔ***	*adi**ʔ***	*adɛi**ʔ***	*adi**k***	'younger sibling'
*aka**ʔ***[a]	*kakɔ**ʔ***	*kakɔ**ʔ***	*kakɔ**ʔ***	*kaka**k***	'older sister'
*bali**ʔ***	*bale**ʔ***	*bale**ʔ***	*balɛi**ʔ***	*bali|**k***	'to return'
*datu**ʔ***[b]	*to**ʔ***	*to**ʔ***	*təʊ**ʔ***	*datu**k***	'grandfather'
*ən|da**ʔ***	*dɔ**ʔ***	*dɔ**ʔ***	*dɔ**ʔ***	*ti|da|**k***	'no, not'
*ini**ʔ***	*nɛnɛ**ʔ***	*nɛnɛ**ʔ***	*nɛi**ʔ***	*nɛnɛ**k***	'grandmother'
*pinta**ʔ***	*mitɔ**ʔ***	*mitɔ**ʔ***	*mitɔ**ʔ***	*pinta**k***[c]	'to request'
Iban -*ʔ* : KM -*ʔ* : CTM -*ʔ* : Dusun -*ʔ* : SM Ø					
*asi**ʔ***	*nasi**ʔ***	*nasi**ʔ***	*nasi**ʔ***	*nasi*	'cooked rice'
*tai**ʔ***	*tai**ʔ***	*tai**ʔ***	*tai**ʔ***	*tahi*	'excrement'
Iban -*ʔ* : KM Ø : CTM Ø : Dusun Ø : SM Ø					
*anti**ʔ***	*n|nati*	*n|nati*	–	*nanti*	'to wait'
*aku**ʔ***	*ŋŋ|aku*	*ŋŋ|aku*	*ŋŋ|akəʊ*	*məŋ-aku*	'to confess'
*bəri**ʔ***	*buwi*	*buwi*	*buwɛi*	*bəri*	'to give'
*buka**ʔ***	*bukɔ*	*bukə*	*bukɛ*	*buka*	'to open'
*buta**ʔ***	*butɔ*	*butə*	*butɛ*	*buta*	'blind'
*ucu**ʔ***	*cucu*	*cucu*	*cucəʊ*	*cucu*	'grandchild'
*dagu**ʔ***	*dagu*	*dagu*	*dagəʊ*	*dagu*	'chin'
*sa**ʔ***	*sɔ*	*sə*	*sɛ*	*satu*	'one'
*duri**ʔ***	*duɣi*	*duɣi*	*duɣɛi*	*duri*	'thorn'

Iban	KM	CTM	Dusun	SM	Gloss
garuʔ	*gaɣu*	*gaɣu*	*gaɣəʊ*	*garu(k)*	'to scratch'
liaʔ	*haliyɔ*	–	*haliyɛ*	*(h)alia*	'ginger'
iuʔ	*(i)yu*	–	*(i)yəʊ*	*(h)iu, yu*	'shark'
isiʔ	*isi*[d]	*isi*	*isɛi*	*isi*	'meat; content'
kayuʔ	*kayu*	*kayu*	*kayəʊ*	*kayu*	'wood'
kənaʔ	*kənɔ*	*kənə*	*kənɛ*	*kəna*	'to hit'
kitaʔ	*kitɔ*	*kitə*	*kitɛ*	*kita*	'1PL'
lagiʔ	*lagi*	*lagi*	*lagɛi*	*lagi*	'again'
lamaʔ	*lamɔ*	*lamə*	*lamɛ*	*lama*	'long (time)'
əmpəlawaʔ	*llabɔ*	*llabə*	*glabɛ*	*laba-laba*	'spider'
limaʔ	*limɔ*	*limə*	*limɛ*	*lima*	'five'
pandiʔ	*mandi*	*mandi*	*mandɛi*	*mandi*	'to bathe'
mudaʔ	*mudɔ*	*mudə*	*mudɛ*	*muda*	'young; unripe'
naŋkaʔ	*nakɔ*	*nakə*	*nakɛ*	*naŋka*	'jackfruit'
pəɲuʔ	*pəɲoŋ*[e]	*pəɲoŋ*	*pəɲuŋ*	*pəɲu*	'turtle'
rusaʔ	*ɣusɔ*	*ɣusə*	*usɛ*	*rusa*	'deer'
saguʔ	*sagu*	*sagu*	*sagəʊ*	*sagu*	'sago'
sawaʔ	*sawɔ*	*sawə*	*sawɛ*	*sawa*	'python'
tadiʔ	*taʔdi*	*taʔdi*	*taʔdɛi*	*tadi*	'just now'
taɲaʔ	*taɲɔ*	*taɲə*	*taɲɛ*	*taɲa*	'to ask'
taŋgaʔ	*taŋgɔ*	*taŋgə*	*taŋgɛ*	*taŋga*	'ladder'
tumaʔ[f]	*tuŋa*	*tuŋə*	*tuŋɔ*	*tuma*	'body louse'

[a] Iban *akaʔ* 'older sibling'.

[b] Iban *datuʔ* 'nobleman, chief'.

[c] Archaic, also *pinta* and *minta* without *-k*.

[d] Also *isiʔ* 'to fill'.

[e] Final *-ŋ* in NEPMs is a later innovation, see §7.3.2.

[f] Might not be a valid cognate set. Correspondence of Iban/SM *-m-* : NEPM *-ŋ-* is unexplained, and Iban *-aʔ* : KM *-a* : CTM *-ə* : Dusun *ɔ* : SM *a* is also irregular. See more discussions in §7.3.2.

In the first set of correspondences, namely Iban *-ʔ* : NEPM *-ʔ* : SM *-k*, all varieties have a final consonant. This is mostly found in words belonging to certain semantic categories: out of the seven examples listed in Table 7.9, four are kinship terms ('younger sibling', 'older sister', 'grandfather' and 'grandmother') and one is a negator. Blust (1979) posits a vocative marker *-q in Proto Western-Malayo-Polynesian (PWMP) and considered SM *-k* in these kinship terms as a fossilised form of the vocative *-q. Final *q in PMP/PWMP would have been regularly reflected as *-h* in SM, but *-q in this particular morpheme was reintroduced and reanalysed as a phonemic *-k* in order to keep the vocative function. SM *-k* in *ti|da|k* 'no, not' probably has a similar origin. Altogether, Adelaar (1992: 119) suggests that *-k* in these cases can "maybe be interpreted as a syntactic device used for words in isolation (including vocatives, negations and greetings)". The same explanation might apply to *-ʔ* in the cognates of these terms in NEPMs, but *-k* in SM *bali|k* 'to return' and *pintak* 'to request', as well as *-ʔ* in their NEPM cognates is unexplained.

In all other sets, Iban *-ʔ* corresponds to SM *Ø*. If NEPMs have a corresponding *-ʔ*, then this correspondence could hint at an inherited origin of these *-ʔ*. Such a correspondence is found in the cognate sets in the second class: *-ʔ* in NEPM *nasiʔ* 'cooked rice' and *taiʔ* 'excrement' shows agreement with Iban *-ʔ*, but SM cognates have *Ø*. However, in a total of 33 cognate sets, NEPMs agree with Iban in having *-ʔ* in merely two cases, and there is a much larger number of cognate sets where NEPMs agree with SM in having *Ø* (the third class in Table 7.9). Interesting as they are, the two instances displaying the correspondence of Iban *-ʔ* : NEPM *-ʔ* : SM *Ø* provide insufficient evidence for the reconstruction of PM *-ʔ. Two more cognate sets showing the correspondence of KM, CTM (and Dusun) *-ʔ* : SM *Ø* are worth mentioning, but their Iban cognates are not known:[52]

1) KM *pulɔʔ* : CTM *pulɔʔ* : Dusun *pulɔʔ* : SM *pula* 'furthermore'
2) KM *jugɔʔ* : CTM *jugɔʔ* : Dusun *jugɛ* : SM *juga* 'also'

Note that in 2), there is disagreement in final segments within NEPMs: KM and CTM *jugɔʔ* ends in *-ʔ* whereas Dusun *jugɛ* in vowel. The inconsistency in fact suggests that *-ʔ* is a secondary development.[53]

[52] Neither is found in any of the following Iban dictionaries: Scott (1956), Richards (1981), Sutlive & Sutlive (1994).

[53] Similarly, *-ʔ* in some discourse particles has been reported in Tioman Malay (spoken on the Tioman island southeast off the coast of Terengganu), e.g., *lagiʔ* 'again', *jugaʔ* 'also' and *pulaʔ* 'furthermore'. Collins (1985) suggests that it might have originated from an earlier phrase marker.

Another noteworthy observation is that KM has a unique instance of noun-verb alternation signalled by *Ø* : *-ʔ*, as in *isi* 'meat; content' and *isiʔ* 'to fill'. This phenomenon, however, has been found solely in KM and has not been identified in any other instances (cf. CTM *isi* and Dusun *isɛi*, meaning both 'meat, content' and 'to fill'). KM *isi* 'meat; content' : *isiʔ* 'to fill' resembles the phenomenon reported in Iban, suggesting that *-ʔ* in these cases might be a fossilised grammatical marker. Nevertheless, given that this is the only example, it provides only a limited piece of evidence.

To wrap up, while the reconstruction of PM *-ʔ is still a debatable topic without definite answers, the poor correlation between NEPM *-ʔ* and Bornean *-ʔ* does not offer sufficient evidence in support of *-ʔ. While the possibility of reconstructing *-ʔ cannot ruled out entirely, on the basis of the findings discussed above, I conclude that *-ʔ* in NEPMs is unlikely to be inherited. Some *-ʔ* are probably reflexes of PWMP vocative *-q, typically preserved in kinship terms. Some could be fossilised grammatical markers. The history *-ʔ* in a few other instances, such as *nasiʔ* 'cooked rice' and *taiʔ* 'excrement', remains unclear.[54] In the following discussions, I adhere to Adelaar's reconstruction of *-ʔ, but it should be borne in mind that except in some kinship terms and the negator as mentioned above, *-ʔ in most cases is irrelevant to subsequent sound changes, and the development of vowels in putative *-Vʔ is indistinguishable from those in *-V (compare PM *bukaʔ 'to open' > KM *bukɔ*, CTM *bukə*, Dusun *bukɛ*, with PM *mata 'eye' > KM *matɔ*, CTM *matə*, Dusun *matɛ*). In other words, even if PM *-ʔ can be reconstructed, the loss of *-ʔ in the majority part of NEPM lexicon must have taken place at a very early stage.

[54] A possible explanation for the sporadic occurrences of these *-ʔ* is that they might have been borrowed from another adjacent Malayic variety, Tioman Malay being one possible candidate of the donor language. Based on the rather limited data in Collins (1985), it appears that Tioman Malay also has content words with *-ʔ*, which occurs more frequently than in NEPMs. To what extent Tioman Malay *-ʔ* : Iban *-ʔ* is systematic still needs to be investigated, but there are at least eleven examples illustrating such a correspondence (although exceptions apparently exist, also see Adelaar 1992: 67), and it is a much better correlation than NEPM *-ʔ* : Iban *-ʔ*. On the whole the history of *-ʔ* in Tioman Malay is also not clear, and a more thorough examination is needed.

7.3.2 PM nasals

PM had four nasals *m, *n, *ɲ and *ŋ. All of them could occur initially and intervocalically, but only *-m, *-n and *-ŋ were allowed finally. Nasals in initial and intervocalic positions are retained, whereas final nasals have been merged and neutralised.

The retention of initial and intervocalic nasals is illustrated in Table 7.10.

Table 7.10: Reflexes of PM initial and intervocalic nasals

PM	KM	CTM	Dusun	SM	Gloss
Initial nasals					
***m**ata	***m**atɔ*	***m**atə*	***m**atɛ*	***m**ata*	'eye'
***n**aiʔ	***n**aiʔ*	***n**aiʔ*	***n**aiʔ*	***n**aik*	'to go up'
***ɲ**aŋa(ʔ)	***ŋ**aŋɔ*	***ŋ**aŋə*	***ŋ**aŋɛ*	***ŋ**aŋa*	'to open wide'
***ɲ**amuk	***ɲ**amoʔ*	***ɲ**amoʔ*	***ɲ**amuʔ*	***ɲ**amuk*	'mosquito'
Intervocalic nasals					
*ru**m**ah	*ɣu**m**ɔh*	*ɣu**m**ɔh*	*u**m**ɔh*	*ru**m**ah*	'house'
*ta**n**ah	*ta**n**ɔh*	*ta**n**ɔh*	*ta**n**ɔh*	*ta**n**ah*	'earth, soil'
*ta**ɲ**aʔ	*ta**ɲ**ɔ*	*ta**ɲ**ə*	*ta**ɲ**ɛ*	*ta**ɲ**a*	'to ask'
*bu**ŋ**a(ʔ)	*bu**ŋ**ɔ*	*bu**ŋ**ə*	*bu**ŋ**ɛ*	*bu**ŋ**a*	'flower'

In one instance, PM *-m- appears to be reflected as *-ŋ-*, as in *tumaʔ 'body louse' > KM *tuŋa*, CTM *tuŋə*, Dusun *tuŋɔ*. The correspondence of final vowels in this set is also irregular: CTM *-ə* could only reflect earlier ⁺-a(ʔ), but KM *-a* and Dusun *-ɔ* point to earlier ⁺-aw or ⁺-ar (see §7.4.3). Thus, it seems likely that KM *tuŋa* and Dusun *tuŋɔ* are not directly related to PM *tumaʔ. A number of cognates reflecting ⁺tuŋaw are actually attested in languages within and outside of the Malay Peninsula: Sambas Malay (northwest Broneo), Berau Malay (eastern Borneo), Ulu Kapuas Malay (central-north Borneo) and Duano (Malacca strait) all have *tuŋaw*, and Ulu Pahang (central-east Malay Peninsula) and Malayic Dayak in Sekadau (northwest Borneo) have *tuŋa* (Anderbeck & Cooper 2017). I suggest *tuŋaw to be reconstructed in PM (alongside *tumaʔ), from which KM *tuŋa* and Dusun *tuŋɔ* have developed. The history of CTM *tuŋə* is less clear.

In final position, PM nasals have merged to *-ŋ* in all three varieties, as demonstrated in Table 7.11.

Table 7.11: Reflexes of PM final nasals

PM	KM	CTM	Dusun	SM	Gloss
Final nasals following *a, *ə					
*tana**m**	*tanɛ*	*tanaŋ*	*tanaŋ*	*tana**m***	‘to plant’
*hitə**m**	*itɛ*	*itaŋ*	*itaŋ*	*hita**m***	‘black’
*huja**n**	*ujɛ*	*ujaŋ*	*ujaŋ*	*huja**n***	‘rain’
*tahə**n**	*tɛhɛ*	*tahaŋ*	*tahaŋ*	*taha**n***	‘to tolerate’
*binta**ŋ**	*bitɛ*	*bitaŋ*	*bitɔŋ*	*binta**ŋ***	‘star’
*pətə**ŋ**[a]	*pətɛ*	*pətaŋ*	*pətɔŋ*	*pəta**ŋ***	‘afternoon’
Final nasals following *i, *u					
*kiri**m**	*kiɣiŋ*	*kiɣiŋ*	*kiɣiŋ*	*kiri**m***	‘to send’
*jaru**m**	*jaɣoŋ*	*jaɣoŋ*	*jaɣuŋ*	*jaru**m***	‘needle’
*aŋi**n**	*aŋiŋ*	*aŋiŋ*	*aŋiŋ*	*aŋi**n***	‘wind’
*tahu**n**	*taoŋ*	*taoŋ*	*tauŋ*	*tahu**n***	‘year’
*kəri**ŋ**	*kəɣiŋ*	*kəɣiŋ*	*kəɣɛiŋ*	*kəri**ŋ***	‘dry’
*buru**ŋ**	*buɣoŋ*	*buɣoŋ*	*buɣəʊŋ*	*buru**ŋ***	‘bird’

[a] PM *pətəŋ ‘dark, obscure’.

The merger of final nasals is straightforward in CTM and Dusun, as the outcome can be observed regardless of the quality of final-syllable vowels. In KM, final nasals following high vowels *i and *u have merged to *-ŋ*, but earlier *-aN and *-əN are reflected as *-ɛ*. The loss of nasals following non-high vowels presumably occurred after the merger and neutralisation of final nasals, accompanied by the raising of *a and *ə to *ɛ* (see §7.4.3).

NEPMs also show nasal accretion in earlier forms with a final high vowel preceded by a nasal onset. A velar nasal was added to these original final open syllables, resulting from the carryover of the nasality of final-syllable onsets, as exemplified in Table 7.12. When the final vowel was a low vowel *-a, no accretion is attested, as seen in *buŋa(ʔ) ‘flower’ > KM *buŋɔ*, CTM *buŋə* and Dusun *buŋɛ*. In KM and CTM, some variation is attested between forms with *-ŋ* and those with Ø, suggesting that nasal accretion is still an ongoing process.

Table 7.12: Final velar nasal accretion

PM	KM	CTM	Dusun	SM	Gloss
*bini	*bini~biniŋ*	*bini~biniŋ*	*biniŋ*	*bini*[a]	'wife'
*kami	–	–	*kamiŋ*	*kami*	'1PL.EXCL'
*(i)ni(ʔ)	*ni~niŋ*	*ni~niŋ*	*nɛiŋ*	*ini*	'DEM.PROX'
*pəɲuʔ	*pəɲoŋ*	*pəɲoŋ*	*pəɲuŋ*	*pəɲu*	'turtle'
*kamu(ʔ)	*moŋ*	*moŋ*	*məʊŋ*	*kamu*	'2SG'

[a] Coarse, commonly *istəri* < Sanskrit.

7.3.3 PM fricatives

Table 7.13 and Table 7.14 illustrate the developments of PM fricatives *s and *h. Initial and intervocalic *s are retained, whereas final *s was underwent lenition to become *-h*. Initial *h was lost in all three varieties, as was intervocalic *h between non-identical vowels. Final *h is retained; hence the merger of *-s, *-h > *-h*.

Table 7.13: Reflexes of PM *s

PM	KM	CTM	Dusun	SM	Gloss
*sayap	*sayaʔ*	*sayaʔ*	*sayaʔ*	*sayap*	'wing'
*səmpit	*səpeʔ*	*səpeʔ*	*səpiʔ*	*səmpit*	'narrow'
*asəp	*asaʔ*	*asaʔ*	*asaʔ*	*asap*	'smoke'
*bəsar	*bəsa*	*bəsɔ*	*bəsɔ*	*bəsar*	'big'
*atas	*ata**h***	*ata**h***	*ata**h***	*atas*	'top'
*nipis	*nipi**h***	*nipi**h***	*nipi**h***	*nipis*	'thin'

Note that final *h* following a high front vowel *i* exhibits a tendency to be palatalised and become [ç] in NEPMs, e.g., *nipih* → [nipih]~[nipiç] 'thin', as explained in §2.2.1.3, §3.2.1.3 and §4.2.1.3. However, the palatalisation process represents only synchronic phonetic variation; diachronically, the phonemic change of *-s > *-h* still holds following a high vowel *i.

Table 7.14: Reflexes of PM *h

PM	KM	CTM	Dusun	SM	Gloss
***h**antu	*atu*	*atu*	*atəʊ*	***h**antu*	'ghost'
***h**iduŋ	*idoŋ*	*idoŋ*	*idəʊŋ*	***h**iduŋ*	'nose'
*ta**h**iʔ	*taiʔ*	*taiʔ*	*taiʔ*	*tahi*	'excrement'
*ta**h**un	*taoŋ*	*taoŋ*	*tauŋ*	*ta**h**un*	'year'
*da**h**an	*dɛ**h**ɛ*	*da**h**aŋ*	*da**h**aŋ*	*da**h**an*	'branch'
*pa**h**a(ʔ)	*pɔ**h**ɔ*	*pə**h**ə*	*pa**h**ɛ*	*pa**h**a*	'thigh'
*pəca**h**	*pəcɔ**h***	*pəcɔ**h***	*pəcɔ**h***	*pəcɔ**h***	'to break'
*jatu**h**	*jato**h***	*jato**h***	*jatəʊ**h***	*jatu**h***	'to fall'

The retention of *-h- between identical vowels does not always result in an intervocalic *h* between identical vowels at the synchronic level, as the vowels surrounding *-h- could have undergone different changes. This is evidenced by *paha(ʔ) > Dusun *pahɛ* 'thigh'. In this case, *-h- is retained between two identical *a, but ultimate *-a(ʔ) is raised to *ɛ* whereas the penultimate *a remains intact. Therefore, synchronically, Dusun has intervocalic -*h*- between non-identical vowels. KM *pɔhɔ*, CTM *pəhə* likely have a similar history, but in these two varieties, penultimate *a was affected by vowel harmony and assimilated to the ultimate vowel: *paha(ʔ) > +pahɔ > KM *pɔhɔ*, and *paha(ʔ) > +pahə > CTM *pəhə*. Assimilation as such only took place when the earlier intervocalic consonant was *-h-. Compare *dahan > +dahɛ > KM *dɛhɛ* 'branch', with *jalan > KM *jalɛ* 'road' and *tanam > KM *tanɛ* 'to plant' without assimilation.

In a few cases, intervocalic *-h- seems to have been retained between two non-identical vowels, e.g., *jahət 'bad' > NEPM *jahaʔ*, *tahən 'to tolerate' > KM *tɛhɛ*, CTM/Dusun *tahaŋ*. Both examples in PM contained *-ahə-. It is possible that PM ultimate *ə merged with *a at an early stage prior to other changes. It is also likely that the reconstruction for two items should be revised as *jahat and *tahan, as the reconstruction of ultimate *ə in PM is debatable (see §7.4.3).

7.3.4 PM liquids

PM had two liquids *l and *r. Initial and intervocalic liquids are mostly retained, except that initial *r- was (sporadically) lost preceding the high vowels *u and *i in Dusun. Final liquids were lost in all three varieties, except in a few cases where *-l is unexpectedly reflected as *-ʔ*. Reflexes of PM liquids are exemplified in Table 7.15.

Table 7.15: Reflexes of PM liquids

PM	KM	CTM	Dusun	SM	Gloss
Initial liquids					
*lamaʔ	*lamɔ*	*lamə*	*lamɛ*	*lama*	‘long (time)’
*libar	*lɛba*	*lɛbɔ*	*libɔ*	*lɛbar*	‘to throw’
*rambut	*ɣamboʔ*	*ɣamboʔ*	*ɣambuʔ*	*rambut*	‘hair’
*rantay	*ɣata*	*ɣata*	*ɣata*	*rantai*	‘necklace; chain’
*rumah	*ɣumɔh*	*ɣumɔh*	*umɔh*	*rumah*	‘house’
*rumput	*ɣupuʔ*	*ɣupuʔ*	*upuʔ*	*rumput*	‘grass’
...	*ɣɛʔŋɛ*	*ɣɛŋaŋ*	*iŋaŋ*	*riŋan*	‘light (weight)’
*ribu	*ɣibu*	*ɣibu*	*ɣibəʊ*	*ribu*	‘thousand’
Intervocalic liquids					
*tali	*tali*	*tali*	*talɛi*	*tali*	‘rope’
*bulan	*bulɛ*	*bulaŋ*	*bulaŋ*	*bulan*	‘moon; month’
*hari	*aɣi*	*aɣi*	*aɣɛi*	*hari*	‘day’
*surat	*suɣaʔ*	*suɣaʔ*	*suɣaʔ*	*surat*	‘letter’
Final liquids					
*təbəl	*təba*	*təba*	*təba*	*təbal*	‘thick’
*paŋgil	*paŋge*	*paŋge*	*paŋgi*	*paŋgil*	‘to call’
*ambil	*ambiʔ*	*ambeʔ*	*ambɛiʔ*	*ambil*	‘to take’
*kəcil[a]	*kəcɛ̃ʔ*	*kəcĩʔ*	*kəcĩʔ*	*kəcil*	‘small’
*akar	*aka*	*akɔ*	*akɔ*	*akar*	‘root’
*ikur	*ɛkɔ*	*ɛkɔ*	*iku*	*ɛkɔr*	‘tail; CLF’

[a] *kəcik was also reconstructed as a doublet.

The loss of *r- preceding *u in Dusun appears to be regular. It can be explained by the similar articulation between a velar *r- and a back vowel *u, that is, the feature [+back] in both sounds led to assimilation and reduction. The loss of *r- before *i, on the other hand, is of a sporadic nature; compare Dusun *iŋaŋ* 'light (weight)' and *iya* 'ringgit' with SM *riŋan* and *riyal*, but Dusun ***ɣibu*** 'thousand' with SM ***ɣibu***. There is, however, a tendency to restore the initial *ɣ*- in Dusun forms which originally had *ɣ-, especially among younger speakers and in careful speech. For instance, both *ukuʔ* and *ɣukuʔ* 'to smoke; cigarette' may be heard (cf. SM *rɔkɔk*), and *ikah* and *ɣikah* 'concise' also exhibit variation (cf. SM *riŋkas*).

In KM and CTM, the loss of final liquids was accompanied by the lowering of preceding high vowels, as can be seen in *paŋgil > *paŋge* 'to call', *air > *ae* 'water', *alur > *alo* 'groove' and *jəmur > *jəmo* 'to dry in the sun'. The lowering of high vowels in these cases must have preceded the loss of final liquids, as final high vowels are retained (§7.4.4.1).

In one example, an intervocalic *r is reflected as *w on the surface, and the schwa preceding *-r- changed to *u* accordingly, as in *bəriʔ 'to give' > KM/CTM *buwi*, Dusun *buwɛi*. The history of this word displays a rather irregular development, as by rule intervocalic *r in disyllables is preserved, even when the penultimate vowel was a schwa, e.g., *bəras 'uncooked rice' > NEPM *bəɣah*, *kəriŋ 'dry' > KM/CTM *kəɣiŋ*, Dusun *kəɣɛiŋ*. Yet, parallel developments of *-ər- > *-u-* can be found in disyllables containing a tautosyllabic ⁺-ər- directly preceding another consonant, i.e., ⁺-ər.C- > -u.C-. Examples include *t**ər**bit 'to emerge' > KM/CTM ***tu**beʔ*, Dusun ***tu**biʔ* and *k**ər**baw 'buffalo' > KM/CTM ***ku**ba*, Dusun ***ku**bɔ*.[55] In all likelihood, the development of *b**əri**ʔ 'to give' > ⁺b**əri** > KM/CTM ***buwi***, Dusun ***buwɛi*** has a similar trajectory with an intermediate stage ⁺bui, which also reflects *-ər- > ⁺-u-. The sound change presumably began with the weakening of the liquid *r, followed by a merger with the preceding schwa. An epenthetic glide in ⁺bui [buwi] was later reinterpreted as phonemic, followed by diphthongisation of ultimate *i in Dusun, hence KM/CTM *buwi* and Dusun *buwɛi* (see §7.4.4.2). What is unusual in the case of *bəriʔ 'to give' is that the sound change *-ər- > ⁺-u- affected a disyllabic word with an intervocalic *-r-, and it only affected this particular word.

---

[55] These two reconstructions are mine. Adelaar (1992: 92) reconstructs *tVr(ə)bit 'to emerge' and *kAr(ə)baw 'buffalo' respectively. See the arguments in §7.3.6.

7.3.5 PM glides

PM had two glides, *w and *y. These glides did not occur initially, and their distributions in intervocalic position were subject to specific phonotactic constraints: *w only occurred between *a's, and *y occurred between vowels other than *i and *ə (Adelaar 1992: 102). Additionally, glides were arguably present in final position following *a, forming *-aw and *-ay, which are often referred to as diphthongs.

Both PM intervocalic glides and (disputable) final glides require extensive discussions, which will be divided into two subsections. In §7.3.5.1, I show that NEPMs have developed more phonemic glides in intervocalic position. In §7.3.5.2, I argue that *-aw and *-ay are VC sequences rather than diphthongs, and that final glides *-w and *-y are comparable to other final consonants.

7.3.5.1 PM intervocalic glides

All PM glides reconstructed in intervocalic position have been preserved, as shown in Table 7.16.

Table 7.16: Reflexes of PM intervocalic glides

PM	KM	CTM	Dusun	SM	Gloss
*sawaʔ	*sawɔ*	*sawə*	*sawɛ*	*sawa*	'python'
*ɲawa	*ɲawɔ*	*ɲawə*	*ɲawɛ*	*ɲawa*	'life, soul'
*layar	*laya*	*layɔ*	*layɔ*	*layar*	'sail'
*kayuʔ	*kayu*	*kayu*	*kayəʊ*	*kayu*	'wood, tree'

According to the reconstructed phonotactic constraints, PM intervocalic glides did not occur following corresponding high vowels. Words such as *buah 'fruit', *dua(ʔ) 'two', *ia '3SG' and *tiup 'to blow' were reconstructed with VV sequences, rather than intervocalic phonemic glides (ˣbuwah, ˣduwa(ʔ), ˣiya or ˣtiyup). There are two reasons to reconstruct these constraints in PM. First, materials from the six varieties on which Adelaar's reconstruction was based invariably have a non-phonemic glide in these environments, or at least analysed as so. For instance, *y* in SM "does not occur adjacent to schwa or *i/e*, nor does *w* occur adjacent to schwa or *u/o*, although non-phonemic glides are heard" (Adelaar 1992: 11). In Iban, "semivowels do not occur adjacent to schwa or to a vowel of the same colouring (the [y] heard between *u* and *a/u*, and the [w] heard between *u* and *a/i*, are non-phonemic glides)" (Adelaar 1992: 26). Second, glides in similar environments were not reconstructed as phonemic in higher-order proto languages, and VV sequences were favoured. Consider PMP *buaq 'fruit' > PM *buah (PMP *q > PM *h), PMP *duha 'two' > PM *dua(ʔ) (PMP *h > PM *Ø), PMP *ia '3SG' > PM *ia and PMP *tiup 'to blow' > PM *tiup.[56] Both internal and external evidence therefore seem to give grounds for the analysis of non-phonemic glides following corresponding high vowels in PM.

In NEPMs, however, intervocalic glides following high vowels must be analysed as phonemic. The reasoning was briefly mentioned in the phonological description (see 2.4.3.1 and §4.3), and it will be further elaborated below. I will show that these phonemic glides in such environments arise from the reinterpretation of original epenthetic glides.

First of all, the phonemic status of glides following high vowels is substantiated by the synchronic shortening of some disyllabic words, which results in monosyllabic forms with initial glides. Examples in Table 7.17 illustrate this process.

Table 7.17: Shortening of disyllables with intervocalic glides

KM	CTM	Dusun	Shortened form	Gloss
buwɔh	*buwɔh*	*buwɔh*	*wɔh*	'fruit'
buwaʔ	*buwaʔ*	*buwaʔ*	*waʔ*	'to do; to make'
buwi	*buwi*	*buwɛi*	KM/CTM *wi*; ITM *wɛi*	'to give'
ɣiya	*ɣiya*	*iya*	*ya*	'ringgit'

Similar shortening can be found in words with intervocalic consonants other than glides, e.g., KM/CTM *dudoʔ* 'to sit; to stay' → *doʔ*, KM *dəŋɛ* 'with' → *ŋɛ* and CTM/Dusun *dəŋaŋ* 'with; and' → *ŋaŋ*. In view of these parallel developments, the most plausible and straightforward analysis for the initial

[56] Adelaar (1992: 204) gave PMP *hiup > PM *t|iup 'to blow', in which PM *t|iup has a fossilised prefix. Blust & Trussel (2023) reconstruct PMP *tiup, but suggest *taR-Səyup could be a possible doublet.

glides in monosyllabic forms, such as *wɔh* 'fruit' and *ya* 'ringgit', is that their corresponding full disyllabic counterparts have phonemic intervocalic glides. These glides are the onsets of final syllables, and the shortening process in all these forms simply deletes the penultimate syllables.

The second piece of evidence comes from the diphthongisation process of earlier high vowels in Dusun and the specific conditions under which this sound change occurred. Diphthongisation only took place when an ultimate high vowel was preceded by an oral consonant and followed by *Ø or a back consonant (*-k, (*-ʔ), *-h and *-ŋ).[57] The exact mechanism of this change will be discussed in depth in §7.4.4, but the relevant conditions are explained below. The following examples illustrate the diphthongisation of PM ultimate high vowels in Dusun:

(1) Diphthongisation of PM ultimate *i > Dusun *ɛi*

*hati	>	*atɛi*	'liver'
*tasik	>	*tasɛiʔ*	'lake'
*putih	>	*putɛih*	'white'
*caciŋ	>	*cacɛiŋ*	'worm'

(2) Diphthongisation of PM ultimate *u > Dusun *əʊ*

*təbu	>	*təbəʊ*	'sugarcane'
*duduk	>	*dudəʊʔ*	'to sit'
*tujuh	>	*tujəʊh*	'seven'
*hiduŋ	>	*idəʊŋ*	'nose'

If the PM word had *Ø or *h as the onset of a final syllable (which is regularly lost), diphthongisation did not occur. In (3), *i and *u remain plain monophthongs. Diphthongisation was also blocked following a nasal onset, but this aspect is not relevant for the current discussion.

(3) Retention of PM ultimate *i and *u

*dahi	>	*dai*	'forehead'
*baik	>	*baiʔ*	'good'
*tahu(ʔ)	>	*tau*	'to know'
*lauk	>	*lauʔ*	'dish'
*jauh	>	*jauh*	'far'

[57] (*-ʔ) is put in parentheses because the reconstruction of this phoneme is uncertain, as discussed in §7.3.1. The same convention applies throughout the rest of this chapter.

From the comparisons between PM words in (1) to (3), it is evident that final-syllable onsets conditioned the diphthongisation process. As a consequence, the outcomes of diphthongisation only appear following a consonant at the synchronic level, and Dusun diphthongs never occur in a VV sequence (see §4.3 for the distribution of Dusun diphthongs at the syllable level, and §4.7 for vowel sequences in Dusun).

Now let us consider the following words attested in Dusun: *buwεi* 'to give', *buwεih* 'foam' and *iyəʊ* 'shark'. Synchronically, intervocalic glides in these words are better analysed as phonemic. The alternative analysis ˣbuεi, ˣbuεih and ˣiəʊ would form VV sequences containing a diphthong, which are not found elsewhere. Diachronically, a phonemic intervocalic glide must have been present pre-Dusun before diphthongisation took place; otherwise, diphthongs would not have been attested in these forms.

The evidence presented above demonstrates that intervocalic *w* following *u* and *y* following *i* are synchronically phonemic, and these glides were already interpreted as phonemic at an earlier stage (at least clearly in Dusun). Given that *w and *y in similar environments were not allowed in PM, NEPMs must have reinterpreted earlier non-phonemic glides as phonemic. Following this analysis, the phonological histories of Dusun *buwεi* 'to give', *buwεih* 'foam' and *iyəʊ* 'shark' are laid out as follows. Adelaar reconstructed PM *bəriʔ 'to give', from which Dusun *buwεi* must have developed. The development from PM *bəriʔ > Dusun *buwεi* has been detailed in §7.3.5, namely *bəriʔ > ⁺bəri > ⁺bui > ⁺buwi > *buwεi*, which evinces the reanalysis of an earlier non-phonemic glide. The word for 'foam' is not reconstructed in PM, but *buqiq is reconstructed in PMP, which would be reflected as *buhih in PM following the sound change of PMP *q > PM *h. From PM *buhih to Dusun *buwεih* (and KM/CTM *buweh*), a three-stage process must be posited: first the loss of *-h-, subsequently the reinterpretation of ⁺-ui- to ⁺-uwi-, followed by the diphthongisation of ⁺i > *εi*. Again, it shows that the genesis of the present-day (and pre-Dusun) phonemic glide involved the reinterpretation of a non-phonemic glide. PM *hiu(ʔ) 'shark' is reconstructed with a VV sequence. Without having to add an intervocalic *y to the reconstruction, *hiu(ʔ) > ⁺iu > ⁺iyu > *iyəʊ* can be suggested.

To sum up, PM intervocalic glides have been preserved in NEPMs. Additionally, these varieties have reinterpreted earlier epenthetic *[w] following *u and *[y] following *i as phonemic.

7.3.5.2 PM final glides

The reason why PM *-aw and *-ay were treated as diphthongs rather than VC sequences was not clearly stated in Adelaar's reconstruction.[58] "Diphthongs" in contemporary languages which constitute evidence for the reconstruction of *-aw and *-ay are only found in final position, and their phonological properties are often ambiguous. For instance, SM *-aw* and *-ay* are treated as diphthongs in Yunus Maris (1980: 41–43) and Teoh (1994: 23), but as VC sequences in Zaharani Ahmad (1993). Presumably the consideration of taking PM *-aw and *-ay as diphthongs has its root in the Austronesian scholarly history, as PM *-aw and *-ay were continuations of PAn/PMP diphthongs *-aw and *-ay. However, "diphthongs" reconstructed at higher levels are also open to different interpretations, and the usage of this term has been questioned. Drawing on abundant synchronic and diachronic evidence, Clynes (1997, 1999) argued that the so-called PAn "diphthongs" do not behave like single complex vowel phonemes; instead, the final glides served as syllable codas at the phonological level. Replying to Clynes (1997), Blust (1998) defended the proposition of PAn diphthongs, while admitting that PAn/PMP *-aw, *-ay, *-uy and *-iw were indeed VC sequences (also see Blust 1990: 235–236). The principal reason for labelling these vowel-glide sequences "diphthongs" lies in their distinctive behaviour when compared with other VC sequences. As Blust (1998: 359–361) asserted, these "diphthongs" were VC sequences "in which the C had special vocalic properties", sequences that "have a marked tendency to monophthongise", and because of their unique behaviour in phonological change, a separate class of diphthongs needs to be organised. In other words, Blust acknowledged that PAn/PMP "diphthongs" were not diphthongs from the standard view (i.e., they were not single phonemic complex vowels), but the label was maintained because they were considered a special type of VC sequences, especially viewed from a diachronic perspective, marked by the tendency of monophthongisation. The same view was held in Blust (2009: 584). For PM *-aw and *-ay, a similar argument might be put forward, but whether this argument actually holds requires examination. In the following discussion, I argue that PM *-aw and *-ay do not necessarily exhibit a stronger propensity to monophthongise; and when monophthongisation is

[58] Adelaar (1992: 100) wrote "there were two diphthongs (*-ay and *-aw): both occurred lexeme finally only, and both are analysable as *a + a semivowel".

attested, *-aw and *-ay did not form a special class on their own. *-aw and *-ay should be taken as vowel-glide sequences which are no more special than other VC sequences, and no separate label is needed.

Examples illustrating the development of PM *-aw and *-ay are provided in Table 7.18. Both *-aw and *-ay are reflected as *a* in KM and CTM, whereas in Dusun, *-aw is reflected as *ɔ* and *-ay as *a*.

Table 7.18: Reflexes of PM vowel-glide sequences

PM	KM	CTM	Dusun	SM	Gloss
*pul**aw**	*pul**a***	*pul**a***	*pul**ɔ***	*pul**aw***	'island'
*hij**aw**	*ij**a***	*ij**a***	*ij**ɔ***	*hij**aw***	'green'
*lant**ay**	*lat**a***	*lat**a***	*lat**a***	*lant**ay***	'floor'
*suŋ**ay**	*suŋ**a***	*suŋ**a***	*suŋ**a***	*suŋ**ay***	'river'

At first glance, the diachronic paths of *-aw and *-ay seem to confirm Blust's observation: the glide elements in both vowel-glide sequences were lost (in other words, diphthongs were monophthongised), leaving a non-high vowel *-a* or *-ɔ*. However, as has been demonstrated so far, all final consonants have a tendency to be neutralised or lost in the course of history in NEPMs, and the deletion of *-w and *-y is neither surprising nor special. In fact, "monophthongisation" of *-aw and *-ay can be easily compared with the evolution of other *-aC sequences.

In KM, not only final glides were lost, final liquids in *-ar and *-al were also deleted with the same result of maintaining ultimate *a, as shown in (4). Only one phonological change needs to apply here – deleting all final approximants (also see Clynes 1997: 356).

(4) Loss of final approximants in KM

*pula**w**	>	*pula*	'island'
*hija**w**	>	*ija*	'green'
*lanta**y**	>	*lata*	'floor'
*suŋa**y**	>	*suŋa*	'river'
*aka**r**	>	*aka*	'root'
*laya**r**	>	*laya*	'sail'
*jua**l**	>	*juwa*	'to sell'
*tuŋga**l**	>	*tuŋga*	'single'

In CTM, a similar diachronic process can be observed, but the outcomes of final approximant deletion are slightly different. While *-a- preceding *-w, *-y and *-l is reflected as *a*, whereas *-ar is reflected as *ɔ*, as illustrated in (5).

(5) Loss of final approximants in CTM

*pulaw	>	*pula*	'island'
*hijaw	>	*ija*	'green'
*lantay	>	*lata*	'floor'
*suŋay	>	*suŋa*	'river'
*akar	>	*akɔ*	'root'
*layar	>	*layɔ*	'sail'
*jual	>	*juwa*	'to sell'
*tuŋgal	>	*tuŋga*	'single'

The development of *-aw and *-ay actually follows a pattern similar to that of *-al, leaving *-ar an outlier, which seems to suggest that *-ar is more special than *-aw, *-ay and *-al. Yet in fact, raising and rounding of ultimate *-a- > ɔ is also found before a few other consonants, namely *-k, (*-ʔ) and *-h:

(6) Raising and rounding of PM ultimate *-a- in CTM

*anak	>	*anɔʔ*	'child'
*miɲak	>	*miɲɔʔ*	'oil'
*kakaʔ	>	*kakɔʔ*	'older sibling'
*darah	>	*daɣɔh*	'blood'
*pəcah	>	*pəcɔh*	'to break'

The reason why *-ar patterns with *-ak, (*-aʔ), and *-ah is not immediately clear. While [+back] is a common feature of these final consonants, ultimate *-a- before other C[+back], such as *-ŋ and *-w, was exempted from raising and rounding. In any case, a two-stage process can be inferred from the changes in (5): ultimate *-a- was raised and backed before *-r (as before some other back consonants), and subsequently final approximants were deleted. *-aw and *-ay were thus not necessarily more susceptible to monophthongisation, and they did not form a special class in their diachronic behaviour.

The development of *-aw > *ɔ* and *-ay > *a* in Dusun presents a particularly interesting case. As in KM and CTM, final liquids were also deleted in Dusun. In (7), it is evident that the diachronic development of *-aw aligns with that of *-ar, whereas *-ay aligns with *-al.

(7) Loss of final approximants in Dusun

*pula**w**	>	*pulɔ*	'island'
*hija**w**	>	*ijɔ*	'green'
*lanta**y**	>	*lata*	'floor'
*suŋa**y**	>	*suŋa*	'river'
*aka**r**	>	*akɔ*	'root'
*laya**r**	>	*layɔ*	'sail'
*jua**l**	>	*juwa*	'to sell'
*tuŋga**l**	>	*tuŋga*	'single'

A closer look reveals that *-aw not only aligns with *-ar, but with all *-aC sequences in which *C was [+back], namely *-k, (*-ʔ), *-ŋ and *-h. In contrast, *-ay aligns with all *-aC sequences with [-back] codas, namely *-p, *-t, *-m, *-n, *-s and *-l.

(8) Ultimate *-a- > *ɔ* preceding *C[+back]

*pul**aw**	>	*pul**ɔ***	'island'
*ak**ar**	>	*ak**ɔ***	'root'
*an**ak**	>	*an**ɔʔ***	'child'
*kak**aʔ**	>	*kak**ɔʔ***	'older sibling'
*dar**ah**	>	*day**ɔh***	'blood'
*bint**aŋ**	>	*bit**ɔŋ***	'star'

(9) Ultimate *-a- > *a* preceding *C[-back]

*lant**ay**	>	*lat**a***	'floor'
*ju**al**	>	*juw**a***	'to sell'
*say**ap**	>	*say**aʔ***	'wing'
*bər**at**	>	*bəɣ**aʔ***	'heavy'
*hit**am**	>	*it**aŋ***	'black'
*jal**an**	>	*jal**aŋ***	'road'
*at**as**	>	*at**ah***	'top'

Once again, final *-aC sequences in Dusun underwent a clear two-stage change: first, ultimate *-a- was raised and backed to *ɔ* preceding all *C[+back], and retained as *a* preceding any *C[-back] (also see §7.4.3). Second, all final approximants were deleted indiscriminately. The diachronic paths of *-aw and *-ay in Dusun also show that vowel-glide sequences were no more likely to monophthongise than other vowel-approximant sequences, and *-aw and *-ay did not pattern any closer than they did with other consonants. The development of ultimate *-a- in final *-aC sequences was only sensitive to the place feature of final consonants, and the "special vocalic properties" of glides played no role.

It is worth noting that NEPMs are by no means unique in this respect. In Perak Malay, another Peninsular Malayic variety, earlier vowel-glide sequences also followed the same diachronic path as vowel-liquid sequences. Citing Zaharani Ahmad (1993: 13), Clynes (1997: 357) presents the following cognate sets between SM and Perak Malay (SM forms can be taken as reflecting an earlier stage of Perak Malay forms):

(10)

SM	Perak Malay	Gloss
jual	*juɛ*	'to sell'
biar	*biɔ*	'to let'
kəday	*kədɛ*	'shop'
pisaw	*pisɔ*	'knife'

These examples further demonstrate that *-aw and *-ay did not form a distinct class, and *-aw patterns with *-ar, whereas *-ay patterns with *-al.

It must be conceded that there are Malayic varieties in which earlier *-aw and *-ay were monophthongised while other *VC sequences are preserved, indicating that vowel-glide sequences were grouped based on certain properties. In Jakarta Malay, for instance, *-aw and *-ay have been monophthongised to *o* and *e* respectively, as in *pul**o*** 'island' and *ram**e*** 'crowded' (Wallace 1976: 122, cf. SM *pul**aw*** and *ram**ay***), whereas final *C in most other *VC sequences are preserved (except for *-ah > *ɛ*). One could argue that the assumption of vowel-glide sequences being more likely to monophthongise over time has a general phonetic basis, since [aw] and [ay] are indistinguishable from diphthongs [au] and [ai] at the phonetic level, whereas other VC sequences are not diphthong-like phonetically, hence less prone to monophthongise. However, if the inherent tendency to be monophthongised is generalised at the phonetic level, it would imply that the phenomenon is cross-linguistically universal and probably common in other languages and language families. If that is the case, it also makes little sense to label *-aw and *-ay in Malayic and Austronesian languages alone as "diphthongs". These vowel-glide sequences can be simply treated as VC while having their special properties acknowledged. After all, each consonant phoneme is unique, and it is not surprising that phonological changes are sensitive to different features, therefore targeting groups of consonants in different ways. Labelling vowel-glide sequences as "diphthongs" is unnecessarily confusing. PM *-w and *-y should be considered as nothing more than phonemic glides in final position.

7.3.6 PM consonant sequences

Consonant sequences reconstructed in PM consisted of a nasal + a homorganic stop, or a velar nasal + *s.[59] Nasal + voiceless stop sequences, as well as sequences of a velar nasal + *s, have been reduced to their obstruent components in NEPMs, whereas sequences of a nasal + a voiced stop are preserved. Examples illustrating the development of PM consonant sequences are provided in Table 7.19.

Table 7.19: Reflexes of PM consonant sequences

PM	KM	CTM	Dusun	SM	Gloss
*ru**mp**ut	*ɣupuʔ*	*ɣupuʔ*	*upuʔ*	*rumput*	'grass'
*bi**nt**aŋ	*bitɛ*	*bitaŋ*	*bitɔŋ*	*bintaŋ*	'star'
*ci**ɲc**in	*ciciŋ*	*ciciŋ*	*ciciŋ*	*ciɲciŋ*	'ring'
*bə**ŋk**ak	*bəkɔʔ*	*bəkɔʔ*	*bəkɔʔ*	*bəŋkak*	'to swell'
*la**ŋs**uŋ	*lasoŋ*	*lasoŋ*	*lasəʊŋ*	*laŋsuŋ*	'directly'
*ra**mb**ut	*ɣamboʔ*	*ɣamboʔ*	*ɣambuʔ*	*rambut*	'hair'
*di**nd**iŋ	*dindiŋ*	*dindiŋ*	*dindɛiŋ*	*dindiŋ*	'wall'
*pa**ɲj**aŋ	*paɲjɛ*	*paɲjaŋ*	*paɲjɔŋ*	*paɲjaŋ*	'long'
*ta**ŋg**aʔ	*taŋgɔ*	*taŋgə*	*taŋgɛ*	*taŋga*	'ladder'

In one KM instance, an earlier nasal + voiced stop sequence has been reduced to its nasal component: *a**ɲj**iŋ 'domestic animal' > *aɲiŋ* 'dog'. The deletion or weakening of voiced stops following nasals has been described as a more general phenomenon in earlier studies on KM (Nik Safiah 1965, 1967; Abdul Hamid 1994), but in the data I collected, such stops are always

[59] They were termed as "consonant clusters" in Adelaar (1992), but they only occurred intervocalically and belong to two different syllables. I distinguish heterosyllabic consonant sequences from tautosyllabic consonant clusters, and reserve the term "consonant clusters" to refer to tautosyllabic onsets or codas.

audible except in *aɲiŋ* 'dog'. It is possible that there are regional variations within KM in this respect. One may also speculate that the voiced stops were once reduced or weakened but restored more recently under the influence of SM, but there is only indirect evidence for this suggestion. In Patani Malay, a close relative to KM with little SM influence, all nasal + voiced stop sequences seem to be reduced to the nasal components (Ruslan Uthai 2011). The reduction of voiced stops is also seen in some ITM subvarieties; for instance, in Tanjung Baru, the stop components in nasal + voiced stop sequences are only weakly audible. The following examples are best transcribed with stops in superscripts: [ɣam^{b}əʊ̯ʔ] 'hair', [tan^{d}aʊ̯ʔ] 'horn', [aɲ^{j}ai̯ŋ] 'dog', [puŋ^{g}aʊ̯ŋ] 'buttock' (cf. SM *rambut*, *tanduk*, *aɲjiŋ* and *puŋguŋ*).[60]

In addition to nasal + obstruent sequences, PM probably had a *-rC- sequence, which was reconstructed as *-r(ə)C-. After reevaluating Adelaar's reconstruction and taking new data into account, I suggest *-rC- to be reconstructed instead of the more ambiguous *-r(ə)C-.

Consonant sequences comprising an *r* directly followed by another C are commonly attested in SM, as in words like *tə**rj**un* 'to jump down', *tə**rb**it* 'to emerge', *bə**rs**ih* 'clean' and *cə**rm**in* 'mirror'. The difficulty in reconstructing these *-rC-* sequences to PM lies in their irregular correspondences in other Malayic varieties, which can 1) have no traces of *r*, 2) have a suspicious epenthetic vowel breaking up the sequence, or 3) have a possibly inherited vowel between *r* and C. Even in Adelaar's SM material, one lexeme was sometimes written in several ways with either *-rC-* or *-rVC-*.[61] Conflicting data in the original reconstruction mainly came from Banjar Hulu (BH) and Iban. Adelaar (1992: 87–89) tried to resolve the problem in the following ways: if one or more varieties have a V following *r* (henceforth post-*r* V) that is likely inherited (in this case SM *ə*, BH *a* and Iban *ə* are probably epenthetic), *-rVC- was reconstructed. Otherwise, it was impossible to determine whether PM had a post-*r* V, and thus *-r(ə)C- was reconstructed. Table 7.20 summarises some of Adelaar's SM, BH and Iban material, reorganised into three classes. Available NEPM cognates have been added, and their relevance is demonstrated below.

---

[60] Similar sounds have been referred to as post-occluded nasals in Jambi Malay (Yanti 2010), postploded nasals in Mualang (Tjia 2007) and "funny" nasals in Acehnese (Durie 1985).

[61] Adelaar's main source for SM is Wilkinson (1959)'s dictionary, in which the variable spellings probably represent regional differences of the recorded Malay varieties.

Table 7.20: Correspondences with SM *-rC-* sequences (adapted from Adelaar 1992: 88)

SM	BH	Iban	KM	CTM	Dusun	Gloss
Class 1						
bərnas, bərənas	*barunas*	–	*bɣənah*	*wwənah*	*bunah*	'rice ears'
bərsih, bərisi, bərisih	*barasih*	*bərəsi*	–	–	–	'clean'
kərdut, kədut, kərut, kərudut	–	*kədut*	*kəduʔ*	*kəduʔ*	*kəduʔ*	'crease, wrinkle'
pərcik, pərcit	*puracit*	*pəraɲcit, pərəɲcit*	*mməciʔ*	*ŋŋəciʔ*	*mmuciʔ*	'to squirt'
tərjaŋ, tərajaŋ	*tirajaŋ, tarajaŋ*	*tərajaŋ*	*tɣajaŋ*	*tɣajaŋ*	*tɣajɔŋ*	'to kick'
Class 2						
bərkas	–	*bərəkas*	–	*wwəkah*	*bukah*	'bundle'
bərsin	–	*bərəsin*	*bɣəsiŋ*	*bɣəsiŋ*	*busiŋ*	'to sneeze'
cərmin, cərəmin	*caramin*	*cərəmin*	*cɣəmiŋ*	*cɣəmiŋ*	*cɣəmiŋ*	'mirror'
jərnih, jənih	*jaranih*	–	*jɣəneh*	*jɣəneh*	*jɣənih*	'clear'
kərbaw	–	*kərəbo, kərəbaw*	*kuba*	*kubɔ*	*kubɔ*	'buffalo'
tərbaŋ, tərəbaŋ	*tarabaŋ*	*tərəbay*	*tɣəbɛ*	*tɣəbaŋ*	*tubɔŋ*	'to fly'
tərbit	–	*tərəbit*	*tubeʔ*	*tubeʔ*	*tubiʔ*	'to emerge'
Class 3						
kərbat, kərəbat, kəbat	*kabat*	*kəbat*	–	*xxəbaʔ*	–	'to bind'
tərjun	*tajun*	*tərəjun*	*tujoŋ~tɣəjoŋ*	*tɣəjoŋ*	*tujuŋ*	'to jump down'
kərdil, kərədil	–	*kədil*	–	*xxidiʔ~kkidiʔ*	*xxədi*	'stunted'

For the five cognate sets in the first class, SM, BH or Iban has a post-*r* V that seems inherited, namely BH *barunas*, SM *bərisih*, SM *kərudut*, Iban *pəraɲcit*, SM *tərajaŋ* and Iban *tərajaŋ*, which yielded the reconstructions of *bArunas, *bArisih, *kArudut, *pura(ɲ)ci(kt) and *tirajaŋ (the reconstruction of antepenultimate vowels is not relevant for now). In the second class, both BH and Iban have *-rVC-* corresponding to SM *-rC-*, but the post-*r* Vs in these cases are presumably the results of epenthesis. In the third class, BH and/or Iban has no traces of *r*, instead having *-VC-* corresponding to SM *-ərC-*. For both second and third classes, *-r(ə)C- was reconstructed, e.g., *bVr(ə)kas 'bundle', *bVr(ə)sin 'to sneeze', *kAr(ə)bat 'to bind' and *tVr(ə)jun 'to jump down'. Adelaar (1992: 89) implicitly opted for the reconstruction of *-rəC-, suggesting that SM *-rC-* could result from the syncope of post-*r* V; but given that no direct evidence was available, an ambiguous *-r(ə)C- was proposed.

I accept Adelaar's reconstructions for the cognate sets in the first class, as post-*r* V in BH and/or Iban in these sets cannot be sufficiently explained otherwise. However, as I will show, data from NEPMs suggest that *bərnas and *pərcit should probably be reconstructed as doublets alongside *bArunas and *pura(ɲ)ci(kt). Such doublet forms are also suggested by the large range of variation attested in SM. For the items in the second and third classes, I argue that the reconstructed forms in PM can only have *-rC-, and the possibility of having *-rəC- can be ruled out based on reflexes in NEPMs.

From the comparisons in Table 7.20, an important sound correspondence between NEPMs and SM can be established, namely NEPM *-u-* : SM *-ər(ə)-*. Though not completely regular, a large number of cognate sets exemplify this correspondence, especially between Dusun and SM, as presented in Table 7.21.

Table 7.21: Correspondences of NEPM *-u-* : SM *-ər(ə)-*

KM	CTM	Dusun	SM	Gloss
kuba	*kubɔ*	*kubɔ*	*kərbaw*	‘buffalo’
tubeʔ	*tubeʔ*	*tubiʔ*	*tərbit*	‘to emerge’
tujoŋ, tɣəjoŋ	*tɣəjoŋ*	*tujuŋ*	*tərjun*	‘to jump down’
–	*wwəkah*[a]	*bukah*	*bərkas*	‘bundle’
bɣənah	*wwənah*	*bunah*	*bər(ə)nas*	‘rice ears’
mməciʔ[b]	*ŋŋəciʔ*	*mmuciʔ*	*pərcik~pərcit*	‘to squirt’
bɣəsiŋ	*bɣəsiŋ*	*busiŋ*	*bərsin*	‘to sneeze’
tɣəbɛ	*tɣəbaŋ*	*tubɔŋ*	*tər(ə)baŋ*	‘to fly’

[a] CTM *ww-* reflects earlier ⁺bɣ-, see §7.5.2.3.
[b] KM *mməciʔ*, CTM *ŋŋəciʔ* and Dusun *mmuciʔ* contain a petrified *NN-* prefix, which was presumably attached to the bases *pəciʔ*, *kəciʔ* and *puciʔ* respectively. The initial consonant *k-* in CTM *kəciʔ* is unexplained.

Three more sets of correspondences may be added:

1) KM *bɣənɛ* : CTM *bɣənaŋ* : Dusun *bunɔŋ* : SM *bərənaŋ* ‘to swim’
2) KM *buwi* : CTM *buwi* : Dusun *buwɛi* : SM *bəri* ‘to give’ < PM *bəriʔ
3) KM *puyuʔ* : CTM *puyuʔ* : Dusun *puyəʊʔ* : SM *pəriuk* ‘cooking pot’ < PMP *pariuk

The correspondences in these examples illustrate a sound change of ⁺-ər- > *-u-*, which is best exemplified in the development of PM *bəriʔ ‘to give’ > KM/CTM *buwi*, Dusun *buwɛi*, SM *bəri*, as discussed in §7.3.4. PMP *pariuk ‘cooking pot’ > KM/CTM *puyuʔ*, Dusun *puyəʊʔ*, SM *pəriuk* must have had a similar trajectory with the intermediate stages of ⁺pəriuk (antepenultimate schwa neutralisation) and ⁺pəryuk (reinterpretation of ⁺-i- to ⁺-y-). It is safe to conclude that when NEPMs have a *u* corresponding to SM *-ər(ə)-*, this *u* must have developed from earlier ⁺-ər-. More importantly, all NEPM *-u-* corresponding to SM *-ər(ə)-* are found in the penultimate syllable immediately preceding a consonant. Based on this environment alone, it can be suggested that the earlier form that gave rise to NEPM *-uC-* had an ⁺-ərC- sequence, and it was presumably disyllabic.

Now let us consider another piece of evidence from the perspective of syllable structure. The uncertainty in the reconstruction of *-r(ə)Cessentially concerns the status of the post-*r* schwa. The ambiguous reconstruction of *tVr(ə)bit 'to emerge' entails two possibilities: either a disyllabic form *tVr.bit or a trisyllabic form *tV.rə.bit. The significance of NEPM data in supporting *-rC- and against *-rəC- also lies in that PM disyllables and trisyllables have divergent histories. All PM trisyllables have been reduced to disyllables, as seen in the following examples (see more discussions in §7.5.2):

Table 7.22: Reduction of PM trisyllables

PM	KM	CTM	Dusun	SM	Gloss
*bAlakaŋ	*blakɛ*	*blakaŋ*	*blakɔŋ*	*bəlakaŋ*	'back'
*tAliŋa(ʔ)	*tliŋɔ~lliŋɔ*	*lliŋə*	*tliŋɛ*	*təliŋa*	'ear'
*tiŋgələm	*tgəlɛ*	*tgəlaŋ~ggəlaŋ*	*tŋəlaŋ*	*təŋgəlam*	'to sink'
*tirajaŋ	*tɣajaŋ*	*tɣajaŋ*	*tɣajɔŋ*	*tərajaŋ, tərjaŋ*	'to kick'

In the process of reduction, all antepenultimate vowels were deleted, whereas all penultimate vowels are retained regardless of vowel quality. A hypothetical PM trisyllabic form with *-rəC-, for instance *tV.rə.bit 'to emerge', would be reflected as KM/CTM ˣ*tɣəbeʔ* and Dusun ˣ*tɣəbiʔ* with the deletion of the antepenultimate vowel and the retention of the penultimate vowel. The expected forms are nevertheless contradicted by the attested forms KM/CTM *tubeʔ* and Dusun *tubiʔ*, which reflect ⁺-ərC- > *-uC-*. From this reasoning, the reconstruction of PM trisyllabic forms with ˣ*-rəC- can be ruled out. The established sound change ⁺-ərC- > *-uC-* could only take place in an earlier disyllabic word; hence PM must have had *-ərC-, i.e., a penultimate schwa and a *-rC- sequence.

As mentioned earlier, the correspondence between NEPM *-u-* and SM *-ər(ə)-* is not always regular. In some instances, KM and CTM have *-ɣə-* corresponding to Dusun *-u-* and SM *-ər-*, e.g., KM/CTM ***bɣə**siŋ* : Dusun ***bu**siŋ* : SM ***bər**sin* 'to sneeze', and in a few other cases, it also appears that Dusun fails to reflect ⁺-ər- as *-u-*, e.g., Dusun ***cɣə**miŋ* : SM ***cər(ə)**min* 'mirror'. I suggest that these NEPM forms with unexpected *-ɣəC-* are the results of more recent borrowings, or simply an adaptation of the pronunciation of SM words. There are several reasons for this suggestion. First, *-ɣəC-* is found in some apparent loanwords, e.g., KM *bɣəseh*, Dusun *bɣəsɛih* 'clean' < SM *bərsih* (native terms are KM/CTM *cuci* and Dusun *cucɛi*), and NEPM *xxətah* 'paper' < $^{+}$kɣətah < SM *kərtas* < Arabic *qirṭas* ($^{+}$kɣ- > *xx*- is a regular reciprocal assimilation, see §7.5.2.3). Second, -*ɣəC*- also appears in the sound adaptation of toponyms, e.g., Dusun *pɣəlih* < SM *Pərlis* (name of a northern Malay state). Third, variation between -*uC*- and -*ɣəC*- is sometimes attested, as in KM *tujoŋ~tɣəjoŋ* 'to jump down'. Older speakers tend to prefer *tujoŋ*, while younger speakers often use *tɣəjoŋ*. Some Dusun speakers also occasionally pronounce *tubɔŋ* 'to fly' as [tɣəbɔŋ] and *bunɔŋ* 'to swim' as [bɣənɔŋ] (cf. SM *tərbaŋ* and *bərənaŋ*), but when asked for confirmation, they would correct the pronunciations to *tubɔŋ* and *bunɔŋ*. These observations all indicate that -*ɣəC*- has a shallower history, and -*u*- bears the more authentic pronunciation. In essence, there is a two-layer reflexes of PM *-ərC-. The original outcome is -*uC*-, but the sound change of *-ərC- > -*uC*- must have ceased to operate at a relatively early stage. Later borrowings of SM -*ərC*- have been adapted to -*ɣəC*-, which presumably underwent an intermediate stage of $^{+}$-əɣəC-, e.g., SM *bərsih* 'clean' > $^{+}$bəɣəsih > KM *bɣəseh*, Dusun *bɣəsɛih* (deletion of antepenultimate schwas). It is not surprising that Dusun attests more forms with -*uC*- whereas KM and CTM often have -*ərC*-, since Dusun is the most conservative variety among NEPMs with least external influence.

Based on the cognate sets discussed above, I reconstruct:

(11)	*kərbaw	'buffalo'
	*tərbit	'to emerge'
	*tərjun	'to jump down'
	*bərkas	'bundle'
	*bərnas	'rice ear'
	*pərcit[62]	'to squirt'
	*bərsin	'to sneeze'
	*tərbaŋ	'to fly'
	*bərnaŋ	'to swim'

Since *bərnas 'rice ear' and *pərcit 'to squirt' cannot account for BH and Iban reflexes, I suggest that they are best viewed as doublets alongside Adelaar's original reconstructions *bArunas and *pura(ɲ)ci(kt). Other forms such as $^{\times}$*cərmin 'mirror' and $^{\times}$*jərnih 'clear' cannot be reconstructed given

62 *pərcit with final *-t instead of *$^{\times}$-k is reconstructed based on Dusun -*iʔ*. *$^{\times}$-ik would have been reflected as *-ɛiʔ in Dusun, see §7.4.4.2.

the lack of any material reflecting *-ər- as *-u-*.[63]

Some additional notes are necessary for the reconstructions in (11). First, *kərbaw 'buffalo' is presumably a loanword from Mon-Khmer, as previously suggested by Thurgood (1999: 322), but I suspect that the borrowing from Mon-Khmer to Malayic predated PM. While cognates of SM *kərbaw* are attested in Chamic and Malayic, they only have limited presence elsewhere in the Austronesian family. In Mon-Khmer languages, however, related forms are widespread in all branches.[64] This distribution suggests that the ultimate origin of this word is Mon-Khmer. Nonetheless, NEPM forms are unlikely to be borrowings from SM or other present-day neighbouring Mon-Khmer languages,[65] and the correspondence of NEPM *-u-* : SM *-ər-* still points to a PM reconstruction *kərbaw, which might be ultimately a pre-PM loanword.

Second, a PMP reconstruction *bəRkəs 'bundle, package' can be found in ACD, with reflexes attested in various Philippine languages and Central-Eastern Malayo-Polynesian languages. This high-order reconstruction lends extra credence to the reconstruction of PM *bərkas 'bundle', rather than raising suspicions about a schwa breaking up the original *-Rk- sequence.

Third, it is beyond trivial that other items typically have either *tər- or *bər-, both of which have reflexes as common prefixes in present-day Malayic varieties. This observation alludes to the possibility that these words could have a morphologically complex structure. The speculation is further encouraged by Blust's note in ACD on *-baŋ 'to fly' being an Austronesian root, which supports the reconstruction of *tər-baŋ, instead of ˣ*tərəbaŋ with a penultimate schwa without a clear source.

Fourth, I revise Adelaar's *(mb)A-rənaŋ 'to swim' to *bərnaŋ. Adelaar (1992: 138) suggested the following phonological history for SM *bərənaŋ* 'to swim': PMP *(ln)aŋuy 'to swim' > PM *(mb)Ar- + *(ln)aŋi > ⁺bər-naŋi > ⁺bər(ə)naŋ-i > SM *bərənaŋ*. In fact, from Adelaar's own interpretation, there is no reason to reconstruct a penultimate schwa to PM, and it appears that *-naŋ in *bərnaŋ can also be seen as an earlier root. I consider the penultimate schwa in SM *bərənaŋ* 'to swim' as resulting from post-PM accretion, and

[63] Blust & Trussel (2023) suspect an ultimate Indic origin for SM *cərmin* 'mirror'.

[64] Shorto (2006: 93) reconstructs Proto Mon-Khmer *krpiʔ, *krpiiw, *krpuʔ, *[kr]puh 'buffalo'.

[65] Words for 'buffalo' in Aslian languages are often borrowed from SM *kərbaw*, e.g., Jahai *krbɔw* (Burenhult 2005) and Temiar *kəɹbau* (Means 1998).

SM root *rənaŋ* 'to swim' from the backformation of *bər-naŋ to ⁺bə-rənaŋ.

Lastly, except for *bərkas 'bundle', all other reconstructions only have a handful of cognate correspondences, if any, outside Malayic. Only PMP *bəRkəs 'bundle, package' and PWMP *burəsin 'to sneeze' are reconstructed in a higher-order proto language, and for the latter reconstruction, only Makassarese *burassiŋ* is listed as a reflex outside Malayic. Forms related to *kərbaw* 'buffalo' outside Malayic and Chamic are often borrowed from SM. The limited distribution provokes the question of how words with *-ərC- ended up in Malayic. This remains a subject of further discussion, but at least from a bottom-up reconstruction based on available evidence, I consider items in (11) reconstructable to PM for now.

The irregularity in the original material cited in Adelaar (1992: 87–89) (see Table 7.20) can also be better explained by the reconstructions proposed in (11). In any case, it seems that PM *-ərC- was not stable, displaying a tendency to undergo changes in various directions. Several paths can be identified: 1) the pre-C **r* was lost, as in *tərjun 'to jump down' > BH *tajun* (PM penultimate *ə > BH *a* is regular), 2) a secondary V was inserted between *-rC-, as in *tərjun > Iban *tərəjun*, *bərnaŋ 'to swim' > SM *bərənaŋ*, and 3) *-ər- preceding *C became *-u-*, as in *tərjun > Dusun *tujuŋ*. On the contrary, if PM had a penultimate schwa in ˣ*-ərəC-, it is improbable that a hypothetical ˣ*tərəjun could have given rise to either Dusun *tujuŋ* or BH *tajun*. The irregular correspondences between BH, Iban and SM might also be attributed to secondary borrowing, but a more comprehensive treatment of data from each variety is required to clarify this issue.

7.4 Reflexes of PM vowels

The four vowels reconstructed in PM, namely *a, *ə, *i and *u, were allowed in either syllable of a disyllabic word, but *ə could not occur in an absolute final position, nor can it be followed by *ʔ or *h. From PM to NEPMs, substantial changes have taken place as regards ultimate vowels, and the direction of changes is closely associated with final segments. In comparison, PM penultimate vowels are relatively stable, except that some high vowels were sporadically lowered to *ɛ* and *ɔ* in KM and CTM.

Table 7.23 provides an overview of the reflexes of PM vowels in NEPMs. The symbol "σ" stands for a syllable.

Table 7.23: Overview of reflexes of PM vowels

PM	Environment	KM	CTM	Dusun	SM
*a	σ̲σ#	*a*	*a*	*a*	*a*
	_(*ʔ)#	*ɔ*	*ə*	*ɛ*	*a*
	_*p, *t, *s, *l#	*a*	*a*	*a*	*a*
	_*k, *h#	*ɔ*	*ɔ*	*ɔ*	*a*
	_*m, *n#	*ɛ*	*a*	*a*	*a*
	_*ŋ#	*ɛ*	*a*	*ɔ*	*a*
	_*r#	*a*	*ɔ*	*ɔ*	*a*
*ə	σ̲σ#	*ə*	*ə*	*ə*	*ə*
	#(*h)_	*Ø*	*Ø*	*Ø*	*ə*
	_*p, *t, *s, *l#	*a*	*a*	*a*	*a*
	_*k#	*ɔ*	*ɔ*	*ɔ*	*a*
	_*m, *n#	*ɛ*	*a*	*a*	*a*
	_*ŋ#	*ɛ*	*a*	*ɔ*	*a*
	_*r#	*a*	*ɔ*	*ɔ*	*a*
*i	σ̲σ#	*i/ɛ*	*i/ɛ*	*i*	*i/ɛ*
	_(*ʔ)#	*i*	*i*	*ɛɪ*[a]	*i*
	_*p, *t, *s#	*i/e*	*i/e*	*i*	*i*
	_*k#	*i/e*	*i/e*	*ɛɪ*[a]	*i*
	_*m, *n#	*i*	*i*	*i*	*i*
	_*ŋ#	*i*	*i*	*ɛɪ*[a]	*i*
	_*h#	*e*	*e*	*ɛɪ*[a]	*i*
	_*l, *r#	*e*	*e*	*i*	*i*
*u	σ̲σ#	*u/ɔ*	*u/ɔ*	*u*	*u/ɔ*
	_(*ʔ)#	*u*	*u*	*əʊ*[b]	*u*
	_*p, *t, *s#	*u/o*	*u/o*	*u*	*u*
	_*k#	*u/o*	*u/o*	*əʊ*[b]	*u*
	_*m, *n#	*o*	*o*	*u*	*u*
	_*ŋ#	*o*	*o*	*əʊ*[b]	*u*
	_*h#	*o*	*o*	*əʊ*[b]	*u*
	_*l, *r#	*o*	*o*	*u*	*u*

[a] Only following oral onsets, otherwise *i*.
[b] Only following oral onsets, otherwise *u*.

The general changes of PM vowels can be summarised as follows:

1) Penultimate *a and *ə are retained, except for the loss of schwa in word-initial position and schwa preceded by *h;
2) Penultimate *i and *u were sporadically lowered to *ɛ* and *ɔ* in KM and CTM, but not in Dusun;
3) Ultimate *a and *ə were merged (if *ə is reconstructable in this position), followed by changes in various directions depending on the word-final segment;
4) Ultimate *i and *u in closed syllables were often lowered to *e* and *o* in KM and CTM, and the lowering was regular preceding certain consonants but sporadic preceding other consonants;
5) Ultimate *i and *u in open syllables and closed syllables with a back coda consonant (*k, (*ʔ), *ŋ and *h) were diphthongised to *ɛi* and *əʊ* in Dusun, and diphthongisation only took place following oral onsets.

The following discussions are divided into four subsections, elaborating on the developments of PM non-high and high vowels in penultimate and final syllables separately.

7.4.1 PM penultimate *a and *ə

Table 7.24 displays the reflexes of PM non-high vowels in penultimate position. *a and *ə are retained unless *ə occurred in an onsetless syllable, in which case the word-initial schwa was lost and the original disyllabic word was shortened to a monosyllabic one (see §7.5.1).

Table 7.24: Reflexes of PM penultimate *a and *ə

PM	KM	CTM	Dusun	SM	Gloss
*mata	*matɔ*	*matə*	*matɛ*	*mata*	‘eye’
*kaki	*kaki*	*kaki*	*kakɛi*	*kaki*	‘foot; leg’
*bərat	*bəɣaʔ*	*bəɣaʔ*	*bəɣaʔ*	*bərat*	‘heavy’
*təbu	*təbu*	*təbu*	*təbəʊ*	*təbu*	‘sugarcane’
*əmpat	*ppaʔ*	*ppaʔ*	*ppaʔ*	*əmpat*	‘four’
*ənəm	*nnɛ*	*nnaŋ*	*nnaŋ*	*ənam*	‘six’

Penultimate *ə preceded by *h was also deleted following the regular loss of initial *h (see §7.3.3). Compare NEPM *mbuh* ‘to blow’ with SM *həmbus*, and KM *ttɛ* ‘to punch’ with SM *həntam* (no PM reconstructions are available).

7.4.2 PM penultimate *i and *u

PM penultimate high vowels are mostly retained, but they have been lowered to *ɛ* and *ɔ* in certain words in KM and CTM. The lowering of high vowels was of a sporadic nature without clear conditions. In contrast, all penultimate high vowels retain their height in Dusun. The retention and sporadic lowering of PM penultimate *i and *u are presented in Table 7.25.

Table 7.25: Reflexes of PM penultimate *i and *u

PM	KM	CTM	Dusun	SM	Gloss
Retention					
*pisaŋ	*pisɛ*	*pisaŋ*	*pisɔŋ*	*pisaŋ*	‘banana’
*ikan	*ikɛ*	*ikaŋ*	*ikaŋ*	*ikan*	‘fish’
*kulit	*kuleʔ*	*kuleʔ*	*kuliʔ*	*kulit*	‘skin’
*puluh	*puloh*	*puloh*	*puləʊh*	*puluh*	‘ten’
Sporadic lowering in KM/CTM					
*ikur	*ɛkɔ*	*ɛkɔ*	*iku*	*ɛkɔr*	‘tail’
*libar	*lɛba*	*lɛbɔ*	*libɔ*	*lɛbar*	‘wide’
*cucuk	*cɔcɔʔ*	*cɔcɔʔ*	*cucəʊʔ*	*cucuk*	‘to prick’
*uraŋ	*ɔɣɛ*	*ɔɣaŋ*	*uɣɔŋ*	*ɔraŋ*	‘person; human’

The exact mechanism behind the sporadic lowering of *i and *u is unclear, but this change is quite common in Malayic varieties in general. For SM, Wolff (2010: 478) vaguely suggests that the sporadic lowering might be attributed to contact with other (Austronesian) languages that have regularly lowered penultimate high vowels. It seems that the splits of *i > *i, e* and *u > *u, o* have occurred multiple times independently across Malayic varieties, as they did not affect the exact same set of words. For instance, the penultimate *u in *cucuk ‘to prick’ has been lowered to *ɔ* in KM/CTM *cɔcɔʔ*, but it has been retained as *u* in SM *cucuk*.

7.4.3 PM ultimate *a and *ə

In final syllables, Adelaar reconstructed *ə alongside *a in closed syllables before consonants other than the glottals *h and *ʔ. The reconstruction is on the basis of Jakarta Malay ultimate *ə*, which was considered a retention from PMP. Multiple lines of evidence supporting ultimate *ə were discussed in Adelaar (1992: 33–39), but this reconstruction remains contested for several reasons. First, most contemporary Malayic varieties, including otherwise conservative ones, do not show *a : ə* distinction in final syllables (Anderbeck in print). Second, *a : ə* distinction is also not found in Old Malay inscriptions (Blust 1988: 13). Third, all Malayic varieties that seem to retain ultimate *ə are in the vicinity of Java (Anderbeck 2019). It is likely that ultimate *ə* in Jakarta Malay and a few other geographically-adjacent Malayic varieties has resulted from contact with Javanese, Sundanese and/or Balinese, which regularly retain PMP ultimate *ə.

In NEPMs, there is no trace of ultimate *ə, and both PM ultimate *a and *ə have the same reflexes. If PM did retain ultimate *ə from PMP, *a and *ə must have merged to ⁺a before other sound changes took place. Subsequently ⁺a underwent changes in different directions, depending on the following segment. In open syllables and closed syllables with untraceable *-ʔ, ⁺a changed to *ɔ*, *ə* and *ɛ* in KM, CTM and Dusun respectively. In other closed syllables (including those with possibly retained *-ʔ), ⁺a was often raised and backed to become *ɔ*. Ultimate *a, *ə > ⁺a > *ɔ* is found:

1) in KM before *-k, (*-ʔ) and *-h;
2) in CTM before *-k, (*-ʔ), *-h and *-r;
3) in Dusun before *-k, (*-ʔ), *-h, *-r, *-ŋ and *-w.

All final consonants that triggered the raising and backing of ⁺a are back consonants (both dorsal and laryngeal), and ⁺a > *ɔ* in this environment can be seen as the result of coarticulation and a form of assimilation. Before final nasals, ⁺a is raised and fronted to *ɛ* in KM. This process presumably began with the nasalisation of ⁺a, followed by the raising of ⁺ã to ⁺ɛ̃ (which is a cross-linguistically common process, see Beddor 1983) and the later loss of vowel nasality.[66] In other environments, ⁺a is retained as *a*. Examples illus-

[66] This *ɛ* is often transcribed as a nasal vowel *ɛ̃* in previous literature (Nik Safiah 1965; Abdul Hamid 1994; Teoh 1994). Ajid (1997) noted that there are regional variations in terms of whether *ɛ* is realised as a nasal vowel.

trating the reflexes of PM ultimate *a and *ə are given in Table 7.26.

Table 7.26: Reflexes of PM ultimate *a and *ə

PM	KM	CTM	Dusun	SM	Gloss
Open syllables					
*dad**a**	*dadɔ*	*dadə*	*dadɛ*	*dada*	'chest'
*buŋ**a**(ʔ)	*buŋɔ*	*buŋə*	*buŋɛ*	*buŋa*	'flower'
*lim**a**ʔ	*limɔ*	*limə*	*limɛ*	*lima*	'five'
Closed syllables					
*an**a**k	*anɔʔ*	*anɔʔ*	*anɔʔ*	*anak*	'child'
*ləm**ə**k	*ləmɔʔ*	*ləmɔʔ*	*ləmɔʔ*	*ləmak*	'grease'
*kak**a**ʔ	*kakɔʔ*	*kakɔʔ*	*kakɔʔ*	*kakak*	'older sister'
*pəc**a**h	*pəcɔh*	*pəcɔh*	*pəcɔh*	*pəcah*	'to break'
*ak**a**r	*aka*	*akɔ*	*akɔ*	*akar*	'root'
*ul**ə**r	*ula*	*ulɔ*	*ulɔ*	*ular*	'snake'
*pul**a**w	*pula*	*pula*	*pulɔ*	*pulaw*	'island'
*bint**a**ŋ	*bitɛ*	*bitaŋ*	*bitɔŋ*	*bintaŋ*	'star'
*pət**ə**ŋ	*pətɛ*	*pətaŋ*	*pətɔŋ*	*pətaŋ*	'afternoon'
*hay**a**m	*ayɛ*	*ayaŋ*	*ayaŋ*	*ayam*	'chicken'
*hit**ə**m	*itɛ*	*itaŋ*	*itaŋ*	*hitam*	'black'
*jal**a**n	*jalɛ*	*jalaŋ*	*jalaŋ*	*jalan*	'road'
*simp**ə**n	*sipɛ*	*sipaŋ*	*sipaŋ*	*simpan*	'to save'
*say**a**p	*sayaʔ*	*sayaʔ*	*sayaʔ*	*sayap*	'wing'
*as**ə**p	*asaʔ*	*asaʔ*	*asaʔ*	*asap*	'smoke'
*sur**a**t	*suɣaʔ*	*suɣaʔ*	*suɣaʔ*	*surat*	'letter'
*bul**ə**t	*bulaʔ*	*bulaʔ*	*bulaʔ*	*bulat*	'round'
*bər**a**s	*bəɣah*	*bəɣah*	*bəɣah*	*bəras*	'uncooked rice'
*bal**ə**s	*balah*	*balah*	*balah*	*balas*	'to reply'
*ju**a**l	*juwa*	*juwa*	*juwa*	*jual*	'to sell'
*təb**ə**l	*təba*	*təba*	*təba*	*təbal*	'thick'
*lant**a**y	*lata*	*lata*	*lata*	*lantay*	'floor'

The mutation of *a in final open syllables is prevalent in Malayic and many other western Austronesian languages. It has been suggested as an areal feature which originated in the Indianised speech of Java and spread with the socio-cultural influences of Javanese courts (Tadmor 2003). However, as Blust (2017: 332–341) pointed out, languages showing final *a mutation apparently transcend the boundary of Javanese influence, and a borrowed areal feature is not a satisfactory explanation. In any case, the change itself lacks an explanatory phonetic motivation. In contrast, the development of ultimate ⁺a in closed syllables has a clear phonetic and phonological basis. Final consonants were grouped in different classes, and a number of features were at play. In KM, back obstruents were treated as one class, before which ⁺a was raised and backed. Nasals were grouped as a separate class, before which ⁺a was raised and fronted. In CTM, ⁺a > *ɔ* took place before back obstruents and the back liquid *-r. In Dusun, ⁺a became *ɔ* before all back consonants regardless of their manner of articulation, which means that only the place of articulation was a significant feature in grouping final consonants. It is evident, therefore, that the changes of ultimate *a (and *ə) in open syllables and in closed syllables are two independent phenomena.

In a few other instances, an earlier ⁺a is reflected as a nasal vowel *ã* or *ɔ̃*. Compare the following cognate sets between NEPMs and SM:

Table 7.27: Correspondences of NEPM *ã* or *ɔ̃* : SM *a*

KM	CTM	Dusun	SM	Gloss
puwãʔ	–	*suwɔʔ*	*suak*	'to part hair'
ttuwãʔ	*ttuwaʔ*	*ttuwaʔ*	*kətuat*	'warts'
dɔʔɔ̃	*dɔʔã*	*duʔã*	*dɔa* [*dɔʔa*]	'to pray'
–	*sɔʔã*	*suʔã*	*sɔal* [*sɔʔal*]	'to question'

No reconstruction is available for the first two cognate sets, but a comparison between KM and SM suggests that *ã* in KM *puwãʔ* 'to part hair' and *ttuwãʔ* 'warts' reflects an earlier ⁺a.[67] The other two sets of cognates are loanwords: KM *dɔʔɔ̃*, CTM *dɔʔã*, Dusun *duʔã* 'to pray' < Arabic *duʿāʾ*, and CTM *sɔʔã*, Dusun *suʔã* 'to question' < Arabic *suʾāl*. However, neither Arabic origins

[67] The correspondence of initial *p-* : *s-* in KM *puwãʔ* : SM *suak* is unexplained.

nor SM cognates have nasal vowels, indicating that *ã* or *ɔ̃* in these instances reflects the nasalisation of an original oral vowel.

Note that nasal vowels in NEPMs typically occur adjacent to a glottal stop. In the cases of Arabic loanwords, the original forms almost always have a voiced pharyngeal fricative /ʕ/ (transliterated as ⟨ʿ⟩) preceding the vowel that has been nasalised, except in *su'āl* 'to ask', in which a glottal stop (transliterated as ⟨'⟩) nasalised the following vowel. Such a connection between nasality and laryngeal articulation (including both glottal and pharyngeal) has been recognised as a more general phenomenon cross-linguistically, for which Matisoff (1975) coined the term *rhinoglottophilia*. The observation is that in many languages, oral vowels in laryngeal environments have nasalised allophones, or have been nasalised diachronically (see Ohala 1975 for explanations from an acoustic perspective). However, it remains unclear why vowel nasalisation only affected a small number of words in NEPMs, as most oral vowels occurring in a laryngeal environment are retained. Coupled with the changes of final consonants, this unconditioned vowel nasalisation eventually led to phonemic contrasts between nasal vowels and oral vowels. A tentative explanation may be that vowel nasalisation initially arose from the borrowing of Arabic loanwords with a laryngeal consonant (typically the pharyngeal fricative /ʕ/), during which process oral vowels adjacent to the laryngeals were perceived as nasalised.[68] This nasality further spread to native words so as to maintain meaning distinctions between pairs that would otherwise be homophones. The spread might have also been facilitated by by the presence of vowel nasality in some ideophones and interjections, such as KM *məcãʔ* 'sound of chewing', *səɣɔ̃ʔ* 'sound of sucking', *wãʔ*, an interjection when one smells something stinky, and NEPM *hɔ̃* 'AFF'. However, this explanation cannot account for all instances of vowel nasalisation, especially the nasalisation of some historical high vowels (see §7.4.4.3).

7.4.4 PM ultimate *i and *u

PM ultimate high vowels are often lowered in KM and CTM. In Dusun and ITM in general, they underwent diphthongisation in certain environments,

[68] Similar vowel nasalisation is sometimes described in the adaptation of Arabic loanwords in SM, as in *saat* [saʔãt] 'second' < Arabic *sāʿa* 'hour, time', and *taat* [taʔãt] 'loyal' < Arabic *ṭāʿa* (Yunus Maris 1980: 8–9).

conditioned by the presence of an onset and its nasality, as well as the presence of a coda and its place and manner of articulation. There are also some idiosyncratic cases where ultimate high vowels are nasalised, as will be discussed below.

7.4.4.1 Lowering in KM and CTM

In KM and CTM, ultimate high vowels are retained in open syllables, but in closed syllables, they have often been lowered to *e* and *o*, and occasionally to *ɛ* and *ɔ*. This lowering process was partially conditioned by the final consonant, but in certain environments, it appears to be irregular. Table 7.28 and Table 7.29 summarise the reflexes of ultimate *i and *u in KM and CTM.

Table 7.28: Reflexes of PM ultimate *i in KM and CTM

PM	KM	CTM	SM	Gloss
Retention in open syllables				
*kak**i**	*kak**i***	*kak**i***	*kak**i***	'foot; leg'
*ap**i**	*ap**i***	*ap**i***	*ap**i***	'fire'
Retention in closed syllables				
*nas**i**ʔ	*nas**i**ʔ*	*nas**i**ʔ*	*nas**i***	'cooked rice'
*kir**i**m	*kiɣ**i**ŋ*	*kiɣ**i**ŋ*	*kir**i**m*	'to send'
*aŋ**i**n	*aŋ**i**ŋ*	*aŋ**i**ŋ*	*aŋ**i**n*	'wind'
*cac**i**ŋ	*cac**i**ŋ*	*cac**i**ŋ*	*cac**i**ŋ*	'worm'
Regular lowering in closed syllables				
*a**i**r	*a**e***	*a**e***	*a**i**r*	'water'
*paŋg**i**l	*paŋg**e***	*paŋg**e***	*paŋg**i**l*	'to call'
*pil**i**h	*pil**e**h*	*pil**e**h*	*pil**i**h*	'to choose'
Irregular lowering in closed syllables				
*kul**i**t	*kul**e**ʔ*	*kul**e**ʔ*	*kul**i**t*	'skin'
*tas**i**k	*tas**e**ʔ*	*tas**e**ʔ*	*tas**i**k*	'lake'
*tul**i**s	*tul**e**h*	*tul**e**h*	*tul**i**s*	'to write'
*tum**i**t	*tum**i**ʔ*	*tum**i**ʔ*	*tum**i**t*	'heel'
*(b)is**i**k	*b|bis**i**ʔ*	*b|bis**i**ʔ*	*bis**i**k*	'to whisper'
*nip**i**s	*nip**i**h*	*nip**i**h*	*nip**i**s*	'thin'

Table 7.29: Reflexes of PM ultimate *u in KM and CTM

PM	KM	CTM	SM	Gloss
Retention in open syllables				
*bat**u**	*bat**u***	*bat**u***	*bat**u***	'stone'
*mal**u**	*mal**u***	*mal**u***	*mal**u***	'shamed'
Retention in closed syllables				
*kay**u**ʔ	*kay**u***	*kay**u***	*kay**u***	'wood'
Regular lowering in closed syllables				
*al**u**r	*al**o***	*al**o***	*al**u**r*	'groove'
*tump**u**l	*tup**o***	*tup**o***	*tump**u**l*	'dull'
*bun**u**h	*bun**o**h*	*bun**o**h*	*bun**u**h*	'to kill'
*jar**u**m	*jaɣ**o**ŋ*	*jaɣ**o**ŋ*	*jar**u**m*	'needle'
*da**u**n	*da**o**ŋ*	*da**o**ŋ*	*da**u**n*	'leaf'
*bur**u**ŋ	*buɣ**o**ŋ*	*buɣ**o**ŋ*	*bur**u**ŋ*	'bird'
Sporadic lowering in closed syllables				
*hid**u**p	*id**o**ʔ*	*id**o**ʔ*	*hid**u**p*	'to live'
*mul**u**t	*mul**o**ʔ*	*mul**o**ʔ*	*mul**u**t*	'mouth'
*dud**u**k	*dud**o**ʔ*	*dud**o**ʔ*	*dud**u**k*	'to sit'
*rat**u**s	*ɣat**o**h*	*ɣat**o**h*	*rat**u**s*	'hundred'
*t|i**u**p	*tiy**u**ʔ*	*tiy**u**ʔ*	*ti**u**p*	'to blow'
*rump**u**t	*ɣup**u**ʔ*	*ɣup**u**ʔ*	*rump**u**t*	'grass'
*gəm**u**k	*gəm**u**ʔ*	*gəm**u**ʔ*	*gəm**u**k*	'fat'
*tik**u**s	*tik**u**h*	*tik**u**h*	*tik**u**s*	'rat'

Ultimate *i is regularly retained in open syllables and closed syllables with a final glottal stop or a final nasal. It is regularly lowered to *e* before liquids and *-h. No example reflecting an earlier *-ip is found, but this is explained by a general tendency of avoiding final labials following *i in PM (Adelaar 1992: 107–108). Before other final segments, namely *t, *k and *s, *i is also often lowered to *e*, but there are irregular outcomes. Ultimate *u retains its height in open syllables and closed syllables with *-ʔ. It has been lowered to *o* before liquids, *-h and nasals. Before *-p, *-t, *-k and *-s, the lowering of *u to *o* was not clearly conditioned.

Given the irregular nature of the lowering of high vowels before certain final segments (especially stops), *i* and *u* are rarely in phonemic contrast with *e* and *o* in final closed syllables. So far only one minimal pair contrasting a high vowel and a mid-high vowel has been found in this particular position, namely *nasiʔ* 'cooked rice' vs. *naseʔ* 'destiny' (< Arabic *naṣīb*). On the other hand, contrasts in final open syllables are common, resulting from the regular retention of high vowels in earlier open syllables and the regular lowering before liquids followed by the loss of final liquids, e.g., *alu* 'pestle' vs. *alo* 'groove' (cf. SM *alu* vs. *alur*).

Some PM words had high vowels in both syllables, and the penultimate high vowel has been irregularly lowered to mid-low (§7.4.2). In these cases, the ultimate high vowel has often been lowered to mid-low accordingly in harmony, as shown in the following examples: *ik**u**r 'tail' > KM/CTM *ɛkɔ*, *cuc**u**k 'to prick' > KM/CTM *cɔcɔʔ*. Vowel harmony is not realised if PM had a final nasal, e.g., *puhun 'tree' > KM/CTM *pɔhoŋ*, also KM *pɛkoŋ* 'to throw' and *gɔɣeŋ* 'to fry'. In one unexplained instance *təlur 'egg' > KM/CTM *təlɔ*, ultimate *u is lowered to *ɔ*, despite the penultimate vowel being a schwa.

7.4.4.2 Diphthongisation in ITM

High vowels in final syllables underwent a divergent history in ITM. As previously reported by Collins in a number of publications (Collins & Naseh Hassan 1981; Collins 1983a,b), earlier ultimate high vowels are often diphthongised in ITM, and subvarieties of ITM spoken in different villages manifest different outcomes of diphthongisation. Importantly, not all ultimate high vowels underwent diphthongisation, as this change only took place in certain environments, conditioned by both the onset and the coda of original final syllables.

Based on the reflexes in Dusun, the environments in which PM ultimate high vowels occurred can be categorised as follows: the onset could be oral, *Ø or *h- (labelled as "vocalic" since *h is regularly lost), or nasal; and the coda could be *Ø, (*-ʔ), *-k, *-ŋ or *-h, or another consonant. A schematic representation of Dusun reflexes of PM ultimate *i and *u is provided in Table 7.30. It shows that a PM ultimate high vowel could occur in six different environments, but diphthongisation only took place when both of the following two conditions are met: 1) the onset was an oral consonant, and 2) the coda was *Ø or one of the back consonants *-k, (*-ʔ), *-ŋ and *-h. In all other environments, ultimate *i and *u are retained as high monophthongs.

Table 7.30: Reflexes of ultimate *i and *u in Dusun

onset \ coda	*Ø, (*-ʔ), *-k, *-ŋ, *-h	others
oral	*ɛi, əʊ*	*i, u*
vocalic	*i, u*	*i, u*
nasal	*i, u*	*i, u*

In what follows, I present examples illustrating the reflexes of ultimate *i and *u in various environments.

First of all, the diphthongisation of ultimate *i > *ɛi* and *u > *əʊ* in Dusun are exemplified by the examples in (12). The changes represented here are phonemic changes; the phonemes /ɛi/ and /əʊ/ show complex allophonic variation at the synchronic level, see §4.2.2.3.4. In one unexplained example *nas**i**ʔ 'cooked rice' > *nasiʔ*, the ultimate *i was exempted from diphthongisation.

(12)	*hat**i**	>	*at**ɛi***	'liver'
	*bəl**i**	>	*bəl**ɛi***	'to buy'
	*dur**i**ʔ	>	*duɣ**ɛi***	'thorn'
	*tas**i**k	>	*tas**ɛi**ʔ*	'lake'
	*put**i**h	>	*put**ɛi**h*	'white'
	*cac**i**ŋ	>	*cac**ɛi**ŋ*	'worm'
	*bat**u**	>	*bat**əʊ***	'stone'
	*kuk**u**	>	*kuk**əʊ***	'nail'
	*dag**u**ʔ	>	*dag**əʊ***	'chin'
	*dud**u**k	>	*dud**əʊ**ʔ*	'to sit; to stay'
	*tuj**u**h	>	*tuj**əʊ**h*	'seven'
	*hid**u**ŋ	>	*id**əʊ**ŋ*	'nose'

In contrast, when the criterion for onset was met but the criterion for coda was not (i.e., the coda was one of *-p, *-t, *-m, *-n, *-s, *-r or *-l), *i and *u are retained as high vowels, as in (13).

(13)	*kul**i**t	>	*kul**i**ʔ*	'skin'
	*kir**i**m	>	*kiɣ**i**ŋ*	'to send'
	*ciɲc**i**n	>	*cic**i**ŋ*	'ring'
	*tul**i**s	>	*tul**i**h*	'to write'
	*pas**i**r	>	*pas**i***	'sand'
	*paŋg**i**l	>	*paŋg**i***	'to call'
	*hid**u**p	>	*id**u**ʔ*	'to live'
	*mul**u**t	>	*mul**u**ʔ*	'mouth'
	*jar**u**m	>	*jaɣ**u**ŋ*	'needle'
	*tur**u**n	>	*tuɣ**u**ŋ*	'to go down'
	*rat**u**s	>	*ɣat**u**h*	'hundred'
	*ik**u**r	>	*ik**u***	'tail'
	*tump**u**l	>	*tup**u***	'dull'

A comparison between (12) and (13) shows that the diphthongisation of ultimate high vowels was sensitive to the place feature of final consonants, a pattern similar to the raising of ultimate *a and *ə in Dusun. The difference is that following *a and *ə, all final consonants including liquids and glides were grouped into two sets, namely front and back, whereas following *i and *u, final liquids were treated as an independent class. Even before a [+back] liquid *r, a high vowel was not diphthongised. The final consonants which triggered diphthongisation can be characterised by a not-so-elegant term "back non-approximants". It should also be pointed out that the diphthongisation of ultimate high vowels is attested before *historical* back non-approximants, which means diphthongisation must have taken place prior to the merger of final stops and final nasals, as well as the lenition of *-s > *-h*.

Next, example (14) shows that when the criterion for coda was met but the onset was either not present or a glottal fricative *h, no diphthongisation took place. *h in the onset position in fact played no role as it is regularly lost, effectively creating an environment akin to *Ø. Additionally, no examples of PM high vowels followed by *-ŋ and preceded by *h- or *Ø were found.

(14)	*dahi	>	*dai*	'forehead'
	*baik	>	*baiʔ*	'good'
	*tahu(ʔ)	>	*tau*	'to know'
	*lauk	>	*lauʔ*	'dish'
	*jauh	>	*jauh*	'far'

Note that all examples in (14) had an earlier *a in the penultimate syllable. In comparison, when the original penultimate vowel was a high vowel, diphthongisation of the ultimate high vowel is still observed. This can be explained by the reinterpretation of an original epenthetic glide between two high vowels as phonemic, effectively changing the onset environment from vocalic to oral, thereby triggering the diphthongisation process. As discussed in §7.3.5.1, Dusun has diphthongs in *buwɛi* 'to give', *buwɛih* 'foam' and *iyəʊ* 'shark', and the following phonological histories are suggested:

1) *bəriʔ > ⁺bəri > ⁺bui > ⁺buwi > *buwɛi* 'to give'
2) PMP *buqiq > PM *buhih > ⁺buih > ⁺buwih > *buwɛih* 'foam'
3) *hiuʔ > ⁺iu > ⁺iyu > *iyəʊ* 'shark'

Lastly, diphthongisation is not attested when the onset was a nasal consonant, as shown in (15). The coda environment was irrelevant to the retention of *i and *u following nasal onsets. Even in words that had final *Ø, (*ʔ), *k, *h or *ŋ, high vowels are still retained as monophthongs.

(15)	*bini	>	*biniŋ*	'wife'
	⁺manik[69]	>	*man**i**ʔ*	'bead'
	*bən**i**h	>	*bən**i**h*	'seed'
	⁺kun**i**ŋ	>	*kun**i**ŋ*	'yellow'
	*laŋ**i**t	>	*laŋ**i**ʔ*	'sky'
	*aŋ**i**n	>	*aŋ**i**ŋ*	'wind'
	*man**i**s	>	*man**i**h*	'sweet'
	*pəɲ**u**ʔ	>	*pəɲ**u**ŋ*	'turtle'
	*ɲam**u**k	>	*ɲam**u**ʔ*	'mosquito'
	*bun**u**h	>	*bun**u**h*	'to kill'
	⁺tən**u**ŋ	>	*tən**u**ŋ*	'to stare'
	*m|in**u**m	>	*min**u**ŋ*	'to drink'
	*tən**u**n	>	*tən**u**ŋ*	'to weave'

Exceptions are found in a few function words such as *(i)ni(ʔ) > *nɛiŋ* 'DEM.DIST' and *kamu > *məʊŋ* '2SG', where ultimate high vowels have been diphthongised following nasal onsets. Yet overall, nasal onsets blocked the

[69] ⁺manik 'bead' is not reconstructable in PM, but a loanword from Tamil *maṇi* (Jones 2007: 193) (*-k* unexplained). In this case, a SM cognate is taken as reflecting an earlier stage of the Dusun form. The same applies for ⁺kuniŋ 'yellow' and ⁺tənuŋ 'to stare' below.

diphthongisation of following high vowels. While it is known that the nasality of consonants may not only nasalise adjacent vowels but can also affect vowel height, which is often manifested as a centralisation effect (Beddor 1983; Beddor et al. 1986; Arai 2004), onset nasality as a conditioning factor for the diphthongisation of following vowels is highly unusual.[70]

To further complicate the matter, diphthongisation is also observed as a synchronic phenomenon in Dusun. As described in §4.2.2.3.3, phonemic high vowels in final closed syllables with an onset have a tendency to be diphthongised, for instance *kuliʔ* 'skin' → [kuliʔ]~[kulɪi̯ʔ] and *biniŋ* 'wife' → [binĩŋ]~[binɛ̃ĩ̯ŋ]. The tendency to diphthongise does not apply to ultimate high vowels in open syllables or those preceded by Ø: ultimate *i* and *u* in *pasi* 'sand', *tupu* 'dull' and *jauh* 'far' are consistently pronounced as monophthongs. This suggests that diphthongisation continues to operate and affect more high vowels, particularly in closed syllables with an onset (both oral and nasal). A two-stage diphthongisation can be posited, as has also been suggested by Collins (1983b). It can be observed more clearly in other subvarieties of ITM, notably in ITM spoken in the village of Payang Kayu (henceforth PK). Data from PK present a more complicated and exceptional case on how onsets could affect the development of following vowels, as will be discussed below.

As in Dusun, the development of ultimate high vowels in PK is conditioned by both the onset and the coda of original final syllables. A more detailed distinction is evident in the coda environment, as shown by the reflexes of ultimate *i and *u summarised in Table 7.31.[71]

[70] In Collins' various publications on ITM diphthongisation, the coda condition was rightfully pointed out, but the onset condition was overlooked.

[71] Data from Payang Kayu are given in broad phonetic transcriptions, as no systematic phonemic analysis has been conducted.

Table 7.31: Reflexes of ultimate *i and *u in Payang Kayu

coda / onset	*Ø, (*-ʔ), *-ŋ	*-k, *-h	others
oral	*aɪ, aʊ*	*a*	*eɪ, oʊ*
vocalic	Ø	Ø	*i, u*
nasal	*eɪ, oʊ*	*eɪ, oʊ*	*eɪ, oʊ*

Following oral onsets, *i and *u before *Ø, (*-ʔ) and *-ŋ are diphthongised to *aɪ* and *aʊ* respectively in PK. However, before historical *-k and *-h, both *i and *u are reflected as a plain low vowel *a*, which is apparently the result of diphthongisation followed by monophthongisation through offglide deletion, i.e., *i > +aɪ > *a* and *u > +aʊ > *a*. Examples illustrating this pattern are given in (16) and (17).

(16)	*kak**i**	>	*kak**aɪ***	'foot; leg'
	*dag**i**ŋ	>	*dag**aɪ**ŋ*	'meat'
	*bat**u**	>	*bat**aʊ***	'stone'
	*kay**u**ʔ	>	*kay**aʊ***	'wood'
	*hid**u**ŋ	>	*id**aʊ**ŋ*	'nose'

(17)	*tas**i**k	>	*tas**a**ʔ*	'lake'
	*put**i**h	>	*put**a**h*	'white'
	*dud**u**k	>	*dud**a**ʔ*	'to sit'
	*tuj**u**h	>	*tuj**a**h*	'seven'

In the same oral onset environment, ultimate high vowels followed by other codas are also diphthongised, but the outcomes differ. Specifically, *i and *u are diphthongised to *eɪ* and *oʊ* respectively, as shown in (18).

(18)	*kul**i**t	>	*kul**eɪ**ʔ*	'skin'
	+lic**i**n	>	*lic**eɪ**ŋ*	'smooth'
	*pas**i**r	>	*pas**eɪ***	'sand'
	*hid**u**p	>	*id**oʊ**ʔ*	'to live'
	*mul**u**t	>	*mul**oʊ**ʔ*	'mouth'
	*jar**u**m	>	*jaɣ**oʊ**ŋ*	'needle'
	*kab**u**s	>	*kab**oʊ**h*	'fog'
	*ik**u**r	>	*ik**oʊ***	'tail'
	*tump**u**l	>	*tup**oʊ***	'dull'

When the final-syllable onsets were vocalic (*Ø or *h), *i and *u before *Ø or back non-approximants were deleted, resulting in the reduction of original disyllables to monosyllables, as illustrated by examples in (19). These words originally had VV sequences *-a.i- or *-a.u- (also $^{+}$-a.u- < *-ahu-), to which a similar offglide deletion applied, i.e., $^{+}$ai, $^{+}$au > *a*.[72]

(19)	*baik	>	*baʔ*	'good'
	$^{+}$bahu	>	*ba*	'shoulder'
	*tahu(ʔ)	>	*ta*	'to know'
	*jauh	>	*jah*	'far'

High vowels between vocalic onsets and other codas are generally retained as monophthongs, as shown in (20). There is an unexplained exception *tahun 'year' > *taŋ*, where the high vowel *u was deleted before a nasal coda.

(20)	*jah**i**t	>	*ja**i**ʔ*	'to sew'
	*ma**i**n	>	*ma**i**ŋ*	'to play'
	*a**i**r	>	*a**i***	'water'
	*la**u**t	>	*la**u**ʔ*	'sea'
	*da**u**n	>	*da**u**ŋ*	'leaf'

Exceptions also applied when the vocalic onsets were preceded by another high vowel in the penultimate syllable, in which cases the glide between two high vowels was interpreted as an oral onset, triggering diphthongisation of the ultimate high vowel:

(21)	*ti**u**p	>	*tiy**oʊ**ʔ*	'to blow'
	*ci**u**m	>	*ciy**oʊ**ŋ*	'to sniff'

Finally, high vowels preceded by nasal onsets are diphthongised to *ɛɪ* and *əʊ* (phonetically nasalised) regardless of coda, as demonstrated in (22).

[72] Due to the limited data, it is unclear how ultimate high vowels would change in earlier high-high vowel sequences, such as *-i.u- and *-u.i-.

(22)				
	*bin**i**	>	*bin**eɪ**ŋ*	‘wife’
	*bən**i**h	>	*bən**eɪ**h*	‘seed’
	$^{+}$kun**i**ŋ	>	*kun**eɪ**ŋ*	‘yellow’
	*laŋ**i**t	>	*laŋ**eɪ**ʔ*	‘sky’
	*aŋ**i**n	>	*aŋ**eɪ**ŋ*	‘wind’
	*man**i**s	>	*man**eɪ**h*	‘sweet’
	*ɲam**u**k	>	*ɲam**oʊ**ʔ*	‘mosquito’
	*bun**u**h	>	*bun**oʊ**h*	‘to kill’
	*tən**u**n	>	*tən**oʊ**ŋ*	‘to weave’
	*m|in**u**m	>	*min**oʊ**ŋ*	‘to drink’

The examples discussed above reveal that PM ultimate high vowels have six different sets of reflexes in PK, each determined by the shape of original final syllables. These complex outcomes can be interpreted as the result of an ordered set of changes affecting various ultimate vowels in different stages. The proposed history of ultimate *i and *u from PM to PK can be outlined as follows.

In the first stage, as in Dusun, diphthongisation in PK affected high vowels between oral onsets and *Ø or back non-approximant codas (*-k, (*-ʔ), *-ŋ, *-h). The results of diphthongisation were $^{+}$aɪ and $^{+}$aʊ, and high vowels remained unchanged in other environments, as illustrated in Table 7.32.

Table 7.32: Ultimate *i and *u in Payang Kayu after initial diphthongisation

coda / onset	*Ø, (*-ʔ), *-ŋ	*-k, *-h	others
oral	$^{+}$aɪ, $^{+}$aʊ	$^{+}$aɪ, $^{+}$aʊ	$^{+}$i, $^{+}$u
vocalic	$^{+}$i, $^{+}$u	$^{+}$i, $^{+}$u	$^{+}$i, $^{+}$u
nasal	$^{+}$i, $^{+}$u	$^{+}$i, $^{+}$u	$^{+}$i, $^{+}$u

The two diphthongs $^{+}$aɪ and $^{+}$aʊ subsequently underwent monophthongisation to *a* with offglide deletion, but only $^{+}$aɪ and $^{+}$aʊ before *-k and *-h were affected (see example 16). Simultaneously, VV sequences *-a.i- and *-a.u- before *Ø or back non-approximants were affected by similar changes, resulting in the deletion of ultimate high vowels (see example 19). At this point, the reflexes of ultimate *i and *u in PK should resemble what is presented in Table 7.33. *i and *u following nasal onsets and/or preceding other codas were still retained as monophthongs.

Table 7.33: Ultimate *i and *u in Payang Kayu after initial diphthongisation and subsequent monophthongisation

onset \ coda	*Ø, (*-ʔ), *-ŋ	*-k, *-h	others
oral	*aɪ, aʊ*	*a*	⁺i, ⁺u
vocalic	*Ø*	*Ø*	⁺i, ⁺u
nasal	⁺i, ⁺u	⁺i, ⁺u	⁺i, ⁺u

Presumably after the initial diphthongisation and subsequent monophthongisation, changes took place in final segments: stops were merged to *-ʔ*, nasals were merged to *-ŋ*, *-s was merged with *-h, and liquids were lost. As a consequence, high vowels which originally preceded non-back consonants or liquids were then followed by one of the back consonants (*-ʔ*, *-ŋ* and *-h*) or *Ø* – a coda environment in which the second phase of diphthongisation was triggered, e.g., *kulit 'skin' > ⁺kuliʔ > *kuleɪʔ*, and *ikur 'tail' > ⁺iku > *ikoʊ*. Diphthongisation in this phase operated in a slightly different onset environment: not only high vowels following oral onsets were diphthongised, those following nasal onsets were also affected by the same change, e.g., *laŋit 'sky' > ⁺laŋiʔ > *laŋeɪʔ*. High vowels following vocalic onsets were still exempted from diphthongisation, e.g., *laut 'sea' > *lauʔ*. Eventually, the second phase of diphthongisation culminated in what is seen in present-day PK:

Table 7.34: Reflexes of ultimate *i and *u in Payang Kayu (repeating Table 7.31)

onset \ coda	*Ø, (*-ʔ), *-ŋ	*-k, *-h	others
oral	*aɪ, aʊ*	*a*	*eɪ, oʊ*
vocalic	*Ø*	*Ø*	*i, u*
nasal	*eɪ, oʊ*	*eɪ, oʊ*	*eɪ, oʊ*

Data from PK illustrate a clearer two-stage diphthongisation process and the complex conditioning factors in the evolution of ultimate high vowels. Several crucial points should be reiterated. First, in both phases of diphthongisation, only high vowels before *Ø or a back non-approximant were affected. Second, high vowels following *Ø or *h were not affected by diphthongisation, except when the penultimate vowel was also high. Third, the nasality of onset was a conditioning factor for the diphthongisation of following vowels in the first phase, but not in the second phase.

Comparing the historical development of ultimate high vowels in Dusun with that in PK, it is evident that both varieties underwent a similar phase of initial diphthongisation with the same conditioning factors. Additionally, both varieties share a common history as regards the changes of final consonants. However, Dusun was only affected by the first phase of diphthongisation (with an incipient second phase at the synchronic level), whereas PK was further affected by monophthongisation of some diphthongs, as well as the second phase of diphthongisation, which differentiated the two subvarieties of ITM in a remarkable way.

7.4.4.3 Nasalisation

Similar to the nasalisation of some earlier low vowels, ultimate high vowels were also sporadically affected by vowel nasalisation, as shown by the examples in Table 7.35. One more cognate set might be added: KM/CTM *ɛsɔ̃ʔ* and Dusun *kisɔ̃ʔ* 'to scoot over' vs. SM *kɛsɔt*, for which no PM reconstruction is available, but NEPM *ɔ̃* and SM *ɔ* likely reflect an earlier *u.

Table 7.35: Nasalisation of PM ultimate *i and *u

PM	KM	CTM	Dusun	SM	Gloss
*kəc**i**l/*kəc**i**k	*kəc**ɛ̃ʔ***	*kəc**ĩʔ***	*kəc**ĩʔ***	*kəc**il***	'small'
PMP *kaw**i**l	*ka**ɛ̃***	*ka**ĩ***	*ka**ĩ***	*ka**il***	'fishhook'
*bus**u**k	*bus**ũʔ***	*bus**ũʔ***	*bus**ũʔ***	*bus**uk***	'rotten'

In §7.4.3, I proposed some general explanations for vowel nasalisation. As far as high vowels are concerned, vowel nasalisation also mostly occurred in laryngeal environments, a phenomenon reminiscent of rhinoglottophilia.

However, this explanation falls short in explaining the nasal vowels *ɛ̃* and *ĩ* in KM *kaɛ̃* and CTM/Dusun *kaĩ* 'fishhook', in which cases the nasal vowels were/are not historically or synchronically adjacent to a laryngeal consonant.

It is also worth noting that reflexes of ultimate high vowels in the examples above do not only deviate from the expected results respecting nasality, but also vowel height and quality. As discussed in §7.4.4.1, when high vowels are lowered in KM, the outcomes are mid-high vowels unless the penultimate vowel is also lowered. In *kəcɛ̃ʔ* 'small' and *kaɛ̃* 'fishhook', however, an earlier high vowel *i is nasalised and lowered to a mid-low vowel *ɛ̃*. In Dusun, ultimate *u in *busuk 'rotten' would have been diphthongised preceding *-k, yet no diphthongisation is seen in this instance.

7.5 Syllable reduction

In addition to changes at the segment level, syllable reduction represents another drastic change in the phonological history of NEPMs. This process involved the reduction of both PM disyllables and trisyllables. PM disyllables of certain shapes have been reduced to monosyllables due to the working of initial schwa deletion, as will be discussed in §7.5.1. PM trisyllables, on the other hand, have been reduced to disyllables as a rule. The processes affecting the reduction of PM trisyllables are elaborated in §7.5.2.

7.5.1 Syllable reduction in PM disyllables

As mentioned in §7.4.1, initial *ə or *hə was deleted in NEPMs, resulting in the reduction of original disyllables to monosyllables. This reduction is evident in the following examples:

Table 7.36: Syllable reduction in PM disyllables

PM	KM	CTM	Dusun	SM	Gloss
*ənəm	*nnɛ*	*nnaŋ*	*nnaŋ*	*ənam*	'six'
*əmas	*mmah*	*mmah*	*mmah*	*əmas*	'gold'
*əmpat	*ppaʔ*	*ppaʔ*	*ppaʔ*	*əmpat*	'four'
*əmbun	*mboŋ*	*mboŋ*	*mbuŋ*	*əmbun*	'dew'
*həmbus	*mbuh*	*mbuh*	*mbuh*	*həmbus*	'to blow'
...	*llɛ*	*llaŋ*	*llɔŋ*	*həlaŋ*	'eagle'
...	*ŋgɛ*	–	–	*əŋgaŋ*	'hornbill'

As can be seen, the reduced monosyllabic forms in Table 7.36 always have initial CC clusters, and the original final-syllable onset is often reflected as a geminate cluster. Take *əmas > *mmah* 'gold' as an example: the word-medial *m was presumably geminated following the penultimate *ə, which is a common developmental path in Austronesian languages (Blust 1995: 127). The initial schwa was then deleted, hence generating *mmah*. As for *əmpat > *ppaʔ* 'four', the intermediate stages presumably involved the reduction of the consonant sequence *-mp- to ⁺p, the gemination of ⁺p to *pp*, and the loss of initial schwa, i.e., *əmpat > ⁺əpat > ⁺əppat > *ppaʔ*. In comparison, the sequence *-mb- in *həmbus 'to blow' is retained, and the reflexes have an initial non-geminate cluster *mb-*.

7.5.2 Syllable reduction in PM trisyllables

A more common type of syllable reduction is found in PM trisyllables. Trisyllabic simple words, prefixed derivatives, as well as early trisyllabic loanwords from Sanskrit, Arabic and Portuguese, have been reduced to disyllables, as shown by the examples in Table 7.37. The reduction of trisyllables as such was triggered by either vowel contraction (§7.5.2.1) or the syncope of antepenultimate vowels (§7.5.2.2). The latter process was followed by cluster assimilation (§7.5.2.3) under certain conditions, which formed word-initial geminates. Notably, the same sets of changes affected both earlier simple words and prefixed derivatives indiscriminately, eventually leading to the restructuring of the morphological systems of NEPMs (see more discussions in Chapter 8).

Table 7.37: Syllable reduction in earlier trisyllables

PM	KM	CTM	Dusun	SM	Gloss
*buhaya	*bɔyɔ*	*bɔyə*	*bwayɛ*	*buaya*	'crocodile'
*biawak	*bɛwɔʔ*	*bɛwɔʔ*	–	*biawak*	'monitor lizard'
*bAlakaŋ	*blakɛ*	*blakaŋ*	*blakɔŋ*	*bəlakaŋ*	'back'
*tiŋgələm	*tgəlɛ*	*tgəlaŋ* ~*ggəlaŋ*	*tŋəlaŋ*	*təŋgəlam*	'to sink'
*hArimaw	*ɣima*	*ɣima*	*ɣimɔ*	*harimaw*	'tiger'
*(mb)Ar-jalan	*j-jalɛ*	*j-jalaŋ*	*b-jalaŋ* ~*j-jalaŋ*	*bər-jalan*	'to walk'
*mAN-alir	*ŋŋ|ale*	*ŋŋ|ale*	*ŋŋ|ali*	*məŋ-alir*	'to flow'
–	*ppalɔ*	*ppalə*	*ppalɛ*	*kəpala*[a]	'head'
–	*xxusi*	*xxusi*	*xusɛi*	*kərusi*[b]	'chair'
–	*tbaka*	*tbaka*	*tmakɔ* ~*mmakɔ*	*təmbakau*[c]	'tobacco'

[a] From Sanskrit *kapāla*.
[b] From Arabic *kursī*.
[c] From Portuguese *tabaco*.

7.5.2.1 Vowel contraction

Vowel contraction affected sequences of vowels across the original antepenultimate and penultimate syllables. In trisyllables with a $*C_1V_1.(C)_2V_2.(C)V(C)$ shape in which the penultimate onset $*C_2$ was *Ø or *h, adjacent $*V_1$ and $*V_2$ were often contracted, as illustrated in Table 7.38. In KM and CTM, adjacent *i and *a were contracted to *ɛ*, while *u and *a were contracted to *ɔ*. In Dusun, only earlier *-a.i- was affected, which was contracted to *i*. Original antepenultimate high vowels in sequences like *-i(h)a- and *-u(h)a- have been reinterpreted as glides, forming part of initial C + glide clusters, e.g., *buhaya > *bwayɛ* 'crocodile' and ⁺biasa > *byasɛ* 'usual', which presumably blocked the vowel contraction.

Table 7.38: Vowel contraction in earlier trisyllables

PM	KM	CTM	Dusun	SM	Gloss
*b**ia**wak	*bɛwɔʔ*	*bɛwɔʔ*	–	*b**ia**wak*	‘monitor lizard’
–	*bɛsɔ*	*bɛsə*	*byasɛ*	*b**ia**sa*[a]	‘usual’
*m**a**-**i**rah	*mɛɣɔh*	*mɛɣɔh*	*miɣɔh*	*mɛrah*	‘red’
*b**ai**k-i[b]	*bɛʔk|i*	*bɛʔk|i*	*b**i**ʔk|ɛi*	*b**ai**k-i*	‘to repair’
*b**uha**ya	*bɔyɔ*	*bɔyə*	*bwayɛ*	*b**ua**ya*	‘crocodile’
–	*sɔɣɔ*	*sɔɣə*	*swaɣɛ*	*s**ua**ra*[c]	‘voice’
...	*kɔli*	*kuali*	*kwalɛi*	*k**ua**li*	‘wok’
–	*pɔsɔ*	*pɔsə*	*pwasɛ*	*p**ua**sa*[d]	‘to fast’

[a] From Sanskrit *abhyāsa* ‘habit’.
[b] PM *baik-i ‘good-APPL’, but fossilised in NEPMs (§5.3.5).
[c] From Sanskrit *svara*.
[d] From Sanskrit *upavāsa*.

7.5.2.2 Antepenultimate vowel syncope

In historical trisyllables with other shapes, antepenultimate vowels have been lost. The result of *V_1 syncope in *$C_1V_1.C_2V(C).(C)V(C)$ is typically a disyllable with an initial C_1C_2 cluster or a geminate C_2C_2 cluster. In words with an initial *h-, the antepenultimate syllable *C_1V_1 was lost altogether. Examples illustrating antepenultimate vowel syncope in earlier trisyllables are presented in Table 7.39. Given that antepenultimate vowels have often been neutralised to schwa in many Malayic varieties (Adelaar 1992: 49–50), antepenultimate vowel syncope in NEPMs was presumably also preceded by schwa neutralisation.

If *C_1 and *C_2 in *$C_1V_1.C_2V(C).(C)V(C)$ happened to be historically identical, a geminate cluster was formed after antepenultimate vowel syncope, e.g., *mAN-masak (məmasak) ‘ACT-cook’ > NEPM *mmasɔʔ* ‘to cook’. However, most geminate clusters C_2C_2- can be traced back to dissimilar *C_1 and *C_2 in earlier forms, which subsequently underwent cluster assimilation, as will be discussed in §7.5.2.3.

Table 7.39: Antepenultimate vowel syncope in earlier trisyllables

PM	KM	CTM	Dusun	SM	Gloss
*bAlakaŋ	*blakɛ*	*blakaŋ*	*blakɔŋ*	*bəlakaŋ*	‘back’
*kAluaŋ	*kluwɛ*	*kluwaŋ*	*kluwɔŋ*	*kəluaŋ*	‘flying fox’
–	*skɔlɔh*	*skɔlɔh*	*skulɔh*	*səkɔlah*[a]	‘school’
*tAliŋa(ʔ)	*tliŋɔ~lliŋɔ*	*lliŋə*	*tliŋɛ*	*təliŋa*	‘ear’
–	*ppalɔ*	*ppalə*	*ppalɛ*	*kəpala*	‘head’
*hArimaw	*ɣima*	*ɣima*	*ɣimɔ*	*harimaw*	‘tiger’
*mAN-ajar[b]	*ŋŋ|aja*	*ŋŋ|ajɔ*	*ŋŋ|ajɔ*	*məŋ-ajar*	‘to study’

[a] From Portuguese *escola*.
[b] PM *mAN-ajar ‘ACT-teach’, also inherited in SM *məŋ-ajar*, but fossilised in NEPMs, see §8.3.2.

Some trisyllabic words had consonant sequences spanning the antepenultimate and the penultimate syllables, i.e., they had a *$C_1V_1C_2.C_3V(C).(C)V(C)$ shape. In these trisyllables, not only was the antepenultimate vowel *V_1 deleted, but the sequences of *-$C_2.C_3$- were also reduced. The result of this reduction is commonly a disyllable with a C_1C_2- or a C_1C_3- cluster, but occasionally a C_3C_3- cluster is attested in the outcome, as shown in Table 7.40.

Table 7.40: Antepenultimate vowel syncope and consonant sequence reduction in earlier trisyllables

PM	KM	CTM	Dusun	SM	Gloss
*tiŋgələm	*tgəlɛ*	*tgəlaŋ ~ggəlaŋ*	*tŋəlaŋ*	*təŋgəlam*	‘to sink’
–	*tbaka*	*tbaka*	*tmakɔ ~mmakɔ*	*təmbakau*	‘tobacco’
*hAmpədu	*ppədu*	*ppədu*	*ppədəʊ*	*həmpədu*	‘gall bladder’
*pAr-habis	*pɣ-abih*	*pɣ-abih*	–	–	‘to finish’
*(mb)Ar-hənti[a]	*bɣ|əti*	*bɣ|əti*	*bɣ|ətɛi*	*bər-hənti*	‘to stop’
*tAr-bakar	*t-baka*	*t-bakɔ*	*t-bakɔ*	*tər-bakar*	‘to be burnt’
*(mb)Ar-lari	*b-laɣi*	*b-laɣi*	*b-laɣɛi*	*bər-lari*	‘to run’

[a] The uncertain reconstruction *(mb) is regularly reflected as *b* in NEPMs.

Some important generalisations can be made about the reduction of *-C_2.C_3- sequences. First, when *-C_2.C_3- were sequences of a nasal + a homorganic voiced stop, KM and CTM generally deleted the nasal *C_2, and the reduced disyllables have C_1C_3- clusters. On the contrary, Dusun deleted the stop *C_3, leaving C_1C_2- clusters.[73] This can be seen in the development of *tiŋgələm 'to sink' to KM *tgəlɛ* and CTM *tgəlaŋ*, but to Dusun *tŋəlaŋ*.

Second, when *-C_2.C_3- consisted of a nasal and a homorganic voiceless stop, the nasal was lost in all three varieties. There is only one such example *hAmpədu 'gall bladder', which had an initial *h that was also deleted. The loss of the nasal component in this particular case results in a geminated voiceless stop, i.e., KM/CTM ***pp**ədu* and ITM ***pp**ədəʊ*. It can be inferred that subsequent to the reduction of *-mp- > ⁺p, ⁺p was geminated, presumably following a schwa (parallel to the development of *əmpat > *ppaʔ* 'four'), i.e., *hAmpədu > ⁺həpədu > ⁺(h)əppədu > KM/CTM *ppədu*, ITM *ppədəʊ*. This path also suggests that the reduction of *-C_2.C_3- must have preceded antepenultimate vowel syncope; otherwise, no gemination following schwa would have taken place.

Third, when *C_3 in *-C_2.C_3- was an *h, the regular loss of *h led to the formation of C_1C_2- clusters in all three varieties, e.g., *(mb)Ar-hənti > KM/CTM *bɣəti*, Dusun *bɣətɛi* 'to stop'.

Lastly, when *C_2 was an *r and *C_3 was a consonant other than *h, *r was deleted and a cluster of C_1C_3- was formed, e.g., *(mb)Ar-lari > KM/CTM *blaɣi*, Dusun *blaɣɛi* (MID-run) 'to run'.

In some cases, KM and CTM seem to have retained a C_1C_2- cluster unexpectedly, as in ***s**əm**b**ah-*hiaŋ (worship-divinity) 'pray to the gods' > KM *smayɛ~mmayɛ*, CTM *mmayaŋ*, Dusun *smayɔŋ~mmayɔŋ* 'to pray' (cf. SM ***s**ə**m**bahyaŋ*), *(ə)saʔ a**m**bil-an 'one taken away (from ten)' > KM *smilɛ*, CTM *smilaŋ~mmilaŋ*, Dusun *smilaŋ* 'nine' (cf. SM *sə**m**bilan*). PM reconstructions of these words were actually compounds and not trisyllabic, but comparisons with their SM cognates suggest that an earlier nasal ⁺C_2 is retained in all three varieties, whereas usually the stop ⁺C_3 is expected to be retained in KM and CTM. A speculative explanation is that independent consonant sequence reduction might have taken place in the roots before the compounds were contracted to an unanalysable form, and SM forms

[73] The reduction of *-C_2.C_3- in trisyllables as such thus seems to be independent of the reduction of consonant sequences in disyllables (§7.3.6), as different changes are observed.

in these cases do not reflect an earlier stage of the NEPM cognates. For instance, *-mb- in *səmbah could be reduced to ⁺m first, after which further changes applied, i.e., *səmbah-*hiaŋ 'pray to the gods' > ⁺səmah-hiaŋ > ⁺səmayaŋ > KM *smayɛ~mmayɛ*, CTM *mmayaŋ*, Dusun *smayɔŋ~mmayɔŋ* 'to pray'. Similarly, *(ə)saʔ ambil-an 'one taken away (from ten)' > ⁺sa amilan > ⁺samilan > KM *smilɛ*, CTM *smilaŋ~mmilaŋ*, Dusun *smilaŋ* 'nine'.

To summarise, antepenultimate vowels syncope occurred in earlier trisyllables *$C_1V(C_2).C_3V(C).(C)V(C)$, sometimes accompanied by the reduction of *-$C_2.C_3$- sequences. These changes generated disyllables with an initial C_1C_2- or C_1C_3- cluster, and occasionally with a C_3C_3- cluster.

7.5.2.3 Cluster assimilation

It has been shown that some disyllables reduced from earlier trisyllables have initial geminate clusters. Some geminate clusters arise from earlier trisyllables *$C_1V(C_2).C_3V(C).(C)V(C)$ where *C_1 and *C_3 (or *C_1 and *C_2) happened to be identical. Most other geminates clusters, however, result from the cluster assimilation of earlier non-geminate clusters.

The most compelling evidence supporting cluster assimilation comes from the synchronic variation attested between a non-geminate cluster and a geminate cluster, e.g., *tAliŋa(ʔ) 'ear' > KM *tliŋɔ~lliŋɔ*, *tiŋgələm 'to sink' > CTM *tgəlaŋ~ggəlaŋ*, Portuguese *tabaco* 'tobacco' > SM *təmbakau* > Dusun *tmakɔ~mmakɔ*. The phonological conditions for cluster assimilation can be deduced from the patterns of synchronic non-geminate clusters, which represent the types of clusters that did not undergo assimilation or have not completed the assimilatory process.

Non-geminate clusters attested in NEPMs typically involve combinations of an obstruent + a liquid, an obstruent + an obstruent or an obstruent + a nasal (§2.5, §3.5 and §4.5). Obstruent + obstruent clusters usually consist of fricative *s* + a stop or a voiceless stop + a voiced stop. Recall that most non-geminate clusters follow the SSP, with the two components having different places of articulation (with the exception of *s* + stop clusters). Diachronically, it signifies that all other non-geminate clusters, which either violated the SSP or had two consonants with the same place of articulation, were assimilated to become geminate clusters.[74] This assimilation also often occurred for non-geminate clusters with two non-identical segments of

[74] Except for nasal + obstruent clusters such as *mb-* and *ŋg-*, which have a different origin

the same sonority. For instance, the Sanskrit loanword *kapāla* or SM *kəpala* 'head' is expected to have an initial ⁺kp- cluster after antepenultimate vowel syncope, but this cluster apparently underwent further assimilation to become a geminate *pp*- cluster, as seen in KM *ppalɔ*, CTM *ppalə* and Dusun *ppalɛ*.[75] In *g<ər>ahəm 'molar tooth' > KM *ɣɣɛhɛ*, CTM *ɣɣahaŋ* (cf. SM *gəraham*), even though a ⁺gɣ- cluster consisted of an obstruent + a liquid, the same places of articulation of the two segments (both velar) led to further assimilation of ⁺gɣ- > *ɣɣ*-. The same assimilatory process can be seen in the reduction of earlier trisyllabic prefixed forms, e.g., *(mb)Ar-jalan (INTR-road) 'to walk' > ⁺b-jalan > KM *j-jalɛ*, CTM *j-jalaŋ* and Dusun *b-jalaŋ~j-jalaŋ*. Furthermore, at the synchronic level, a number of non-geminate clusters that comply with the SSP also exhibit the tendency of being assimilated to geminate clusters, as in KM *smayɛ~mmayɛ* 'to pray', *tganoŋ~gganoŋ* 'Terengganu', CTM *tgəlaŋ~ggəlaŋ* 'to sink', *smilaŋ~mmilaŋ* 'nine', and Dusun *tmakɔ~mmakɔ* 'tobacco', *smayɔŋ~mmayɔŋ* 'to pray'. Note that the direction of assimilation is typically regressive, i.e., $^*C_1C_2$- > C_2C_2-. More examples illustrating the reduction of trisyllables and the formation of initial geminate clusters are given in Table 7.41.

Table 7.41: Syllable reduction and cluster assimilation in earlier trisyllables

PM	KM	CTM	Dusun	SM	Gloss
*bəŋkaruŋ	*kkaɣoŋ*	*kkaɣoŋ*	*maʔkaɣəʊŋ*[a]	*məŋkaruŋ*	'grass lizard'
...	*ttupaʔ*	*ttupaʔ*	*ttupaʔ*	*kətupat*	'k.o. rice cake'
...	*ccambɔh*	*ccambɔh*	*ccambɔh*	*kəcambah*	'bean sprouts'
–	*nnatɛ*	*nnataŋ*	*nnatɔŋ*	*binataŋ*	'animal'
*b|in|antu	*nnatu*	*nnatu*	*nnatəʊ*	*mənantu*	'child-in-law'
*(mb)Ar-diri	*d-diɣi*	*d-diɣi*	*b-diɣɛi* ~*d-diɣɛi*	*bər-diri*	'to stand'
*mAN-tanək	*nn|anɔʔ*	*nn|anɔʔ*	*nn|anɔʔ*	*mən-<t>anak*	'to cook (rice)'

[a] Dusun *maʔkaɣəʊŋ* is unexplained.

as noted in §7.5.1.

[75] But compare KM/ITM *cpədɔʔ*, CTM *ppədɔʔ* 'cempadak (k.o. fruit)' with SM *cəmpədak*. The ⁺cp cluster underwent assimilation to become *pp*- in CTM, but remains non-geminate in KM and Dusun.

A distinctive type of cluster assimilation can be seen in the genesis of *xx*- and *ww*- clusters, with the latter only attested in CTM. Initial *xx*- is found in words like CTM *xxusi* 'chair', *xxɛtə* 'car' and *xxabaʔ* 'to climb'. The first two items are borrowed from Arabic *kursī* and Portuguese *carreta*, presumably via SM *kərusi* and *kərɛta*. The word *xxabaʔ* 'to climb' can be compared with its SM cognate *kərabat*. These comparisons suggest that *xx*- reflects the assimilation of an earlier ⁺kɣ- cluster. As both ⁺k and ⁺ɣ were velar, cluster assimilation was indeed expected. Yet in this case, assimilation preserves the features of both sounds, namely the voiceless feature of the stop ⁺k and the fricative manner of ⁺ɣ, hence a reciprocal assimilation. The correspondences of NEPM *xx*- : SM *kər*- are further exemplified in Table 7.42.

Table 7.42: Correspondences of NEPM *xx*- : SM *kər*-

KM	CTM	Dusun	SM	Gloss
xxusi	*xxusi*	*xusɛi*	*kərusi*	'chair'
xxɛtɔ	*xxɛtə*	*xitɛ*	*kərɛta*	'car'
xxətah	*xxətah*	*xxətah*	*kərtas*	'paper'
xxijɔ	*xxijə*	*xxəjɛ*	*kərja*	'work'
xxəpoʔ	*xxəpoʔ*	*xxəpuʔ*	*kərɔpɔʔ*	'k.o. cracker'
xxaniŋ	*xxaniŋ*	*xxaniŋ*	*kərani*	'clerk'
xxabaʔ	*xxabaʔ*	*xxabaʔ*	*kərabat*	'to climb'

In Dusun, initial *xx*- underwent reduction preceding high vowels, hence *xusɛi* 'chair' (sometimes further reduced to *usɛi*) and *xitɛ* 'car' (cf. the reduction of *ɣ- preceding high vowels in §7.3.4). In a parallel development, earlier ⁺bɣ- is sometimes assimilated to *ww*- in CTM; compare CTM *wwapə* 'how much', *wwəkah* 'to arrest' and *wwənah* 'rice ears' with SM *bərapa*, *bərkas* and *bərnas*.

In a few other instances, the assimilation of earlier non-geminate clusters appears to be progressive; compare KM *ssəmɔ*, CTM *ssəmə* 'all' with SM *səmua*. KM *ssəmɔ* and CTM *ssəmə* probably had an earlier ⁺sm- cluster, which has been progressively assimilated to *ss*-. Also compare KM *ssəjeʔ*, CTM *mməjeʔ*, Dusun *ssəjiʔ* 'mosque' with SM *masjid*, which ultimately comes from Arabic *masjid*. The origin of this loanword had a non-native medial -*sj*- sequence, which was presumably broken up by a schwa when borrowed into NEPMs, i.e., ⁺masəjid. The reduction of this trisyllable resulted in an initial ⁺ms- cluster, which was regressively assimilated to *ss-* in KM and Dusun, but progressively to *mm-* in CTM.

The reduction of earlier trisyllables and subsequent cluster assimilation account for the origins of most initial geminates in NEPMs. Additionally, initial geminates are also found in some animal names, often corresponding to SM cognates that are reduplicated, as mentioned in §5.3.5. The comparisons are displayed in Table 7.43 (partially repeated from Table 5.9).

Table 7.43: Initial geminates in some animal names

KM	CTM	Dusun	SM	Gloss
kkatɔʔ	*kkatɔʔ*	*kkatɔʔ*	*katak*	'frog'
kkuɣɔ	*kkuɣə*	*kkuɣɛ*	*kura-kura*	'(land) turtle'
ɣɣamɔ	*ɣɣamə*	*maʔamɛ*	*rama-rama*	'butterfly'
llabɔ	*llabə*	*glabɛ*	*laba-laba*	'spider'

I suspect that initial geminates in these forms result from partial reduplication in an earlier disyllabic root (either CV- or Cə- reduplication), followed by the regular reduction of trisyllables, e.g., ⁺katak > ⁺ka-katak/⁺kə-katak > *kkatɔʔ* 'frog'. Dusun *maʔamɛ* 'butterfly' presumably reflects an earlier full reduplicated form whereby initial ⁺ɣa- and medial ⁺-ɣ- were deleted, i.e., ⁺ɣama-ɣama > ⁺ma-ama > *maʔamɛ*. Initial *g-* in Dusun *glabɛ* 'spider' is unexplained.

7.5.3 Interim summary

In summary, syllable reduction has been a prevalent phonological change in the development of NEPMs, affecting words of various shapes. On the one hand, disyllabic words with initial *(h)ə- have been reduced to monosyllables. On the other hand, all trisyllabic words have been reduced to disyllables, driven by vowel contraction or vowel syncope.

Vowel contraction affected vowels across the original antepenultimate and penultimate syllables of a trisyllable (i.e., $*V_1V_2$ in $*C_1V_1.(C)_2V_2.(C)V(C)$, $*C_2$ may be *h), and vowel syncope affected the antepenultimate vowel in all other trisyllables, giving rise to various types of initial consonant clusters.

Non-geminate clusters resulting from antepenultimate vowel syncope may undergo cluster assimilation. The general principle governing assimilation is the SSP, which stipulates that non-geminate clusters violating the SSP should be assimilated to geminate clusters. In addition, clusters with two consonants of the same place of articulation were also subject to assimilation. The direction of assimilation was typically regressive, but reciprocal and progressive assimilation are also attested. Some non-geminate clusters complying with the SSP have also been observed to undergo assimilation or display a tendency to do so synchronically, as evidenced by *bɣ- > CTM *ww-*, as well as some variation between non-geminate and geminate clusters.

Given a PM form with a $*C_1V(C_2).C_3V(C).(C)V(C)$ shape, any $*C_3$ or $*C_2$ can appear geminated at the synchronic level. This diachronic path explains the large inventory of geminates, which was further expanded by geminates resulting from reciprocal assimilation. The absence of geminate glottals and glides (except CTM *ww-*) is also explained by this evolution, as these phonemes either never appeared in the position of $*C_2$ or $*C_3$, or have been regularly deleted. Lastly, it is noteworthy that vowel syncope and cluster assimilation are not uncommon processes of geminate formation from a cross-linguistic perspective (Blust 1995; Blevins 2004: 168–191). What is unique in NEPMs is that these two processes took place successively in an unusual position, namely word-initially, and affected all earlier trisyllables, including simple words, prefixed derivations and loanwords, in the same way.

7.6 Relative chronology of sound changes

The list below summarises the most important sound changes that have been discussed so far, each identified by a number. Many sound changes occurred in all three varieties, whereas some are only attested in one or two varieties. Based on the particular environments in which some sound changes took place, a relative chronological order can be established.

(1) Rounding of ultimate *a, *ə > ⁺a > ɔ;
(2) Lowering of ultimate *i and *u in KM and CTM;
(3) Diphthongisation of ultimate *i and *u in ITM;
(4) Merger of final nasals;
(5) Merger of final stops;

(6) Loss of final approximants;
(7) *-s > *-h*;
(8) *-aN, *-əN > $^{+}$aN > -*ɛ* in KM;
(9) Loss of non-final *h;
(10) Reduction of word-medial sequences;
(11) Sporadic lowering of penultimate *i and *u in KM and CTM;
(12) Loss of initial *ə;
(13) Vowel contraction across antepenultimate and penultimate syllables;
(14) Antepenultimate schwa syncope and cluster assimilation.

(1)–(3) are the most noteworthy sound changes in the vowel systems, all of which took place in final syllables. These three vowel changes were conditioned by historical final segments, which indicates that they must have predated the changes of final consonants (4)–(7). For instance, *a, *ə > $^{+}$a > ɔ in Dusun occurred before historical back consonants *-k, (*-ʔ), *-h, *-r, *-ŋ and *-w. This change must have taken place before the final consonants were deleted or merged. Sound change (8) in KM presumably followed the merger of nasals, as all final nasals were affected in the same way.

Sound changes that took place in non-final syllables, including (9) the loss of non-final *h, (10) the reduction of word-medial sequences and (11) the sporadic lowering of penultimate vowels in KM and CTM, are independent of the changes in final syllables. It cannot be determined whether they took place before or after the changes in final syllables (1)–(8). The loss of initial *ə (12) apparently followed the loss of non-final *h, as both initial *ə and *hə were affected.

The reduction of trisyllables, realised by (13) vowel contraction across the antepenultimate and penultimate syllables, and (14) antepenultimate vowel syncope and cluster assimilation, is also independent of other changes. The general tendency of syllable reduction is common to all three varieties, but the details are not identical.

At first glance, it appears that the common changes attested in all three varieties, namely (4)–(7) and (9)–(10), are shared innovations that might be reconstructable to an immediate ancestral language, namely Proto NEPM. However, the chronological order of sound changes shows that changes in final consonants (4)–(7) must have been preceded by changes of ultimate vowels, which are distinct in each variety. There is no single vowel change that affected all three varieties in the exact same way in this position, except that (2) is shared by KM and CTM. Even though the sound change (1) ultimate *a, *ə > ⁺a > ɔ is recurring, the environment in which the rounding took place differed across NEPMs (see §7.4.3). In other words, it is not possible that (4)–(7) took place in a common ancestral language; instead, they must have recurred independently or resulted from diffusion, after ultimate vowel changes had already affected NEPMs in distinct ways. The only sound changes that might be reconstructed to an earlier stage are (9) and (10), as well as (2) for KM and CTM, but all these changes are fairly common in Malayic varieties in general, therefore offering little value in subgrouping.

In conclusion, while there are some common phonological changes in NEPMs, they cannot be taken as shared innovations inherited from a common ancestral language. In other words, as far as sound changes are concerned, there is no evidence showing that NEPMs form a discrete subgroup within Malayic. This finding contradicts earlier proposals that suggested the existence of a "Northeastern Peninsular Malay(ic)" subgroup (Collins 1989: 253–254; Tadmor 1995: 13–14; Hammarström et al. 2023), and has significant implications for understanding the migration history of NEPM speakers, which will be discussed further in §9.3.

7.7 Summary

This chapter has explored the phonological history of NEPMs through a top-down approach, focusing on examining the reflexes of PM phonemes in the present-day daughter languages and establishing sound changes that have taken place over time. Particular emphasis has been given to syllable reduction, a prevalent process that has affected both disyllables and trisyllables.

Some general trends and important observations in the phonological history of NEPMs are summarised as follows. In the consonant system, changes were primarily observed in word-final position. Final stops have merged to *-ʔ*, and final nasals have been neutralised to *-ŋ*. Fricatives merged to *-h*, and all approximants were lost. Word-medial consonant sequences comprising a nasal + a homorganic voiceless stop or *s were reduced to their obstruent components. Overall, the development of consonants from PM to NEPMs exhibits a trend of reduction. In contrast, the development of vowels may be viewed as a matter of complexification. All three varieties have acquired more vowels compared to their ancestral language, as a result of the lowering of earlier high vowels, the raising of low vowels, and the nasalisation of oral vowels in certain environments. While Dusun (and ITM in general) retains a conservative feature of preserving penultimate high vowels (Anderbeck in print), ultimate high vowels underwent diphthongisation, with complex conditions determined by the presence and the nasality of the onset, as well as the presence and the place feature of the coda. At the syllable level, some PM disyllables have been reduced to monosyllables as the result of losing initial *(h)ə, and earlier trisyllables have been reduced to disyllables through vowel contraction or antepenultimate vowel syncope. The reduced disyllables have word-initial consonant clusters, some of which were further affected by cluster assimilation, resulting in geminate clusters.

Beyond examining the phonological history of NEPMs, this chapter also included a reevaluation of certain PM reconstructions. Based on NEPM data, some modifications to existing reconstructions were proposed, and several new reconstructions were suggested. Among others, I discussed the status of PM *-ʔ and concluded that NEPM material does not support *-ʔ. Furthermore, I proposed the reconstruction of *-rC- sequences in PM, although the origins of words containing such sequences require further investigation.

A comparison across NEPMs reveals that the three varieties share many phonological developments, particularly in the consonant system. However, the establishment of a relative chronology of sound changes reveals that these superficial common changes in the consonant systems must have followed distinct developments in the vowel systems. Therefore, it is not possible to attribute these common changes to shared innovations, or reconstruct them to an immediate common ancestral language. NEPMs thus do not seem to derive from a discrete subgroup within Malayic.