Malayic varieties of Kelantan and Terengganu: description and linguistic history

Wu, J.

CHAPTER 8

Morphological history

8.1 Introduction

This chapter examines the historical development of the morphological system from PM to NEPMs, focusing on two topics: first, the retention of affixes from PM and the innovation of initial gemination as a morphophonological operation; second, the loss of other affixes from PM and the possible mechanisms behind this general tendency of morphological reduction.

This chapter begins with an overview of the affixes reconstructed in PM (§8.2). As described in Chapter 5, NEPMs have notably small inventories of affixes, all of which are prefixes. In §8.3, I demonstrate that all NEPM prefixes can be traced back to PM following regular sound changes. Furthermore, many initial geminated segments may be viewed as regular reflexes of earlier prefixes under certain phonological conditions, but not all of them can be accounted for in this way. I argue that initial gemination is being generalised as a result of analogical change. In §8.4, I turn to PM affixes that are not inherited in NEPMs. As will be shown, NEPMs have lost all PM suffixes and circumfixes, which evidences an overall reduction of morphology. Some affixes are lost without a trace, while others are retained in a few fossilised forms. I propose that the morphological reduction was primarily driven

by internal phonological changes. While the possibility of substratal influences triggering morphological reduction cannot be ruled out, there is little supporting evidence. §8.5 summarises this chapter.

8.2 Affixes reconstructed in PM

In the same vein as reconstructing PM phonology and lexicon, Adelaar (1984, 1992) provided a reconstruction of PM affixes. Table 8.1 summarises the reconstructed affixes and the bases to which they could be attached. The list is by no means exhaustive, as the reconstruction focused on affixation on nouns and verbs, and only affixes that were presumably active are presented here. In addition to these affixes, PM had clitics such as *sA= 'one' and *=ɲa '3SG', and some grammatical morphemes that were likely fossilised, e.g., *b(a)- 'someone who behaves like BASE' and *=ŋ 'LIG' (Adelaar 1994, 2004a); they are not treated in the following discussion.

The sixteen affixes listed in the table include both inflectional and derivational affixes. There are nine verb-forming affixes (*(mb)Ar-, *pAr$_1$-, *tAr-, *mAN-, *-i, *-aʔ, *maka-, *-an$_1$ and *kA- -an$_1$) and seven noun-forming affixes (*-an$_2$, *-An, *kA- -an$_2$, *pAN-, *pAr$_2$-, *pAN- -an and *pAr- -an). Some notes on the reconstruction of certain affixes are necessary here. The prefix *(mb)Ar- was reconstructed with an uncertain initial segment that could be either *m or *b, despite all contemporary Malayic languages having *b*. This uncertainty was on account of the fact that the earliest Old Malay inscription had *mar-*, which was considered the continuation of PMP *maR- 'ACT', and this prefix also appeared as *bar-* in some later Old Malay inscriptions (see an overview in Mahdi 2005: 185). Additionally, *mAN- was reconstructed as one prefix with two distinct functions. It served as an agent-oriented marker when attached to dynamic verbal bases (*mAN$_1$-) and as an intransitive verbal marker when attached to nominal and stative verbal bases (*mAN$_2$-). The suffix *-An might be collapsed with *-an$_2$ as one suffix covering the function of 'LOC; RES' when attached to dynamic transitive verbal bases.[76]

[76] The reason to reconstruct two separate suffixes is that PMP had two distinct suffixes *-an 'LOC' and *-en 'RES', the latter of which would have been reflected as *-en* in Jakartanese (hence PM *-An), which is nevertheless unattested. With a bottom-up reconstruction within Malayic, however, there is no clear evidence for the reconstruction of *-An.

Table 8.1: Affixes reconstructed in PM

		Nouns	Dynamic transitive	Dynamic intransitive	Stative intransitive
*(mb)Ar-	'INTR'	+	-	+	-
*pAr$_1$-	'TR'	+	-	+	+
*tAr-	'NVOL'	-	+	+	?
*mAN-	*mAN$_1$- 'AGT'	-	+	+	-
	*mAN$_2$- 'INTR'	+	-	-	+
*-i	'APPL'	+	+	+	+
*-aʔ	'SUBJ'	-	+	+	?
*maka-	'TR.CAUS'	-	-	+	+
*-an$_1$	'DISTR'	+ ('COL')	+ ('RECP')	+	-
*-an$_2$	'NMLS'	-	+ ('LOC')	-	+ ('ATTR)'
*-An	'NMLS'	-	+ ('RES')	-	-
*kA- -an$_1$	'NVOL'	+ ('ADVS')	+	+	+ ('ADVS')
*kA- -an$_2$	'NMLS'	-	-	+ ('LOC')	+ ('ABST.ATTR')
*pAN-	'NMLS'	-	+ ('INST')	+ ('INST')	+ ('ATTR')
*pAr$_2$-	'NMLS	-	+ ('INST')	+ ('INST')	-
*pAN- -an	'NMLS'	-	+ ('ABST; LOC')	+ ('ABST; LOC')	-
*pAr- -an	'NMLS'	+ ('LOC')	+ ('ABST; LOC')	+ ('ABST; LOC')	-

Many PM affixes could derive new forms with varied meanings depending on the word class of the base. For practical reasons, some coverall glosses are given in the second column in the table, followed by more accurate specifications when deemed necessary. For a comprehensive account of the reconstruction of these affixes and their grammatical functions, see Adelaar (1992: 145–194). A concise summary can be found in Anderbeck (in print).

Needless to say, PM was far from being isolating. In comparison, KM and CTM only have five prefixes, namely *bɣ-* 'INTR; MID', *pɣ-* 'CAUS; FCT', *tɣ-* 'NVOL', *NN$_1$-* 'IPFV' and *NN$_2$-* 'NMLS' (§5.3.1). ITM has an even smaller inventory with four prefixes as it lacks the causative/factitive prefix. In addition to prefixation, however, NEPMs utilise the process of initial gemination to realise certain grammatical functions (§5.3.2). The general evolution of the morphology from PM to NEPMs can thus be characterised as a process of reduction with some traits of innovation.

8.3 Morphological retention and innovation

The five prefixes found in present-day NEPMs are all retentions from PM, as shown in (1).

(1)	*(mb)Ar-	'INTR'	>	*bɣ-*	'INTR; MID'
	*pAr$_1$-	'TR'	>	*pɣ-*	'CAUS; FCT'
	*tAr-	'NVOL'	>	*tɣ-*	'NVOL'
	*mAN-	'AGT; INTR'	>	*NN$_1$-*	'IPFV'
	*pAN-	'NMLS'	>	*NN$_2$-*	'NMLS'

Except for *mAN- 'AGT; INTR' > *NN$_1$-* 'IPFV', other prefixes generally retain the original meanings and functions, despite some analytical differences. For instance, I treat NEPM *bɣ-* as a middle (voice) marker when attached to verbal bases (§5.3.1.2), whereas this function was subsumed as part of the intransitive marker for PM *(mb)Ar-. Similarly, PM *pAr$_1$- was considered a prefix forming transitive verbs, but a distinction was made in NEPM *pɣ-* between a causative marker when prefixed to verbal bases and a factitive marker when prefixed to nominal bases (§5.3.1.4). The reason behind the semantic shift from *mAN- 'AGT; INTR' to *NN$_1$-* 'IPFV' is unclear, and it may be considered an innovation. However, as pointed out in §5.3.1.5, aspectual functions of cognates to *NN$_1$-* appear to have a wider distribution in Malayic

languages, and it is not unlikely that such an aspectual meaning was already present in PM *mAN-.

Formally, the changes reflected in the phonological forms of these prefixes follow regular sound changes. In §7.5, I demonstrated that PM trisyllables underwent syllable reduction and became disyllables in NEPMs, commonly through antepenultimate vowel syncope and subsequent cluster assimilation. Importantly, these sound changes affected both simple words and prefixed derivatives in the same way. Since the canonical shape of PM roots was disyllabic, prefixes typically fell on the antepenultimate syllables, which were the targets of syllable reduction. The phonological evolution of these prefixes retained from PM is described in the following sections. I first consider *(mb)Ar-, *pAr$_{l}$- and *tAr- in §8.3.1. *mAN- and *pAN- are treated in §8.3.2.

8.3.1 PM *(mb)Ar-, *pAr$_{l}$- and *tAr-

To illustrate the sound changes reflected in PM *(mb)Ar-, *pAr$_{l}$- and *tAr-, some examples are given in (2) to (4).

(2) PM > KM

***(mb)Ar**-anak	>	***bɣ**-*anɔʔ*	(INTR-child)	‘to give birth’
***(mb)Ar**-layar	>	***b**-*laya*	(INTR-sail)	‘to sail’
***(mb)Ar**-lari	>	***b**-*laɣi*	(MID-run)	‘to run’
***pAr$_{l}$**-habis	>	***pɣ**-*abih*	(CAUS-finished)	‘to finish’
***pAr$_{l}$**-hati	>	***pɣ**-*ati*	(FCT-liver)	‘to observe’
$^{+}$**pAr$_{l}$**-lumat	>	***p**-*lumaʔ*	(CAUS-crushed)	‘to crush’
$^{+}$**tAr**-iŋat	>	***tɣ**-*iŋaʔ*	(NVOL-think)	‘to remember’
***tAr**-bakar	>	***t**-*baka*	(NVOL-burn)	‘to be burnt’

(3) PM > CTM

***(mb)Ar**-anak	>	***bɣ**-*anɔʔ*	(INTR-child)	‘to give birth’
***(mb)Ar**-lari	>	***b**-*laɣi*	(MID-run)	‘to run’
***pAr$_{l}$**-habis	>	***pɣ**-*abih*	(CAUS-finished)	‘to finish’
$^{+}$**pAr$_{l}$**-hancur	>	***pɣ**-*aco*	(CAUS-crushed)	‘to crush’
***pAr$_{l}$**-hati	>	***pɣ**-*ati*	(FCT-liver)	‘to observe’
$^{+}$**tAr**-iŋat	>	***tɣ**-*iŋaʔ*	(NVOL-think)	‘to remember’
***tAr**-bakar	>	***t**-*bakɔ*	(NVOL-burn)	‘to be burnt’

(4) PM > ITM (*pAr$_1$- is not inherited)

***(mb)Ar**-anak	>	***bɣ**-anɔʔ*	(INTR-child)	'to give birth'
⁺**(mb)Ar**-asal	>	***bɣ**-asa*	(INTR-origin)	'to originate'
***(mb)Ar**-lari	>	***b**-layɛi*	(MID-run)	'to run'
***tAr**-ambil	>	***tɣ**-ambɛiʔ*	(NVOL-take)	'to take (by mistake)'
***tAr**-bakar	>	***t**-bakɔ*	(NVOL-burn)	'to be burnt'

In all examples, the antepenultimate vowel (reconstructed as an ambivalent *A) was deleted. Recall that the immediate result of antepenultimate vowel syncope in a PM trisyllable with a *C$_1$V(C$_2$).C$_3$V(C).(C)V(C) shape is a disyllable with an initial C$_1$C$_3$- or C$_1$C$_2$- cluster (§7.5.2.2). For the PM forms listed above, when the initial segment of the base was a vowel or *h (which was regularly deleted), the liquid *r in the prefixes was in the position of *C$_3$ with an empty *C$_2$. Consequently, the reflexes of these derivatives have initial *bɣ-*, *pɣ-* and *tɣ-*, as seen in examples such as KM *bɣ-anɔʔ*, *pɣ-abih* and *tɣ-iŋaʔ*. When the initial segment of the base was a consonant other than *h, it occupied the position of *C$_3$, with *r in the prefix occupying *C$_2$. Since *r in *C$_2$ position was regularly deleted, PM *(mb)Ar-, *pAr$_1$- and *tAr- are reflected as single segments *b-*, *p-* and *t-*, as seen in KM *b-layi*, *p-luma*ʔ and *t-baka*. Synchronically, these sing-segment prefixes can be analysed as allomorphs of *bɣ-*, *pɣ-* and *tɣ-* before consonant-initial bases, as detailed in §5.3.1.1.

Following antepenultimate vowel syncope and *r deletion, the reduced single-segment prefix *b-*, *p-* or *t-* essentially forms a consonant cluster with the base-initial consonant. When the prefix is identical to the base-initial consonant, a geminate cluster is formed, as illustrated in (5) and (6).

(5) PM > KM/CTM

***(mb)Ar**-buah	>	***b**-buwɔh*	(INTR-fruit)	'to bear fruit'
***(mb)Ar**-baris	>	***b**-bayih*	(INTR-line)	'to queue'
***tAr**-tidur	>	***t**-tido*	(NVOL-sleep)	'to fall asleep'
⁺**tAr**-tiŋgal	>	***t**-tiŋga*	(NVOL-leave)	'to be left behind'

(6) PM > ITM

***(mb)Ar**-bini	>	***b**-biniŋ*	(INTR-wife)	'to marry (a wife)'
***(mb)Ar**-bau	>	***b**-bau*	(INTR-smell)	'smelly'
***tAr**-tidur	>	***t**-tidu*	(NVOL-sleep)	'to fall asleep'
⁺**tAr**-tiŋgal	>	***t**-tiŋga*	(NVOL-leave)	'to be left behind'

In other cases, the prefix would first form a non-geminate cluster with the base-initial consonant. This non-geminate cluster, like other non-geminate clusters resulting from the reduction of morphologically simple trisyllables, was subject to further cluster assimilation. As discussed in §7.5.2.3, clusters violating the SSP were typically assimilated regressively to become geminates. For prefixed forms with *(mb)Ar-, *pAr$_{l}$- or *tAr-, it is expected that *b-*, *p-* and *t-* were assimilated to base-initial segments that were equally or less sonorous. The function of these original prefixes is thus realised by an initial geminated segment on the surface. This development of PM prefixes is illustrated in (7) to (9).

(7) PM > KM/CTM

***(mb)Ar**-jalan	>	***j-jalɛ***	(INTR-road)	'to walk'
***(mb)Ar**-jəmur	>	***j-jəmo***	(MID-jəmo)	'to sunbathe'
***pAr$_{l}$**-kəriŋ	>	***k-kəɣiŋ***	(CAUS-dry)	'to dry s.th.'
***pAr$_{l}$**-tidur	>	***t-tido***	(CAUS-sleep)	'to put s.o. to sleep'
⁺**tAr**-kəjut	>	***k-kəjuʔ***	(NVOL-startle)	'to be startled'
⁺**tAr**-saŋkut	>	***s-sakoʔ***	(NVOL-hang)	'to be hung'

(8) PM > CTM

***(mb)Ar**-diri	>	***d-diɣi***	(INTR-self)	'to stand'
***(mb)Ar**-jəmur	>	***j-jəmo***	(MID-jəmo)	'to sunbathe'
⁺**pAr$_{l}$**-kuat	>	***k-kuwaʔ***	(CAUS-strong)	'to strengthen'
***pAr$_{l}$**-tidur	>	***t-tido***	(CAUS-sleep)	'to put s.o. to sleep'
⁺**tAr**-kəjut	>	***k-kəjuʔ***	(NVOL-startle)	'to be startled'
⁺**tAr**-saŋkut	>	***s-sakoʔ***	(NVOL-hang)	'to be hung'

(9) PM > ITM

***(mb)Ar**-cabaŋ	>	***c-cabɔŋ***	(INTR-branch)	'branched'
***(mb)Ar**-darah	>	***d-daɣɔh***	(INTR-blood)	'to bleed'
⁺**(mb)Ar**-sandar	>	***s-sandɔ***	(MID-lean)	'to lean (oneself)'
⁺**tAr**-kəjut	>	***k-kəjuʔ***	(NVOL-startle)	'to be startled'
⁺**tAr**-pijak	>	***p-pijɔʔ***	(NVOL-step.on)	'to step on (unintentionally)'

As a result, following regular sound changes, PM *(mb)Ar-, *pAr$_{l}$- and *tAr- are expected to have three sets of reflexes under different phonological conditions: they are reflected as *bɣ-*, *pɣ-* and *tɣ-* before bases with initial vowels or *h, as *b-*, *p-* and *t-* before bases with a more sonorous initial consonant,

and as a segment identical to the base-initial consonant elsewhere. This anticipated evolution of PM prefixes is presented in a schematic form in (10).

(10) Expected reflexes of *(mb)Ar-, *pAr$_1$- and *tAr-

*(mb)Ar-, *pAr$_1$-, *tAr- > { *bɣ-, pɣ-, tɣ-* / __*(h)V; *b-, p-, t-* / __*C with higher sonority; base-initial C / __other *C }

The summary in (10) carries two important implications. First, the schema demonstrates a shared historical connection between NEPM prefixes and complex geminates; both can be traced back to original PM prefixes. Second, the three types of reflexes are expected to occur in complementary distributions, which suggests that they may be viewed as allomorphic alternations of underlying prefixes at the synchronic level. These two implications can be substantiated to a large extent. Many complex geminates arise as the results of regular sound changes, and they can be seen as deriving from an underlying prefix synchronically.

However, it is essential to note that not all complex geminates can be satisfactorily analysed in this way. A number of unexplained irregularities were already noted in §5.3.2.2. For instance, KM has several instances of complex geminate liquids, e.g., *l-luma?* (CAUS-crushed) 'to crush', *l-luwah* (CAUS-wide) 'to widen' and *ɣ-ɣayɔ* (INTR-Eid.al-Fitr) 'to celebrate Eid al-Fitr'. According to the generalisation in (10), the causative marker and the intransitive marker should have appeared as *p-* and *b-* respectively before bases with an initial liquid, as there is no clear phonological motivation for the assimilation of *bl-* > *ll-* or *bɣ-* > *ɣɣ-*. Similarly, CTM and ITM have examples in which a geminated voiced obstruent functions as a non-volitional marker, e.g., CTM *b-bukə*/ITM *b-bukɛ* (NVOL-open) 'opened; to open (on its own)', CTM *g-gatoŋ* (NVOL-hang) 'to be hung' and ITM *j-jatəʊh* (NVOL-fall) 'to fall (unintentionally)'. If the non-volitional marker were a retention of *tAr-, it should have been reflected as *t-* before a more sonorous voiced obstruent. The complex geminates in these examples cannot be straightforwardly derived from underlying prefixes or traced back to earlier prefixes following regular sound changes. I argue that they must have directly derived from the gemination of base-initial segments, i.e., C_x- → C_xC_x- (see §5.3.2.2). This initial gemination, in my view, is an innovation that likely developed as a result of reanalysis and analogical change.

Examples (7) to (9) demonstrate that regular sound changes can give rise to complex geminates at the phonetic/phonological level (irrespective of their underlying morphological structure), leading to the emergence of minimal pairs that only display contrasts in the length of initial consonants. Within a linguistic system with numerous pairs like these, it would not be surprising that the speakers associate these initial geminated segments with the marking of certain grammatical functions. Consequently, a new rule of initial gemination is generalised, and it may be extended to other bases. To elaborate on this idea, consider the examples in (11).

(11) KM

jalɛ	'road'	vs.	***j**-jalɛ*	(INTR-road)	'to walk'
pəɣɛ	'war'	vs.	***p**-pəɣɛ*	(INTR-war)	'to be at war'
diɣi	'self'	vs.	***d**-diɣi*	(INTR-self)	'to stand'
saiŋ	'friend'	vs.	***s**-saiŋ*	(INTR-friend)	'to befriend'
...			...		
ɣayɔ	'Eid.al-Fitr'	vs.	*X* = ***ɣ**-ɣayɔ*	(INTR-Eid.al-Fitr)	'to celebrate Eid al-Fitr'

The first four pairs illustrate how phonemic contrasts between bases with an initial singleton and derivatives with an initial geminate can result from regular sound changes. The geminated segments *j-*, *p-*, *d-* and *s-* all reflect *(mb)Ar- 'INTR'. Synchronically, they may be analysed as surface realisations of an underlying prefix *bɣ-*, which undergoes *ɣ* deletion preceding a consonant-initial base, followed by the assimilation of the reduced prefix *b-* to the respective base-initial consonant. The allomorphic alternation is nevertheless rather opaque. Given those four pairs and many others, a more transparent association can be established between initial geminated segments (or abstractly, an empty morphological template with its phonemic content copied from the base-initial consonant) and the function of an intransitive verbal marker on nominal bases. This association can be seen as a process of reinterpretation, where the surface realisation remains the same, but the underlying grammatical apparatus producing those surface forms changed (Joseph 2001: 357). By analogy, this association can be extended to new contexts, such as other nouns like *ɣayɔ* 'Eid.al-Fitr', to derive *ɣ-ɣayɔ* 'INTR-Eid.al-Fitr', which would otherwise have had the form ×*b-ɣayɔ*.

In a similar manner, many regular reflexes of PM *pAr$_1$- 'TR' and *tAr- 'NVOL' must have also been reanalysed as the outcomes of the operation of initial gemination. As a result, initial gemination synchronically covers various grammatical functions including an intransitive verbal marker, a causative marker and a non-volitional marker, and it has become a rather productive process.

To briefly recap, PM *(mb)Ar-, *pAr$_1$- and *tAr- have regular reflexes of *bɣ-* 'INTR; MID', *pɣ-* 'CAUS; FCT' and *tɣ-* 'NVOL' when preceding vowels, which take on allomorphic variants *b-*, *p-* and *t-* when preceding more sonorous consonants. While many initial geminated segments are also regular reflexes of these PM prefixes, initial gemination has generalised to become a synchronic process with diverse morphological functions. This evolution has been driven by the reanalysis of existing linguistic material and analogical processes.

8.3.2 PM *mAN- and *pAN-

The formal changes reflected in *mAN- 'AGT; INTR' > *NN$_1$-* 'IPFV' and *pAN- 'NMLS' > *NN$_2$-* 'NMLS' follow a similar trajectory, which is also closely connected to the process of syllable reduction. However, it is noteworthy that these two PM prefixes are only inherited in restricted phonological environments.

In PM, the nasal element *N in *mAN- and *pAN- already exhibited morphophonological alternations depending on the initial segment of the base. The following alternations of *N are summarised based on Adelaar (1992: 160–163):

1) *N was realised as a homorganic nasal before bases with an initial stop;
2) in cases where the initial stop was voiceless (*p, *t and *k), the stop was substituted by the homorganic nasal;
3) *N was realised as palatal before base-initial *s, substituting the *s;
4) *N was deleted before initial nasals and liquids.

Furthermore, *N was presumably realised as *ŋ before bases with an initial vowel, although this generalisation was not included in Adelaar's summary.

PM *mAN- and *pAN- are most clearly inherited in NEPMs before bases with an initial vowel or a voiceless obstruent, as illustrated in (12) to (14).

(12) PM > KM

***mAN**-aŋkat	>	***ŋŋ**-*aka?*	(IPFV-lift)	‘lifting’
***mAN**-paŋgil	>	***mm**-<*p*>*aŋge*	(IPFV-call)	‘calling’
***pAN**-tutup	>	***nn**-<*t*>*uno?*	(NMLS-close)	‘lid, cover’
⁺**mAN**-cukur	>	***ɲɲ**-<*c*>*uko*	(IPFV-shave)	‘shaving’
⁺**mAN**-kəmas	>	***ŋŋ**-<*k*>*əmah*	(IPFV-tidy)	‘tidying’
***pAN**-sakit	>	***ɲɲ**-<*s*>*ake?*	(NMLS-sick)	‘disease’

(13) PM > CTM

***pAN**-pəgaŋ	>	***mm**-<*p*>*əgaŋ*	(NMLS-hold)	‘handle’
***mAN**-tanəm	>	***nn**-<*t*>*anaŋ*	(IPFV-plant)	‘planting’
⁺**mAN**-kəmas	>	***ŋŋ**-<*k*>*əmah*	(IPFV-tidy)	‘tidying’
⁺**pAN**-sangkut	>	***ɲɲ**-<*s*>*ako?*	(NMLS-hang)	‘hanger’

(14) PM > ITM

***mAN**-putuŋ	>	***mm**-<*p*>*utəʊŋ*	(IPFV-cut)	‘cutting’
***mAN**-tulis	>	***nn**-<*t*>*ulih*	(IPFV-write)	‘writing’
⁺**mAN**-kutip	>	***ŋŋ**-<*k*>*uti?*	(IPFV-pick)	‘picking’
***pAN**-sakit	>	***ɲɲ**-<*s*>*aki?*	(NMLS-sick)	‘disease’

Following the aforementioned alternations, the phonological realisations of PM forms in (12) were presumably *mA**ŋ**aŋkat, *mA**m**aŋgil, *pA**n**utup, ⁺mAɲukur, ⁺mA**ŋ**əmas and *pA**ɲ**akit. These trisyllables were also affected by syllable reduction, whereby the antepenultimate vowel *A was deleted. In the case of ***m**A**m**aŋgil, the result of antepenultimate vowel syncope was a geminate cluster *mm-*, as the two consonants surrounding *A were identical. In the other examples, vowel syncope generated clusters such as ⁺mŋ-, ⁺pn-, ⁺mn- and ⁺pɲ-, which were then regressively assimilated to become geminate nasals.[77] At the synchronic level, all these geminate nasals can be analysed as a prefix *NN-* occurring before bases with an initial vowel or a voiceless obstruent with corresponding morphophonological alternations (§5.3.1.1). These sound changes account for the identical shape of the reflexes of *mAN- and *pAN-, both being *NN-*. The distinction between NN_1- ‘IPFV’ and NN_2- ‘NMLS’ is established based on their differing grammatical functions.

[77] The nasal ⁺n and ⁺ɲ are in fact more sonorous than the stop *p*. The assimilation of ⁺pn- > *nn-* and ⁺pɲ- > *ɲɲ-* further illustrates that clusters complying with the SSP may also undergo assimilation.

The retention of *mAN- and *pAN- is less clear in other environments. As mentioned in §5.3.1.1, *NN$_1$-* 'IPFV' and *NN$_2$-* 'NMLS' are generally not attested before bases with an initial voiced obstruent, an initial liquid or an initial nasal, except in a few instances. PM *mAN- appears to be reflected as *m-* before a base with an initial liquid, namely KM *m-lamboŋ* (IPFV-bump) 'bumping' < $^{+}$mAN-lambuŋ ($^{+}$mAlambuŋ). In another example, an initial *p-* may be seen as the reflex of *pAN-, namely NEPM *p-lawɔʔ* (NMLS-lie) 'liar, the habit of lying' < $^{+}$pAN-lawak (cf. SM *pəlawak*). In view of these two examples, it is possible to argue that *mAN- and *pAN- are reflected as *m-* and *p-* before bases with an initial liquid. If this holds true, it would necessitate an expansion of the synchronic inventory of NEPM affixes. However, given the limited number of instances, these *m-* and *p-* cannot be explained satisfactorily for now. It is also likely that these forms are nonce borrowings.

One may suspect that *mAN- and *pAN- have been transformed into initial gemination in some environments, but this suggestion is implausible upon closer examination. On the one hand, there is no attestation of initial gemination as a nominaliser inherited from *pAN-. Initial gemination as an imperfective marker (presumably a reflex of *mAN-), on the other hand, is indeed attested in a few instances, such as NEPM *juwa* 'to sell' → *j-juwa* 'to trade' and ITM *bəlɛi* 'to buy' → *b-bəlɛi* 'to go shopping' (see §5.3.2.2). However, these geminated segments are unlikely to be inherited from *mAN-, as they do often not reflect the results of regular sound changes.

Consider the anticipated development of *mAN- before bases with an initial voiced obstruent in ITM. Since *N was realised as a homorganic nasal without nasal substitution in this environment, PM forms prefixed with *mAN- should have taken on a *C$_1$VC$_2$.C$_3$V(C).(C)V(C) structure with a *-C$_2$.C$_3$- cluster consisting of a nasal + a voiced obstruent, e.g., *mAN-bəli 'ACT-buy' → *mAmbəli. Following regular sound changes, *mAmbəli is expected to be reflected as *mməlɛi* in ITM with the deletion of *b in the *-mb- cluster;[78] that is, *mAN- should have been reflected as *NN-* triggering nasal substitution before voiced obstruents. Nevertheless, this pattern is not attested. ITM forms like *b-bəlɛi* 'to go shopping' and *j-juwa* 'to trade' presumably derive via initial gemination, although the precise source of

[78] Compare with parallel development in morphologically simple words: *tiŋgələm > ITM *tŋəlaŋ* 'to sink', $^{+}$təmbakaw > *tmakɔ~mmakɔ* 'tobacco', see §7.5.2.2.

this imperfective marker is not entirely clear.[79]

It is notable that *mAN- has often become fossilised before bases with an initial nasal, reflected as a geminated nasal segment. This is exemplified in (15) and (16).

(15) PM > KM/CTM

***mAN**-mimpi	(*mAmimpi)	>	***m**|*mipi*	'to dream'
⁺**mAN**-napas[80]	(⁺mAnapas)	>	***n**|*napah*	'to breathe'
***mAN**-(nt)antiʔ	(*mAnantiʔ)	>	***n**|*nati*	'to wait'
⁺**mAN**-ɲaɲi	(⁺mAɲaɲi)	>	***ɲ**|*ɲaɲi*(ŋ)	'to sing'

(16) PM > ITM

***mAN**-mimpi	(*mAmimpi)	>	***m**|*mipɛi*	'to dream'
⁺**mAN**-ɲaɲi	(⁺mAɲaɲi)	>	***ɲ**|*ɲaɲiŋ*	'to sing'
***mAN**-ɲawa	(*mAɲawa)	>	***ɲ**|*ɲawɛ*	'to breath'

In all these cases, only the historical derivatives with a geminate nasal is inherited, whereas the original bases have been lost. There is thus no contrast between forms like ˣ*ɲaɲi*(ŋ) vs. *ɲŋaɲi*(ŋ) at the synchronic level, which means the geminate nasal cannot be analysed as derived by morphological means synchronically. Moreover, the geminate nasals do not seem to carry any grammatical function. They are therefore analysed as fossilised (§5.3.5). Similar fossilisation is also attested before bases with some other segments, as illustrated in (17) and (18).

(17) PM > KM/CTM

***mAN**-alir	(*mAŋalir)	>	***ŋŋ**|*ale*	'to flow'
⁺**mAN**-pikir[81]	(*mAmikir)	>	***mm**|*ike*	'to think'
⁺**mAN**-tari	(*mAnari)	>	***nn**|*aɣi*	'to dance'
⁺**mAN**-kuap	(⁺mAŋuap)	>	***ŋŋ**|*uwaʔ*	'to yawn'

(18) PM > ITM

***mAN**-alir	(*mAŋalir)	>	***ŋŋ**|*alɛi*	'to flow'
⁺**mAN**-pikir	(*mAmikir)	>	***mm**|*iki*	'to think'
***mAN**-taŋis	(*mAnaŋis)	>	***nn**|*aŋih*	'to cry'
⁺**mAN**-kantuk	(⁺mAŋuntuk)	>	***ŋŋ**|*atəʊʔ*	'drowsy'

[79] NEPM *j-juwa* 'to trade' may have developed from *(mb)Ar-jual 'INTR-sell' (cf. SM *jual* 'to sell' → *bər-jual* 'to trade'), in which case the prefix *(mb)Ar- or *bər-* may be viewed as having an imperfective meaning.

[80] Ultimately from Arabic *nafas*, cf. SM *nafas~napas* 'breath'.

To sum up, PM *mAN- 'AGT; INTR' and *pAN 'NMLS' are reflected as *NN$_1$-* 'IPFV' and *NN$_2$-* 'NMLS' in NEPMs, which are only clearly inherited before bases with an initial vowel or a voiceless obstruent. It is also noteworthy that the categories of bases with which *mAN- can occur have been narrowed down. PM *mAN- occurred on nouns and verbs from all categories (both dynamic and stative verbs), but NEPM *NN$_1$-* is only prefixed to dynamic verbs. *NN$_1$-* 'IPFV' is therefore presumably a continuation of *mAN$_1$- 'AGT' (though the distinction between *mAN$_1$- 'AGT' and *mAN$_2$- 'INTR' is essentially an analytical one). *pAN- is almost never inherited before bases with an initial segment other than a vowel or a voiceless obstruent. Considering the restricted productivity of *NN$_2$-* 'NMLS', it may be argued that this prefix is on its way of being fossilised.

8.4 Morphological reduction

Except for the five affixes discussed above, all other PM affixes are lost or no longer active in NEPMs. Some affixes survived in a few fossilised forms, as in the case of *-i 'APPL', *-an$_1$ 'DISTR' and *-an$_2$ 'NMLS'. The fossilisation of these suffixes is discussed in §8.4.1. I make an attempt to distinguish inherited words with fossilised suffixes from recent borrowings, and I show that *-i and *-an are only fossilised in a handful of very particular instances. In §8.4.2, I examine the loss of other affixes from PM to NEPMs, proposing that the reduction can be explained as the result of internal phonological changes. In §8.4.3, I consider possible external causation for the morphological reduction in NEPMs, and I conclude that given the lack of evidence, contact-induced change and substratal interference cannot be convincingly established.

8.4.1 Fossilisation of PM *-i and *-an

The fossilisation of PM *-i 'APPL', *-an$_1$ 'DISTR' and *-an$_2$ 'NMLS' in NEPMs is illustrated in (19) to (21). *-i is regularly reflected as *-i* in KM and CTM, and as *-ɛi* in ITM. *-an is reflected as *-ɛ* in KM, and as *-aŋ* in CTM and ITM. These words are considered as having fossilised suffixes because the putative bases

[81] Ultimately from Arabic *fikr*, cf. SM *pikir~fikir* 'to think'.

are not attested independently; compare the putative bases to the left of the "|" with inherited bases listed on the rightmost columns. Furthermore, as evident from these examples, NEPMs have a similar set of words in which earlier suffixes have become fossilised.

(19) PM > KM

*baik-i	(good-APPL)	>	*bɛʔk|i*	'to repair'	cf. *baiʔ*
⁺main-an$_2$	(play-NMLS)	>	*mɛn|ɛ*	'game, toy'	cf. *maiŋ*
*buat-an$_2$	(do-NMLS)	>	*bɔʔt|ɛ*	'action'	cf. *buwaʔ*
*manis-an$_2$	(sweet-NMLS)	>	*nnis|ɛ*	'palm sugar'	cf. *manih*
*duri-an$_1$	(thorn-DISTR)	>	*duɣ|ɛ*	'durian'	cf. *duɣi*
⁺kasi(h)-an$_1$	(love-DISTR)	>	*ssɨy|ɛ*	'pitiful'	cf. *kaseh*
*rambut-an$_1$	(hair-DISTR)	>	*mɔʔt|ɛ*	'rambutan'	cf. *ɣamboʔ*

(20) PM > CTM

*baik-i	(good-APPL)	>	*bɛʔk|i*	'to repair'	cf. *baiʔ*
⁺main-an$_2$	(play-NMLS)	>	*mɛn|aŋ*	'game, toy'	cf. *maiŋ*
*manis-an$_2$	(sweet-NMLS)	>	*nnis|ɛ*	'palm sugar'	cf. *manih*
*duri-an$_1$	(thorn-DISTR)	>	*dɨy|aŋ*	'durian'	cf. *duɣi*
⁺kasi(h)-an$_1$	(love-DISTR)	>	*ssɨy|aŋ*	'pitiful'	cf. *kaseh*
*rambut-an$_1$	(hair-DISTR)	>	*mɔʔt|aŋ*	'rambutan'	cf. *ɣambuʔ*

(21) PM > ITM

*baik-i	(good-APPL)	>	*biʔk|ɛi*	'to repair'	cf. *baiʔ*
⁺main-an$_2$	(play-NMLS)	>	*main|aŋ*	'game, toy'	cf. *maiŋ*
*manis-an$_2$	(sweet-NMLS)	>	*manis|aŋ* ~*nnis|aŋ*	'palm sugar'	cf. *manih*
*duri-an$_1$	(thorn-DISTR)	>	*duɣ|aŋ*	'durian'	cf. *duɣɛi*
⁺kasi(h)-an$_1$	(love-DISTR)	>	*sɨy|aŋ*	'pitiful'	cf. *kasɛih*[82]
*rambut-an$_1$	(hair-DISTR)	>	*ambut|aŋ*	'rambutan'	cf. *ɣambuʔ*

Let us first consider the KM examples in (19). The majority of these examples demonstrate a phonological history that parallels that of morphologically simple words with similar shapes, following regular sound changes that reduced PM trisyllables to disyllables. In the first three examples, the vowel sequences *-ai- and *-ua- in *b**ai**k-i, ⁺m**ai**n-an and *b**ua**t-an were contracted

[82] This form is uncommon. The more common word for 'to love' is *bɣahɛi*, but compare ITM *kkasɛih* 'lover' with SM *kəkasih*.

to *ɛ* and *ɔ* respectively, triggering syllable reduction (§7.5.2.1). The accretion of *ʔ* in *bɛʔk|i* 'to repair' and *bɔʔt|ɛ* 'action' is unexpected. It could be that PM *baik-i was pronounced with an epenthetic glottal stop with the suffixation of *-i, i.e., [baiʔki], which is retained in KM *bɛʔki*. *duri-an$_1$ > *duy|ɛ* 'durian' presumably has the following history, whereby the syllable reduction was realised by $^{+}$-ər- > $^{+}$u (see §7.3.4):

(22) *duri-an$_1$ > $^{+}$dərian (antepenultimate schwa neutralisation)
> $^{+}$duian ($^{+}$-ər- > $^{+}$u)
> $^{+}$duyan (reanalysis of $^{+}$i > $^{+}$y)
> *duy|ɛ* ($^{+}$-an > *ɛ*)

The sound changes reflected in *manis-an$_2$ > *nnis|ɛ* 'palm sugar' and $^{+}$kasi(h)-an$_1$ > *ssiy|ɛ* 'be pitiful' are also regular, involving antepenultimate vowel syncope (> $^{+}$mnisan, $^{+}$ksi(h)an) and subsequent cluster assimilation ($^{+}$mn- > *nn-*, $^{+}$ks- > *ss-*). The exact path from *rambut-an$_1$ > *mɔʔt|ɛ* 'rambutan' is less clear. There was probably an intermediate stage of $^{+}$maut|an which directly gave rise to *mɔʔtɛ* (parallel to *buat-an$_2$ > *bɔʔt|ɛ* 'action'), but how *rambut-an$_1$ developed into $^{+}$maut|an remains obscure.

The histories of corresponding CTM forms in (20) are largely comparable, except that the penultimate *i* in *diyaŋ* 'durian' is unexplained. ITM forms in (21) also typically reveal similar histories reflecting the reduction of trisyllables to disyllables, but there are more irregularities. ITM *biʔk|ɛi* 'to repair', *nnis|aŋ* 'palm sugar', *duy|aŋ* 'durian' and *siy|aŋ* 'pitiful' presumably have the same histories as their KM and CTM cognates, but *siy|aŋ* 'pitiful' appears to have undergone further reduction of $^{+}$ss- > *s-*, and *nnis|aŋ* 'palm sugar' has a trisyllabic variant *manis|aŋ*. The other two examples *main|aŋ* 'game, toy' and *ambut|aŋ* 'rambutan' also retain their trisyllabic shapes without undergoing syllable reduction.

In addition to the examples presented above, NEPMs have a number of trisyllabic words with *-ɛ* or *-aŋ*, corresponding to SM *-an* 'NMLS' and reflecting PM *-an$_2$ 'NMLS':

(23) KM trisyllables with *-ɛ* corresponding to SM *-an*

balas|ɛ	'response'	vs.	*balas-an*	(reply-NMLS)
pilih|ɛ	'choice'	vs.	*pilih-an*	(choose-NMLS)
pakay|ɛ	'clothes'	vs.	*pakai-an*	(wear-NMLS)
jawap|ɛ	'answer'	vs.	*jawap-an*	(answer(v.)-NMLS)
harap|ɛ	'hope'	vs.	*harap-an*	(hope(v.)-NMLS)

(24) CTM trisyllables with *-aŋ* corresponding to SM *-an*

balas|*aŋ*	'response'	vs.	*balas-an*	(reply-NMLS)
tanam|*aŋ*	'plantation'	vs.	*tanam-an*	(plant-NMLS)
ukum|*aŋ*	'penalty'	vs.	*hukum-an*	(law-NMLS)
pakay|*aŋ*	'clothes'	vs.	*pakai-an*	(wear-NMLS)
makan|*aŋ*	'food'	vs.	*makan-an*	(eat-NMLS)

(25) ITM trisyllables with *-aŋ* corresponding to SM *-an*

baŋun|*aŋ*	'building'	vs.	*baŋun-an*	(build-NMLS)
makan|*aŋ*	'food'	vs.	*makan-an*	(eat-NMLS)
ukum|*aŋ*	'penalty'	vs.	*hukum-an*	(law-NMLS)
pakay|*aŋ*	'clothes'	vs.	*pakai-an*	(wear-NMLS)
tulis|*aŋ*	'writing'	vs.	*tulis-an*	(write-NMLS)

A comparison between the examples in (23) to (25) and those in (19) to (21) raises several questions. First, are NEPM trisyllables ending in *-ɛ/-aŋ* also inherited with fossilised suffixes? If yes, why do some PM suffixed forms have disyllabic reflexes, whereas others retain a trisyllabic shape? If not, why did *-i and *-an only survive in the examples in (19) to (21), but not elsewhere? To answer these questions, I suggest that NEPM words in (23) to (25) are in fact recent loanwords from SM, marked by their unexpected trisyllabic shapes. ITM trisyllables in (21) are presumably also borrowed. PM *-i and *-an are generally lost, except in a few special cases which either showed phonological peculiarity or semantic idiosyncrasy. The following scenarios are proposed.

It is likely that PM suffixed derivatives were also affected by a process of syllable reduction, just like prefixed derivatives and trisyllabic simple words. In the case of prefixed derivatives or trisyllabic simple words, the reduced syllables were typically the antepenultimate syllables, presumably because of precedent antepenultimate schwa neutralisation. The antepenultimate syllables in suffixed derivatives, on the other hand, were likely exempted from neutralisation to schwa as they were integral parts of the bases. Syllable reduction therefore affected the suffixes, which could be considered the weakest syllables.[83] Exceptions to this general rule of syllable reductions are attested in a handful of very particular cases, in which *-i and *-an have become fossilised: either the original trisyllables had been reduced to

[83] Other factors such as stress assignment might have also been at work here.

disyllables by other means, or the suffixed forms were lexicalised in PM and treated as if they were morphologically simple. Four out of seven examples in (19) were phonologically special: *baik-i 'to repair', ⁺main-an$_2$ 'game, toy', *buat-an$_2$ 'action' and *duri-an$_1$ 'durian' were affected by vowel contraction, leading to the prior reduction of PM trisyllables to disyllables. There was therefore no further phonological motivation for the loss of suffixes. As for the other three examples *manis 'sweet', *rambut 'hair' and ⁺kasi(h) 'love', the suffixing of *-an derived complex words with rather idiosyncratic meanings ('palm sugar', 'rambutan' and 'be pitiful'). It could be that in PM they were already petrified and no longer conceived as deriving from corresponding bases by the speakers.[84] They were subsequently affected by antepenultimate schwa neutralisation, vowel syncope and cluster assimilation like morphologically simple words, e.g., *manis|an > ⁺mənisan > ⁺mnisan > *nnisɛ* 'palm sugar'.

Following this reasoning, I assume that trisyllabic words like *balasɛ* 'response' and *pilihɛ* 'choice' are not inherited, but borrowed from SM with sound adaptations.[85] This hypothesis is supported by some apparent non-native sound patterns, e.g., an initial *h* and a tap *ɾ* in *harapɛ* 'hope', which are likely direct influences of SM *harapan*. It also explains why there are only a few examples of trisyllables with *-ɛ/-aŋ* in NEPMs. Similar explanations may be applied to trisyllabic ITM forms in (21). If inherited, ⁺main-an$_2$ 'game, toy' should have been reflected as ˣmin|aŋ in ITM (reflecting *-ai- > *i*), and it is probable that *nnis|aŋ* 'palm sugar' is the inherited form, whereas its variant *manis|aŋ* is a recent loanword. ITM *ambut|aŋ* may be borrowed from SM *rambutan* with the deletion of *r-*.

8.4.2 Loss of other affixes

All other PM affixes are lost without a trace in NEPMs. Two questions are explored in this section: first, which and what kind of affixes are lost, and second, what drove the loss of affixes. I draw attention to the observation

[84] As an analogy, consider English *health* and *heal*. While *health* originally derived from *heal* and still does so analytically, it is generally not perceived as "complex" by native speakers.

[85] It could also be that they are nonce borrowings or merely instances of code-switching, but it is not possible to make a distinction between these categories with the data available so far.

that the loss of affixes followed a pattern with uniform outcomes, and I suggest that the morphological reduction had a phonological motivation.

The sixteen affixes reconstructed in PM are repeated in Table 8.2 and arranged according to their positional categories. There were seven prefixes, five suffixes and four circumfixes.

Table 8.2: Affixes reconstructed in PM (Adelaar 1984, 1992)

*(mb)Ar-	'INTR'	*-i	'APPL'	*kA- -an$_1$	'NMLS'
*pAr$_1$-	'TR'	*-aʔ	'SUBJ'	*kA- -an$_2$	'NMLS'
*tAr-	'NVOL'	*-an$_1$	'DISTR'	*pAN- -an	'NMLS'
*mAN-	'ACT; INTR'	*-an$_2$	'NMLS'	*pAr- -an	'NMLS'
*maka-	'TR.CAUS'	*-An	'NMLS'		
*pAN-	'NMLS'				
*pAr$_2$-	'NMLS'				

As described earlier, NEPMs only have reflexes of *(mb)Ar-, *pAr$_1$-, *tAr-, *mAN- and *pAN-. A striking pattern can be revealed from a closer examination of the retention and loss of affixes: prefixes tend to be retained, whereas all suffixes and circumfixes are lost, except the few instances of fossilised *-i and *-an.[86] Out of seven prefixes reconstructed to PM, only *maka- 'TR.CAUS' and *pAr$_2$- 'NMLS' are not inherited. Moreover, *pAr$_1$- 'TR' is also not inherited in ITM. The divergent histories between prefixes on the one hand and suffixes/circumfixes on the other hand suggest that the morphological reduction was not random. Furthermore, it should be emphasised again that derivatives in NEPMs, either derived synchronically or historically, have a

[86] It is worth noting that the generalisation is at odds with some previous reports. Ras (1970: 439–411) takes note of a productive nominaliser *-ɛ* in KM as a continuation of *-an$_2$ 'NMLS', and a few examples of an applicative suffix *-kɛ* which corresponds to SM *-kan* and reflects a PM preposition *akAn. In my KM corpus, however, no instances of *-kɛ* or *akɛ* are attested. Their equivalent *-kaŋ* or *akaŋ* is also not found in CTM or ITM. However, as Ras himself points out, some of these suffixed formed cited in his study might have derived from the written standard language. Abdul Hamid (1994) dedicates a whole chapter to describing various affixes in KM, including nominalising affixes *pə-*, *pə- -ɛ̃* and *kə- -ɛ̃*, which supposedly reflect PM *pAr$_2$-/*pAN-, *pAN- -an/*pAr- -an and *kA- -an$_2$. These affixes are also not attested in my data. My analysis of NEPM morphological history nevertheless accords with Collins' early documentation of ITM morphology (1983: 52–55), which only includes three prefixes, roughly transcribed as *NN-*, *bəɣ-* and *tə-*, and no suffixes.

canonical disyllabic shape (see §5.2.2, §5.3.5 and §8.4.1). Once this canonical disyllabic structure is recognised, the loss of affixes from PM to NEPMs is not hard to understand. I argue that the morphological reduction was mainly motivated by the structural pressure of disyllabisation.

There were essentially three types of derivatives in PM: PREFIX-BASE, BASE-SUFFIX and CIRCUMFIX-BASE-CIRCUMFIX. Given a disyllabic base, prefixed and suffixed forms typically had a trisyllabic shape, whereas circumfixed forms had a quadrisyllabic shape. As I detailed in §7.5 and §8.3, trisyllabic simple words and prefixed forms were reduced to disyllables due to the workings of vowel contraction or antepenultimate vowel syncope (which may be followed by cluster assimilation, but this is irrelevant here). Prefixes, which generally fell on the antepenultimate syllables, were the targets of vowel syncope. Consequently, PM prefixes which took up a full syllable became subsyllabic, consisting of consonants only and showing morphophonological alternations depending on the following consonant. For trisyllabic suffixed forms and quadrisyllabic circumfixed forms, syllable reduction apparently worked in a different way. Following the suggestions put forward in §8.4.1, I assume that it was the suffixes that were deleted in trisyllabic suffixed forms in order to reach disyllabic targets. The fossilisation of some suffixes in a specific set of words also indicates that the loss of suffixes was not a wholesale process, but phonologically conditioned. Similarly, circumfixes in quadrisyllabic forms were likely lost under the same pressure. The circumfix was scrapped as the most straightforward means to reach disyllabism, perhaps also mediated by stress assignment and prosodic prominence, as well as semantic transparency.

A tendency towards disyllabicity is not uncommon among Austronesian languages. It is well known that Austronesian roots and bases are predominantly disyllabic (Chrétien 1965). In addition, monosyllables or derived trisyllables in many Austronesian languages exhibit a tendency to restore and maintain disyllabicity through various processes (Blust 2007, 2013: 682–686; Himmelmann 2005: 116). While disyllabicity is observed as a unified outcome along parallel paths throughout the family, the motivations behind this tendency are not always clear. In the case of NEPMs, I have shown that the tendency towards disyllabicity has a phonological basis, at least partially. The reduction of trisyllables to disyllables was realised by vowel contraction or vowel syncope, with the latter likely being preceded by neutralisation to schwa – all of which are common and natural sound

changes. The history of retained affixes (prefixes and fossilised suffixes) aligns well with the general direction of phonological evolution, indicating that affixes did not simply shear away. Morphological reduction went hand in hand with phonological erosion, and presumably was driven by it.

A few remarks can be made about the loss of two PM prefixes *maka- 'TR.CAUS' and *pAr$_2$- 'NMLS'. Unlike other monosyllabic prefixes, *maka- took up two syllables, which would have formed quadrisyllabic derivatives. It is therefore unsurprising that its development followed a different path. In any case, reflexes of *maka- are rare, which seem to be only found in Kendayan varieties and Old Malay (Adelaar 1992: 165). The loss of *pAr$_2$- might be due to its competition with *pAN- 'NMLS' in forming nouns. Both *pAr$_2$- and *pAN- had a similar function, and they were originally in a paradigmatic relationship with corresponding verbal derivations: *pAr$_2$- formed deverbal nouns on verbal bases that had *(mb)Ar- 'INTR' or *pAr$_1$- 'TR', whereas *pAN- formed deverbal nouns on the basis of other verbs. It is likely that this paradigmatic relationship was eroded (as in many Malayic varieties), and only one nominaliser *pAN- is retained in NEPMs.

The foregoing proposition admittedly has some weaknesses, and several questions are left unanswered. First, the loss of *pAr$_1$- 'TR' in ITM is unexplained from a phonological perspective; there is no clear reason why this prefix is not inherited, as it is in KM and CTM. Second, if phonological changes are taken as the internal driving force behind the loss of affixes, it still needs to be explained why such changes and subsequent morphological reduction took place in NEPMs, but not more recurrently in other Malayic and Austronesian languages. It is reasonable to speculate that phonological evolution alone might not fully account for the observed morphological reduction, and there might be some external causation at play, which I will examine below.

8.4.3 Contact-induced change?

The morphological reduction in the history of NEPMs has sometimes been ascribed to contact-induced change, a substratal influence in particular. A preliminary version of this idea was first put forward by Winstedt (1923: 96), who suggested that "it is possible that aboriginal, Mon and Siamese influences have clipped and shaped the speech of these States [referring to north states including Kelantan, Kedah, Pinang and Perak]." A few decades later,

Benjamin (1987) was outspoken in arguing for such a prehistory for KM specifically. In the article titled *Ethnohistorical perspectives on Kelantan's prehistory*, he makes several claims regarding the linguistic history of KM, as summarised below:

1) the population of the Isthmian parts of the Malay Peninsula (including Kelantan) was Mon-speaking prior to the arrival of Malay;
2) the local population shifted from Mon-speaking to Malay-speaking in a "replacement-from-above" manner, as evidenced by the homogeneity of modern KM;
3) the linguistic shift presumably took place at some time around the twelfth century AD when the area was under the control of the ancient kingdom known as Tambralinga. The linguistic shift was the result of Tambralinga's submission to the Srivijaya empire.

In short, Benjamin (1987: 126–127) contends that "northern Malay [KM and Kedah Malay] might well repay investigations as being Malay spoken with a Mon accent". The claims above are, to a large extent, the corollaries of archaeological and historical data. There is archaeological evidence indicating the existence of Mon(-Khmer) kingdoms dating back to the sixth to the thirteenth century, excavated in present-day Sathing Phra (southern Thailand), which is the probable location of Tambralinga (Stargardt 1983: 32, also see §1.4.1.2). Moreover, several inscriptions found in the vicinity, the latest of which dates back to the thirteenth century, are purportedly written in Mon. While no concrete linguistic data are presented to sustain these hypotheses (apparently because there was little available at that time), Benjamin draws attention to the observation that the absence of suffixes in KM agrees with the pattern in Mon-Khmer languages. Similar ideas are also alluded to in Benjamin (1997: 85). Interestingly, Benjamin (1987: 129) suggests that KM and CTM/ITM have divergent (pre)histories. He emphasises that Terengganu falls within the territory of traditional "Malay world", whereas Kelantan lies beyond it to the north; accordingly, the Malay varieties spoken in Terengganu reflect a more "normal" uninterrupted evolution. The inland variety of Terengganu is assumed to be a "Low" variety of Malay that developed *in situ*, and the coastal variety is a direct offshoot of court Malay.

Generally speaking, it is not unreasonable to presume that NEPMs were in contact with non-Malayic languages for at least parts of their histories.

There is general consensus that the prehistorical homeland of Malayic languages is in West Borneo, and PM began to disperse approximately 2,000–2,500 years ago. The coastal distribution of the Malayic varieties on the Malay Peninsula indicates that their settlements have a shorter history than those in Sumatra or Borneo (see §1.2). The inland areas of the peninsula, on the other hand, still host the Aslian languages whose speakers must have settled much earlier, probably some 4,000 years ago (Diffloth 2005; Benjamin 2012; Dunn et al. 2013). The presence of non-Malayic Austronesian loanwords in Aslian languages also suggests that there were likely pre-Malayic Austronesian languages on the peninsula which are now extinct (Skeat & Blagden 1906: 435–438; Blench 2006). Archaeological and historical evidence, as summarised in §1.4.1.2, also shows that the Isthmian parts of the Malay Peninsula must have witnessed a number of Mon-Khmer civilisations in the first millennium, whose influences probably persisted until the arrivals of the Malay from the south and the Thai from the north. All these are grounds for assuming that there are several layers of Austro-Asiatic (AA, including Aslian, Mon and Khmer), pre-Malayic Austronesian and Malayic presence on the peninsula, leading to inevitably complex contact histories between languages from these different layers.

It should be noted, however, that Benjamin's inferences about the linguistic history of KM were made in the 1980s, and upon reexamination with our current knowledge, several imprecise interpretations have come to light. While historians and archaeologists generally agree on the location of Tambralinga being around contemporary Nakhon Si Thammarat, with Sathing Phra being one of its most important trade centres (Wolters 1958; Wales 1974; Welch & McNeill 1989; Jacq-Hergoualc'h 2002), there is no evidence indicating that Kelantan was part of the same political regime. More importantly, the presence of Mon inscriptions in the region appears to be misinformation (Bauer 1992). The inscriptions referred to by Benjamin are written either in Old Khmer (the Grahi inscription, 1183 AD, Chaiya) or Sanskrit in Old Khmer script (1230 AD, Nakhon Si Thammarat) (Jacq-Hergoualc'h 2002: 421–425). On the whole, while it is not unlikely that Kelantan has an early history associated with Mon-Khmer population and cultures, solid evidence supporting this inference is still lacking. Also, Benjamin's suggestion regarding KM having a divergent history in contrast to CTM and ITM does not hold. As I have shown, the typological profiles and morphological histories of NEPMs are largely comparable, and there

is no indication that KM underwent linguistic shift whereas CTM and ITM reflect regular uninterrupted evolution.

The question now is, do NEPMs show traces of contact-induced change in their structures? Can morphological reduction be attributed to substratal influences or early language shift? Giving satisfactory answers to these questions requires a comprehensive inspection of the grammars of NEPMs, Aslian languages, (Old) Mon and Khmer, and I can only scrape the surface of these issues here. By briefly examining the manifestation of possible outcomes of contact-induced change in NEPMs, I suggest that language contact might have played a role in the evolution of NEPMs, but as it stands, there is not much evidence speaking in favour of it. The difficulty in attributing morphological reduction in NEPMs to external causation is twofold. On the one hand, there is no clearly identifiable source language driving this change. On the other hand, there is no apparent structural interference from a potential substrate language in other aspects of the grammar.

If the linguistic histories of NEPMs did involve contact-induced change, it can be inferred that the speech communities existing before the Malayic expansion must have eventually shifted their original language(s) to NEPMs, as present-day NEPM speakers are not bi/multilingual in any local non-Austronesian language. The morphological reduction of NEPMs may be seen as a form of simplification, which suggests a scenario of adult language shift involving imperfect second language acquisition (Thomason & Kaufman 1988; Thomason 2001a,b, 2010; Trudgill 2010; Ross 2013). There might have been an abrupt linguistic shift from the local (AA?) languages to the incoming Malayic varieties. The speech communities could have failed to acquire the suffixes and circumfixes in Malayic as these categories are absent in their first languages, hence generating new morphologically-reduced Malayic varieties. However, the observation of NEPMs having undergone morphological reduction alone does not make a solid case of contact-induced change. If there was contact, there should be at least some other indications of interference in either lexicon, phonology, morphology or syntax. These indications are hard to find.

Let us begin by examining lexical evidence. A small number of AA loanwords in northern Peninsular Malayic varieties have been cited in the literature. Benjamin (1987: 133) draws attention to Kedah Malay *bəndaŋ* 'paddy field', which is assumed to be a loanword from (old) Mon *bnaŋ* 'unit of paddy land'. KM *bəndɛ* 'paddy field' is an apparent cognate, with final *-ɛ* reflecting

earlier *-aŋ* (but cf. CTM *sawɔh padi* and ITM *umɛ*). A few other toponyms are believed to have Mon or Khmer etyma, including Sungai *Lebir* in Kelantan, deriving from Old Mon *lbir* 'sea; river' (Benjamin 1987: 139). Andaya (2001: 319) mentions another word *glong* meaning 'irrigation canals' in a northern Malay dialect (Kedah Malay?), suggesting that these specific cultural terms "may indicate the Melayu on the Peninsula learned wet rice cultivation techniques from the early Mon population in the area". The significance of these few reported loanwords is difficult to evaluate, but it is worth pointing out that overall, the lexicon of NEPMs is overwhelmingly Malayic.[87] An examination of the etyma of 260 words in an extended Swadesh list (see appendix A) reveals that almost all basic vocabularies in NEPMs have cognates in SM. I found only the following words in NEPMs that do not appear to have an apparent cognate in SM, as listed in Table 8.3.

Table 8.3: NEPM basic lexical items without SM cognates

KM	CTM	ITM	SM	Gloss
kəkɔh	*xxəkɔh*	*kəkɔh*	*gigit*	'to bite'
plaka	–	*litɔ*	*guruh*	'thunder'
kɛcɛʔ	–	–	*cakap*	'to say'
–	*bahaŋ*	*bahaŋ*	*pukul*	'to hit'
–	–	*mikɛ*	*kamu*	'2PL'
dɛmɔ	–	*dimɛ*	*mərɛka*	'3PL'

It is worth noting that the inferred cognates of some words in Table 8.3 are actually included in the Malay dictionary *Kamus Dewan* (Sheikh Othman 2007), e.g., *kəkah* 'to bite' (> KM/ITM *kəkɔh*), *kərəkah* 'to bite' (> CTM *xxəkɔh*), *pəlakar* 'thunder' (> KM *plaka*), *lintar* 'thunder' (> ITM *litɔ*), *kecek* 'to

[87] Here, a word being considered Malayic implies that it has cognates available in SM and/or other Malayic varieties outside the peninsula. Some probable AA loanwords have been noted for SM, such as *kətam* 'crab', *həlaŋ* 'eagle', *səmut* 'ant' and *cucu* 'grandchild' (Benjamin 2012: 152). They typically have cognates in NEPMs, i.e., KM *səmoʔ*, ITM *səmuʔ* 'ant', KM/CTM *cucu*, ITM *cucəʊ* 'grandchild', and their cognates are also widespread in other Malayic languages. The wide distribution of these cognate sets suggests that the contact between AA and Malayic may be of great antiquity. It is an important observation, but not one that is directly relevant for the scenario of a more recent linguistic shift from AA to Malayic, as discussed in this context.

say' (> KM *kɛcɛʔ*) and *dɛma* '3PL' (> KM *dɛmɔ* and ITM *dimɛ*). However, they are marked as dialect-specific, so I assume they are not genuine cognates in SM. On the other hand, some other NEPM words do not have cognates that are commonly used in SM, e.g., KM/CTM *kɔpɛʔ*, ITM *pɛiʔ* 'breast' (cf. SM *buah dada*) and CTM *gɔɲɔh* 'to rub' (cf. SM *gɔsɔk*), yet corresponding forms like *kɔpɛk* and *gɔɲɔh*) are included in *Kamus Dewan* and not marked as dialect-specific. In these cases, I assume that cognates are available in SM. Altogether, it can be concluded that the lexical compositions of NEPMs are almost purely Malayic (in its broad sense). This stands in contrast with other proposed scenarios of linguistic shift to Malayic, Jakun being a case in point, which includes a fair number of Aslian words in the basic vocabulary (Skeat & Blagden 1906; Seidlitz 2005; Anderbeck 2012).

Second, the phonological patterns of NEPMs, including their phoneme inventories and the general preference for disyllabicity, are typical of Malayic languages. While some drastic sound changes have taken place, no clear foreign segments or sound patterns have been added to their phonologies. This can be compared with Urak Lawoi', a Malayic language spoken off the coast of southern Thailand, which displays more evident contact-induced interference in its phonology. Urak Lawoi' has undergone final denasalisation: compare Urak Lawoi' *kiri**p*** 'to send', *turo**t*** 'to descend', *bita**k*** 'star' with SM cognates *kiri**m***, *turu**n*** and *bintaŋ*. It is likely that the denasalisation took place via an intermediate stage of nasal preplosion (i.e., *-m* > *-^{p}m* > *p*, *-n* > *-^{t}n* > *-t* and *-ŋ* > *-^{k}ŋ* > *-k*), a cross-linguistically unusual sound pattern that is commonly found in Aslian languages (and further afield in various Bornean languages) (Adelaar 1995: 87–89; Blust 1997: 154–169). Urak Lawoi' has also developed a set of aspirated stops, presumably resulting from more recent contact with Thai (Hogan 1988: 15). One aspect in the phonologies of NEPMs that might be indicative of foreign influences is the genesis of contrastive vowel nasality, which is a common feature in Aslian languages (Matisoff 2003: 14–15; Benjamin 2012: 179; Kruspe et al. 2015: 424–425). However, none of the words with phonemic nasal vowels seems to have an AA origin. It also remains unclear whether vowel nasality arose after (the ancestors of) NEPMs came in contact with AA languages, as it can be largely explained as the result of internal sound changes (§7.4.3).

Moving towards morphology, I have shown that all affixes in NEPMs are inherited from PM; none is borrowed from another (unknown) source. More importantly, as pointed out earlier in §8.4.1, PM affixes are not

stripped entirely: prefixes are typically retained, and some suffixes have become fossilised under certain circumstances. The regularities reflected in morphological reduction and the interconnection between phonological and morphological changes suggest that the developments are internally-motivated, rather than driven by a general process of simplification in imperfect second-language acquisition. Alternatively, one may suggest that the morphological reduction in NEPMs arose from the convergence of the Malayic morphology with an AA pattern where suffixing and circumfixing are lacking. However, given the lack of borrowing of lexical and grammatical materials, it is unlikely that there was intense contact which could have led to the convergence of morphological patterns.

Lastly, while I have not yet been able to closely examine the possible interference in syntactic patterns, NEPMs do not seem to exhibit strikingly non-Malayic syntactic features that might be attributed to substratal influences. Considering the overall similarity in syntactic structures between AA and Malayic languages, diagnosing possible syntactic transfer might prove to be a challenging endeavour.

To sum up, given the absence of an identifiable contact language(s) and clear traces of contact in other aspects of the grammar, there is a lack of concrete evidence for contact-induced change. While it is impossible to rule out the possibility of contact, and further investigation might uncover more evidence demonstrating substrate influences, based on the data available at present, I take an agnostic stand and conclude that all three varieties of NEPMs reflect a rather "normal" evolution from PM.

8.5 Summary

This chapter has investigated the morphological history of NEPMs as developed from PM. Three primary aspects of this evolution have been examined: the retention of PM prefixes, the innovation of initial gemination and the overall tendency of morphological reduction.

All affixes in NEPMs are retentions from PM. The prefixes *bɣ-* 'INTR; MID', *pɣ-* 'CAUS; FCT' and *tɣ-* 'NVOL' are reflexes of PM *(mb)Ar- 'INTR', *pAr$_1$- 'TR' and *tAr- 'NVOL' respectively. *NN$_1$-* 'IPFV' and *NN$_2$-* 'NMLS' developed from *mAN- 'AGT; INTR' and *pAN- 'NMLS'. Not only are the functions of these prefixes broadly retained, their formal evolution also fits well into the general

phonological history. Since PM prefixes typically fell on the antepenultimate syllables which were affected by vowel syncope and subsequent cluster assimilation, their phonological shapes have been reduced.

Following regular sound changes, PM prefixes are sometimes reflected as a segment identical to the base-initial consonant, creating morphologically complex geminates. While these geminates can often be seen as allomorphic alternations of underlying prefixes, numerous pairs of a base and a derivative contrasting an initial singleton with a geminate at the surface level have led to the reinterpretation of how complex geminates are derived. I have proposed that the process of initial gemination has become associated with realising certain grammatical functions in a more transparent manner, leading to its extension as a morphophonological operation more generally as a result of analogy. The evolution from prefixing to initial gemination in the history of NEPMs can be seen as a prime exemplification of how non-concatenative morphology such as the manipulation of consonant length can arise from the concatenation of morphemes.

Lastly, it is evident that NEPMs have undergone significant morphological reduction. I highlighted that the morphological reduction was neither random nor a wholesale process. PM prefixes, for the most part, are generally retained, whereas all suffixes and circumfixes are lost or fossilised. I suggested that the morphological reduction was primarily driven by internal phonological motivations, mainly due to the structural pressure of disyllabisation. When the disyllabic targets were achieved by other phonological changes, or when the original suffixed forms were lexicalised, PM suffixes such as *-i 'APPL', *-an$_1$ 'DISTR' and *-an$_2$ 'NMLS' became fossilised in a handful of instances. The fossilisation of these suffixes further illustrates that even the loss of suffixes was phonologically conditioned. To give a fuller account of possible mechanisms behind the morphological reduction, I discussed the hypothesis of substratal influences and potential contact-induced change. The current linguistic landscape of the Malay Peninsula and relevant archaeological and historical evidence suggest a complex contact history of Peninsular Malayic varieties. However, given the lack of an identifiable contact language(s) and clear traces of contact-induced change in lexicon, phonology and morphology, I argued that the hypothesis of substratal influences driving morphological reduction does not find favour with linguistic data.