Malayic varieties of Kelantan and Terengganu: description and linguistic history

Wu, J.

CHAPTER 5

Morphology

5.1 Introduction

This chapter presents a comparative description of the morphology of NEPMs, on account of the considerable similarities observed in their morphological systems. It is organised into two parts.

Part one (§5.2) outlines the defining criteria and characteristics of basic morphological units such as words, affixes, bases, roots and clitics in NEPMs, aiming at providing the reader with a proper understanding of the building blocks of the morphological systems. It begins with a discussion of wordhood in §5.2.1, followed by an examination of the internal structure of words in §5.2.2. Clitics, which share properties of both affixes and words, are discussed in §5.2.3.

Part two (§5.3) delves into how basic morphological units combine to form complex words. It covers various word-formation processes, including prefixation (§5.3.1), initial gemination (§5.3.2), compounding (§5.3.3) and reduplication (§5.3.4). Some fossilised complex words are addressed in §5.3.5.

Finally, §5.4 provides a summary of this chapter.

5.2 Morphological units

5.2.1 Words

While the notion of "word" is often assumed in morphosyntactic descriptions, discussions on wordhood are seldom found for Malayic varieties, with exceptions such as Gil (2020) for Riau Indonesian and McDonnell (2016) for Besemah. Following Dixon & Aikhenvald (2002) and Aikhenvald et al. (2020), I propose that it is possible to distinguish phonological words from grammatical words in NEPMs, whereby the former category is identified based on phonological criteria, and the latter on morphosyntactic criteria.

Relevant properties for identifying phonological words (or prosodic words) in NEPMs can be drawn from segmental features and phonological rules. As described in §2.4, §3.4 and §4.4, a set of phonotactic constraints applies at a level that can be considered as phonological wordhood. There are constraints on the number of consonants permitted in an initial cluster, permissible segments at both edges of a word, and the distribution of vowels within a word. Taking KM as an example, the following diagnostic criteria can be used to identify the boundaries of phonological words:

1) If a string of utterances has three consecutive consonants, there must be a word boundary between the first consonant and the following two consonants. For example, in *tumboʔ lluaʔ* (pound crush) 'pound to crush', the string of three consonants /-ʔll-/ has a word boundary between /ʔ/ and /ll/.
2) Geminate clusters indicate a word boundary to the left, as they only occur word-initially. In the same example of *tumboʔ lluaʔ* 'pound to crush', the word boundary is also signalled by /ll-/.
3) A coda /h/ signals a word boundary to the right. In *ayɔh makɛ* (father eat) 'father eats', the coda /h/ in *ayɔh* 'father' indicates the right edge of a phonological word.
4) A nucleus /ə/ signals a word boundary before the syllable in which it occurs. In *moŋ kənɔ* (2SG must) 'you must', the schwa in *kənɔ* 'must' indicates the word boundary to the left.
5) The mid-high vowels /e, o/ signal a word boundary following the syllable in which they occur. In the same example of *moŋ kənɔ* 'you must', /o/ in *moŋ* '2SG' indicates the word boundary to the right.

These criteria are applicable in the native lexicon, but not necessarily in loanwords, toponyms or person names. For example, *klene?* 'clinic' circumvents criterion 5), and *ehsɛ* 'Ihsan' (a person name) circumvents criteria 3) and 5).

Phonological words can be further identified as the units in which phonological processes such as vowel nasalisation and nasal spreading take place. Nasal onsets nasalise following vowels, and the nasality spreads across glides and glottals, affecting vowels in the subsequent syllables within the same phonological words (see §2.2.2.1, §3.2.2.1 and §4.2.2.1). Across word boundaries, however, nasal spreading is blocked. As shown in the CTM example in (1), *nnawɔ?* [nnãwɔ̃ʔ] 'to lie' attests nasal spreading, but *wwapə* [wwapə] 'how many, how much' following *ɔmɔ* [ɔmɔ̃] 'age' is not affected by nasalisation, indicating that it constitutes a separate phonological word.

(1) CTM

nnawɔ?	[nnãwɔ̃ʔ]	'to lie'
ɔmɔ wwapə	[ɔmɔ̃ wwapə]	'how old' (age how.much)

Furthermore, phonological words are prosodically independent in the sense that they can be preceded and followed by pauses or intonation breaks. There are typically no such pauses in the middle of a phonological word. Most phonological words also have the ability to stand freely, e.g., as an answer to questions. This is, however, not a necessary criterion. Function words such as prepositions like KM *dəŋɛ* 'with; and' and *kalu* 'if; TOP' do not occur in complete isolation, but they fit all other criteria of a phonological word.

Other prosodic features such as stress assignments are not clearly applicable to NEPMs. There also does not seem to be a minimality constraint for phonological words. While phonological words in NEPMs are typically disyllabic, monosyllabic structures are attested for both content words and function words, as summarised in Table 5.1 for KM. Even words with a monomoraic CV shape such as *ni* 'DEM.PROX' and *tu* 'DEM.DIST' can be uttered in isolation as single-word answers to questions, therefore qualifying as phonological words.[29]

[29] Bimoraic word minimality appears to hold for surface phonological words, but this constraint only seems applicable to content words. There are two observations that sug-

Table 5.1: Examples of phonological words in KM

Word shape	Word type			
	Content word		Function word	
Disyllabic	*makɛ*	'to eat'	*padɔ*	'from; at'
	budɔʔ	'kid'	*dəŋɛ*	'with; and'
	tbuwɛ	'hornet'	*ləpah*	'after; then'
	ppalɔ	'head'	*kalu*	'if; TOP'
Monosyllabic	*jɛ*	'hour'	*ni*	'DEM.PROX'
	cɔʔ	'hoe'	*tu*	'DEM.DIST'
	ɣɔʔ	'bush'	*hɔ̃*	'AFF'
	nnɛ	'six'	*dɔʔ*	'NEG'
	mmah	'gold'	*moŋ*	'2SG'

Grammatical words (or morphosyntactic words, syntactic words) are defined as a number of grammatical elements which always occur together in a fixed order and have conventionalised coherence and meaning (Dixon & Aikhenvald 2002: 19, 35). They can be moved, replaced or deleted by syntactic operations, and they are the smallest units on which syntactic rules can apply (Kroeger 2005: 318; Haspelmath & Sims 2010: 203). In most cases, grammatical words coincide with phonological words in NEPMs: all examples in Table 5.1 are both phonological words and grammatical words. Nevertheless, there are some instances where these two types of words do not match.

---

gest the requirement of this word minimality. First, when pronounced in isolation, content words with an underlying CV(C) shape are almost always accented with an initial geminate at the phonetic level, e.g., /jɛ/ → [jjɛ] 'hour', /la/ → [lla] 'sheet'. Second, when the numeral clitic *s=* 'one' is attached to these CV(C) words, no gemination is found; instead, an epenthetic schwa is inserted between the clitic *s=* and the following consonant, as in *s=jɛ* /sjɛ/ → [səjɛ] 'one hour', *s=la* /sla/ → [səla] 'one sheet'. The second observation also supports analysing the underlying forms of *jɛ* 'hour' and *la* 'sheet' as having initial singletons rather than geminates. Two claims can be made to explain these observations: first, initial geminates are moraic, but initial non-geminate clusters are not; second, surface content words need to respect the bimoraic requirement of word minimality. For words with a subminimal monomoraic CV(C) shape, the licit minimal word status of the surface is guaranteed by initial gemination. For proclitic + host groups with a $C_xC_yV(C)$ shape (which are recursive phonological words, see §5.2.3 below), the augmentation is achieved by schwa epenthesis.

On the one hand, some grammatical words consist of two independent phonological words, as in cases of full reduplication and compounds, e.g., CTM *kkatɔʔ-kkatɔʔ* (RDP-frog) 'frogs' and *buɣoŋ-atu* (bird-ghost) 'owl'. Their status as single grammatical words is justified by their non-permutability and conventionalised meanings: they cannot be interrupted by other material while keeping their semantics intact, and compounds like *buɣoŋ-atu* 'owl' have idiomatic meanings that cannot be entirely determined from their constituents, which differentiates them from noun-noun juxtaposition phrases. Furthermore, grammatical words as such display morphological cohesion by undergoing derivation as one morphological unit, as illustrated by the KM example *t-*[*kəleh-kəleh*] (NVOL-RDP-look) 'to look casually' in (2).

(2) KM
diyɔ doʔ ***t-kəleh-kəleh*** *kɔ tɯwɛ tu.*
3 PROG NVOL-RDP-look to owner DEM.DIST
'He was peeping at the owner.' (KM_180814_n01_20)

On the other hand, some grammatical words can be reduced to become phonologically dependent, thus coalescing with neighbouring phonological words. For instance, prepositions like KM *di* 'LOC' and *kɔ* 'to; AGT' can be reduced to single-segment grammatical words *d=* and *k=* respectively, as illustrated in (3).

(3) KM

a.

di skɔlɔh	(LOC school)	'at school'
di tuboh	(LOC body)	'on the body'
kɔ moŋ	(to 2SG)	'to you'
kɔ jiɣɛ	(AGT neighbour)	'by the neighbour'

b.

datah	*d=atah*	(LOC=top)	'on top'
dumɔh	*d=<ɣ>umɔh*	(LOC=house)	'in the house'
kaku	*k=aku*	(to=1SG)	'to me'
kɔɣɛ	*k=ɔɣɛ*	(AGT=person)	'by someone'

In terms of syntactic constructions, *d=* 'LOC' and *k=* 'to; AGT' in (3b) serve as heads of the prepositional phrases, occupying the same positions as their full forms in (3a). Yet, phonologically, *d=* and *k=* form an inseparable unit with the following words (which become the hosts), i.e., *d=atah* [da.tah] 'on top' and *k=aku* [ka.ku] 'to me'. Cases like these resemble the classic instances

of simple clitics (cf. English *'s* and *is*, see Zwicky 1977). For more discussions on clitics, see §5.2.3.

5.2.2 Internal structure of words

Having established wordhood in NEPMs, this section examines the internal structure of words. The usage of the term "words" from now on generally refers to grammatical words, unless otherwise specified.

Depending on whether a word can be segmented into smaller morphemes, a distinction can be made between simple words and complex words. Simple words are free morphemes on their own, and complex words are composed of two or more morphemes, often with an affix attached to a base. This type of complex words is referred to as "derivatives".[30] For example, KM *t-kəjuʔ* (NVOL-startle) 'to be startled' and *t-kəleh-kəleh* (NVOL-RDP-look) 'to peep' have a prefix *t-* 'NVOL' marking non-volitionality (an allomorph of *tɣ-*, see §5.3.1.3), attached to the bases *kəjuʔ* 'to startle' and *kəleh-kəleh* 'RDP-look' respectively. When the base itself is a morpheme, it is also a root; that is, *kəjuʔ* 'to startle' in *t-kəjuʔ* is both a base and a root, whereas *kəleh-kəleh* 'RDP-look' is a base containing two roots. Examples like *t-kəleh-kəleh* 'to peep' are nevertheless rare; thus bases and roots are equivalent in most cases. In addition to derivatives, complex words may also consist of multiple roots, as in full reduplication and compounds.

Example sentences from each NEPM variety are given in (4) to (6). Various types of words can be observed in these examples. There are derivatives such as KM *t-kəjuʔ* (NVOL-startle) 'to be startled' in (4), CTM *ɲŋ-<s>alɔʔ* (IPFV-bark) 'barking' in (5) and ITM *j-jalaŋ* (INTR-road) 'to walk' in (6). Compounds are also present, such as KM *buɣoŋ-atu* (bird-ghost) 'owl' in (4), and full reduplication is seen in ITM *cakaʔ-cakaʔ* (RDP-speak) 'to speak (continuously)' in (6). For most words, however, there is a one-to-one correspondence between a morpheme and a word. The internal structure of complex words is also rather simple; they are generally bimorphemic.

[30] The distinction between inflectional and derivational morphology is not clear-cut in NEPMs. Since NEPMs do not mark grammatical categories like gender, number or case, inflectional morphology is generally absent. However, some word-formation processes may be viewed as inflectional. For instance, the nasal prefix NN_1- 'IPFV' is analysed as an imperfective aspectual marker (§5.3.1.5), and full reduplication of nouns overtly expresses plurality and diversity (§5.3.4), thus showing some degree of inflectional characteristics.

(4) KM
masɔ diyɔ d=atah pɔkɔʔ tu, diyɔ t-kəjuʔ tɛŋɔʔ
when 3 LOC=top tree DEM.DIST 3 NVOL-startle see
buɣoŋ-atu, diyɔ poŋ jatoh.
bird-ghost 3 also fall
'When he was on top of the tree, he was startled seeing an owl. Then he fell.' (KM_180812_n01_16)

(5) CTM
aɲjiŋ tu təɣuh lagi, diyə ɲŋ-<s>alɔʔ agi.
dog DEM.PROX continue again 3 IPFV-bark again
'The dog kept on barking.' (CTM_181023_n02_24)

(6) ITM
budɔʔ təʊ dɔʔ cakaʔ-cakaʔ, j-jalaŋ təɣuh=jə.
kid DEM.DIST NEG RDP-speak, INTR-road directly=just
'The kids didn't say a word and just went on their way.' (ITM_180919_n01_46)

In fact, it is also common to have a whole sentence consisting of simple words only, as shown by (7). Suffice it to say, NEPMs are fairly isolating.

(7) KM
kalu tumih diyɔ tu, nɔʔ makɛ ... kalu waʔ aɣi ni,
TOP sauté 3 DEM.DIST want eat ... if make day DEM.PROX
ɛsɔʔ baɣu leh makɛ.
tomorrow only.then can eat
'The sauté she makes ... if it's made today, it can only be eaten the next day.' (KM_180820_cv03_02)

Affixes are bound morphemes that cannot stand on their own. NEPMs have a small number of affixes, all of which are prefixes, as listed in Table 5.2. The exact functions and usage of these prefixes will be discussed in more detail in §5.3.1. Additionally, some suffixes or circumfixes may be identified, but they are analysed as either fossilised or borrowed, see §5.3.5.

Table 5.2: Affixes in NEPMs

KM	CTM	ITM	Gloss
bɣ-	*bɣ-*	*bɣ-*	'INTR; MID'
tɣ-	*tɣ-*	*tɣ-*	'NVOL'
pɣ-	*pɣ-*	–	'CAUS; FCT'
NN$_1$-	*NN$_1$-*	*NN$_1$-*	'IPFV'
NN$_2$-	*NN$_2$-*	*NN$_2$-*	'NMLS'

The overwhelming preference for prefixing in NEPMs is somewhat surprising from a typological perspective. While this pattern goes against the general tendency of favouring suffixing in the world's languages (Greenberg 1957; Bybee et al. 1990; Aikhenvald 2007), it appears to be the regional norm. In addition to NEPMs, neighbouring Aslian languages also strongly favour prefixes (Kruspe et al. 2015). Even further afield, Iban (Borneo), Rejang (Sumatra) and Chamic (coastal Mainland Southeast Asia), as well as Mon-Khmer languages neighbouring Chamic, can be added to the list of languages that exclusively have prefixes (Richard 1982; Thurgood 1999).

Also worthy of note is that all prefixes in NEPMs are smaller than a full syllable, consisting of consonants only. The distinction between affixes and words is therefore also reflected clearly in their phonological properties.

Roots in NEPMs are essentially simple words, as bound roots are difficult to motivate. For instance, KM/CTM *bɣ|əti* and ITM *bɣ|ətɛi* 'to stop' may seem to have the roots -(ɣ)*əti* or -(ɣ)*ətɛi* prefixed with *b*(ɣ)- 'MID' (cf. SM *bər-hənti* 'MID-stop'). However, the putative roots are not only unattested as independent words, but they are also not attested anywhere else in the lexicon with a recurrent meaning. Therefore, there is no evidence for treating them as roots synchronically (see more discussions in §5.3.5).

On a last note, it should be emphasised that derivatives and roots/simple words are subject to the same phonotactic rules, and together they constitute the domain of phonological wordhood. This prosodic pattern has several consequences in the morphological system, one of which is manifested as the constraint on the prefixation process. As all phonological words can have maximally two consonants in the initial cluster, and all prefixes consist of consonants only, prefixes may undergo morphophonological alternations in order to respect the phonological well-formedness

in the derivatives. When prefixes like *bɣ-* 'INTR; MID' and *tɣ-* 'NVOL' are attached to C-initial roots, the liquid *ɣ* in the prefix is deleted so that the derivatives have an initial CC cluster, e.g., KM *b-laɣi* 'MID-run' and *t-baka* 'INTR-burn'. With roots that already have initial CC clusters, prefixation is generally not allowed.[31] Further details are provided in §5.3.1.1.

5.2.3 Clitics

Between affixes and words is the intermediate category of clitics. Generally speaking, clitics resemble affixes in that they lack phonological independence, but at the same time they are similar to independent words in that they show higher mobility and lower degree of host selectivity (see Zwicky & Pullum 1983; Zwicky 1985). In NEPMs, a number of elements may be conveniently labelled as clitics, characterised as word-like forms that are prosodically dependent or deficient. Three subtypes of clitics can be further distinguished on account of their heterogeneous properties, and they can be viewed as existing on a cline between affixes and full-fledged phonological words.

The first type of clitics is represented by shortened variants of prepositions such as *d=* ← *di* 'LOC' and *k=* ← *kɔ* 'to; AGT' in KM, as already shown earlier in example (3). When reduced, these prepositions are integrated with their hosts prosodically: they cannot be uttered in isolation or be interrupted by other material or pause. Cliticisation as such (optionally) occurs before vowel-initial hosts, producing single-segment proclitics *d=* and *k=*, as illustrated in (8).

(8) KM

a. *masɔ diyɔ* ***d=atah*** *pɔkɔʔ tu* ...
when 3 LOC=top tree DEM.DIST ...
'When he was on top of the tree ...' (KM_180812_n01_16.1)

b. *diyɔ ŋŋ-aja=kɛ, diyɔ tɔʔ tɛŋɔʔ* ***kɔ*** *budɔʔ, diyɔ tɛŋɔʔ* ***k=atah.***
3 IPFV-teach=TAG, 3 NEG look to kid 3 look to=top
'When she teaches, you know, she doesn't look at the kids, but looks above.' (KM_180820_cv03_110.2)

[31] Those roots may be reduplicated, or form a compound with another root.

c. *ɲɔ supɔh* ***k*=*ɔɣɛ***=*lah.*
ANAPH curse AGT=person=SFP
'It would get cursed by people.' (KM_180820_cv03_142)

Similar to KM, CTM has *di* 'LOC; AGT' and *kə* 'to', and ITM has *də* 'LOC; AGT' and *kə* 'to', which can be reduced to the clitics *d*= and *k*= respectively, as exemplified in (9) and (10).

(9) CTM

a. *gagɔh hɔ̃ yə, jɛmbeŋ yə bako, buboh atah basika,*
strong AFF 3 carry 3 basket put top bike
d*=*atah *paɣə dəpaŋ.*
LOC=top rack front
'He was strong; he carried the basket and put it on the bike, on the front rack.' (CTM_181025_n02_32)

b. *aku m-maɣɔh* ***k*=*anɔʔ*** *aku.*
1SG MID-angry to=child 1SG
'I am angry at my kid.' (CTM_220927_e02_30)

(10) ITM

a. *diyɛ tiŋuʔ lubɔŋ* ***d*=*ujəʊŋ*** *kayəʊ nuŋ.*
3SG see hole LOC=end wood there
'He saw a hole at the tip of the tree there.' (ITM_180907_n02_19)

b. *akəʊ nɔʔ caɣɛi nn-<t>uləʊŋ* ***k*=*akəʊ*** *s=uɣɔŋ.*
1SG want look.for NMLS-help to=1SG one=CLF
'I want to look for a helper for myself.' (ITM_180921_e01_31)

NEPMs also have a proclitic *s*= 'a; one; same', as in KM *s=ɔɣɛ*, CTM *s=ɔɣaŋ* and ITM *s=uɣɔŋ* 'one=person; one=CLF'. The clitic *s*= also occurs before consonant-initial bases, forming various consonant clusters with initial *s*, including the geminate cluster *ss*-, e.g., KM/CTM *s=bako* and ITM *s=baku* 'one=basket', as well as ITM *s=sikaʔ* 'one=comb (of banana)'. Historically *s*= can be seen as the reduced form of corresponding numerals for 'one' (KM *sɔ*, CTM *sə* and ITM *sɛ*), although synchronically, the cliticised forms and the free forms exhibit different properties, see §6.2.6.1.

The phonological features of proclitics like *d*=, *k*= and *s*= are extremely similar to those of affixes, for which reason I call them "affixal clitics" (a term

borrowed from Selkirk 1995, also see Anderson 2005: 46). Both affixal clitics and affixes in NEPMs consist of consonants only. More importantly, the proclitic + host group forms a recursive phonological word, and it is subject to the same phonotactic constraints as combinations of a prefix + a root.

The second type of clitics is referred to as "free clitics", represented by discourse markers such as KM *=lah* 'FOC; SFP', *=kɛ* 'TAG' and *=kɔ* 'Q'.[32] Their usage is illustrated in (11). These discourse markers express a wide range of functions, see more discussions in §6.2.12.

(11) KM

a. *pah tumih=**lah** bawɛ kitɔ iɣih, tumih tumih ...*
then sauté=FOC onion 1PL slice sauté sauté ...
'Then just sauté the onions we sliced, sauté, sauté ...'
(KM_180820_cv03_171.1)

b. *baɲɔʔ kkaɣɔ dɔh=**kɛ**ʔ*
many item already=TAG
'That's already a lot of stuff, right?' (KM_180820_cv03_256)

c. *moŋ doʔ təŋɔh pɣ-aco batu=**kɔ**ʔ*
2SG PROG middle CAUS-crushed stone=Q
'Are you crushing the stone?' (KM_180827_e01_30)

Similar to *d=* and *k=*, these discourse markers are prosodically dependent on their hosts, as no pause is possible between the host and the clitic (in these cases the host precedes the clitic). The differences between these two types of clitics is, on the one hand, reflected in their shapes, and on the other hand, reflected in the prosodic structure of the host + clitic group. Unlike affixal clitics, these free clitics take up full syllables, and the host + enclitic group does not form a phonological word.

The equivalents of KM *=lah* 'FOC; SFP', *=kɛ* 'TAG' and *=kɔ* 'Q' in CTM and ITM are *=lah*, *=kaŋ* and *=kə* and respectively, illustrated in (12) and (13). In these two varieties, the tag marker *=kaŋ* (which derives from the non-verbal negator *bukaŋ*) has a special type of usage that may be referred to as a double tag marker. As shown in (12b) and (13c), *kaŋ* 'TAG' occurs twice, both before and after the main clauses *tɣabo* 'scatter' and *paka skəʔ mulɛ* 'wore skirts back then'.

[32] The question marker *=kɔ* 'Q' needs to be distinguished from the preposition *kɔ* 'to; AGT'.

(12) CTM

a. *lalu=**lah*** *s=ɔɣaŋ budɔʔ llaki ŋə basika.*
pass.by=FOC one=CLF kid male with bike
'A boy with a bike passed by.' (CTM_181025_n02_20)

b. *ikaʔ mɔlɛʔ, **kaŋ**=tɣabo=**kaŋ**?*
hold good TAG=scatter=TAG?
'Hold well, (otherwise) it will scatter, you know?'
(CTM_220927_e02_114)

c. *budɔʔ ni anɔʔ mɔʔciʔ=**kə**?*
kid DEM.PROX child auntie=Q?
'Is this kid auntie's child?' (CTM_181029_e02_17)

(13) ITM

a. *gɛi j-jalaŋ=**lah*** *tigɛ uɣɔŋ tah.*
go INTR-road=FOC three CLF DEM.DIST
'The three guys went on walking.' (ITM_180919_n01_32)

b. *kaʔ umɔh diyɛ adɛ=**kaŋ**?*
near house 3SG EXIST=TAG
'He has them at his place, no?' (ITM_180930_cv01_15)

c. *nɛiŋ kuciʔ ... kuciʔ bajəʊ nɛh, kaŋ=paka*
DEM.PROX pocket ... pocket shirt DEM.PROX TAG=wear
skəʔ mulɛ=kaŋ, gɛi skulɔh paka skəʔ.
skirt(ENG) beginning=TAG go school wear skirt(ENG)
'This pocket ... this pocket, we used to wear skirts back then, you know? We wore skirts to school.' (ITM_220920_cv01_173)

d. *nɔʔ ɣasɛ manih=**kə** dɔʔ?*
want feel sweet=Q NEG
'Does it taste sweet or not?' (ITM_180917_cv01_62)

The last type of morphemes that may be classified as clitics are phonologically deficient words, which I refer to as "weak words". In addition to cliticised *k=*, the preposition *kɔ* in KM has another variant *kə*. It can be seen as an intermediate stage in the cliticisation process of *kɔ* → *kə* → *k=*, but its wordhood status is somewhat ambiguous. On the one hand, elements like this enjoy some prosodic autonomy: as shown in (14), *kə* can be separated from

the following word by hesitation and filler material, which sets it apart from typical clitics like *k=*. On the other hand, these weak words are not quite like phonological words in their segmental structure: *kə* does not conform to the phonotactic constraints in KM, as phonological words never end in schwas.

(14) KM
diyɔ poŋ gi ***kə*** *... ggapɔ ... təpi ... ggapɔ təpi*
3 also go to ... whatchamacallit ... edge ... whatchamacallit edge
utɛ tu.
forest DEM.DIST
'He then went to the edge of the forest.' (KM_180812_n01_10)

However, except for the deficiency in phonological shapes, the clitic status of these weak words cannot be justified on other grounds. Hence, in transcriptions I treat these morphemes as words surrounded by space.

Also in ITM, the prepositions *də* 'LOC; AGT' and *kə* 'to' are not prototypical phonological words since they end in schwa. ITM also has an anaphoric marker *ŋə* with a final schwa. Examples with these weak words in ITM are given in (15).

(15) ITM

a. *diyɛ naiʔ atah batəʊ bəsɔ təh, pəgɔŋ* ***də*** *dahaŋ kayəʊ.*
3SG go.up top stone big DEM.DIST, hold LOC branch wood
'He climbed up the big stone, grabbing the branches of a tree.'
(ITM_180907_n01_19)

b. *diyɛ wɛi buwɔh pɛ tɔh tigɛ buti* ***kə*** *budɔʔ llakɛi tigɛ*
3SG give fruit pear DEM.DIST three CLF to kid male three
uɣɔŋ təʊ.
CLF DEM.DIST
'He gave three pears to the three boys.' (ITM_180907_n01_36)

c. *diyɛ iŋaʔ nɔʔ ambɛiʔ s=buti=jə buwɔh təʊ.*
3SG think want take one=CLF=just fruit DEM.DIST
tuwaŋ ***ŋə*** *dɔʔ iɣɔ diyɛ tiŋuʔ,* ***ŋə*** *ambɛiʔ s=baku.*
owner ANAPH NEG notice 3SG look ANAPH take one=basket
'He wanted to take only one pear. But seeing that the owner didn't take notice, he just took the whole basket.'
(ITM_180919_n01_18–19)

Moreover, weakened forms of the demonstratives in ITM may fit into the category of weak words (see more on demonstratives in §6.2.5). Table 5.3 shows that the full forms of ITM demonstratives *nɛiŋ* 'DEM.PROX' and *təʊ* 'DEM.DIST' have diphthongs, but they can be weakened to *nVh~nV?* or *tVh~tV?* respectively, the V being any non-high monophthong. The factors determining the choice of *V* in *nVh~nV?* or *tVh~tV?* and the choice of the final consonant remain unclear, but *-h* appears to be more common than *-?*. These weakened demonstratives are considered phonologically deficient, as there are no other words in ITM ending in *-əh*, *-ə?*, *-ɛh* or *-ɛ?*.

Table 5.3: Demonstratives in ITM

Full forms	Weakened forms	Gloss
nɛiŋ	*nah~na?* *nɛh~nɛ?* *nɔh~nɔ?* *nəh~nə?*	'DEM.PROX'
təʊ	*tah~ta?* *tɛh~tɛ?* *tɔh~tɔ?* *təh~tə?*	'DEM.DIST'

To sum up, three broad categories of clitics or clitic-like elements can be distinguished in NEPMs, and there are three parameters in which they differ from each other, as outlined in Table 5.4. Affixal clitics and free clitics cannot be separated from their hosts with pauses, but only affixal clitics are well integrated with the hosts to form phonological words. Weak words, on the other hand, are characterised by their phonological deficiency. These three types of clitics form a continuum, with affixal clitics exhibiting most affix-like properties, and weak words behaving almost like full-fledged words.

Table 5.4: Types of clitics in NEPMs

	Possibility of pauses	Phonological integration	Phonological deficiency	Examples in KM
Affixal clitics	-	+	+	*d*= 'LOC' *k*= 'to; AGT' *s*= 'a; one; same'
Free clitics	-	-	(+)	=*lah* 'FOC; SFP' =*kɛ* 'TAG' =*kɔ* 'Q'
Weak words	+	-	+	*kə* 'to'

5.2.4 Interim summary

The previous sections have outlined the characteristics of basic morphological units in NEPMs, including words, affixes, bases and clitics. In addition to offering a more detailed description of the building blocks of the morphological systems, the foregoing examination is paramount for understanding what it means to be an isolating language, which will be a crucial theme in the discussion of the morphological history of NEPMs (Chapter 8). Traditionally, isolating languages are associated with a low morpheme per word ratio, but this definition only holds on the basis of a proper comprehension of wordhood and other bound morphemes.

The discussions above also highlighted the interplay between phonology and morphology in NEPMs. This interplay is evident from the defining criteria of phonological wordhood, the distinct phonological shapes between words and affixes, and the varied phonological proprieties of clitics. Furthermore, the examination aims to fill a gap in the general descriptive literature, where concepts like words and clitics are often taken for granted without further explanation. We now shift the focus to the examination of how affixes and simple words can combine to form complex words in NEPMs.

5.3 Word-formation

The traditional categorisation of word-formation processes includes two primary types: derivation and compounding (Aikhenvald 2007; Booij 2007; Štekauer et al. 2012). Derivation involves the use of bound morphemes or morphological processes, whereas compounding involves the combination of free morphemes.

NEPMs have little derivational morphology, which is limited to prefixation and initial gemination. These two processes are described in §5.3.1 and §5.3.2. While both processes involve adding a segment to the left of the base, they differ in that initial gemination involves a templatic segment which copies its phonemic content from the initial consonant of the base, hence representing a type of non-concatenative morphological process (Davis & Tsujimura 2014: 191; Spencer 2001: 125).

Compounding is somewhat productive in NEPMs. Reduplication takes the form of full reduplication and in a few instances of echo reduplication, yielding complex words composed of two roots. Reduplicated forms thus show more formal similarities to compounds than to derivatives; for this reason, reduplication can be seen as a special type of compounding (also see Fabb 2001; Inkelas & Zoll 2005). Compounding and reduplication are discussed in §5.3.3 and §5.3.4 respectively.

5.3.1 Prefixation

NEPMs have a relatively small inventory of affixes when compared to other Malayic varieties (cf. McDonnell et al. in print). KM and CTM have five prefixes, four of which are verbalising prefixes, namely *bɣ-* 'INTR; MID', *tɣ-* 'NVOL', *pɣ-* 'CAUS; FCT' and NN_1- 'IPFV'. Additionally, there is one homophonous nominalising prefix NN_2- 'NMLS'. ITM has one prefix less as it lacks the causative/factitive marker. These prefixes exhibit morphophonological alternations, with allomorphs occurring in different phonological environments, as explained in §5.3.1.1.

5.3.1.1 Morphophonological alternations

The prefixes *bɣ-* 'INTR; MID', *tɣ-* 'NVOL' and *pɣ-* 'CAUS; FCT' have several allomorphs, with their shape determined by the initial segment of the base to

which they are attached. They occur in their full forms before vowel-initial bases, whereas before consonant-initial bases, the liquid *ɣ* is deleted, and the prefixes occur as *b-*, *t-* and *p-* respectively. Examples illustrating the prefixation of *bɣ-*, *tɣ-* and *pɣ-* in KM are given in (16).

(16) Prefixation of *bɣ-* 'INTR; MID', *tɣ-* 'NVOL' and *pɣ-* 'CAUS; FCT' in KM

Before vowel-initial bases				
anɔʔ	'child'	→	*bɣ-anɔʔ*	'to give birth; to be born'
iŋaʔ	'to think'	→	*tɣ-iŋaʔ*	'to remember; to miss'
ilɛ	'to disappear'	→	*pɣ-ilɛ*	'to lose'
Before consonant-initial bases				
ɣasɔ	'to taste; to feel'	→	*b-ɣasɔ*	'to feel'
bakɔ	'to burn'	→	*t-bakɔ*	'to be burnt'
lumaʔ	'crushed'	→	*p-lumaʔ*	'to crush'

The deletion of *ɣ* in these prefixes when attached to consonant-initial bases can be explained by the phonotactic constraint that limits the number of initial consonants in phonological words to a maximum of two. Also importantly, the CC- clusters resulting from prefixation need to be phonologically well-formed. In other words, these clusters should comply with the SSP, which is applied with varying strictness in different NEPM varieties (see §2.5, §3.5 and §4.5). In KM and CTM, the prefixes *b-*, *t-* and *p-* only appear when the initial consonant of the base is minimally as sonorous as the prefix (see examples in (18), (22), (26) and (34) below). In cases where the base-initial consonant is identical to the prefix, initial geminates are produced at the phonetic level, which may be alternatively viewed as deriving from the process of initial gemination (see §5.3.2), as in KM *biniŋ* 'wife' → *b-biniŋ* 'to marry a wife' or CTM *tido* 'to sleep' → *t-tido* 'to fall asleep (non-volitionally)'. In ITM, where the SSP applies less strictly, the prefix *bɣ-* 'INTR; MID', which has an initial voiced stop *b-*, occasionally appears before a base with a voiceless stop, e.g., *kabuh* 'fog' → *b-kabuh* 'foggy'. However, such clusters of a voiced stop + a voiceless stop tend to be unstable and often alternate with geminate clusters, as in *b-kabuh~k-kabuh* (INTR-fog) 'foggy'.

The same phonotactic constraint generally prevents bases with initial consonant clusters from undergoing prefixation, but there are a few exceptions where *bɣ-* 'INTR; MID' is attached to bases with an initial CC cluster, appearing as *bə-*. Examples include KM/CTM *bə-tyabo* (MID-scattered)

'cluttered', ITM *bə-slimu?* (INTR-blanket) 'to cover (oneself) with a blanket', *bə-glisɔh* (MID-anxious) 'to feel anxious', and *bə-s=buti* (INTR-one=CLF) 'to have one' (in which the base surprisingly has a clitic *s=*). It is worth noting that the initial clusters in these bases typically consist of an obstruent and a liquid, and only the prefixation of *bɣ-* 'INTR; MID' is attested before CC-initial bases.

The other two prefixes *NN$_1$-* 'IPFV' and *NN$_2$-* 'NMLS' are geminate nasals. The capital *N* represents an underspecified nasal that is subject to nasal assimilation and nasal substitution, which are common morphophonological alternations in languages in West Indonesia (Blust 2004, 2013: 242–244). In NEPMs, these nasal prefixes only occur before disyllabic bases whose initial segment falls into one of the following categories: vowels, the liquid *ɣ*, or the voiceless obstruents *p, t, c, k* and *s*. Their morphophonological alternations are illustrated by KM examples in (17), with the underlying initial consonants that are deleted or substituted being indicated in angle brackets <>.[33]

(17) Morphophonological alternations of *NN-* in KM

NN$_1$-aka?	(IPFV-lift)	→	*ŋŋ-aka?*	'lifting'
NN$_1$-ɣukah	(IPFV-climb)	→	*ŋŋ-<ɣ>ukah*	'climbing'
NN$_1$-paŋge	(IPFV-call)	→	*mm-<p>aŋge*	'calling'
NN$_1$-tanɛ	(IPFV-plant)	→	*nn-<t>anɛ*	'planting'
NN$_2$-cɛtɔ?	(NMLS-print)	→	*ɲɲ-<c>ɛtɔ?*	'printer'
NN$_1$-kute?	(IPFV-pick)	→	*ŋŋ-<k>ute?*	'picking'
NN$_2$-sapuh	(NMLS-sweep)	→	*ɲɲ-<s>apuh*	'broom'

These nasal prefixes take the default realisation of velar *ŋŋ-* when occurring before vowel-initial bases. When they occur before bases with initial *ɣ*, the *ɣ* is deleted, and *NN-* also takes the form of velar *ŋŋ-*.[34] Before bases with initial

[33] There is also one KM example where *NN$_1$-* seems to occur before a base with initial *l* and surface as a singleton *m-*, i.e., *?NN$_1$-lamboŋ* (IPFV-bump) → *m-lamboŋ* 'bumping'. However, the allomorphic alternation of *NN$_1$-* → *m* is phonologically implausible, and from a diachronic perspective, *m-* does not reflect an underspecified *N, see §8.3.2.

[34] Another example is CTM *NN$_1$-ɣacoŋ* (IPFV-poison) → *ŋŋ-acoŋ* 'poisoning'. A parallel pattern of initial *ɣ* deletion can be seen in the cliticisation of *di* 'LOC' and *kɔ* 'to; AGT', as in KM *di ɣumɔh* → *d=umɔh* 'LOC=house' or *kɔ ɣumɔh* → *k=umɔh* 'to=house'. The hosts are usually not affected by preceding prepositions, but initial *ɣ* in nouns like *ɣumɔh* 'house' is deleted in the cliticisation process. Just like the prefixation of *NN-* before bases with initial *ɣ*, the cliticisation of prepositions before hosts with initial *ɣ* also appears as if it takes places before vowel-initial hosts.

voiceless obstruents, *NN-* undergoes nasal assimilation and nasal substitution, whereby its place of articulation is assimilated to that of the base-initial obstruent. Essentially, initial voiceless obstruents in the bases are replaced by homorganic geminate nasals: *p-* is replaced by *mm-*, *t-* by *nn-*, *c-* by *ɲɲ-* and *k-* by *ŋŋ-*. Exceptions apply for bases with initial *s-*, which is usually replaced by *ɲɲ-* instead of the expected *nn-*. In some examples, both *ɲɲ-* and *nn-* are attested, e.g., KM/CTM *NN$_1$-susuʔ* (IPFV-hide) → *ɲɲusuʔ/nnusuʔ* 'hiding'.

The following sections provide a more detailed description of the functions of each prefix.

5.3.1.2 Prefix *bɣ-* 'INTR; MID'

The prefix *bɣ-* derives intransitive verbs, with two more specific functions depending on the word class of the bases it is attached to.

First, when attached to nominal bases, *bɣ-* is a category-changing prefix, deriving intransitive verbs with the general meaning of 'to have, to produce BASE', as illustrated in (18) to (21). Note that semantic adjectives are considered a type of intransitive verbs called stative verbs (as opposed to dynamic verbs), see §6.2.2.

(18) Intransitive verbaliser *bɣ-* 'INTR'

KM				
anɔʔ	'child'	→	*bɣ-anɔʔ*	'to give birth; to be born'
ae	'water'	→	*bɣ-ae*	'watery'
isi	'content'	→	*bɣ-isi*	'fat'
ɔbaʔ	'medicine'	→	*bɣ-ɔbaʔ*	'to receive treatment'
lapih	'layer'	→	*b-lapih*	'layered'
laya	'sail'	→	*b-laya*	'to sail'
ɣəgɔ	'price'	→	*b-ɣəgɔ*	'pricy'
CTM				
anɔʔ	'child'	→	*bɣ-anɔʔ*	'to give birth; to be born'
ae	'water'	→	*bɣ-ae*	'watery'
aŋiŋ	'wind'	→	*bɣ-aŋiŋ*	'windy'
isi	'content'	→	*bɣ-isi*	'fat'
lubaŋ	'hole'	→	*b-lubaŋ*	'to have a hole'
layɔ	'sail'	→	*b-layɔ*	'to sail'
ɣayə	'Eid al-Fitr'	→	*b-ɣayə*	'to celebrate Eid al-Fitr'

ITM

anɔʔ	'child'	→	*bɣ-anɔʔ*	'to give birth; to be born'
ai	'water'	→	*bɣ-ai*	'watery'
aŋiŋ	'wind'	→	*bɣ-aŋiŋ*	'windy'
upɔh	'wage'	→	*bɣ-upɔh*	'to work'
dəbu	'splash'	→	*b-dəbu*	'to swash'
likəʊ	'twist'	→	*b-likəʊ-likəʊ*	'tortuous'
ɣayɛ	'Eid al-Fitr'	→	*b-ɣayɛ*	'to celebrate Eid al-Fitr'
slimuʔ	'blanket'	→	*bə-slimuʔ*	'to wear a blanket'
cləbũʔ	'plop'	→	*bə-cləbũʔ*	'to make a plop sound'

(19) KM
mɔʔ aku ***bɣ-anɔʔ*** *k=aku di kapoŋ kusia-bayu.*
mother 1SG INTR-child to=1SG LOC village Kusial-Bharu
'My mother gave birth to me in the village of Kusial Bharu.'
(KM_180825_e01_28)

(20) CTM
bɣ-aŋiŋ *mɔlɛʔ pətaŋ ni.*
INTR-wind nice afternoon DEM.PROX
'The wind blows nicely this afternoon.' (CTM_220927_e02_67)

(21) ITM
kudih məʊŋ ***bɣ-ai*** *ah akəʊ tiŋuʔ.*
scabies 2SG INTR-water INTERJ 1SG look
'I see that your scabies have suppurated.' (ITM_220915_e03_24)

Second, when attached to verbal bases, *bɣ-* functions as a middle (voice) marker, which signifies that the action denoted by the verb is imposed on the actor itself (Kemmer 1993).[35] For instance, KM/CTM *ato* 'to arrange (s.th.)' is a transitive verb, and the prefixation of *bɣ-* derives an intransitive verb *bɣ-ato* meaning 'to line up', which can be conceptualised as 'to arrange oneself'. Similarly, ITM *bɣ-alɛih* 'to change position' is derived from *alɛih* 'to move (s.th.)', and *bɣ-alɛih* can be conceptualised as 'to move oneself'.

[35] It is acknowledged that "middle voice" is a controversial notion (see Inglese 2022), and its application in NEPMs is less satisfactory, especially given that NEPMs do not have active/passive voice marking morphology. The term is adopted loosely here, and I opt for using "middle marker" over "middle voice marker".

KM/CTM *layi* and ITM *layɛi* 'to run' are used in imperative mood to order someone to run, whereas *b-layi*/*b-layɛi* means someone is running. The prefix *bɣ-* 'MID' is also found on stative verbs denoting feelings and emotions, in which cases the difference in meaning between the base and the derived form is often subtle, as in KM/CTM *sɣabuʔ* 'upset' → *bə-sɣabuʔ* 'to feel upset', and ITM *glisɔh* 'anxious' → *bə-glisɔh* 'to feel anxious'. More examples illustrating the prefixation of *bɣ-* 'MID' are provided in (22), along with example sentences presented in (23) to (25).

(22) Middle prefix *bɣ-* 'MID'

KM

aŋjɔʔ	'to move'	→	*bɣ-aŋjɔʔ*	'to change position'
iŋaʔ	'to think'	→	*bɣ-iŋaʔ*	'to take care'
ubɔh	'to change (s.th.)'	→	*bɣ-ubɔh*	'to become different'
lateh	'to train (s.o.)'	→	*b-lateh*	'to exercise'
ɣasɔ	'to taste; to feel'	→	*b-ɣasɔ*	'to feel'
tɣabo	'scattered'	→	*bə-tɣabo*	'cluttered'
sɣabuʔ	'upset'	→	*bə-sɣabuʔ*	'to feel upset'
glabɔh	'sad'	→	*bə-glabɔh*	'to feel sad'

CTM

ato	'to arrange'	→	*bɣ-ato*	'to line up'
aleh	'to move'	→	*bɣ-aleh*	'to change position'
iŋaʔ	'to think'	→	*bɣ-iŋaʔ*	'to take care'
ubɔh	'to change (s.th.)'	→	*bɣ-ubɔh*	'to become different'
laŋgɔ	'to hit'	→	*b-laŋgɔ*	'to collide'
tɣabo	'scattered'	→	*bə-tɣabo*	'cluttered'

ITM

atu	'to arrange'	→	*bɣ-atu*	'to line up'
alɛih	'to move'	→	*bɣ-alɛih*	'to change position'
ubɔh	'to change (s.th.)'	→	*bɣ-ubɔh*	'to become different'
jəɣəʊʔ	'wet'	→	*b-jəɣəʊʔ*	'to become wet'
laŋgɔ	'to hit'	→	*b-laŋgɔ*	'to collide'
ɣasɛ	'to taste; to feel'	→	*b-ɣasɛ*	'to feel'
susuŋ	'to arrange'	→	*b-susuŋ*	'to pile up'
glisɔh	'anxious'	→	*bə-glisɔh*	'to feel anxious'
tɣabu	'scattered'	→	*bə-tɣabu*	'cluttered'

(23) KM

a. *aku doʔ* ***lateh*** *anɔʔ-muɣiʔ aku ni.*
1SG PROG train child-pupil 1SG DEM.PROX
'I'm training my student.' (KM_221026_e01_49)

b. *aku gi* ***b-lateh*** *dəŋa saiŋ aku.*
1SG go MID-train with friend 1SG
'I'm going to exercise with my friend.' (KM_221026_e01_50)

(24) CTM

a. *moŋ* ***laŋgɔ*** *məndə taʔdiʔ*
2SG hit what just.now
'What did you hit just now?' (CTM_220927_e01_80)

b. *pagi aʔdi aku* ***b-laŋgɔ*** *dəŋaŋ kkɛtə.*
morning just.now 1SG MID-hit with car
'This morning I collided with a car.' (CTM_220927_e02_79)

(25) ITM

a. ***susuŋ*** *mulɛiʔ gaʔ ikĩʔ kaiŋ tah.*
arrange good EMPH little cloth DEM.DIST
'Put away these clothes.' (ITM_220923_e01_10)

b. *kaiŋ batĩʔ diyɛ baŋɔʔ a,* ***b-susuŋ-susuŋ****.*
cloth batik 3 much INTERJ MID-RDP-arrange
'He has a lot of batik, all piling up.' (ITM_220923_e01_11)

5.3.1.3 Prefix *tɣ-* 'NVOL'

The prefix *tɣ-* 'NVOL' is attached to dynamic verbal bases that are either transitive or intransitive (for the distinction between stative and dynamic verbs, see §6.2.2). It is used to mark non-volitionality or unintentionality of the action denoted by the base verb, as illustrated by the examples in (26).

(26) Non-volitional verbal prefix *tɣ-* 'NVOL'

KM

atoʔ	'to collide'	→	*tɣ-atoʔ*	'to bump against'
aleh	'to move'	→	*tɣ-aleh*	'to change position'
iŋaʔ	'to think'	→	*tɣ-iŋaʔ*	'to remember; to miss'
igaʔ	'to catch'	→	*tɣ-igaʔ*	'to get caught'
baka	'to burn'	→	*t-baka*	'to be burnt'
bukɔ	'to open'	→	*t-bukɔ*	'to open (on its own)'
kəjuʔ	'to startle'	→	*t-kəjuʔ*	'to be startled'
jatoh	'to fall'	→	*t-jatoh*	'to fall (unintentionally)'

CTM

atoʔ	'to collide'	→	*tɣ-atoʔ*	'to bump against'
akaʔ	'to lift'	→	*tɣ-akaʔ*	'to be lifted'
ambeʔ	'to take'	→	*tɣ-ambeʔ*	'to take (by mistake)'
iŋaʔ	'to think'	→	*tɣ-iŋaʔ*	'to remember'
bakɔ	'to burn'	→	*t-bakɔ*	'to be burnt'
tido	'to sleep'	→	*t-tido*	'to fall asleep (non-volitionally)'
gatoŋ	'to hang'	→	*t-gatoŋ*	'to be hung'
laŋgɔ	'to hit'	→	*t-laŋgɔ*	'to get hit; to hit (accidentally)'

ITM

akaʔ	'to lift'	→	*tɣ-akaʔ*	'to be lifted'
ambɛiʔ	'to take'	→	*tɣ-ambɛiʔ*	'to take (by mistake)'
iŋaʔ	'to think'	→	*tɣ-iŋaʔ*	'to remember; to miss'
igaʔ	'to catch'	→	*tɣ-igaʔ*	'to get caught'
bukɛ	'to open'	→	*t-bukɛ*	'to open (on its own); be left open'
bakɔ	'to burn'	→	*t-bakɔ*	'to be burnt'
gatəʊŋ	'to hang'	→	*t-gatəʊŋ*	'to be hung'
sakuʔ	'to hook'	→	*t-sakuʔ*	'to be hooked'
laŋgɔ	'to hit'	→	*t-laŋgɔ*	'to get hit; to hit (accidentally)'

With transitive bases, the prefixation of *tɣ-* is often a valency-decreasing device, whereby the derived forms become intransitive. Compare KM *baka* 'to burn' with the prefixed form *t-baka* 'to be burnt' in (27), CTM *akaʔ* 'to lift' with *tɣ-akaʔ* 'to be lifted' in (28), and ITM *bukɛ* 'to open' with *t-bukɛ* 'to be left open' in (29). The bases in all three examples are transitive, while the corresponding derivatives are intransitive.

(27) KM

a. *jiɣɛ diyɔ* ***baka*** *ɣumɔh tu.*
neighbour 3 burn house DEM.DIST
'His neighbour burnt the house.' (KM_180825_e01_39)

b. *ɣumɔh tu* ***t-baka****.*
house DEM.DIST NVOL-burn
'The house was burnt.' (KM_180825_e01_38)

(28) CTM

a. *bɛloŋ kuniŋ tu* ***akaʔ*** *s=ɔɣaŋ budɔʔ ppɯwaŋ ni.*
balloon yellow DEM.DIST lift one=CLF kid female DEM.PROX
'The yellow balloon is lifting a girl.' (CTM_181028_e01_51)

b. *kuwaʔ aŋiŋ malaŋ,* ***tɣ-akaʔ*** *abih ataʔ.*
strong wind night NVOL-lift finished roof
'The wind at night was strong, and the roof was all blown away.' (CTM_220927_e02_108)

(29) ITM

a. *...* ***bukɛ*** *pitəʊ-malɛiŋ paŋgi kkatɔʔ ŋə.*
... open door-thief call frog ANAPH
'He opened the window, calling his frog.' (ITM_180907_n01_6.2)

b. *akəʊ gɛi k=umɔh məʊŋ, məʊŋ taʔdɔʔ, pitəʊ məʊŋ* ***t-bukɛ****.*
1SG go to=house 2SG 2SG NEG.EXIST door 2SG NVOL-open
'I went to your house and you were not there, but your door was left open.' (ITM_220915_e03_45)

There are also some instances where the prefixation of *tɣ-* does not decrease the valency of the transitive verbal base, but instead, it derives a form that highlights the unintentionality of the action, as shown by the contrast between (30a) and (30b).

(30) ITM

a. *nah* ***ambɛiʔ*** *kwalɛi nah.*
DEM.PROX take wok DEM.PROX
'Take this wok.' (ITM_180917_cv01_63)

b. *akəʊ* ***tɣ-ambɛiʔ*** *bukəʊ məʊŋ, maʔãh.*
1SG NVOL-take book 2SG sorry
'I took your book by mistake, sorry.' (ITM_220915_e03_39)

Similarly, when attached to intransitive bases, the prefix *tɣ-* indicates an involuntary or uncontrolled action without affecting the valency of the verbal base. In examples (31) to (33), both the bases and the corresponding derived forms are intransitive.

(31) KM
diyɔ doʔ ***t-kəleh-kəleh*** *kɔ tuwɛ tu.*
3 PROG NVOL-RDP-look to owner DEM.DIST
'He was peeping at the owner.' (KM_180814_n01_20)

(32) CTM
aku ***t-tido*** *dalaŋ kəlah taʔdi.*
1SG NVOL-sleep inside class just.now
'I fell asleep in the class just now.' (CTM_181029_e02_30)

(33) ITM
akəʊ laləʊ bawɔʔ kayəʊ təh, ***tɣ-atəʊʔ*** *ppalɛ akəʊ,*
1SG pass.by bring wood DEM.DIST NVOL-collide head 1SG
bəŋɔŋ tliŋɛ-tliŋɛ.
buzzing RDP-ear
'I was passing by carrying the wood, then my head bumped against it, and my ears are buzzing.' (ITM_2200915_e03_42)

It is noteworthy that the cognate of *tɣ-* in some other Malayic varieties can be attached to stative intransitive verbs, marking a superlative, comparative or excessive degree, as attested in SM, Minangkabau, Banjar Hulu and Besemah (Adelaar 1992: 151–155; McDonnell 2016: 42). This usage is, however, not attested in NEPMs.

5.3.1.4 Prefix *pɣ-* 'CAUS; FCT'

The prefix *pɣ-* derives transitive verbs, and it is (historically) in a paradigmatic relation with *bɣ-*, which derives intransitive verbs (Adelaar 1984). Among NEPM varieties, *pɣ-* is only attested in KM and CTM. It can be attached to both intransitive verbs and nouns, serving two functions.

First, when attached to intransitive verbal bases, *pɣ-* derives causative verbs with the meaning of 'to make BASE, to cause BASE', as illustrated in (34) to (36).

(34) Causative prefix *pɣ-* 'CAUS'

KM

abih	'finished'	→	*pɣ-abih*	'to finish'
aŋaʔ	'warm'	→	*pɣ-aŋaʔ*	'to warm up'
ilɛ	'to disappear'	→	*pɣ-ilɛ*	'to lose'
lumaʔ	'crushed'	→	*p-lumaʔ*	'to crush'
ləsaʔ	'to disappear'	→	*p-ləsaʔ*	'to steal'
lɯwah	'wide'	→	*p-lɯwah*	'to expand'
ɣəbɔh	'to fall'	→	*p-ɣəbɔh*	'to bring down'

CTM

abih	'finished'	→	*pɣ-abih*	'to finish'
aco	'crushed'	→	*pɣ-aco*	'to crush'
aŋaʔ	'be warm'	→	*pɣ-aŋaʔ*	'to warm up'
iŋaʔ	'to remember'	→	*pɣ-iŋaʔ*	'to remind'
ɣəbɔh	'to fall'	→	*p-ɣəbɔh*	'to bring down'

(35) KM

a. *ɣayɔ* ***abih*** *dɔh.*
Eid.al-Fitr finished already
'Hari Raya (Eid al-Fitr) is already finished.' (KM_180825_e01_29)

b. *aku* ***pɣ-abih*** *dɔh xxijɔ aku.*
1SG CAUS-finished already work 1SG
'I already finished my work.' (KM_180825_e01_30)

(36) CTM

a. *yə poŋ b-laŋgɔ* ***ɣəbɔh.***
3 then MID-crash fall
'He crashed and fell.' (CTM_181025_n02_39.1)

b. *yə* ***p-ɣəbɔh*** *basika yə, yə ambeʔ.*
3 CAUS-fall bike 3 3 take
'He dropped his bike and took (the pears).'
(CTM_181025_n02_26)

Second, when attached to nominal bases, *pɣ-* derives factitive verbs that convey a general meaning of 'to use BASE on the object' or 'to treat object as BASE', as shown by the examples in (37) and (38). More broadly speaking, the derived form signals that the subject makes the object be in a certain condition that involves the BASE. I refer to this function of *pɣ-* as a factitive marker (Lyons 1977: 491). For instance, the prefixation of *pɣ-* on the noun *ati* 'liver' derives the verb *pɣ-ati* 'to observe', which can be interpreted as 'to treat s.th./s.o. as a liver (i.e., the locus of emotion)'. When attached to *uba?~ɔba?* 'medicine', *pɣ-* derives *pɣ-uba?~pɣ-ɔba?* 'to cure', essentially 'to make s.th./s.o. in a state that involves medicine' or 'to use medicine on s.th./s.o.'. Cross-linguistically, it is not uncommon to have one prefix that serves to derive both causatives and factitives, as is also the case in Kambera (Klamer 1998: 178–184) and Boumaa Fijian (Dixon 1988: 181–191).

(37) Factitive prefix *pɣ-* 'FCT'

KM

ati	'liver'	→	*pɣ-ati*	'to observe'
uba?	'medicine'	→	*pɣ-uba?*	'to cure'
isi	'content'	→	*pɣ-isi*	'to clean (fish)'
ɣəgɔ	'price'	→	*p-ɣəgɔ*	'to set a price for'

CTM

ati	'liver'	→	*pɣ-ati*	'to observe'
isi	'content'	→	*pɣ-isi*	'to clean (fish)'

(38) KM

to? *bɔmɔh tu=lah* ***pɣ-uba?*** *aku.*
mister witch DEM.DIST=FOC FCT-medicine 1SG
'It was that traditional healer who cured me.' (KM_221025_e02_63)

The causative/factitive prefix *pɣ-* is not found in ITM. Overall, its equivalent is attested to a limited extent within Malayic; in addition to KM and CTM, it is found in a few other Peninsular Malayic varieties including Kedah Malay and Jakun Malay (Adelaar 1984), as well as in Mualang (Tjia 2007: 44–45). In ITM, causative constructions are formed periphrastically with the auxiliary verbs *wɛi* 'CAUS' and *wa?* 'CAUS', which are also used as content verbs meaning 'to give' and 'to do; to make' respectively. Some examples of causative constructions in ITM are provided in (39).[36]

[36] The histories of *wɛi* 'CAUS' and *wa?* 'CAUS' reflect the common grammaticalisation

(39) ITM

a. *adɛiʔ* ***wɛi*** ***ilɔŋ*** *bəʊʔ təʊ.* .
younger.sibling CAUS disappear book DEM.DIST
'The younger kid lost the book.' (ITM_180921_e01_01)

b. *kitɛ* ***waʔ*** ***bəsɔ*** *apɛi.*
1PL.INCL CAUS big fire
'We raise the heat.' (ITM_180917_cv01_76)

In KM and CTM, causative constructions can also be formed with *waʔ* 'to make; CAUS', as in (40). The causative prefix *pɣ-* may also cooccur with *waʔ*, as shown in (41) and (42).

(40) KM
aku pisɔh ***waʔ*** *jatoh bɔla tu* *daɣi* *mɛjɔ.*
1SG deliberately CAUS fall ball DEM.DIST from table
'I deliberately made that ball fall from the table.'
(KM_180827_e02_20)

(41) KM
kaʔ *loŋ* ***waʔ*** ***pɣ-ilɛ*** *buku tu.*
sister eldest make/CAUS CAUS-disappear book DEM.DIST
'Sister (referring to the eldest child) lost the book.'
(KM_180827_e02_30)

(42) CTM
waʔ ***pɣ-aco*** *ladə tu* *sbəloŋ gunə.*
make/CAUS CAUS-be.crush chilli DEM.DIST before use
'Crush the chilli before use.' (CTM_181029_e02_81)

5.3.1.5 Prefix *NN$_1$-* 'IPFV'

The prefix *NN$_1$-* is a category-preserving morpheme found on verbal bases. While its equivalent in other Malayic varieties is often assumed to be an act-

paths of 'to give' > 'CAUS' and 'to make' > 'CAUS' (Heine & Kuteva 2004: 117–118, 152). The difference between these two causative markers is not entirely clear; *wɛi* is typically used before a dynamic verb, whereas *waʔ* is used before a stative verb, but sometimes they show free variation, as in *wɛi kəɣɛiŋ* and *waʔ kəɣɛiŋ* (CAUS dry) 'to dry'. There are also some counterexamples, such as *waʔ laɣi* (CAUS run) 'to make run' and *wɛi cai* (CAUS liquid) 'to make liquid'.

ive or actor-oriented voice marker (cf. SM *məN-*, see Cole et al. 2008; Sneddon 2010: 255; McDonnell et al. in print), *NN$_1$-* in NEPMs is not associated with voice alternation. Instead, it serves as an imperfective aspect marker, describing situations that are habitual, continuous or progressive while paying special attention to the internal structure of the situation (Comrie 1976: 16; de Swart 2012: 757). The following discussion starts with a focus on the usage of *NN$_1$-* in KM, for which the largest amount of data is available, before extending the analysis to CTM and ITM.

Example (43) illustrates the presence and absence of the prefix *NN$_1$-* on the base *tanɛ* 'to plant' in KM. At first sight, the prefixed form *nn-<t>anɛ* appears to have two different aspectual readings: habitual in (43b) and progressive in (43c). In (43d), however, *nn-<t>anɛ* further combines with an auxiliary *doʔ* 'PROG'. Though not mandatory, the cooccurrence of *doʔ* 'PROG' with a verb prefixed with *NN$_1$-* is very common, which casts doubt on the analysis of *NN$_1$-* as a progressive aspect marker.

(43) KM

a. *moŋ* ***tanɛ*** *pɔkɔʔ ggapɔ?*
2SG plant tree what
'What tree(s) do/did you plant?' (KM_180824_fn)

b. *ayɔh aku* ***nn-<t>anɛ.***
father 1SG IPFV-plant
'My father is a farmer.' (Lit. 'My father plants.')
(KM_180830_e01_13)

c. *puwɔʔ moŋ* ***nn-<t>anɛ*** *ggapɔ tuʔ*
group 2SG IPFV-plant what DEM.DIST
'What are you planting over there?' (KM_180824_fn)

d. *moŋ doʔ* ***nn-<t>anɛ*** *pɔkɔʔ ggapɔ?*
2SG PROG IPFV-plant tree what
'What tree(s) are you planting?' (KM_180824_fn)

An alternative view that unifies (43b), (43c) and (43d) is that they all focus on the incompleteness of the event "with no information about its endpoints" (Smith 1997: 73), either by suggesting its ongoing state or its durative nature, as opposed to (43a). Such a contrast between verbs with and without *NN$_1$-* can also be seen in example (44). In (44a), the unmarked verb form

ɣukah 'to climb' implies the completion of the event in its entirety (translated as he *climbed up* the tree), whereas *ŋŋukah* in (44b) suggests the event of climbing is ongoing.

(44) KM

a. *diyɔ poŋ* ***ɣukah*** *pɔkɔʔ, nɔʔ laɣi padɔ tbuwɛ, diyɔ* ***ɣukah****.*
3 also climb tree want run from hornet 3 climb
'He climbed up the tree, wanting to run from the hornets.'
(KM_180812_n01_15)

b. *ɔɣɛ doʔ* ***ŋŋ****-<****ɣ****>****ukah*** *nɔʔ kuteʔ buwɔh pɛ.*
person PROG IPFV-climb want pick fruit pear
'Someone is climbing (up the tree) to pick up pears.'
(KM_180814_n01_01)

Examples in (45) illustrate a more elaborated aspectual distinction between the bare form *tumih* 'to sauté' and the prefixed form *nn-<t>umih* 'IPFV-sauté'. All sentences in (45) are taken from a conversation where the speakers discussed Kelantanese cooking.

(45) KM

a. *ɔɣɛ gətiŋ tɔʔ* ***tumih****.*
person Geting NEG sauté
'People from Geting do not sauté.' (KM_180820_cv03_5)

b. *pah* ***tumih****=lah bawɛ kitɔ iɣih,* ***tumih tumih*** *... biya ...*
then sauté=FOC onion 1PL slice, sauté sauté ... let ...
bawɛ tu biya gaɣiŋ napɔʔ kɔko.
onion DEM.DIST let crispy look brown
'Then sauté the onions we sliced, sauté, sauté, until the onions are crispy and brown.' (KM_180820_cv03_171)

c. *diyɔ* ***tumih*** *... diyɔ waʔ awah* ***nn****-<****t****>****umih*** *tu,*
3 sauté ... 3 make ingredient IPFV-sauté DEM.DIST
ladɔ ija dəŋa asɛ-lima, diyɔ tɔʔ capo ae.
chilli green and sour-citrus 3 NEG mix water
'They sauté ... (while) they sauté the ingredients, (they only use) green chilli and lime, they don't add water.'
(KM_180820_cv03_60)

d. *kalu kitɔ* ***nn-<t>umih*** *ni,*
TOP 1PL IPFV-sauté DEM.PROX
diyɔ ssəɣɔ bawɛ di miɲɔʔ tu.
3 feel onion LOC oil DEM.DIST
'As for the kind we saute, we taste the garlic in the oil.'
(KM_180820_cv03_41)

e. *mɛmɛ masɔ kitɔ blɛndə tu kɯwaʔ baũ,*
indeed when 1PL blend DEM.PROX strong smell
kitɔ ***nn-<t>umih*** *poŋ bakeʔ tapi mɛmɛ səda?.*
1PL IPFV-sauté also rise but indeed delicious
'Indeed when we blend (the ingredients), the smell is strong; when we sauté, (the smell) also rises, but it's really delicious.'
(KM_180820_cv03_78)

In (45a), the event of 'sauté' is viewed as a whole from the outside; *tɔʔ tumih* 'not sauté' is presented as an observation. In (45b), *tumih* first occurs in an imperative form signalled by the focus clitic *=lah*, and the second and third *tumih* describe an event with an endpoint, namely 'until the onions are crispy and brown'. In (45c)–(45e), in contrast, 'sautéing' is viewed as a process with an internal temporal constituency, with the focus placed on something that is involved or happens during the process of sautéing, rather than the beginning or the endpoint of the situation. Also notable is that all instances of *nn-<t>umih* in (45c)–(45e) occur in subordinate clauses, providing background statements. Those clauses can be roughly translated as 'while/when sautéing', even though the temporal conjunctions are not expressed. From these examples, it is clear that what differentiates *nn-<t>umih* 'IPFV-sauté' from *tumih* 'to sauté' is the viewpoint towards the situation being described, and *nn-<t>umih* 'IPFV-sauté' is typically associated with imperfective viewpoints.

In another example presented in (46), the same situation is first referred to with a bare form *salɔʔ* 'to bark', then with a prefixed form *ɲɲ-<s>alɔʔ* 'IPFV-bark'. The change in the choice of verbal forms presumably reflects the change in the speaker's viewpoint, which first took the event of 'barking' as a complete whole, then shifted the focus to its interior composition, during which 'the dog disturbed the beehive'.

(46) KM
aɲiŋ diyɔ gi ***salɔʔ*** *... gi* ***ɲŋ****-<****s****>****alɔʔ*** *di pɔkɔʔ tu,*
dog 3 go bark ... go IPFV-bark LOC tree DEM.DIST

doʔ itu kaca saɣɛ tbuwɛ.
PROG ? disturb nest hornet

'His dog went barking ... while it was barking at the tree, it disturbed the beehive.' (KM_180812_n01_12)

The same analysis can be extended to CTM and ITM. Based on similar examples with *NN$_1$-* in my CTM and ITM corpora, I assume *NN$_1$-* functions in the same way. Some examples are given in (47) and (48).

(47) CTM

a. *masə diyə* ***ɲŋ****-<****s****>****alɔʔ*** *tu, tbuwaŋ tu poŋ tubeʔ*
when 3 IPFV-bark DEM.DIST hornet DEM.DIST also go.out
ɣama-ɣama daɣipadə saɣaŋ diyə.
RDP-many from nest 3

'While the dog was barking, many hornets flew out of their nest.' (CTM_181023_n01_25)

b. *ayɔh moŋ* ***nn****-<****t****>****anaŋ*** *padi=kəʔ*
father 2SG IPFV-plant paddy=Q

'Is your father a rice farmer?' (CTM_181029_e02_15)

(48) ITM

a. *mɔʔciʔ gɛi* ***mm****-<****p****>****utəʊŋ*** *gətɔh.*
auntie go IPFV-cut rubber

'I went to cut rubber/I used to cut rubber (for a living).' (ITM_180923_n01_20)

b. *duwɛ tigɛ iku nɛh* ***nn****-<****t****>****uŋgəʊ*** *iŋaʔ apɛi atah bukiʔ,*
two three CLF DEM.PROX IPFV-wait light.up fire top hill
dɔʔ jupɛ caɣɛi.
NEG meet find

'The few others waited and lit up fire on the hill, as they didn't find (the civet and the chicken).' (ITM_180927_n01_14)

c. *kaiŋ təʊ basɔh, cəʔ məʊŋ gɛi* ***ŋŋ-<k>əɣɛiŋ*** *kaiŋ təʊ*
cloth DEM.DIST wet IMP 2SG go IPFV-dry cloth DEM.DIST
sikĩʔ lagɛi.
little again
'The cloth is still wet, try to dry it a bit more.'
(ITM_180921_e01_06)

The prefixed form *ɲŋ-<s>alɔʔ* 'IPFV-bark' in (47a) conveys a similar meaning as its counterpart in (46). In (47b) and (48a), *nn-<t>anaŋ* and 'IPFV-cut' *mm-<p>utəʊŋ* 'IPFV-cut' both have a habitual reading (the latter appears in a narrative about the speaker's experience in the past). In (48b), *nn-<t>uŋgəʊ* 'IPFV-wait' has a temporal constituency, during which the event of *iŋaʔ apɛi* 'light up fire' took place. In (48c), *ŋŋ-<k>əɣɛiŋ* 'IPFV-dry' focuses on the process of 'drying' without information about its endpoint.

One potential complication with the analysis of *NN$_1$-* as an imperfective marker is that *NN$_1$-* would be the only aspectual affix in NEPMs, and also the only inflectional affix. The bare verbal forms seem to be underspecified for aspectual interpretations, but to what extent this observation can be upheld needs further examination. At the same time, however, the aspectual function of *NN$_1$-* in NEPMs is not entirely exceptional when compared to its equivalents in other Malayic varieties. The interpretation motivated above is in line with recent proposals concerning the aspectual functions of *məN-* in SM (Soh & Nomoto 2009, 2015; Nomoto 2013; Soh 2013). Soh and Nomoto show that *məN-* in SM has a progressive aspectual effect which makes it generally incompatible with stative verbs, and situations described by verbs with *məN-* are always eventive and atelic. Many characteristics of an imperfective marker have also been reported for *məN-* in Kuala Lumpur Malay, including its tendency to occur in subordinate rather than main clauses, and to describe ongoing or durative activities as opposed to completed or punctual activities (Gil 2002: 273).

5.3.1.6 Prefix *NN$_2$-* 'NMLS'

The last prefix in NEPMs is *NN$_2$-* 'NMLS', which is homophonous to *NN$_1$-* 'IPFV'. It is attached to dynamic verbal bases to derive nouns, and it is the only nominalising prefix attested in NEPMs. Some examples with *NN$_2$-* 'NMLS' are presented in (49).

(49) Nominaliser NN_2- 'NMLS'

KM

pəgɛ	'to hold'	→	*mm-<p>əgɛ*	'handle'
tutoʔ	'to close'	→	*nn-<t>utoʔ*	'cap'
cɛtɔʔ	'to print'	→	*ɲɲ-<c>ɛtɔʔ*	'printer'
cuke	'to pick'	→	*ɲɲ-<c>uke* (*gigi*)	'(tooth)pick'
sapuh	'to sweep'	→	*ɲɲ-<s>apuh*	'broom'
sakoʔ	'to hang'	→	*ɲɲ-<s>akoʔ*	'hanger'
sakeʔ	'to hurt; sick'	→	*ɲɲ-<s>akeʔ*	'disease'

CTM

pəgaŋ	'to hold'	→	*mm-<p>əgaŋ*	'handle'
tutoʔ	'to close'	→	*nn-<t>utoʔ*	'cap'
timbaŋ	'to weight'	→	*nn-<t>imbaŋ*	'scale'
kisɔ	'to blend'	→	*ŋŋ-<k>isɔ*	'blender'
sakoʔ	'to hang'	→	*ɲɲ-<s>akoʔ*	'hanger'
sapuh	'to sweep'	→	*ɲɲ-<s>apuh*	'broom'
sakeʔ	'to hurt; sick'	→	*ɲɲ-<s>akeʔ*	'disease'

ITM

pəgɔŋ	'to hold'	→	*mm-<p>əgɔŋ*	'handle'
tutuʔ	'to close'	→	*nn-<t>utuʔ*	'cap'
timbɔŋ	'to weight'	→	*nn-<t>imbɔŋ*	'scale'
kisɔ	'to blend'	→	*ŋŋ-<k>isɔ*	'blender'
sapəʊ	'to sweep'	→	*ɲɲ-<s>apəʊ*	'broom'
sakiʔ	'to hurt; sick'	→	*ɲɲ-<s>akiʔ*	'disease'

Unlike its SM equivalent *pəN-*, which productively forms nouns referring to the actor of a performance or the instrument with which the action is performed, NN_2- is typically restricted to forming nouns referring to instruments. Actors of performances are often expressed by compounds with a nominal head meaning 'person', 'expert' of 'craftsman', as shown by examples in (50) to (52).

(50) Compounds denoting actors in KM

joŋ nnuleh	(expert write)	'writer'
joŋ nnaɣi	(expert dance)	'dancer'
ɔɣɛ ŋŋaji	(person study)	'student, researcher'
ɔɣɛ xxijɔ	(person work)	'worker'
ɔɣɛ gaji	(person wage)	'maid'

(51) Compounds denoting actors in CTM

tukaŋ lukih	(craftsman paint)	'painter'
ɔɣaŋ jjuwa	(person sell)	'seller'
ɔɣaŋ sake?	(person sick)	'patient'
ɔɣaŋ gaji	(person wage)	'maid'

(52) Compounds denoting actors in ITM

tukɔŋ nnulih	(craftsman write)	'writer'
uɣɔŋ xxəjɛ	(person work)	'worker'
uɣɔŋ saki?	(person sick)	'patient'

A compound like *joŋ nnuleh* 'writer' in KM can be analysed as *joŋ nn-<t>uleh* (expert *NN*-write) which contains a *NN-* prefix, but the exact meaning of this prefix is ambiguous. One the one hand, *nn-* in *nn-<t>uleh* can be interpreted as NN_1- 'IPFV'. *Joŋ nnuleh* is thus someone who habitually writes, following the idea of NN_1- being an imperfective aspect marker. On the other hand, *nn-* in *nn-<t>uleh* may be interpreted as NN_2- 'NMLS', which derives an attributive modifier for the nominal head *joŋ* 'expert'. The second interpretation might reflect the diachronic path more accurately (as a continuation of the wider application of the PM nominaliser *pAN-, see Adelaar 1992: 183–184), but this attributive use of nouns with NN_2- must have been fossilised, as *nnuleh* cannot be used as a noun meaning 'writer' on its own.

Some words for instruments are formed periphrastically in a similar way, e.g., KM *ala? nnənoŋ* (tool weave) 'loom', *ala? ŋŋuko* (tool measure) 'measuring tools', and ITM *jaɣoŋ ŋŋai?* (needle knit) 'knitting needle'. English words have also been borrowed to fill the gaps created by the restricted usage of the nominaliser, e.g., KM *pɛsəɲjɛ* 'passenger', KM/CTM *lɔya* 'lawyer', KM/CTM *pɔsmɛn* and ITM *pusmɛiŋ* 'postman', as well as KM *gɛ* and CTM/ITM *gaŋ* 'glue, gum' (cf. SM *pən-<t>umpang* (NMLS-ride) 'passenger', *pə-guam* (NMLS-dispute) 'lawyer', *pəŋ-hantar* (NMLS-deliver) 'postman, delivery person', and *pə-ləkat* (NMLS-stick) 'glue').

5.3.1.7 Interim summary

As an interim summary, Table 5.5 provides an overview of the prefixes in NEPMs and the bases to which they can be attached. Among the five prefixes, *bɣ-* and *tɣ-* each has two distinct functions when attached to bases from different word classes.

Table 5.5: Overview of prefixation in NEPMs

		Nouns	Dynamic transitive verbs	Dynamic intransitive verbs	Stative intransitive verbs
bɣ-	'INTR'	+	-	-	-
	'MID'	-	+	+	+
tɣ-	'NVOL'	-	+	+	-
pɣ-	'CAUS'	-	-	+	+
	'FCT'	+	-	-	-
NN$_1$-	'IPFV'	-	+	+	-
NN$_2$-	'NMLS'	-	+	+	-

(+ : attested, - : not attested)

5.3.2 Initial gemination

It has been observed that NEPMs have geminates that contrast with their singleton counterparts in word-initial position (§2.2.1.2, §3.2.1.2 and §4.2.1.2). Many singleton-geminate pairs are not only related in their phonological shapes but also in their semantics, suggesting that geminates can be analysed as morphologically complex. The gemination of an initial singleton consonant, i.e., C_x- → C_xC_x-, can be proposed as a single morphophonological process to explain this pattern.

Initial gemination serves various grammatical functions, as will be described in §5.3.2.1. It will then become clear that the initial gemination resembles prefixation in many ways; the relationship between these two derivational processes will be explored in §5.3.2.2. Next, §5.3.2.3 takes a closer at another type of initial gemination, which essentially results from the cliticisation of prepositions.

5.3.2.1 Grammatical functions of initial gemination

First, initial gemination can serve as an intransitive verbaliser that operates on nominal bases. This is illustrated by the examples in (53) through (56), where the derived forms have the general meaning of 'to have, to produce BASE' or 'to have the quality of, to be engaged in BASE'.

(53) Gemination as an intransitive verbaliser

KM				
pəɣɛ	'war'	→	***p-p***əɣɛ	'to be at war'
bɯwɔh	'fruit'	→	***b-b***ɯwɔh	'to bear fruit'
kɯwɔh	'gravy'	→	***k-k***ɯwɔh	'to become gravy'
jalɛ	'road'	→	***j-j***alɛ	'to walk'
saiŋ	'friend'	→	***s-s***aiŋ	'to befriend'
ɣayɔ	'Eid al-Fitr'	→	***ɣ-ɣ***ayɔ	'to celebrate Eid al-Fitr'
CTM				
bʊɲiŋ	'sound'	→	***b-b***ʊɲiŋ	'to make sound'
bau	'smell'	→	***b-b***au	'smelly'
diɣi	'self'	→	***d-d***iɣi	'to stand'
jalaŋ	'road'	→	***j-j***alaŋ	'to walk'
kawaŋ	'friend'	→	***k-k***awaŋ	'to befriend'
ITM				
biniŋ	'wife'	→	***b-b***iniŋ	'to marry (a wife)'
daɣɔh	'blood'	→	***d-d***aɣɔh	'to bleed'
cabɔŋ	'branch'	→	***c-c***abɔŋ	'branched'
kəbuŋ	'farm'	→	***k-k***əbuŋ	'to farm'
gunɛ	'use'	→	***g-g***unɛ	'useful'

(54) KM

puwɔʔ tu *doʔ* ***p-pəɣɛ***.
tribe DEM.DIST PROG INTR-war
'Those tribes are at war.' (KM_180827_e01_51)

(55) CTM

sɛpaʔ batu, yə poŋ ***j-jalaŋ***.
kick stone 3 also INTR-road
'He kicked the stone, and went on his way.' (CTM_181025_n02_48.1)

(56) ITM

waʔpɛ kakɛi məʊŋ ***d-daɣɔh****?*
why leg 2SG INTR-blood
'Why is your leg bleeding?' (ITM_220918_e01_25)

Second, initial gemination can operate on verbal bases and function as a middle marker, as demonstrated in (57). The bases are typically transitive, and the derived forms become intransitive. In the case of ITM *sandɔ* 'to lean' > *s-sandɔ* 'to lean (oneself) against', the base *sandɔ* is also an intransitive verb which has an inanimate subject (e.g., a ladder *sando* on the wall), whereas *ssandɔ* is used when a person leans (oneself) against something/someone else. Example sentences are provided in (58) to (60).

(57) Gemination as a middle marker

KM

***t*əpoh**	'to hit'	→	***t*-*t*əpoh**	'to collide'
***j*əmo**	'to dry (clothes)'	→	***j*-*j*əmo**	'to sunbathe'
***s*usoŋ**	'to arrange'	→	***s*-*s*usoŋ**	'to pile up'
***g*uliŋ**	'to roll'	→	***g*-*g*uliŋ**	'to lie down'

CTM

***j*əmo**	'to dry (clothes)'	→	***j*-*j*əmo**	'to sunbathe'
***s*usoŋ**	'to arrange'	→	***s*-*s*usoŋ**	'to pile up'
***ɲ*alə**	'to light up s.th.'	→	***ɲ*-*ɲ*alə**	'to light up'

ITM

***t*əmuŋ**	'to meet'	→	***t*-*t*əmuŋ**	'to meet'
***j*əɣəʊʔ**	'wet'	→	***j*-*j*əɣəʊʔ**	'to become wet'
***s*andɔ**	'to lean'	→	***s*-*s*andɔ**	'to lean (oneself) against'
***ɲ*alɛ**	'to light up s.th.'	→	***ɲ*-*ɲ*alɛ**	'to light up'

(58) KM

baʔpɔ moŋ ***j*-*j*əmo** *dəŋa panah?*
why 2SG MID-dry with hot
'Why are you sunbathing with this heat?' (KM_221026_e01_64)

(59) CTM

baɲɔʔ ah buku ***s*-*s*usoŋ**.
much INTERJ book MID-arrange
'There are a lot of books piling up.' (CTM_220927_e02_103)

(60) ITM

budɔʔ təʊ ***s*-*s*andɔ** *d=ayɔh ah.*
kid DEM.DIST MID-lean LOC=father INTERJ
'The kid is leaning on his father.' (ITM_180921_e03_2)

Third, when applied to dynamic verbal bases (either transitive or intransitive), initial gemination can also derive verbs denoting non-volitional or unintentional events. Examples of this derivation are listed in (61), with sentences in context given in (62) to (64). Initial gemination in these examples typically reduces the valency of transitive verbs, or downplays the actors of intransitive verbs while highlighting the non-volitionality of the actions.

(61) Gemination as a non-volitional marker

KM

t*əpoh*	'to hit'	→	***t-t****əpoh*	'to hit (unintentionally)'
t*ido*	'to sleep'	→	***t-t****ido*	'to fall asleep'
k*əjuʔ*	'to startle'	→	***k-k****əjuʔ*	'to be startled'
s*akoʔ*	'to hook'	→	***s-s****akoʔ*	'to be hooked'

CTM

c*abuʔ*	'to pull out'	→	***c-c****abuʔ*	'to be pulled out'
k*əjuʔ*	'to startle'	→	***k-k****əjuʔ*	'to be startled'
s*akoʔ*	'to hook'	→	***s-s****akoʔ*	'to be hooked'
sɛpaʔ	'to kick'	→	*s-sɛpaʔ*	'to kick (unintentionally)'

ITM

p*ijɔʔ*	'to step on'	→	***p-p****ijɔʔ*	'to step on (unintentionally)'
b*akɔ*	'to burn'	→	***b-b****akɔ*	'to be burnt'
b*ukɛ*	'to open'	→	***b-b****ukɛ*	'to open (on its own)'
j*atəʊh*	'to fall'	→	***j-j****atəʊh*	'to fall (unintentionally)'

(62) KM

diyɔ ***k-kəjuʔ****=****lah****,* *tɛŋɔʔ tibɔ-tibɔ kkatɔʔ tu tubeʔ*
3 NVOL-startle=SFP look suddenly frog DEM.DIST come.out
daɣipadɔ lubɛ, diyɔ ***k-kəjuʔ****.*
from hole 3 NVOL-startle
'He was startled seeing a frog suddenly coming out from the hole, he was startled.' (KM_180812_n01_13)

(63) CTM

yə b-laŋgɔ ɣəbɔh hɔ̃, ***c-cabuʔ*** *tɔpi yə.*
3 MID-crash fall AFF NVOL-pull.out hat 3
'He crashed (with the stone) and fell, and his hat was blown away.'
(CTM_181025_n02_39.2)

(64) ITM
adɛiʔ ***t-tiŋga*** *də skulɔh.*
younger.sibling NVOL-leave LOC school
'The younger kid was left behind at school.' (ITM_180921_e01_46)

Fourth, gemination can apply to intransitive verbs (both stative and dynamic) to derive transitive verbs with a causative meaning. This is only attested in KM and CTM, as shown in (65) to (67). The absence of initial gemination as an causative marker in ITM is not surprising, as it aligns with the lack of any morphological marker for causativity in this variety (see §5.3.1.4).

(65) Gemination as a causative marker

KM

b*əto*	'correct'	→	***b-b****əto*	'to correct s.th.'
t*ido*	'to sleep'	→	***t-t****ido*	'to put s.o. to sleep'
d*əɣah*	'loud'	→	***d-d****əɣah*	'to raise (voice)'
k*əɣiŋ*	'dry'	→	***k-k****əɣiŋ*	'to dry s.th.'
g*adiʔ*	'thick'	→	***g-g****adiʔ*	'to thicken'
s*iyaʔ*	'finished'	→	***s-s****iyaʔ*	'to finish'
l*umaʔ*	'crushed'	→	***l-l****umaʔ*	'to crush'

CTM

b*əsɔ*	'big'	→	***b-b****əsɔ*	'to enlarge'
t*əɣaŋ*	'clear'	→	***t-t****əɣaŋ*	'to clarify'
k*uwaʔ*	'strong'	→	***k-k****uwaʔ*	'to strengthen'
m*asoʔ*	'to enter'	→	***m-m****asoʔ*	'to stuff'
m*andi*	'to bathe'	→	***m-m****andi*	'to bathe s.o.'

(66) KM
diyɔ ***g-gadiʔ*** *satɛ* *diyɔ.*
3 CAUS-thick coconut.milk 3
'It thickens the coconut milk.' (KM_180820_cv03_25)

(67) CTM
aku ***m-mandi*** *anɔʔ aku.*
1SG CAUS-bathe child 1SG
'I am bathing my child.' (CTM_181029_e02_25)

Fifth, there are a few examples where initial gemination derives verbs with an imperfective aspectual meaning, as demonstrated in (68) and (69).

(68) Initial gemination as an imperfective marker

KM/CTM

jəɣiʔ	'to cry'	→	***j-jəɣiʔ***	'crying'
juwa	'to sell'	→	***j-juwa***	'to trade'

ITM

bəlɛi	'to buy'	→	***b-bəlɛi***	'to go shopping'
juwa	'to sell'	→	***j-juwa***	'to trade'

(69) ITM

sukɛ gɛi ***b-bəlɛi,*** *dɔʔ sədɔ abih pitih dalaŋ bɛiʔ.*

like go IPFV-buy NEG realise finished money inside bag

'We liked going shopping, didn't realise the money in the bag was all finished.' (ITM_180923_n01_35)

Finally, in a few examples in KM, initial gemination functions as an intensifier when applied to stative verbs or temporal nouns, as illustrated in (70) and (71).

(70) Initial gemination as an intensifier marker

KM

***p**agi*	'morning'	→	***p-p**agi*	'(in the) early morning'
***j**aŋɔʔ*	'pretty'	→	***j-j**aŋɔʔ*	'very pretty'
***p**əkaʔ*	'thick'	→	***p-p**əkaʔ*	'very thick'

(71) KM

aku m-masɔʔ jaʔ ***p-pagi*** *pah malɛ.*

1SG CAUS-ripe from INTS-morning until night

'I cooked from early morning until the night.' (KM_180825_e01_51)

To summarise, initial gemination is a polyfunctional morphophonological operation that can be applied to bases from various word classes. An overview of its functions is provided in Table 5.6. Note that gemination as a causative marker is only attested in KM and CTM, and gemination as an intensifier is only attested in KM, where the nominal bases are restricted to temporal nouns.

Table 5.6: Overview of functions of initial gemination

	Nouns	Dynamic transitive verbs	Dynamic intransitive verbs	Stative intransitive verbs
'INTR'	+	-	-	-
'MID'	-	+	+	+
'NVOL'	-	+	+	-
'CAUS'	-	-	+	+
'IPFV'	-	+	+	-
'INTS'	(-)	-	-	+

(+ : attested, - : not attested, (-): attested with limited presence)

As can be seen from the table, initial gemination may serve multiple grammatical functions with bases from a given word class, which means that the derived forms may have than more meaning. With a dynamic transitive verb, for example, gemination may serve as a middle marker or denote non-volitionality. In the derivation of KM *təpoh* 'to hit' > *t-təpoh*, the derived form can mean either 'to collide' (which requires a following preposition) or 'to hit (non-volitionally)', as illustrated in (72).

(72) KM

a. *moŋ* ***t-təpoh*** *ŋa* *sapɔ?*
2SG MID-hit with who
'Who did you collide with?' (KM_221026_e01_46)

b. *maʔãh, maʔãh, aku* ***t-təpoh*** *moŋ.*
sorry sorry 1SG NVOL-hit 2SG
'Sorry, sorry, I hit you by mistake.' (KM_221026_e01_96)

With a dynamic intransitive verb, gemination may function as either a non-volitional marker or a causative marker. For instance, KM/CTM *tido* 'to fall' can derive *t-tido* which means either 'to fall asleep' or 'to put s.o. to sleep'.

5.3.2.2 Initial gemination and prefixation

The preceding description shows that initial gemination and prefixation exhibit many similarities. A comparison between Table 5.5 and Table 5.6 in-

dicates a significant overlap in the majority of the grammatical functions they serve: both can mark intransitivity, middle (voice), non-volitionality, causativity and the imperfective aspect.

Three additional observations further highlight the intricate relation between initial gemination and prefixation. First, to some extent, they exhibit a complementary distribution that is determined phonologically. While both processes add segments to the left of the base, initial gemination only takes place before a consonant-initial base, whereas prefixes are often attached to vowel-initial bases; compare KM *anɔʔ* → *bɣ-anɔʔ* (INTR-child) 'to give birth; to be born' with *dayɔh* → *d-dayɔh* (INTR-blood) 'to bleed'. Second, as previously mentioned in §5.3.1.1, the derivation of certain complex forms can be ambiguous. For instance, KM *b-biniŋ* 'to marry (a wife)' can be viewed as either having a prefix *b-* 'INTR' that coincidentally matches the base-initial consonant, or it may stem from the initial gemination of *b-* in *biniŋ* 'wife'. Third, variation is occasionally attested between prefixes and geminated segments (especially in ITM), e.g., ITM *jalaŋ* 'road' → *b-jalaŋ~j-jalaŋ* (INTR-road) 'to walk'. These observations suggest the possibility of unifying prefixation and initial gemination as one single process with allomorphic alternations. However, as I will argue below, the unified analysis cannot be sustained upon closer examination, and initial gemination should be acknowledged as a separate morphophonological operation.

Recall the allomorphic alternations of the prefixes *bɣ-* 'INTR; MID', *tɣ-* 'NVOL' and *pɣ-* 'CAUS; FCT' (§5.3.1.1): these prefixes appear as *bɣ-*, *tɣ-* and *pɣ-* before vowel-initial bases, and the allomorphs *b-*, *t-* and *p-* surface before bases with an initial consonant, which is typically equally or more sonorous than the initial consonant in the prefixes. In a few examples, *bɣ-* 'INTR; MID' also appears as *bə-* before bases with initial CC clusters; those few instances will be ignored for now. It is conceivable to propose that initial geminated segments could be additional allomorphs of the same underlying prefixes, which occur under other phonological conditions. When the bases have initial consonants that are less sonorous than the prefixes, the C_xC_y- clusters derived from prefixation are phonologically ill-formed, hence C_x regressively assimilates to C_y, the result of which appears as geminates. This possible allomorphic alternation of *bɣ-*, *tɣ-* and *pɣ-* is schematised as follows:

(73) Possible allomorphic alternations of *bɣ-, pɣ-, tɣ-*

bɣ-, tɣ-, pɣ- → { *bɣ-, tɣ-, pɣ-* / __V ; *b-, t-, p-* / __C with same or higher sonority ; base-initial C / __other C }

This hypothetical allomorphic alternation between prefixation and gemination provides a plausible explanation for the derivational process in many cases, as illustrated by KM examples in (74). All these complex forms can be analysed as deriving from the prefixation of *bɣ-* 'INTR; MID', which has different surface realisations under different phonological conditions.

(74) An unified analysis for the prefixation of *bɣ-* 'INTR; MID' in KM

bɣ- → *bɣ-* / __V				
anɔʔ	'child'	→	***bɣ-anɔʔ***	'to give birth; to be born'
ae	'water'	→	***bɣ-ae***	'watery'
ubɔh	'to change (s.th.)'	→	***bɣ-ubɔh***	'to become different'
bɣ- → *b-* / __C with same or higher sonority				
laɣi	'to run'	→	***b-laɣi***	'to run'
ɣasɔ	'to feel'	→	***b-ɣasɔ***	'to feel'
buwɔh	'fruit'	→	***b-buwɔh***	'to bear fruit'
bɣ- → geminated segments / __other C				
pəɣɛ	'war'	→	***p-pəɣɛ***	'to be at war'
təpoh	'to hit'	→	***t-təpoh***	'to hit (unintentionally)'
kuwɔh	'gravy'	→	***k-kuwɔh***	'gravy'

The problem with this approach is, however, that not all instances of complex geminates can be satisfactorily explained in this way. It is important to note that all prefixes have an initial obstruent. Following the phonological conditions proposed for the hypothetical allomorphic alternations, *bɣ-* , *tɣ-* and *pɣ-* are expected to surface as *b-*, *t-*, *p-* before bases with initial liquid *l-* and *ɣ-*, which are more sonorous. Consequently, this would predict that complex geminate liquids should not occur. However, this prediction is contradicted by some attested complex forms such as KM *l-luma?* (CAUS-crushed) 'to crush' and *ɣ-ɣayɔ* (INTR-Eid.al-Fitr) 'to celebrate Eid al-Fitr', as shown in (75).[37]

[37] The expected form *p-luma?* (CAUS-crushed) 'to crush' is also attested, but its variation with *l-luma?* cannot be explained phonologically. The other expected form ˣ*b-ɣayɔ* is not attested and not accepted by the consultants.

(75) KM

a. *moŋ kənɔ tumboʔ* ***l-luma?*** *ladɔ tu.*
2SG must pound CAUS-crushed chilli DEM.DIST
'You must pound to crush the chilli.' (KM_180827_e01_31)

b. *baɣu ni saiŋ abah maɣi* ***ɣ-ɣaɣɔ.***
have.just DEM.PROX friend Abah come INTR-Eid.al-Fitr
'Recently Abah's friend came over to celebrate Eid al-Fitr.'
(KM_180816_cv01_18)

Unexplained irregularities are also observed in CTM and ITM. Similarly, since the prefix *tɣ-* 'NVOL' has an initial voiceless obstruent, *t-* is expected to surface before bases with a more sonorous initial voiced obstruent. Contrary to the expectations, however, derivatives with initial geminates are attested, as shown in (76).

(76) Gemination as a non-volitional marker

KM/CTM

b*abah*	'to overturn'	→	***b-b****abah*	'to be overturned'
g*atoŋ*	'to hang'	→	***g-g****atoŋ*	'to be hung'

ITM

b*akɔ*	'to burn'	→	***b-b****akɔ*	'to be burnt'
b*ukɛ*	'to open'	→	***b-b****ukɛ*	'to open (on its own)'
j*atəʊh*	'to fall'	→	***j-j****atəʊh*	'to fall (unintentionally)'

The initial geminates in derivatives like ITM *b-bukɛ* 'to open (on its own)' and *j-jatəʊh* 'to fall (unintentionally)' cannot be seen as deriving from an underlying prefix that is assimilated to the base-initial consonant, since there is no phonological basis for such assimilation (×*tb-* → *bb-* or ×*tj-* → *jj-*). The more plausible analysis, therefore, is to treat *b-bukɛ* and *j-jatəʊh* as derived from the bases *bukɛ* and *jatəʊh* from a morphophonological operation of initial gemination (*b-* → *bb-* and *j-* → *jj-*).

Lastly, the derivation of KM forms like *pagi* 'morning' → *p-pagi* 'early morning' also indicates that initial gemination is not identical to prefixation, as there is no corresponding prefix that serves the same grammatical function as an intensifier.

In conclusion, while initial gemination resembles prefixation to certain extent (and historically it has indeed originated from prefixation in most

cases, see §8.3), it should be treated as a separate morpho(phono)logical process at the synchronic level.

5.3.2.3 Initial gemination and the cliticisation of prepositions

There are yet another two types of initial gemination that have not been discussed so far: gemination as a locative marker and an agent marker. In both cases, initial gemination operates on nominal bases. Unlike the previous types of initial gemination which are comparable to prefixation, these two processes are more similar to the cliticisation of prepositions, deriving forms that take the same syntactic slots as prepositional phrases.

Examples in (77) illustrate the usage of initial gemination as a locative marker, which derives locative nouns with a general meaning of 'on/at/in/to/from/by BASE'. The double-hyphen "=" is used to indicate the clitic status of the geminated segments. Some examples in contexts are given in (78) to (80).

(77) Gemination as a locative marker

KM

***b**ala*	'police station'	→	***b**=**b**ala*	'at the police station'
***t**aŋɛ*	'hand'	→	***t**=**t**aŋɛ*	'in the hand'
***d**əpɛ*	'front'	→	***d**=**d**əpɛ*	'in the front'
***d**indiŋ*	'wall'	→	***d**=**d**indiŋ*	'on the wall'
***k**əda*	'shop'	→	***k**=**k**əda*	'in the shop'
***k**apoŋ*	'village'	→	***k**=**k**apoŋ*	'in the village'
***s**apɔ*	'who'	→	***s**=**s**apɔ*	'to whom'

CTM

***p**asɔ*	'market'	→	***p**=**p**asɔ*	'at/from the market'
***b**andɔ*	'city'	→	***b**=**b**andɔ*	'in/to the city'
***t**anɔh*	'ground'	→	***t**=**t**anɔh*	'on/to the ground'
***t**əŋɔh*	'middle'	→	***t**=**t**əŋɔh*	'in the middle'
***d**apo*	'kitchen'	→	***d**=**d**apo*	'in/to/from the kitchen'
***k**aki*	'leg'	→	***k**=**k**aki*	'on the leg'
***s**suŋa*	'river'	→	***s**=**s**uŋa*	'in the river'

ITM

***b**ajəʊ*	'shirt'	→	***b**=**b**ajəʊ*	'on the shirt'
***t**iɣɔŋ*	'pole'	→	***t**=**t**iɣɔŋ*	'on the pole'
***t**əkəʊʔ*	'neck'	→	***t**=**t**əkəʊʔ*	'around the neck'
***d**usuŋ*	'Dusun'	→	***d**=**d**usuŋ*	'in Dusun'
***k**apɔʔ*	'axe'	→	***k**=**k**apɔʔ*	'on the axe'
***l**iliŋ*	'candle'	→	***l**=**l**iliŋ*	'on the candle'

As seen in the examples below, the derived forms with initial geminates resemble prepositional phrases: *k=kəda* in (78) means *di kəda* (LOC shop) 'in the shop', *t=tanɔh* in (79) means *kə tanɔh* (to ground) 'to the ground', and *p=pasɔ* in (80) can be replaced with *də kapɔʔ* (LOC axe) 'on the axe'.

(78) KM
*mujo adɔ ɔɣɛ jatɛ **k=kəda** doʔ ɣɔyaʔ, diyɔ ɣɔyaʔ ...*
lucky EXIST person male LOC=shop PROG say 3 say ...
'Luckily there was a man in the shop saying ... he said ...'
(KM_180816_cv01_45.1)

(79) CTM
*tibə-tibə aŋjiŋ tu jatoh **t=tanɔh**.*
suddenly dog DEM.DIST fall LOC=ground
'Suddenly the dog fell to the ground.' (CTM_181023_n02_11)

(80) ITM
*uɣɔŋ llakɛi təʊ t-kəjuʔ tiŋuʔ daɣɔh **k=kapɔʔ**.*
person male DEM.DIST NVOL-startle look blood LOC=axe
'The man was startled when he saw blood on the axe.'
(ITM_180921_e03_19)

Initial gemination may also function as an agent marker in passive constructions (see §6.4.1), as shown in (81) to (83). In these cases, initial geminated segments essentially replace the corresponding agent markers (KM *kɔ~kə*, CTM *di* and ITM *də*).

(81) KM
*anɔʔ aku kənɔ ttɛ **c=ceʔgu** aɣi ni di skɔlɔh.*
child 1SG ADVS hit AGT=teacher day DEM.PROX LOC school
'My child was slapped by the teacher at school today.'
(KM_180816_cv01_45.1)

(82) CTM
ikaŋ hɔʔ aku bəli p=pasɔ taʔdi makaŋ ***k=kuciŋ.***
fish REL ISG buy LOC=market just.now eat AGT=cat
'The fish I bought at the market was eaten by the cat.'
(CTM_220927_e02_65)

(83) ITM
budɔʔ təʊ jahaʔ, kənɛ igaʔ ***p=pulih.***
kid DEM.DIST bad ADVS catch AGT=police
'That kid was bad, and he got caught by the police.'
(ITM_220915_e03_36)

Recall that the basic prepositions *dV* and *kV* in NEPMs can be shortened to single-segment proclitics *d=* and *k=*, which appear before vowel-initial hosts, e.g., KM *d=atah* 'on top' and ITM *k=akəʊ* 'to me' (§5.2.1 and §5.2.3, also see prepositions in §6.2.10). The initial gemination of base-initial consonants in examples (78) to (83) represents a similar type of cliticisation, even though the geminated segments do not directly reflect the shorted form of a preposition. Instead, they follow a template whose phonemic content was copied from the initial consonant of the following host.

Similar to the unified analysis of initial gemination and prefixation as discussed in §5.3.2.2, one may argue that these geminated locative/agent markers essentially result from the cliticisation of prepositions *dV* and *kV*, the outcomes of which further assimilate to the base/host-initial consonant. For example, KM *k=kəda* (LOC=shop) 'in the shop' in (78) may be seen as having an underlying prepositional proclitic *d=*, i.e., *d=kəda*, with the assimilation of *dk-* → *kk-* at the surface level. Similarly, ITM *p=pulih* (AGT=police) 'by the police' in (83) may also be analysed as having an underlying *d=* 'AGT' which assimilates to *p-* in *pulih* 'police', generating *pp-*. Yet again some important observations suggest that this is not the optimal analysis.

Assuming that it is also the SSP that regulates the cliticisation and assimilation processes, one prediction would be that the prepositional proclitics *d=* and *k=* should occur before nouns with a more sonorous initial segment, without undergoing assimilation. However, this prediction is not borne out. On the one hand, hypothetical forms like ˣ*k=budɔʔ* 'to the kid' are not attested. In fact, the proclitics *d=* and *k=* never form non-geminate clusters with the host-initial consonants. On the other hand, geminated locative markers are attested in hosts with an initial liquid, as in (84).

(84) ITM
ibiŋ *təʊ* *ikaʔ* ***l=liliŋ.***
ribbon DEM.PROX tie LOC=candle
'The ribbon is tied on the candle.' (ITM_180921_e02_4)

It therefore appears that initial gemination as a locative/agent marker and the cliticisation of the basic prepositions *dV* and *kV* to *d=* and *k=* are two independent processes. The cliticisation only takes place before vowel-initial hosts; and alternatively, initial gemination of the base-initial consonant can have a general locative meaning or mark the agent in passive constructions. Both processes are optional, as prepositions can also stand on their own in full forms.

5.3.3 Compounding

Compounding is defined as the formation of a new word by adjoining two (or more) words (Bauer 2003: 40). The results are compounds that show lexical integrity, which differ from phrases in that they often have a conventionalised and idiomatic meaning. In the present study, compounds are transcribed using a hyphen '-' linking the two constituents. Depending on the grammatical relationship between the constituents, three types of compounds can be distinguished: attributive, coordinative and subordinative (see Lieber 2010: 46–49).

The most common type of compounds in NEPMs is attributive compounds, in which one element acts as the modifier of the other. These compounds are typically left-headed with a nominal head, and the modifier on the right can be either a noun or a stative verb. Some examples are given in Table 5.7. A few right-headed compounds can be found as exceptions. For example, the head *jaɣi* 'finger' in KM *ibu-jaɣi* (mother-finger) 'thumb' (also cf. ITM *ibəʊ-jaɣɛi*) is modified by *ibu* 'mother' to the left. KM *asɛ-lima* (sour-citrus) 'lime' is also right-headed.

Table 5.7: Attributive compounds in NEPMs

KM	CTM	ITM	Literal translation	Gloss
N + N compounds				
buɣoŋ-atu	*buɣoŋ-atu*	*buɣəʊŋ-atəʊ*	bird-ghost	'owl'
kayu-api	–	*kayəʊ-apɛi*	wood-fire	'firewood'
ubi-kayu	*ubi-kayu*	*ubɛi-kayəʊ*	tuber-wood	'cassava'
–	*tali-pəɣuʔ*	–	rope-stomach	'intestine'
–	*pitu-maliŋ*	*pitəʊ-malɛiŋ*	door-thief	'window'
N + stative V compounds				
lima-nipih	*lima-nipih*	*limɔ-nipih*	citrus-thin	'lime'
bawɛ-bəsa	*bawaŋ-bəsɔ*	*bawɔŋ-bəsɔ*	onion/garlic-big	'onion'
bawɛ-puteh	*bawaŋ-puteh*	*bawɔŋ-putɛih*	onion/garlic-white	'garlic'
timoŋ-cinɔ	*timoŋ-cinə*	*timuŋ-cinɛ*	melon-Chinese	'watermelon'

Coordinative compounds consist of two constituents showing a relation of coordination. Examples include KM *ae-taŋɛ* (water-hand) 'home cooking', KM/CTM *toʔ-nɛnɛʔ* (grandfather-grandmother) 'ancestors', and NEPM *mɔʔ-ayɔh* (mother-father) 'parents'.

The third type of compounds is subordinate compounds, comprising a dynamic verb and a nominal element acting as the argument of the verb. KM *ae-pacuʔ* (water-squirt) 'fountain', KM/CTM *ae-tɣəjoŋ* (water-jump) 'waterfall', and ITM *litɔŋ-pukɔŋ* (cross-crotch) 'helter-skelter' are examples of subordinate compounds.

NEPMs also use compounding to form a particular type of stative verbs meaning 'very BASE', as illustrated by the KM and ITM examples in (85) and (86). I refer to them as "augmented stative verbs". Some of these augmented stative verbs may be classified as coordinate compounds formed by the juxtaposition of two stative verbs, as in ITM *kuɣuh-kəɣɛiŋ* (skinny-dry) 'very skinny' and *səjəʊʔ-siyaʔ* (cold-finished) 'very cold'. In many instances, however, the second constituents in these augmented stative verbs are not independently attested, and their exact semantics are not clear. I tentatively consider these words as compounds with "bound words" (Fabb 2001: 69).

(85) KM

masiŋ	'salty'	*masiŋ-pəɣaʔ*	'very salty'
manih	'sweet'	*manih-lətiŋ*	'very sweet'
pahiʔ	'bitter'	*pahiʔ-ləpɛ*	'very bitter'
masɛ	'sour'	*masɛ-puɣi*	'very sour'
pətah	'spicy'	*pətah-ŋaŋa*	'very spicy'
gəmoʔ	'fat'	*gəmoʔ-gdəpo*	'very fat'
kuɣuh	'thin'	*kuɣuh-kəkɛʔ*	'very thin'
udoh	'ugly'	*udoh-baŋa*	'very ugly'
busũʔ	'smelly'	*busũʔ-kəhoŋ*	'very smelly'

(86) ITM

bəsɔ	'big'	*bəsɔ-daʔɔ*	'very salty'
kəcĩʔ	'small'	*kəcĩʔ-kutẽh, kəcĩʔ-tuwẽʔ*	'very small'
manih	'sweet'	*manih-mlətiŋ*	'very sweet'
masaŋ	'sour'	*masaŋ-gəbaŋ*	'very sour'
paiʔ	'bitter'	*paiʔ-ləpãŋ*	'very bitter'
pədah	'spicy'	*pədah-dəsiʔ*	'very spicy'
udəʊh	'stupid'	*udəʊh-səpaŋ*	'very stupid'
kuɣuh	'skinny'	*kuɣuh-kəɣɛiŋ*	'very skinny'
gəmuʔ	'fat'	*gəmuʔ-dibũʔ*	'very fat'
busũʔ	'smelly'	*busũʔ-bbaŋɔ*	'very smelly'
cumɛ	'pretty'	*cumɛ-lutẽ*	'very pretty'
panah ʔ	'hot'	*panah-klətɛiʔ*	'very hot'
səjəʊʔ	'cold'	*səjəʊʔ-siyaʔ*	'very cold'

5.3.4 Reduplication

Reduplication is broadly defined as the repetition of part or all of a linguistic constituent to form a new constituent with a different function (Inkelas 2014: 169). In NEPMs, reduplication is restricted to full reduplication and echo reduplication, yielding new word forms with two roots. Following Inkelas & Zoll (2005)'s analysis, full reduplication can be viewed as the compounding of two identical words, the outcomes of which are comparable to the coordinate compounds discussed earlier. Echo reduplication is best treated as a subtype of full reduplication, where the prosodic word shape of the reduplicant is retained, but certain segments undergo

slight modifications. Two examples of echo reduplication are attested in KM: *gatɔh-gateh* (RDP-cycle) 'to cycle (continuously)' and *ɣaba-gaba* (RDP-notice) 'to pay attention (casually)'. Since *gateh* 'to cycle' and *gaba* 'to notice' are bases that can occur independently, these examples also demonstrate that reduplicants are copied to the left of the bases. In the first example, the vowel *e* changes to *ɔ*, and in the second example, the consonant *g* becomes *ɣ*.

Reduplication is a productive process that can be applied to words from various word classes. The following discussion provides an overview of the semantics of reduplication. As will be shown, reduplication mostly functions in an iconic way.

First, when applied to nouns, reduplication overtly expresses plurality and diversity, as illustrated by the examples in (87).

(87) Reduplication of nouns: plurality and diversity

a. KM
diyɔ tɛŋɔʔ atah kaʔ batu tu adɔ ***pɔkɔʔ-pɔkɔʔ,***
3 look top near stone DEM.DIST EXIST RDP-tree
diyɔ iŋaʔ kɔ ɣatiŋ kayu.
3 think PREP branch wood
'He saw some trees on the stone; he thought they were branches.' (KM_180812_n01_19)

b. CTM
adə tigə ɛkɔ ayaŋ dudoʔ kaʔ tanɔh tu,
EXIST three CLF chicken sit near ground DEM.DIST
anɔʔ-anɔʔ *yə adə ah.*
RDP-child 3 EXIST INTERJ
'There are three chickens on the ground, and their children were also there.' (CTM_181029_n01_5)

c. ITM
bəʊʔ-bəʊʔ *atah mijɛ nɛh haʔ mikɛ.*
RDP-book top table DEM.PROX REL 3PL
'The books on the table are theirs.' (ITM_180909_e02_20)

The reduplication of temporal nouns often results in an adverbial reading, e.g., KM *mulɔ* 'beginning' → *mulɔ-mulɔ* 'in the beginning' and ITM *pagɛi* 'morning' → *pagɛi-pagɛi* 'in the early morning'.

Second, when applied to stative verbs, reduplication can indicate intensity, as illustrated in (88).

(88) Reduplication of stative verbs: intensity

a. KM
supɔ nasiʔ-ləmɔʔ, diyɔ tɔʔleh api ***dəɣah-dəɣah.***
like rice-grease 3 cannot fire RDP-high
'Like (when making) Nasi Lemak, the heat cannot be very high.'
(KM_180820_cv03_49)

b. CTM
diyə poŋ pəloʔ aŋjiŋ tu ***kuwaʔ-kuwaʔ*** *ah.*
3 also hug dog DEM.DIST RDP-strong INTERJ
'He hugged the dog very tightly.' (CTM_181023_n02_14)

c. ITM
kaləʊ dudəʊʔ umɔh mɔʔ əndɔh tah puŋ ***lamɛ-lamɛ***
if stay house mother Endah DEM.DIST also RDP-long
dɔʔ səda ʔ jugɛ.
NEG nice also
'If we stay at Mrs. Endah's place for too long, that's also not nice.'
(ITM_180926_cv02_37)

Depending on the context, reduplication sometimes signals the reverse semantics of attenuation. In (89a), *manih-manih* does not mean 'very sweet' but 'kind of sweet', and in (89b), *kɔko-kɔko* means 'brownish, a bit brown'.

(89) Reduplication of stative verbs: attenuation

a. KM
ikɛ ***manih-manih*** *ggitu=lah, isi samba.*
fish RDP-sweet like.that=SFP content sambal
'The fish that's kind of sweet, filled with sambal.'
(KM_180820_cv03_74)

b. KM
diyɔ macɛ diyɔ wanə ija ... ija ***kɔko-kɔko.***
3 like 3 colour green ... green RDP-brown
'Its colour is like green, brownish green.' (KM_180820_cv03_75)

When a stative verb functions as an attributive modifier or a predicate, its reduplication can contribute to a plural reading of the head noun or the subject. Examples in (90) illustrate this function.

(90) Reduplication of stative verbs: plurality

a. KM
stai ɔɣɛ ***tɯwɔ-tɯwɔ****=lah.*
style(ENG) person RDP-old=SFP
'It's the old people's style.' (KM_180820_cv03_248)

b. CTM
yə laʔ wɔh pɛ hɔʔ ***kɔtɔ-kɔtɔ*** *... diyə laʔ ŋaŋ kaiŋ yə.*
3 wipe fruit pear REL RDP-dirty ... 3 wipe with cloth 3
'He wiped the dirty pears; he wiped them with his cloth.'
(CTM_180825_n02_9)

c. ITM
ləpah anɔʔ ***bəsɔ-bəsɔ*** *tah, adɛ=lah ɣəzəkɛi sikĩʔ-sikĩʔ.*
after child RDP-big DEM.DIST have=FOC livelihood RDP-little
'After the children grew up, we had a little bit of saving.'
(ITM_180923_n01_27)

Third, with dynamic verbs, reduplication often encodes continuation and iterativity of the actions, as shown in (91).

(91) Reduplication of dynamic verbs: continuation and iterativity

a. KM
pah kitɔ ***gɔlɛʔ-gɔlɛʔ*** *ggitu ah.*
then 1PL RDP-flip like.that INTERJ
'Then we keep flipping (the fish) like that.'
(KM_180820_cv03_83.2)

b. CTM
sakeʔ lutuʔ, sakəʔ məndə hɔ̃, ***uwaŋ-uwaŋ*** *stɔkiŋ, sakeʔ.*
hurt knee hurt what AFF RDP-throw sock hurt
'His knee hurt, something hurt ... then he was dusting his socks.'
(CTM_181025_n02_42.2)

c. ITM
diyɛ puŋ ***lamba-lamba*** *kə haʔ ɣama kkatɔʔ nəh.*
3SG then RDP-wave to REL many frog DEM.PROX
'Then he was waving at the many frogs.' (ITM_180907_n01_31)

Another common function of reduplicating dynamic verbs is to indicate casualness or aimlessness, as illustrated in (92).

(92) Reduplication of dynamic verbs: casualness

a. KM
diyɔ doʔ ***ita-ita*** *tɛŋɔʔ tuwɛ diyɔ tɔʔ gaba,*
3 PROG RDP-peep look owner 3 NEG pay.attention
diyɔ poŋ akaʔ.
3 then lift
'He peeped (casually), seeing that the owner wasn't paying attention, he just took (a basket).' (KM_180814_n01_22)

b. CTM
baleʔ-baleʔ *taʔdi, b-bukə pitu.*
RDP-return just.now NVOL-open door
'When I came back just now, the door was open.'
(CTM_220927_e02_123)

c. ITM
təŋɔh dimɛ ***dudəʊʔ-dudəʊʔ, minuŋ-minuŋ*** *ai,*
middle 3PL RDP-sit RDP-drink water
makaŋ-makaŋ *nɛh ...*
RDP-eat DEM.PROX ...
'While they were sitting around, drinking and eating (casually) ...' (ITM_180927_n03_3.1)

Reduplication can also apply to interrogatives, forming indefinite pronouns or pronominal adverbs with the meanings such as 'anywhere' or 'anything', as in the examples in (93).

(93) Reduplication of interrogatives

CTM

manə	'which; where'	→	*manə-manə*	'anywhere'
bilə	'when'	→	*bilə-bilə*	'anytime'

ITM

məndɛ	'what'	→	*məndɛ-məndɛ*	'anything'
kwanɛ	'to where'	→	*kwanɛ-kwanɛ*	'to anywhere'

Lastly, some reduplicated forms have a conventionalised meaning that cannot be immediately deduced from the base form. For example, KM *ɣupɔ-ɣupɔ* (RDP-appearance) means 'seemingly' (cf. CTM *ɣupə-ɣupə* and ITM *upɛ-upɛ*), and *tibɔ-tibɔ* (RDP-arrive) means 'suddenly' (cf. CTM *tibə-tibə* and ITM *tibɛ-tibɛ*).

A question worth exploring is whether initial gemination in NEPMs can be categorised as a special type of reduplication, i.e., the reduplication of a bare consonant. There are two different views on the more general relationship between gemination and reduplication in the literature. On the one hand, some scholars consider gemination as a type of partial reduplication, with the reduplicant being a single segment (Inkelas 2005, 2014; Rubino 2005, 2013). On the other hand, others argue that the reduplicative template must consist of well-defined prosodic constituents, the smallest of which being a mora (McCarthy & Prince 1986, 1995). The doubling of a segment like a consonant, which is not admitted as a proper prosodic constituent, is therefore excluded from reduplication. Whether initial consonant doubling can be considered as an instance of reduplication depends on how linguistic facts are formalised. In NEPMs, it makes sense to treat these two phenomena as separate morpho(phono)logical processes for the following reasons. For one, as discussed earlier, initial gemination is primarily derivational and related to prefixation and the cliticisation of certain prepositions, while full reduplication is closer to compounding, often carrying iconic meanings. Overall, initial gemination and full reduplication do not concur in their semantics and functions, except in cases where the bases are temporal nouns. Compare KM/CTM *p-pagi* with ITM *pagɛi-pagɛi* '(in the) early morning'; in both, initial gemination and full reduplication indicate intensity with an adverbial reading. Moreover, if initial gemination is seen as a subtype of reduplication, it would be the only type of partial reduplication. This created a pattern where the reduplicative template is either as small as a single consonant, or as big as a full phonological word, leaving a wide gap for all other types of prosodic constituents. Although this pattern is not necessarily problematic, on the whole the analysis seems unfavourable.

5.3.5 Fossilised complex words

The previous sections have examined how complex words are formed in NEPMs through various word-formation processes at the synchronic level. In addition, there are words that initially appear complex but are, in fact, derived historically and are no longer analysable synchronically, either because of the loss of one or more constituents in the original derivative or because of the contraction of earlier compounds or reduplicated forms. In this section, I introduce these fossilised complex words and explain why they are treated as such.

Examples of historical derivatives in NEPMs are presented in Table 5.8. The '|' sign marks the historical morpheme boundary. At first glance, these words appear to have derived from affixation, especially when compared with their SM correspondences: the first two sets may be seen as having the prefix *bɣ-* 'INTR; MID', and the forms with initial nasals could be associated with the prefix *NN$_1$-* 'IPFV'. However, none of the putative bases is attested as isolated words synchronically (e.g., ˣ*əti/ətɛi* or ˣ(*p*)*ike/*(*p*)*ikɛi*); as for the last two sets, neither the bases ˣ*bɛʔk/biʔk*, ˣ*mɔʔt/ambut* nor the suffixes ˣ*-i/-ɛi*, ˣ*-ɛ/-aŋ* are attested.

Table 5.8: Historical derivatives that are synchronically unanalysable

KM	CTM	ITM	SM	Gloss
bɣ|əti	*bɣ|əti*	*bɣ|ətɛi*	*bər-hənti*	'to stop'
b|ɣənɛ	*b|ɣənaŋ*	*b|unɔŋ*	*bə-rənaŋ*	'to swim'
mm|ike	*mm|ike*	*mm|ikɛi*	*məN-<p>ikir*	'to think'
nn|aɣi	*nn|aɣi*	*nn|aɣɛi*	*məN-<t>ari*	'to dance'
ŋŋ|uwaʔ	*ŋŋ|uwaʔ*	*ŋŋ|uwaʔ*	*məN-<k>uap*	'to yawn'
ŋŋ|ale	*ŋŋ|ale*	*ŋŋ|ali*	*məN-alir*	'to flow'
bɛʔk|i	*bɛʔk|i*	*biʔk|ɛi*	*baik-i*	'to repair'
mɔʔt|ɛ	*mɔʔt|aŋ*	*ambut|aŋ*	*rambut-an*	'rambutan'
nnis|ɛ	*nnis|aŋ*	*manis|aŋ*	*manis-an*	'palm sugar'

Also importantly, historical derivatives are indistinguishable from simple words in terms of their phonological properties. As mentioned in §2.4, §3.4, §4.4 and §5.2.2, simple words and (historical) derivatives have similar phonological shapes and are subject to the same phonotactic constraints. Con-

sider the following two sets of words in KM:

(94) KM

a.	*bɣəsiŋ*			‘to sneeze’
	bɣəti	*bɣ|əti*		‘to stop’
	bɣanɔʔ	*bɣ-anɔʔ*	(INTR-child)	‘to give birth; to be born’
b.	*nnatɛ*			‘animal’
	nnaɣi	*nn|aɣi*		‘to dance’
	nnanɛ	*NN$_1$-<t>anɛ*	(IPFV-plant)	‘planting’

In each set, the three forms have different morphological structures: *bɣəsiŋ* ‘to sneeze’ and *nnatɛ* ‘animal’ are simple words, *bɣ|əti* ‘to stop’ and *nn|aɣi* ‘to dance’ are historical derivatives, and *bɣ-anɔʔ* ‘to give birth; to be born’ and *nn-<t>anɛ* ‘planting’ are complex words. However, the phonological structure of words within each set is similar: all three words in (94a) have a CCVCV(C) shape with an initial *bɣ-* cluster, and all three words in (94b) have a CCVCV shape with an initial *nn-* cluster. There is no phonological difference between *bɣəsiŋ* ‘to sneeze’ and *bɣ|əti* ‘to stop’, and when the diachronic view is set aside, their morphological structures are identical. In other words, within the internal system of KM, the only reasons to consider words like *bɣ-anɔʔ* and *nn-<t>anɛ* as complex are the occurrences of their bases *anɔʔ* and *tanɛ*, and the form-meaning association between the derivative and the base.

Some fossilised complex words were originally compounds. A noteworthy example is the word for ‘sun’, namely KM/CTM *ttaɣi* and ITM *mataʔaɣɛi* (cf. SM *mata-hari* ‘sun’, lit. ‘eye-day’). KM/CTM *ttaɣi* apparently developed from the contraction of *mata-hari > $^{+}$matari > *ttaɣi* (involving the loss of *h and the merger of two *a, followed by syllable reduction; see Chapter 7 for more detail on sound changes). Synchronically, KM/CTM *ttaɣi* ‘sun’ cannot be further decomposed. The analysis for ITM *mataʔaɣɛi* is somewhat disputable, and it is perhaps best treated as a compound with a cranberry morpheme *mataʔ-* (cf. *matɛ* ‘eye’, which has a different shape).

Lastly, some NEPM words with initial geminates correspond to SM forms with full reduplication, suggesting that they may be fossilised reduplicated forms. Some examples are given in Table 5.9.

Table 5.9: Fossilised reduplication in NEPMs

KM	CTM	ITM	SM	Gloss
kkatɔʔ	*kkatɔʔ*	*kkatɔʔ*	*katak*	'frog'
kkuɣɔ	*kkuɣə*	*kkuɣɛ*	*kura-kura*	'(land) turtle'
ɣɣamɔ	*ɣɣamə*	*maʔamɛ*	*rama-rama*	'butterfly'
llabɔ	*llabə*	*glabɛ*[a]	*laba-laba*	'spider'
–[b]	*ppaɣu*	*ppaɣəʊ*	*paru-paru*	'lung'

[a] Initial *g-* is unexplained.
[b] KM *pləpoŋ* 'lung'.

These words cannot be analysed as complex given the absence of bases such as ˣ*katɔʔ* and ˣ*kuɣɔ/kuɣə/kuɣɛ*. It is likely they reflect earlier partial reduplication (e.g., ⁺kəkatak, ⁺kəkura) followed by regular deletion of antepenultimate vowels (see §7.5). The evidence is nevertheless circumstantial, only inferred from their correspondence with SM forms.

5.4 Summary

This chapter has provided an overview of the morphological systems of NEPMs, starting with a discussion of wordhood and other morphological units such as affixes and clitics. The examination then moved onto the formation of complex words through various morphological processes, and special attention was paid to fossilised complex words.

Words in NEPMs are primarily defined on phonological grounds, with evidence drawn from segmental features and phonotactics. Grammatical words often, but not always, coincide with phonological words. The overwhelming majority of words in NEPMs consist of only one morpheme, i.e., they are simple words. When a word consists of more than one morpheme, its internal structure is often relatively simple with only one affix. Based on this observation, the general isolating profile of NEPMs was motivated. Prefixes differ from words not only in their morphological boundness but also in their phonological shapes, as they are typically subsyllabic. An intermediate category between affixes and words is formed by clitics, which exhibit a wide range of phonological behaviour. Three subtypes of clitics may be

identified, namely affixal clitics, free clitics and weak words.

In terms of word-formation, complex words in NEPMs can be derived through prefixation and initial gemination, whereby a complex form with a geminate cluster C_xC_x- derives from a base with a singleton consonant C_x-. While initial gemination resembles prefixation and the cliticisation of prepositions with overlapping grammatical functions, a closer examination reveals that it must be recognised as a morphophonological process which does not utilise invariant segmental material. Other word-formation processes include compounding and reduplication. Reduplication in NEPMs can be considered a special type of compounding, as it is restricted to full reduplication and echo reduplication with reduplicants taking the shape of a root.

The morphology of NEPMs showcases several noteworthy features, both within the Malayic group and from a cross-linguistic perspective. First, NEPMs have notably small inventories of affixes (five or four) with a strong prefixing preference. Within the Malayic varieties, such reduced morphology is characteristic of the contact varieties in Eastern Indonesia (Adelaar 2005c; Paauw 2008). Despite being vernacular varieties, NEPMs nevertheless share a similar morphological profile with these contact varieties, which raises questions about the role played by language contact in the evolution of NEPMs (see more discussions in §8.4.3). Second, the prefixation process is severely limited by the phonological conditions on permitted clusters in word-initial position, which further exemplifies the interplay between phonology and morphology (§5.2.4). Lastly, the grammatical functions performed by prefixes are often overtaken by the morphophonological operation of initial gemination (see more discussions in Chapter 7 from a diachronic perspective). While morphological gemination is known in a few languages including Arabic and Alabama (Hardy & Montler 1988; El Zarka 2005), no previous reports of morphological gemination in word-initial position have been documented to my knowledge.