Universiteit Leiden
The Netherlands

Malayic varieties of Kelantan and Terengganu: description and linguistic history

Wu, J.

Citation

Wu, J. (2023, September 28). *Malayic varieties of Kelantan and Terengganu: description and linguistic history*. *LOT dissertation series*. LOT, Amsterdam. Retrieved from https://hdl.handle.net/1887/3642470

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/3642470

Note: To cite this publication please use the final published version (if applicable).

Malayic varieties of Kelantan and Terengganu

Description and linguistic history

Published by

LOT
Binnengasthuisstraat 9
1012 ZA Amsterdam
The Netherlands

phone: +31 20 525 2461

e-mail: lot@uva.nl
http://www.lotschool.nl

Cover illustration: edited version of a sketch map published by the Royal Geographical Society to illustrate the paper *A journey through the Malay states of Trengganu and Kelantan* (Clifford 1897).

ISBN: 978-94-6093-436-0
DOI: https://dx.medra.org/10.48273/LOT0651
NUR: 616

Copyright © 2023 Jiang Wu. All rights reserved.

Malayic varieties of Kelantan and Terengganu

Description and linguistic history

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr.ir. H. Bijl,
volgens besluit van het college voor promoties
te verdedigen op donderdag 28 september 2023
klokke 10.00 uur

door

Jiang Wu

geboren te Rui'an, China
in 1992

Promotores:	Prof. dr. Marian Klamer Prof. dr. Tom Hoogervorst (KITLV & State University of Malang)
Copromotor:	Dr. Alexander Adelaar (University of Melbourne & Palacký University Olomouc)
Promotiecommissie:	Prof. dr. Maarten Kossmann Prof. dr. Stefanie Shamila Pillai (University of Malaya) Dr. Martine Bruil Dr. Alexander D. Smith (National University of Singapore)

The research reported here was supported by the Dutch Research Council (NWO) under project number 322-70-013.

Contents

Acknowledgements xiii

List of abbreviations and symbols xvii

1 Introduction **1**
1.1 Overview 1
1.2 The Malayic languages 4
1.3 Peninsular vernacular Malayic varieties 8
1.4 Malayic varieties in Kelantan and Terengganu 10
1.4.1 Geo-historical settings 10
1.4.1.1 Geography, demographics and livelihoods 10
1.4.1.2 History 15
1.4.2 Basic linguistic facts 21
1.5 Previous research 26
1.5.1 Early studies 27
1.5.2 Linguistic studies from 1960s 29
1.6 Present study 32
1.6.1 Fieldwork locations and native speaker consultants 33
1.6.2 Data collection and processing 36
1.6.3 Transcription conventions and citation codes 39
1.7 Structure of this dissertation 40

2 Phonology of Kelantan Malay **41**
2.1 Introduction 41
2.2 Segment inventory 41

2.2.1 Consonant system 41
2.2.1.1 Consonant inventory 41
2.2.1.2 Contrasts between consonants 43
2.2.1.3 Phonetic realisations of consonants . . . 48
2.2.2 Vowel system 49
2.2.2.1 Vowel inventory 49
2.2.2.2 Contrasts between vowels 50
2.3 Syllable structure 53
2.4 Word structure 56
2.4.1 Structure of simple words 56
2.4.2 Structure of derivatives 59
2.4.3 Phoneme distribution at the word level 61
2.4.3.1 Distribution of consonants 61
2.4.3.2 Distribution of vowels 62
2.5 Consonant clusters 63
2.6 Consonant sequences 69
2.7 Vowel sequences 70
2.8 Stress 71
2.9 Summary 74

3 Phonology of Coastal Terengganu Malay 77
3.1 Introduction 77
3.2 Segment inventory 78
3.2.1 Consonant system 78
3.2.1.1 Consonant inventory 78
3.2.1.2 Contrasts between consonants 79
3.2.1.3 Phonetic realisations of consonants . . . 82
3.2.2 Vowel system 83
3.2.2.1 Vowel inventory 83
3.2.2.2 Contrasts between vowels 84
3.2.2.3 Phonetic realisations of vowels 87
3.3 Syllable structure 88
3.4 Word structure 90
3.4.1 Structure of simple words 90
3.4.2 Structure of derivatives 92
3.4.3 Phoneme distribution at the word level 94
3.4.3.1 Distribution of consonants 94

3.4.3.2 Distribution of vowels 95
3.5 Consonant clusters 96
3.6 Consonant sequences 100
3.7 Vowel sequences 101
3.8 Stress 102
3.9 Summary 104

4 Phonology of Inland Terengganu Malay 107
4.1 Introduction 107
4.2 Segment inventory 108
4.2.1 Consonant system 108
4.2.1.1 Consonant inventory 108
4.2.1.2 Contrasts between consonants 108
4.2.1.3 Phonetic realisations of consonants . . . 112
4.2.2 Vowel system 113
4.2.2.1 Vowel inventory 113
4.2.2.2 Contrasts between vowels 114
4.2.2.3 Phonetic realisations of vowels 116
4.3 Syllable structure 122
4.4 Word Structure 125
4.4.1 Structure of simple words 125
4.4.2 Structure of derivatives 127
4.4.3 Phoneme distribution at the word level 128
4.4.3.1 Distribution of consonants 128
4.4.3.2 Distribution of vowels 129
4.5 Consonant clusters 130
4.6 Consonant sequences 135
4.7 Vowel sequences 137
4.8 Stress 137
4.9 Summary 139
4.10 Comparison of NEPM phonologies 140

5 Morphology 143
5.1 Introduction 143
5.2 Morphological units 144
5.2.1 Words 144
5.2.2 Internal structure of words 148

5.2.3 Clitics 151
5.2.4 Interim summary 157
5.3 Word-formation 158
5.3.1 Prefixation 158
5.3.1.1 Morphophonological alternations 158
5.3.1.2 Prefix *bɣ-* 'INTR; MID' 161
5.3.1.3 Prefix *tɣ-* 'NVOL' 164
5.3.1.4 Prefix *pɣ-* 'CAUS; FCT' 167
5.3.1.5 Prefix NN_1- 'IPFV' 170
5.3.1.6 Prefix NN_2- 'NMLS' 175
5.3.1.7 Interim summary 177
5.3.2 Initial gemination 178
5.3.2.1 Grammatical functions of initial gemination 178
5.3.2.2 Initial gemination and prefixation 184
5.3.2.3 Initial gemination and the cliticisation of prepositions 188
5.3.3 Compounding 191
5.3.4 Reduplication 193
5.3.5 Fossilised complex words 199
5.4 Summary 201

6 Word classes and basic syntax 203
6.1 Introduction 203
6.2 Word classes 204
6.2.1 Nouns 204
6.2.2 Verbs 205
6.2.3 Adverbs 210
6.2.4 Pronouns 213
6.2.4.1 Personal pronouns 213
6.2.4.2 Relative pronouns 216
6.2.5 Demonstratives and deictics 217
6.2.6 Quantifiers and numerals 220
6.2.6.1 Numerals 221
6.2.6.2 Non-numeral quantifiers 224
6.2.7 Classifiers 225
6.2.8 Interrogatives 227

6.2.9 Negators 232
6.2.10 Prepositions 236
6.2.11 Conjunctions 241
6.2.12 Discourse particles 243
6.2.13 Interjections 246
6.3 Noun phrases 248
6.4 Basic clause structure 251
6.4.1 Verbal clauses 251
6.4.1.1 Dynamic verbal clauses 251
6.4.1.2 Stative verbal clauses 257
6.4.2 Non-verbal clauses 260
6.4.2.1 Nominal predicates 260
6.4.2.2 Quantifier/numeral predicates 261
6.4.2.3 Locational clauses 262
6.4.2.4 Existential clauses 264
6.5 Summary 267

7 Phonological history 269
7.1 Introduction 269
7.2 Some notes on the reconstruction of PM 270
7.3 Reflexes of PM consonants 272
7.3.1 PM stops 275
7.3.2 PM nasals 285
7.3.3 PM fricatives 287
7.3.4 PM liquids 289
7.3.5 PM glides 291
7.3.5.1 PM intervocalic glides 291
7.3.5.2 PM final glides 295
7.3.6 PM consonant sequences 300
7.4 Reflexes of PM vowels 308
7.4.1 PM penultimate *a and *ə 310
7.4.2 PM penultimate *i and *u 311
7.4.3 PM ultimate *a and *ə 312
7.4.4 PM ultimate *i and *u 315
7.4.4.1 Lowering in KM and CTM 316
7.4.4.2 Diphthongisation in ITM 318
7.4.4.3 Nasalisation 327

7.5 Syllable reduction 328
7.5.1 Syllable reduction in PM disyllables 328
7.5.2 Syllable reduction in PM trisyllables 329
7.5.2.1 Vowel contraction 330
7.5.2.2 Antepenultimate vowel syncope 331
7.5.2.3 Cluster assimilation 334
7.5.3 Interim summary 337
7.6 Relative chronology of sound changes 338
7.7 Summary 340

8 Morphological history 343
8.1 Introduction 343
8.2 Affixes reconstructed in PM 344
8.3 Morphological retention and innovation 346
8.3.1 PM *(mb)Ar-, *pAr$_1$- and *tAr- 347
8.3.2 PM *mAN- and *pAN- 352
8.4 Morphological reduction 356
8.4.1 Fossilisation of PM *-i and *-an 356
8.4.2 Loss of other affixes 360
8.4.3 Contact-induced change? 363
8.5 Summary 369

9 General discussion and conclusions 371
9.1 Overview 371
9.2 Summary of main findings 371
9.3 Implications for the Malayic migration history 377
9.4 Limitations and directions for future research 380

Appendices 383

A Basic word lists 383

B Sample texts 393
B.1 Sample text of KM 393
B.2 Sample text of CTM 399
B.3 Sample text of ITM 409

Bibliography 419

English summary 445
Samenvatting in het Nederlands 447
Ringkasan dalam Bahasa Melayu 449
Curriculum vitae 451

Acknowledgements

The completion of this dissertation owes a debt of gratitude to the remarkable people who have supported and helped me throughout this journey. First and foremost, my deepest thanks go to my wonderful supervisors Marian Klamer, Sander Adelaar and Tom Hoogevorst, whose guidance and inspiration have been invaluable. Without Marian encouraging me to apply for the NWO grant, this project would have never existed in the first place. From the very beginning of my PhD, she has been giving me the freedom to explore my own research interests, and she has always had confidence in me. I am grateful for Sander's professional input on Malayic languages, and his meticulous feedback on my manuscripts over numerous Skype calls. Tom provided insights on the etymologies of many words, and he offered me many coffees at 't Suppiershuysinghe, which were always accompanied by thought-provoking discussions.

Numerous other individuals have contributed to this dissertation in one way or another. I consider myself fortunate to have been surrounded by a group of like-minded people during the initial years of my PhD at LUCL. I wish to thank all the members of Marian's VICI group: Hanna Fricke, Owen Edwards, Francesca Moro, Gereon Kaiping, George Saad and Yunus Sulistyono. Owen read several chapters of my dissertation and offered valuable comments, and he introduced me to map-making software. Hanna gave me introductory courses on ELAN and FLEx, and she helped me with many technical issues. Additionally, all the feedback I received on my presentations in conferences and symposiums have enriched this work in meaningful ways. My former teacher and classmates in Beijing, Congcong Fu,

Huizhong Wang and Wenxue Jing, have engaged in inspiring chats with me and provided advice on various aspects related to Malay lexicon and grammar. Congcong and Huizhong also checked the Malay summary of this dissertation. I would also like to thank Thom van Hugte for proofreading the final manuscript and reviewing the Dutch summary.

About eight months of my PhD were spent in Malaysia, and those were unforgettable times thanks to the hospitality and companion of many people there. To all the speakers from Kelantan and Terengganu who generously hosted me and dedicated their time to teaching me their fascinating languages: *ribuan terima kasih!* I wish to thank Ihsan, Amirul and Asma and their families in Tanah Merah, Nurudin's family in Kuala Terengganu, and Pakcik Abdullah and Makcik Jamaliah in Ulu Terengganu, as well as all other language consultants whose names I cannot list here. Tarmizi and Karl Anderbeck brought me to the field in Kelantan and Terengganu for the first time in July 2018, which really opened up my fieldwork experience. I am indebted to Prof. Datuk Seri Awang Sariyan, Prof. Stefanie and Prof. Salinah for their assistance in establishing connections with the local communities. Nadia, Asilah and Aisyah assisted me in transcribing some parts of the data, and my friends in Kuala Lumpur, Zhenfeng Li and Wanzhen Zhao, made my stay in Malaysia much more enjoyable.

My PhD journey in the past six years would not have been the same without my fellow linguists at LUCL, many of whom have become much more than just colleagues to me: Zhen Li, Qing Yang, Tingting Zheng, Zhuoyi Luo, Rasmus Puggaard-Rode, Sarah von Grebmer Zu Wolfsthurn, Andrew Wigman, Lis Kerr, Priscilla Lam, Charlie Xu, Matthew Sung, Olga Nozdracheva, Xinyi Wen, Jiaqi Wang, Ruoyu Shi, Jian Sun, Hana Hasanah, Xander Vertegaal, Astrid van Alem, Cesko Voeten, Ami Okabe, Laura Smorenburg, Meike de Boer, Nurenzia Yannuar, Hang Cheng, Han Hu, Menghui Shi, Min Liu, Jinlei Zhou and Dan Yuan, as well as the rest of them. I would like to thank Zhen, Qing, Tingting and Zhuoyi for their camaraderie, for all the lunches and dinners we had, and for putting up with my grumbling and whining. Rasmus was a great officemate, who was always willing to discuss phonetics and phonology with me. I also received much help from Sarah in handling all the intricate details during the final stage of my PhD. To each and every one I have met in LUCL, I extend my gratitude for our conversations, no matter how trivial, and for the enjoyable moments we shared over drinks and social gatherings.

Lastly, I want to express my heartfelt gratitude to my friends and life partners who have provided me with invaluable companionship and made Leiden feel like home. I am deeply thankful to Stefano Mattia, who has been by my side throughout the years, offering unwavering support. To Minyan, Thom, Dannan, Siyun and Yi, my sincere appreciation for our genuine friendship. And to my parents and family, my deepest thanks for your immeasurable understanding and love.

List of abbreviations and symbols

Glosses

1	first person
2	second person
3	third person
ABST	abstractive
ACT	actor
ADVS	adversative
AFF	affirmative
AGT	agent
ANAPH	anaphoric
APPL	applicative
ATTR	attributive
CAUS	causative
CLF	classifier
COL	collective
DEM	demonstrative
DIST	distal
DISTR	distributive
EMPH	emphasis
EXIST	existential
EXCL	exclusive
FCT	factitive

FOC	focus
IMP	imperative
INCL	inclusive
INDEF	indefinitive
INST	instrumental
INTERJ	interjection
INTR	intransitive
INTS	intensifier
IPFV	imperfective
HES	hesitation
HUM	human referent
LIG	ligature
LOC	locative
MID	middle (voice)
NEG	negator
NMLS	nominaliser
NHUM	non-human referent
NVOL	non-volitional
PART	particle
PERS	person
PL	plural
PREP	preposition
PROG	progressive
PROH	prohibitive
PROX	proximal
Q	question
RDP	reduplication
RECP	reciprocal
REL	relativiser
RES	resultative
SFP	sentence final particle
SG	singular
SUBJ	subjunctive
TAG	tag question
TOP	topic
TR	transitive
VF	verum focus

Languages

AA	Austro-Asiatic
BH	Banjar Hulu
CTM	Coastal Terengganu Malay
ENG	English
ITM	Inland Terengganu Malay
KM	Kelantan Malay
NEPM	Northeastern Peninsular Malayic
PAn	Proto Austronesian
PK	Payang Kayu
PM	Proto Malayic
PMP	Proto Malayo-Polynesian
PWMP	Proto Western-Malayo-Polynesian
SM	Standard Malay

Citation codes

a	audio
cv	conversation
d	discussion and interview
e	elicitation
fn	field notes
n	narrative
v	video
wl	word list

Other abbreviations

A	agent
C	consonant
Env.	environment
f.n.	footnote
k.o.	kind of
n.	noun
NP	noun phrase
O	object
P	patient
p.c.	personal communication

PP	prepositional phrase
QP	quantifier predicate
R	recipient
S	subject
s.o.	someone
s.th.	something
SSP	Sonority Sequencing Principle
SVC	Serial verb construction
T	theme
V	vowel
VP	verb phrase

Usage of symbols

σ	syllable
#	word boundary
Ø	zero (phoneme or morpheme)
[ ]	phonetic form; phrase
/ /	phonemic form (also in *italics*)
⟨ ⟩	orthographic representation
< >	segment deleted or substituted during morphophonological processes
→ or ←	synchronic derivation
> or <	diachronic change or loanword origin
*	reconstructed form
+	inferred form at an intermediate stage
×	unreconstructable or unattested form
†	historical forms that are no longer attested
?	tentative or uncertain analysis
-	morpheme boundary
=	clitic boundary
|	historical morpheme boundary
~	free variation between two (or more) forms
≈	approximately equal to

CHAPTER 1

Introduction

1.1 Overview

This dissertation examines the Malayic varieties spoken in Kelantan and Terengganu, two Malaysian states located on the northeast coast of the Malay Peninsula. It focuses on three varieties, namely Kelantan Malay, Coastal Terengganu Malay and Inland Terengganu Malay, all belonging to the Malayic subgroup within the Austronesian language family. The area where these varieties are spoken is indicated in the map in Figure 1.1. The primary objectives of this study are twofold: first, to provide a synchronic description of these languages, and second, to offer a historical account of their development, which could shed light on the migration history of the speakers.

The following abbreviations will be used throughout the dissertation: KM for Kelantan Malay, CTM for Coastal Terengganu Malay and ITM for Inland Terengganu Malay. For ease of reference, the three varieties are also collectively referred to as Northeastern Peninsular Malayic varieties (henceforth NEPMs). The term "variety" is chosen to avoid the fuzzy distinction between "language" and "dialect". As will be discussed in more detail in §1.2, there is no clear differentiation between "non-Malay Malayic

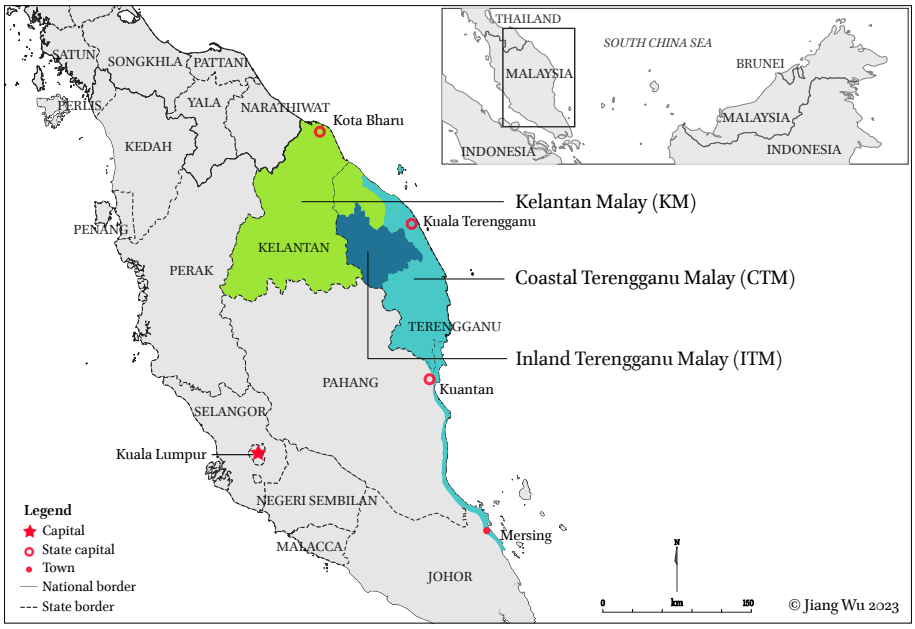

Figure 1.1: Malayic varieties in Kelantan and Terengganu

languages" and "Malay dialects" within the Malayic group. While the labels "non-Malay" and "Malay" can refer to ethnic groups, often based on religious and cultural practices, and sometimes political considerations, the distinction does not necessarily apply to the languages spoken by these ethnic groups. In Malaysia, the speech forms of ethnic Malays are considered dialects of Malay. However, as has been noted repeatedly, some of these "Malay dialects" are distinct enough to be unintelligible to speakers of Standard Malay (henceforth SM) (e.g., Blust 1988; Adelaar 2004b, 2018); those spoken in the northern states such as Kelantan and Terengganu are prime examples. It is likely that NEPMs should be considered separate languages in their own right, and for this reason, they are referred to as "Malayic varieties" rather than "Malay varieties".

In the field of Malayic and Austronesian linguistics, NEPMs, especially KM, are widely recognised for their unique structural features. They have attracted an extensive amount of scholarly interests since the late 19th century, and most Malaysian linguists from Kelantan and Terengganu have written about their own speech varieties. Nonetheless, despite the abundance

of literature that has provided valuable insights, many issues have not been fully addressed, and there remain a number of reasons why NEPMs deserve further investigation in this dissertation (see §1.5 for a more comprehensive appreciation of previous studies).

First of all, previous studies were typically engaged in the comparison between NEPMs and SM, rather than treating NEPMs as linguistic entities on their own. More importantly, focus was often given to the sound system and lexicon alone, with little attention paid to the morphosyntactic aspects. One objective of the present study is to provide a modern linguistic description of NEPMs by adopting a structural approach, covering both their phonology and basic morphosyntax.

Second, the study of NEPMs holds a significant place in Malayic historical linguistics. The Malay Peninsula is generally viewed as a late settlement of the Malayic-speaking people following their migration from the homeland in West Borneo via Sumatra (Blust 1985; Adelaar 2004b). This suggests that Peninsular Malayic varieties have a relatively short history, and they are often considered offshoots of court Malay as documented from the fifteenth century (from which SM is a direct descendant). Contrary to expectations, however, NEPMs exhibit some noteworthy retentions that are not present in other Peninsular Malayic varieties, as previously noted by Collins (1983a) for ITM. The second aim of this study is therefore to establish the diachronic development of NEPMs from Proto Malayic (henceforth PM), which can contribute to a more fine-grained internal classification of the Malayic subgroup and a deeper understanding of the Malayic migration history.

Third, the history of NEPMs is interesting from the perspective of contact linguistics. Along socio-historical lines, NEPMs are categorised as vernacular or "inherited" Malayic varieties (Adelaar & Prentice 1996). Nevertheless, NEPMs share certain similarities with contact varieties or so-called "Pidgin Derived Malay", as will be demonstrated in the current description. The region where present-day NEPMs are spoken is indeed a contact zone, with Aslian languages spoken in the inland of the Malay Peninsula and historical presence of Mon-Khmer languages, both groups belonging to the Austro-Asiatic (henceforth AA) family. Some earlier studies have posited that the peculiarities of NEPMs might be attributed to an AA substratum (e.g., Winstedt 1923: 96; Benjamin 1987, 1997). This hypothesis will be tested in this dissertation.

Finally, the documentation of NEPMs, ITM in particular, is of utmost importance in view of their language vitality and endangerment. ITM is spoken by only approximately 50,000 to 70,000 people in the inland area of Terengganu, and it is not being passed down to younger generations who tend to switch to the more prestigious CTM, which is the *de facto* standard variety in Terengganu (see more in §1.4.2). KM and CTM each have over a million speakers, and they are vigorously spoken across generations as an essential part of the local people's cultural identities. Even so, they are increasingly being influenced by SM.

The remainder of this introductory chapter provides further information on NEPMs and the area where they are spoken. §1.2 presents an overview of the Malayic languages with a focus on their historical background and classifications. §1.3 introduces the vernacular Malayic varieties spoken on the Malay Peninsula. §1.4 takes a closer look at NEPMs, examining the geo-historical settings of Kelantan and Terengganu and providing basic linguistic facts about NEPMs. §1.5 reviews previous studies on NEPMs. §1.6 explains the methodology, data collection and data processing in this study, and offers a summary of the transcription conventions. §1.7 outlines the structure of this dissertation.

1.2 The Malayic languages

The Malayic languages are a group of languages belonging to the Malayo-Polynesian branch of the Austronesian family, spoken primarily in island Southeast Asia. The Malayic subgroup includes Malay proper, the standardised forms of which are the national languages of Malaysia, Indonesia, Brunei and Singapore, a large variety of Malay dialects, and various languages that are sufficiently close to Malay. The total number of Malayic-speaking population is difficult to estimate, but Malay proper alone has almost 280 million speakers (including those who speak Indonesian as a second language, Adelaar 2018: 571). The dispersal and distribution of the Malayic languages are depicted in Figure 1.2. The figure also shows that the core Malayic-speaking areas are West Borneo, Sumatra and the Malay Peninsula.

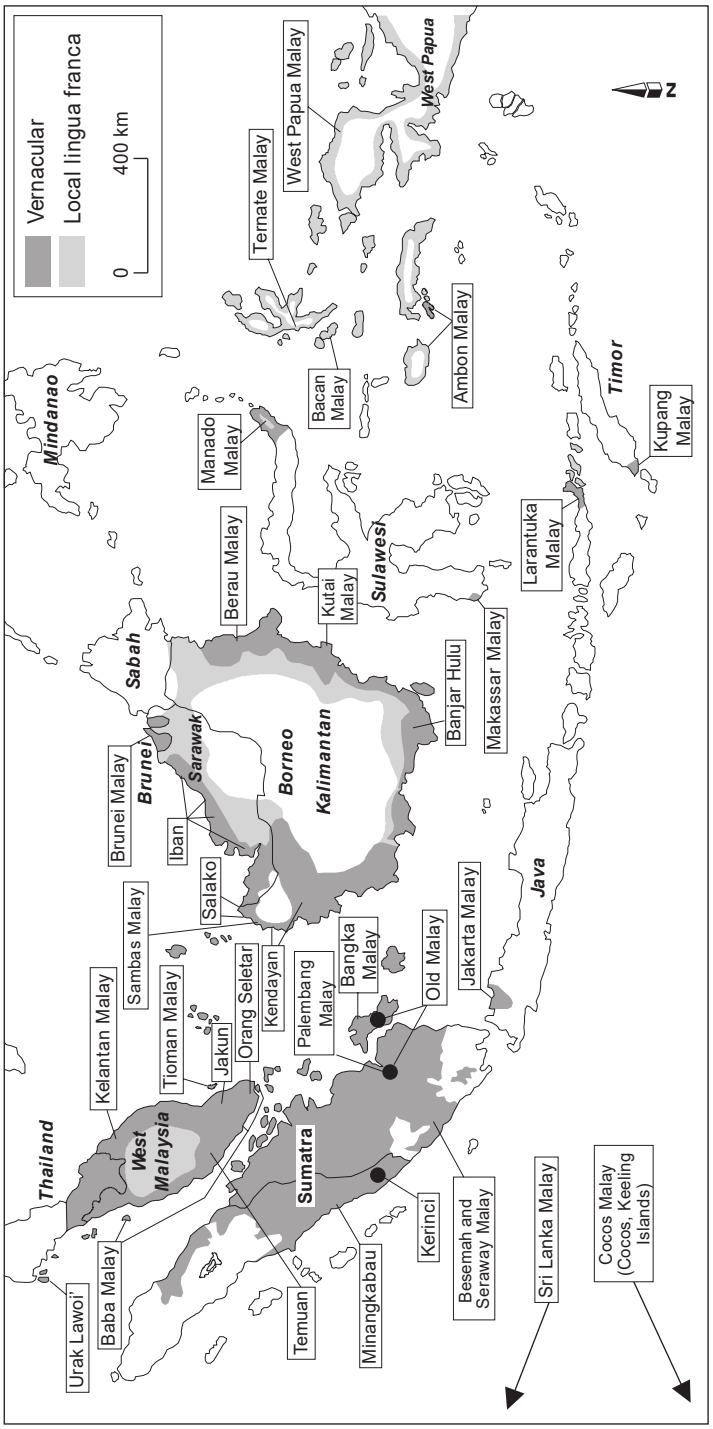

Figure 1.2: The spread of Malayic varieties (modified from Adelaar 2005c: 203)

The establishment of Malayic as a subgroup within the Malayo-Polynesian languages is on the basis of a set of shared innovations that all Malayic languages have undergone since Proto Malayo-Polynesian (henceforth PMP), which can be reconstructed to a common ancestral language, namely Proto Malayic (PM). The reconstruction of PM in Adelaar (1992) was primarily based on six Malayic varieties: SM, Minangkabau (central-west Sumatra), Banjar Hulu (southeast Borneo), Seraway (southwest Sumatra), Iban (northwest Borneo) and Jakarta Malay (Java).

The internal subgrouping of Malayic, on the other hand, has been much disputed. Even before Malayic was well recognised as subgroup, a distinction was often made between Malay dialects and non-Malay languages, usually based on non-linguistic criteria. In Borneo, for instance, Malay dialects refer to the speech forms of ethnic Malays who are Muslims, whereas if the speakers do not consider themselves as Malay on ethnic, cultural or religious grounds, their speech forms are regarded as separate languages (Hudson 1970). This differentiation has its usefulness, but an undesired linguistic implication is that the so-called Malay dialects are perceived as genetically closer to one another, and that they constitute a lower-level group, i.e., a Malay group within Malayic. However, the demarcation between the hypothetical Malay group and the Malayic group, whose members supposedly descend from two distinct proto languages, has never been made explicit, and the scope of languages descending from "Proto Malay" remains unclear. As Blust (1988: 1–5) shows, Banjar Malay is commonly taken as a Malay dialect whereas Iban is not, but on the basis of lexical evidence, Banjar Malay and Iban are equally distinct from SM. Blust (1988: 6–7) further suggests that northern Peninsular Malayic varieties such as Kedah Malay and Terengganu Malay, which are traditionally taken as Malay dialects, might not be much closer to SM than non-Malay languages such as Minangkabau and Kerinci. Asmah (1995) intended to reconstruct Proto Malay (*bahasa Melayu induk*), where all Peninsular Malayic varieties were conveniently included, but the boundary of languages belonging to her Malay group was not well defined, and some varieties showing important retentions such as ITM were overlooked. The labels "Malay dialects" and "non-Malay languages" can still be found in more recent literature, but it should be borne in mind that the distinction is often arbitrary and not indicative of genetic distance.

Various subgrouping proposals based on more convincing linguistic evidence have been put forth by Adelaar (1992, 1993, 2008), Collins (1994),

Nothofer (1996, 1997), Ross (2004), Anderbeck (2012) and Smith (2017), but a detailed classification has not yet been reached (see appraisals in Adelaar 2005d: 17–19 and Anderbeck in print). A general consensus is that West Bornean languages such as Iban and Kendayan (also known as Kanayatn) are distantly related to SM, representing primary branch(es) in the Malayic family tree. The majority of other Malayic varieties, including all Peninsular varieties, cannot be satisfactorily classified into finer-grained groups due to the lack of clear exclusively shared innovations. They are often grouped together as belonging to one single branch, which has been referred to variously as "Nuclear Malayic" (Ross 2004, which serves as the basis for the classification on *Glottolog 4.7*, Hammarström et al. 2023), "other Malayic" (Smith 2017), or simply "Malay" (Anderbeck 2012). Based on the highest linguistic diversity and retentions attested in languages in West Borneo, scholars generally agree that this region is the prehistorical homeland of the Malayic languages (Blust 1985, 1988, 1994; Adelaar 1988, 1992, 1995, 2004b; Nothofer 1996, 1997; Collins 2001, 2006). Additionally, the spread of languages towards the interior in Borneo also suggests a longer period of diffusion. Southeast Sumatra is traditionally taken as the cradle of Malay civilisation and culture, where Malayic speakers founded the maritime empire Srivijaya and developed a separate Malay identity, leaving behind the earliest inscriptions written in Old Malay dating back to the seventh century (Andaya 2001: 317; Andaya & Andaya 2017: 31–32; Adelaar 2004b: 4–5). The Malay Peninsula, on the other hand, is commonly considered as a late settlement of Malayic-speaking population, as evidenced by the demographic pattern where Malays dominate the coasts and push Orang Asli (Malay for 'aboriginal people') further inland (Skeat & Blagden 1906: 434; Bellwood 1993; Adelaar 1988: 74, 2004b: 4).

A supplementary classification has been made along socio-historical lines. Three broad categories were recognised by Adelaar & Prentice (1996): 1) literary Malay, 2) lingua franca Malay and 3) "inherited" Malay. A number of other terms with similar meanings were used in later publications: literary Malay has been referred to as Court Malay, Classical Malay or standard varieties; lingua franca Malay as vehicular Malay, trade Malay or Pidgin Derived Malay; and "inherited" Malay as vernacular varieties (Adelaar 2005c, 2018; Paauw 2008; Anderbeck in print).[1] Originally intended for

[1] The scope of these terms is not always the same, and they are not necessarily mutually

categorising different forms of Malay, this classification was later expanded to include other Malayic languages (see, for instance, Adelaar 2005c). Literary Malay is the predecessor of present-day SM, which developed in Malay courts across the region from the fifteenth century. Vehicular Malay refers to the contact varieties spoken mainly in Eastern Indonesia and other ports throughout island Southeast Asia, which likely arose against a certain socio-historical background with a pidginised form of Malay as a common source. "Inherited" or vernacular Malay(ic) are varieties spoken in traditional Malayic speech communities in Borneo, Sumatra and the Malay Peninsula, practically comprising all other Malayic varieties that appear to have directly inherited from PM, including non-Malay languages such as Iban and Kendayan.

1.3 Peninsular vernacular Malayic varieties

With few exceptions (e.g., SM as the literary variety and Baba Malay, which is a vehicular variety), Peninsular Malayic varieties are vernaculars along socio-historical lines. They are primarily spoken by ethnic Malay groups, hence typically known as Malay dialects. There are also some Malayic-speaking Orang Asli groups such as Temuan and Jakun, and a few groups of Malayic-speaking Orang Laut (Malay for 'sea people') including Orang Seletar and Urak Lawoi' (see Figure 1.2).

This diversity already captured the attention of British lexicographers and grammarians during colonial times. While there had not been dedicated studies on any particular non-standard Malay(ic) variety, notes on regional variation were included in some early Malay dictionaries and grammars in the nineteenth century. The grammar by Crawfurd (1852: 75–76) briefly mentions that Malay dialects often differ in pronunciation and the usage of personal pronouns. The Malay–English dictionary by Clifford & Swettenham (1894: vi) contains a section more specifically on local peculiarities of the Peninsular dialects, where the authors outline the pronunciation of various dialects and note that "the states of Patani and Kelantan are more rich in local words than any other places in the Peninsula and there the low-

exclusive. For example, some vernacular varieties such as Minangkabau and Jakarta Malay are also used as lingua francas. It is therefore best to avoid the term "lingua franca Malay" and restrict the second category to "vehicular Malay".

est form of Malay is spoken." Another English–Malay dictionary (Shellabear 1916: iv) comments that the Malay language is spoken in considerably diverse dialects across the islands in the archipelago, and the varieties spoken on the east coast of the peninsula differ particularly from those on the west coast.

Contemporary Peninsular Malaysia consists of eleven states, and it is often claimed that each state has its own dialect (e.g., Asmah 1977, 1985 and their revised versions published in 1991 and 2008; Nik Safiah et al. 1986: 30–32; Abdul Hamid 1994: 1–2). However, as Collins (1989) points out, this "canon of Malay dialects" corresponding to the state boundary grid is far from the reality. As an example, consider "Terengganu Malay". This name is commonly used to refer to the Malay dialect spoken in the state of Terengganu as if it is a homogeneous variety, but the label is ill-defined for several reasons. First, not all populations in Terengganu speak "Terengganu Malay"; people from the northern district Besut and part of Setiu are predominantly KM-speaking. Second, "Terengganu Malay" is not only spoken in Terengganu, but also in fisherman's villages thinly stretching southwards along the coast to Johor (Ismail 1973; Collins 1989). The term Coastal Terengganu Malay (CTM) is therefore more appropriate. Lastly, the population in the inland area of Terengganu speaks a highly distinct variety which has been referred to as Ulu Terengganu or Inland Terengganu Malay (ITM). While ITM is often considered a subdialect of "Terengganu Malay", it is in fact not necessarily closer to CTM than to KM.

The exact number of distinct Malayic varieties and their boundaries remain to be studied further, but suffice it to say, among all Peninsular varieties, those spoken in the northern states stand out with marked features. This observation was already noted at the turn of the twentieth century (Clifford & Swettenham 1894; Winstedt 1923), and it has been confirmed by later studies on some of the varieties spoken in northern states including Kelantan and Terengganu (see §1.5). There have also been a number of overviews of the diversity of Malay dialects on the Malay Peninsula (Ismail 1973; Farid 1976: 112–132; Teoh 1994: 104–107), or in Malaysia as a whole (Asmah 1977, 1995). From these, it is evident that the varieties spoken in Kelantan and Terengganu are among the most divergent ones, most notably for the remarkable sound changes they have undergone and the specific usage of some local words.

1.4 Malayic varieties in Kelantan and Terengganu

Before outlining the basic linguistic facts about the Malayic varieties spoken in Kelantan and Terengganu, it is crucial to first examine the geo-historical settings of the region, as presented in §1.4.1. This examination is essential as it illuminates how the development and distribution of languages are intrinsically connected to the geographic environment and the historical background within which they have evolved. It also lays the foundation for a deeper understanding of the linguistic characteristics and historical evolution of NEPMs.

1.4.1 Geo-historical settings

1.4.1.1 Geography, demographics and livelihoods

In the Malay Peninsula, as in many other places in Southeast Asia, the most important natural features with which local communities interact are the rivers and the sea (Dobby 1942; Miksic 1978: 170). Prior to the nineteenth century, traditional Malay communities typically depended on the rivers and the sea for their livelihood; the Malays in Kelantan and Terengganu were no exceptions. Villages were established by riverbanks or coastlines before roads were built, where water routes served as the primary means of movements and communication. Another geographical trait characterising Kelantan and Terengganu is the surrounding mountain ranges, which largely isolate these states from the rest of the peninsula and have posed great impediments to trans-peninsular movements until recent times (see, for example, Swettenham 1885; Clifford 1897). The geographical details of Kelantan and Terengganu are provided in greater depth in this section, which also encompasses information about the demographics and livelihoods of the populations in these states.

Kelantan is the largest Malaysian state on the Malay Peninsula, spanning a total area of 15,040 km^2. It is bordered by the Narathiwat Province of Thailand to the north, Perak to the west, Pahang to the south and Terengganu to the southeast (see Figure 1.1). Its geographic boundaries are relatively well defined, with the Golok River marking the Malaysian-Thai border, the jungle-clad Titiwangsa Range extending over the Kelantan-Perak boundary, the Tahan Range delimiting Kelantan from Pahang, and

the Pantai Timur Range covering a large part of the Terengganu frontier. Bounded by these mountain ranges lies a low-lying and flat alluvial plain, with the Kelantan Delta situated at the estuary of the Kelantan River.

The Kelantan River, which is named after the name of the state (or might have given its name to the state), is fed by several major tributaries that originate in the south and southwest of the state, as illustrated in Figure 1.3. The Nenggiri River (also known as the Betis River in its upper reaches) has its headwaters in the Titiwangsa Mountains and flows eastward, merging with the Galas River at Kuala Sungai. From there the Galas River flows northeast wards and merges with the Lebir River. The Galas River and the Lebir River both originate in southern Kelantan near the border with Pahang, and they converge at Kuala Krai to form the Kelantan River.

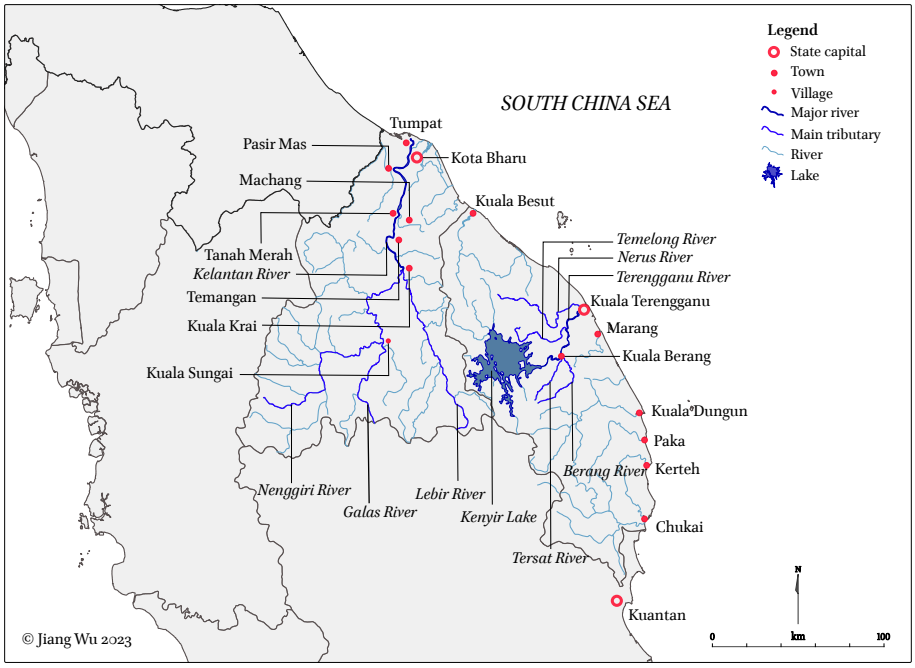

Figure 1.3: River systems of Kelantan and Terengganu

The Kelantan River and its tributaries play a vital role for the local communities, providing water for agriculture and supporting fishing, which are crucial sources of livelihood. The importance of the river system can be seen from the placement of main settlements in the state: from Kuala Krai, ma-

jor towns along the banks of the Kelantan River include Tanah Merah, Pasir Mas, Kota Bharu, the capital city of Kelantan, and Tumpat. The rivers must have been of great importance even in ancient times, as evidenced by the discovery of Gua Cha, one of the most significant archaeological sites on the peninsula, located on the bank of the Nenggiri River (Sieveking 1954; Adi 1985).

According to the 2020 Malaysian Population and Housing Census (available at https://www.dosm.gov.my/v1/), Kelantan has a population of approximately 1.79 million, among which the Bumiputera (Malay for 'sons/daughters of the soil', comprising Malays and indigenous groups including Orang Asli of the peninsula) make up 96.6%. Chinese make up another 2.5% of the population, and Indians constitute 0.3%. The Kelantanese Malays practise Islam. Politically, the Islamic Party of Malaysia has been ruling Kelantan uninterruptedly for over three decades. The party has been pushing for the gradual enforcement of Shari'a laws in the state, leading to the state's reputation as one of the most conservative Malay heartlands alongside Terengganu.[2]

The Kelantanese Malays have a long tradition of practising intensive wet-rice agriculture, and the fertile soil of the Kelantan Delta has made it a major centre of rice production in Malaysia (Dobby 1951; Hill 1951; Cheng 1969). Rice cultivation remains a significant part of the state's economy, along with rubber-tapping, which is another traditional economic activity in the village sphere (Downs 1960; Nash 1974). Other crops grown in Kelantan include oil palms, coconuts, cassava and various vegetables and fruits such as durians, papayas and rambutans. Fishing has also been an important source of livelihood for fisherman's villages along the coastline (Graham 1908: 65; Firth 1943, 1966; Norfatiha & Nor Hayati 2022). In more recent years, Kelantan's economy has become more diversified, with increasing investments in manufacturing activities and tourism. Traditional agriculture is becoming less attractive to young generations, and as the population grows, some parts of the traditional rice paddies have been cleared to make room for housing developments.

[2] The Islamic customs, such as the way of dressing, appear to have been rather different a century ago, see Graham (1908: 24–26, 31–33).

Situated to the southeast of Kelantan, Terengganu (formerly also spelled Trengganu or Tringganu) has an elongated shape covering an area of 13,035 km^2. Its geography is characterised by a long coastline along the South China Sea, stretching over 200 kilometres. The coast strip consists largely of open sandy beaches, which extend further south into Pahang and north into Kelantan. From the east to the west, the flat coastal plain gradually rises to hilly terrain in the interior, where the mountains form Terengganu's borders with Kelantan and Pahang.

Terengganu has several rivers that stem from the mountainous interior and flow towards the sea, each supporting an important town at their estuaries along the coast (see Figure 1.3). This pattern is quite different from the one in Kelantan, where traditional settlements are primarily located along the Kelantan River and its tributaries. The Terengganu River, which flows through the state capital Kuala Terengganu, is the largest and most prominent river in Terengganu. It originates in the highlands in the central part of the state, fed by several main tributaries including the Berang, Tersat, Telemong and the Nerus River. Another major town along the Terengganu River is Kuala Berang, which is near the confluence of the Berang, Tersat and the Terengganu River. In addition to the Terengganu River, other notable rivers in the state include the Besut River in the north, the Marang River, the Dungun River and the Kemaman River in the south, which support the towns of Kuala Besut, Marang, Kuala Dungun and Chukai respectively. Just like in Kelantan, these rivers serve as important waterways for transportation and commerce, as well as providing water for agriculture and other uses.

Figure 1.3 also shows that a large portion of the upstream Terengganu River is now submerged by the Kenyir Lake, which is the largest man-made lake in Southeast Asia. The lake was formed by damming several tributaries of the Terengganu River for the purpose of generating hydroelectric power. The construction of the Kenyir Dam and the creation of the Kenyir Lake between 1978 to 1985 have considerably altered the landscape of the interior of Terengganu, as evident from a comparison between the present-day map of river systems and the depiction in Firth (1943: 194), as shown in Figure 1.4. The project also led to the relocation of several villages, both Malay and Orang Asli ones.

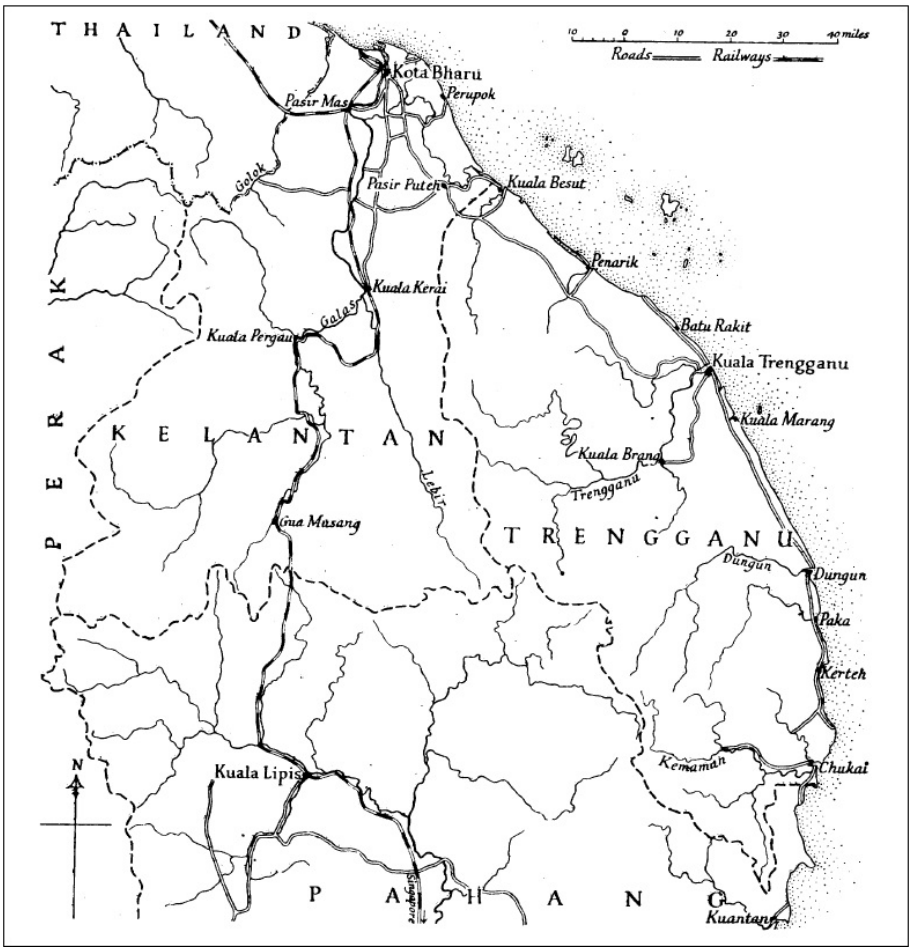

Figure 1.4: A map of Kelantan and Terengganu before the 1980s
(Firth 1943: 194)

As of 2020, Terengganu had a population of around 1.15 million, with the percentage of Bumiputera standing at 97.6%. Chinese and Indians make up a small percentage of the population at 2.1% and 0.2% respectively. Along with Kelantan, Terengganu is one of the Malaysian states with the highest concentration of ethnic Malays who practise Islam. Currently, Terengganu is also ruled by the Islamic Party of Malaysia. The primary economic activities in Terengganu used to be agriculture and fishing. Apart from rice, rubber, oil palms and coconuts, other important crops grown in the state in-

clude corns, peanuts, peppers, cucumbers and watermelons (p.c. with consultants). Thanks to the state's extensive coastline, fishing and other related industry such as boatbuilding continue to play an important role in the economy and cultural heritage (Norfatiha & Nor Hayati 2022, for the historical significance of the fishing industry in Terengganu, see Firth 1943, 1966 and Gosling 1978). However, the discovery of oil wells off the coast in the 1970s, especially in the southern areas of Kerteh and Paka, has significantly transformed Terengganu's economic structure to become heavily reliant on the oil and gas industry. In recent years, Terengganu has also seen growth in the tourism sector, particularly with the popularity of its coastal islands.

1.4.1.2 History

While ethnic Malays dominate contemporary Kelantan and Terengganu, this was likely not the case in the past. The northern part of the Malay Peninsula and the Kra Isthmus further north have been a contact zone for centuries, if not millennia. This region has long been where the speakers of Austronesian languages from the south and east met with the Austro-Asiatic and Thai people passing down to the peninsula from the north. Such interactions often led to conflicts, but during peaceful times, this area benefited from its strategic location along early trade routes connected to the South China Sea, attracting foreign travellers and traders from China, India, the Middle East and Europe. This section provides a concise history of Kelantan and Terengganu, highlighting the interactions and power transitions between various groups of people.

Ancient kingdoms established in the region can be dated back to as early as the first centuries of the Christian era. Ptolemy's map, which was drawn based on the Roman geographer's book *Geography* composed in the second century, shows two ports on the east coast on the Malay Peninsula, marked as Perimula and Coli polis (or Koli polis, Kole polis). The locality of these two ports has been variously identified by historians as corresponding to present-day Nakhon Si Thammarat (Ligor) and Kelantan (Gerini 1909: 105–111), or at the mouths of the Terengganu River and the Kemaman River (Braddell 1936: 37), or somewhere near the Kuantan River in present-day Pahang (Linehan 1951: 94; Wheatley 1955: 16). Chinese historical records also attest to the existence of political entities in this region in the early years of the Christian era. *Han-shu* [The book of the Han Dynasty], which includes a

Treatise on Geography, notes a country named 都元 (now read Du-yuan) in the first century, which is sometimes believed to be related to present-day Dungun in southern Terengganu (Hsu 1961: 97). The seventh-century *Liang-shu* [The book of the Liang Dynasty] recorded a mission to Funan (southern Indochina, present-day Cambodia and southern Vietnam) during the Wu Dynasty in the third century. The record also mentioned that Funan attacked several countries including 屈都昆 (Qu-du-kun), 九稚 (Jiu-zhi) and 典孙 (Dian-sun). As these countries were documented as being situated across the gulf from Funan, they must have been on the east coast of the Malay Peninsula.[3] 九稚 (Jiu-zhi) was also known elsewhere as 拘利 (Ju-li), which, according to Wheatley (1955: 15–16), likely corresponds to Kole in Ptolemy's map. While we know little more than the names and approximate locations of these ancient kingdoms (see a summary of different interpretations in Wheatley 1973: 14–25, 152–155), these early records demonstrate that the northeast coast of the Malay Peninsula was already home to important settlements that attracted travellers from both the west and the east.

The northern Malay Peninsula attained more prominence between the fifth and the seventh century, owing to the decline of Funan's power to the north and the emergence of extensive trading networks. A number of Hindu-Buddhist kingdoms existed in this region around the sixth century, as attested by Chinese dynasty annals such as *Liang-shu* and *Sui-shu* [The book of the Sui Dynasty]. Some of the most important kingdoms include 狼牙修 (Lang-ya-xiu or Langkasuka), 赤土 (commonly transliterated as Chi'tu) and 丹丹 (Dan-dan or Tan-tan), and from various sources it is clear that they lay next to each other from the north to the south along the east coast of the peninsula. Hsu (1947, 1961: 161–166) identifies Chi'tu as in present-day Songkhla and Tan-tan in Kelantan, whereas Wheatley (1973: 36, 55) locates Chi'tu in the upper reaches of the Kelantan River, and Tan-tan in Terengganu. As recorded in *Sui-shu*, Chi'tu was an advanced kingdom, to which an embassy was sent in the year 607. It was described as a Mon-Khmer kingdom founded by the descendants of Funan, where Hinduism was practised. Langkasuka was located in the northern part of the peninsula, generally suggested as the predecessor of the later Patani

[3] 屈都昆 (Qu-du-kun) is probably the same country that was referred to as 屈都乾 (Qu-du-qian), 屈都 (Qu-du) or 都昆 (Du-kun) in several other early works (Wheatley 1973: 21–22). *Jin-shu* [The book of the Jin Dynasty] recorded an event of 屈都乾 (Qu-du-qian) being invaded by Champa in the fourth century.

Kingdom (Wheatley 1956, 1973; Teeuw & Wyatt 1970: 1–3; Bougas 1990; Andaya & Andaya 2017: 74). The archaeological sites discovered in the city of Yarang seem to support this proposal (Wales 1974; Jacq-Hergoualc'h 2002: 166–191). *Liang-shu* mentions that Langkasuka was probably established around the second century, and its king sent an envoy to China in 515. Various names related to Langkasuka recurred in Chinese historical records until the fifteenth century, making it one of the most long-lasting kingdoms in the region. It is likely that the territory of Langkasuka extended to Kelantan and Terengganu in its heydays after Chi'tu and Tan-tan went into decline (Sheppard 1949).

From the seventh century onwards, Srivijaya rose to power as a maritime empire centred in Sumatra, and eventually gained control of much of the Malay Peninsula and surrounding islands. Langkasuka, together with other kingdoms located on the Malay Peninsula at that time, was recorded as paying tributes to Srivijaya in the thirteenth-century *Zhu-fan-zhi* [A description of barbarian nations]. It is also in this record that the names Kelantan (吉蘭丹 Ji-lan-dan, which is currently the state's official Chinese name) and Terengganu (登牙侬 Deng-ya-nong) first appeared, indicating that these states assumed independent identity before the thirteenth century, although still vassals of Srivijaya. The same source also referred to a neighbouring country called 佛罗安 (Fo-luo-an or Fo-lo-an), which is thought to be located in present-day Kuala Berang in the interior of Terengganu, where the Terengganu inscription stone was later discovered (Wheatley 1973: 70, also see below). These countries (or city-states) practised Buddhist culture (just like their suzerain Srivijaya), and produced local goods such as gharuwood, lakewood, sandalwood and ivory.

By the early fourteenth century, Srivijaya had fallen. The Siamese Ayutthaya Kingdom in the north and the Javanese Majapahit Kingdom in the south began to rise and exert influence on the Malay Peninsula. The states on the peninsula likely maintained semi-independence as small principalities. The names 吉蘭丹 (Ji-lan-dan) and 丁家盧 (Ding-jia-lu) occurred in *Dao-yi-zhi-lüe* [A brief account of island barbarians] written around 1339 (Rockhill 1915), and 古蘭丹 (most likely a misprint of 吉蘭丹, i.e., Ji-lan-dan) and 丁架路 (Ding-jia-lu) are shown in the Mao Kun Map which documents the voyages of Admiral Zheng He (or Cheng-ho) between 1403 and 1433. Both states are mentioned as dependencies of Majapahit in the fourteenth century Javanese poem *Nagarakṛtāgama* (Winstedt 1935: 30; Pigeaud

1960: 17), and Terengganu (written as 丁機宜 Ding-ji-yi) is also recorded in *Ming-shi* [The history of the Ming Dynasty] as a vassal of Majapahit.

The fourteenth century also saw the transition from Hindu-Buddhism to Islam in island Southeast Asia. The Terengganu inscription stone, written in Classical Malay in the Jawi script (a writing system based on the Arabic script), symbolises the presence of Malay and Islamic influence on the peninsula. The inscription has a date that can be read variously between 1303 and 1387, and it describes a proclamation by a Terengganu ruler who claimed that Terengganu was the first state to receive Islam and provided basic Shari'a laws for the guidance of his subjects (Paterson 1924; Andaya & Ishii 1992: 514). The introduction of Islam to the east coast of the peninsula therefore predated Malacca's conversion to Islam, which probably took place during the reign of Sultan Megat Iskandar Shah around 1414 (Wake 1964; Cœdès 1968: 246; also see Mills 1930: 49; Teeuw & Wyatt 1970: 4).[4] The Islamisation of Kelantan presumably happened around the same time or somewhat later in the mid-fifteenth century. According to *Ming-shi*, the Maharaja of Kelantan 苦马儿 (Ku-ma-er) sent an embassy to China in 1411 (Rentse 1934: 47), and *Sejarah Melayu* [The Malay annals] describes an event of Malacca invading Kelantan around 1500 for not paying homage and mentions the name Sultan Mansur Shah of Kelantan (Winstedt 1938: 12). The names and the titles of the rulers suggest that Kelantan was still an Indianised state in 1411, but it had already embraced Islam by the end of the fifteenth century.

As Malacca quickly grew in power in the fifteenth century, Terengganu became integrated into the Malacca Sultanate by the time of Sultan Muhammad Shah (ca. 1424–1444). Kelantan also became a vassal of Malacca following the attack around 1500 (Winstedt 1938: 5, 12). According to *Sejarah Melayu*, the ruling family of Terengganu was allegedly murdered by the sultan of Pahang (who was an elder brother of the sultan of Malacca) in 1478, after which a former Pahang governor's family ruled Terengganu for over a century (Linehan 1936: 14–15; Sheppard 1949: 5–6). Following the Portuguese conquest of Malacca in 1511, the last sultan of Malacca retreated to Johor, where his heir established a new ruling dynasty, with which Terengganu maintained close ties. The history of Kelantan in this

[4] Islam mostly likely spread to the peninsula directly from Sumatra, where it had already been introduced by 1281 (Cœdès 1968: 202, 231).

period, on the other hand, was intertwined with that of Patani to the north. Built on the ground of former Langkasuka, Patani emerged around the middle of the fifteenth century and soon adopted Islam (Teeuw & Wyatt 1970: 3–4; Bougas 1990: 115). After the fall of Malacca, Patani became a prosperous and important trade centre that was favoured by Chinese and Muslim merchants. It was also frequently visited by Portuguese, Dutch and English traders. Despite some internal disruption and strong Siamese influences, Patani reached the peak of its prosperity in the early seventeenth century (Teeuw & Wyatt 1970: 5–20; Ibrahim Syukri 1985: 13–38; Andaya & Andaya 2017: 73–75). It appears that the significance of Kelantan diminished dramatically following the rise of Patani and the southward expansion of Siamese pressure, and it is likely that Kelantan was divided into small chiefdoms that subordinated either to Patani or Terengganu (Graham 1908: 38–39; Andaya & Andaya 2017: 73). According to some sources, much of Kelantan had already been incorporated into Patani by the time of the reign of Raja Biru (ca. 1616–1624) (Abdullah Mohamed 1981: 21–22; Andaya & Andaya 2017: 75). This is also testified by the Chinese record *Dong-xi-yang-kao* [Notes on Eastern and Western Oceans], compiled in 1617, in which Kelantan was described as a port of Patani. There was nevertheless a period of Kelantanese rule in Patani in the second half of the seventeenth century or the early eighteenth century, suggesting a close relationship between the two states (Teeuw & Wyatt 1970: 20–22; Andaya & Andaya 2017: 76). Oral traditions, as summarised in Rentse (1934), also tell that the ancestors of the royal family of Kelantan came from overseas and first arrived in Patani. The prosperity of Patani came to an end when the Patani-Siam relationship deteriorated in the late seventeenth century. The city was eventually invaded and destroyed by the Thais in 1786, leading to Patani's complete subjugation to Thai rule (Teeuw & Wyatt 1970: 23; Ibrahim Syukri 1985: 41–44).

The current Terengganu Sultanate was established in 1725 by Sultan Zainal Abidin, a younger brother of a former Johor sultan who took refuge in Terengganu after being expelled (Sheppard 1949: 8–11). Meanwhile, Kelantan was ruled by many local chieftains after the decline of Patani. Following a period of disorder, a local chief named Long Pandak from Kubang Labu came into power. Eventually, Long Yunus, the son of an admiral to Raja Long Pandak, managed to unify Kelantan with the assistance of Sultan Mansur Shah of Terengganu, and was enthroned as the Sultan of Kelantan

in 1793 (Marriott 1916: 17; Rentse 1934: 51–53). Both states were nevertheless struggling to maintain their independence after attaining sultanate status, as the Siamese Kingdom of Rattanakosin once again began to expanded its influence southwards with greater demands. Both Kelantan and Terengganu were sending *bunga mas* (Malay for 'golden tree', a form of tribute) to the Thai king by the late eighteenth century and early nineteenth century, and it was clear that Siam viewed Kelantan and Terengganu as its tributary states (Newbold 1839: 61–65; Rentse 1934: 59; Sheppard 1949: 19; Wyatt 1974; Andaya 1986). In fear of absorbed by the Thais, Kelantan and Terengganu sought assistance from the British, who had already gained a strong hold in the southern part of the peninsula. In 1822, Kelantan petitioned the British to be accepted as a vassal state, but their plea was in vain (Andaya & Andaya 2017: 128–129).

The Thai claim to suzerainty over Kelantan and Terengganu lasted until the beginning of the twentieth century. With the Anglo-Siamese Treaty 1909, Kelantan and Terengganu (together with the Kedah and Perlis on the west coast) were transferred to the British control and became British protectorates known as the "Unfederated Malay States". Each state received a British advisor while keeping their own local ruler. After the Second World War, Kelantan and Terengganu joined the Malayan Union in 1946 and subsequently became part of the Federation of Malaya in 1948, which ultimately gained independence in 1957.

The history of Kelantan and Terengganu summarised above reveals that these states have undergone several distinct phases of development, and they have been populated by different groups of people over time. There is strong evidence indicating that Malays, who adopted Islam in the fourteenth to fifteenth centuries, have dominated both states since then. However, little is known with certainty about the populations or languages spoken in this region before the fourteenth century. Chinese historical records suggest that sixth-century kingdoms such as Chi'tu were possibly Mon-Khmer in their culture and population composition. By the thirteenth century, the northern peninsula came under the influence of Srivijaya, which was a Malay state (Cœdès 1968: 82–83). Nevertheless, it is unclear whether Srivijiya's vassal states on the peninsula were still Mon-Khmer states or if they had already been dominated by Malays. According to the founding legend of Patani, the kingdom developed from a coastal village established by Malays from the southern peninsula and Sumatra, while its

direct predecessor was an inland Mon-Khmer kingdom (Ibrahim Syukri 1985: 13–14, who used the term "Siam-Asli" to refer to Mon-Khmer). The legend may not be far from the truth. As mentioned earlier, the predecessor of Patani was presumably Langkasuka, which lay to the north of Chi'tu in the sixth century. Like its neighbour, Langkasuka was probably a Mon-Khmer kingdom as well, and it might have remained so until it was replaced by the Malay Kingdom of Patani in the fifteenth century. Citing Benjamin (1997), Andaya (2001: 324–328) also concludes that the Malay Peninsula was not considered part of the "Melayu lands" before the time of the Malacca Sultanate, and the northern part of the peninsula received a particularly greater influence from Mon-Khmer culture. More concrete evidence of this influence can be found in present-day Sathing Phra to the north of Patani, where artefacts and the citadel discovered on the archaeological sites are believed to be of Mon character dating back to the sixth to the thirteenth century (Stargardt 1983: 32). However, no such evidence can be found in Patani, let alone further south in Kelantan and Terengganu. In short, historical evidence suggests that the northern part of the peninsula likely underwent a transition from being dominated by Mon-Khmer culture and population to its present-day Malay dominance, yet the exact period of this transition cannot be precisely determined.

1.4.2 Basic linguistic facts

There are at least three sufficiently distinct Malayic varieties spoken in Kelantan and Terengganu. KM is believed to exhibit relative homogeneity across the state (Ismail 1973), and this observation is largely confirmed by my personal experiences visiting various districts in Kelantan including Kota Bharu, Tumpat, Pasir Puteh, Pasir Mas, Tanah Merah and Machang. More recently, it has been pointed out that there are notable variations between the coastal variety and the inland variety spoken along the Nenggiri River (Mohd Tarmizi 2018a, b, c), which calls for further study on the regional variation of KM.[5] The KM-speaking area extends to the northern

[5] A reappraisal of Tarmizi's data gives the impression that both varieties share most of the typical KM features, with minor differences that can be seen in the reflection of two sound changes. First, penultimate high vowels have been sporadically lowered in the coastal variety, but they are retained in some inland varieties, e.g., PM *ikur > coastal [ɛkɔː], inland [ikuː] 'tail'; *uraŋ > coastal [ɔɣɛ̃], inland [uɣaŋ] 'person' (following the author's tran-

districts in Terengganu including Besut and Setiu, and potentially spans across the Kelantan-Pahang and Kelantan-Perak borders. KM also shares a close relationship with Patani Malay spoken in the southernmost provinces of Thailand across the border.[6] To what extent KM and Patani Malay resemble or differ from each other still needs to be demonstrated through systematic comparisons, but the available material suggests that they may be considered the same variety, as they are closely related on the one hand, and sufficiently different from other Malayic varieties on the other hand (Tadmor 1995: 13). The ISO 639-3 code *mfa* is assigned to Kelantan-Patani Malay (Eberhard et al. 2023). It is estimated that there are around two million speakers of KM in Malaysia, while Patani Malay has over one million speakers in southern Thailand.

In Terengganu, as previously noted, the varieties spoken along the coast and in the inland area are highly divergent in many aspects. In fact, throughout this dissertation, it will become clear that CTM is closer to KM than to ITM. Collins (1989: 251) reports that CTM is spoken in a narrow strip of sometimes discontiguous villages along the the east coast, from the north of Kuala Terengganu to at least Mersing in Johor (see Figure 1.1). It is also spoken by dwellers in the middle course of various rivers in Terengganu, likely as a result of the movement of inhabitants from the coastal area towards the inland. In the upper valleys of the Terengganu River and its tributaries, which form the district of Hulu Terengganu, villagers residing along the rivers speak various forms of ITM. However, Kuala Berang, the main town of the Hulu Terengganu district, is primarily CTM-speaking. Another vernacular variety spoken in the upper valleys of the Dungun River, known as the Pasir Raja dialect, appears to be closer to the Ulu Tembeling dialect of Pahang Malay (Mohd Tarmizi 2020). Unfortunately, the scanty data on this variety does not allow further discussion. Neither CTM nor ITM has been recognised as a distinct Malay(ic) variety by *Ethnologue*, and neither has been assigned an ISO code. It is estimated that there are around one million speakers of CTM, while the number of ITM speakers is significantly

scription). Second, *-an, *-am and *-aŋ merged to [-ɛ̃] in the coastal variety, whereas in the inland variety *-an and *-am merged to [-ɛ] but *-aŋ is retained as [-aŋ], e.g., PM *ik**an** > coastal [ikɛ̃], inland [ikɛ] 'fish'; *mal**am** > coastal [malɛ̃], inland [malɛ] 'night'; but *ur**aŋ** > coastal [ɔɣɛ̃], inland [uɣ**aŋ**] 'person'.

[6] There is a variant spelling of Pattani Malay (with two "t"s), especially in English and Thai contexts. In the present study, I use the Malay spelling with one "t" for consistency.

lower, likely ranging from 50,000 to 70,000. Additionally, there is concern for the endangerment of ITM as it is not being passed down to younger generations. People under the age of twenty generally do not speak ITM, or only have passive knowledge of it.

Like other vernacular varieties, NEPMs are considered low varieties or basilects, and they are only used in informal settings. Formal education and administration are conducted in SM, while religious matters are typically handled in Arabic. For example, during a funeral in a village in Kelantan, the official welcome would be given in (colloquial) SM by the master of ceremony, while the eulogy would be delivered in Arabic. Guests would likely converse with each other in KM. A diglossic situation like this has probably persisted for centuries in the region, with commoners using the vernacular variety and the royal courts using some form of literary Malay. Language use in Hulu Terengganu is particularly interesting, as villagers who grow up speaking ITM often acquire both CTM and SM. When a villager goes to the market in Kuala Berang, conversations with vendors typically occur in CTM. CTM is also used when conversing with local police officers or in restaurants outside the village. In comparison, during an elementary school sporting event that I attended in Kampung Dusun, all official announcements were made in SM. Hulu Terengganu therefore represents a triglossic situation where ITM is the basilect or the lowest variety, CTM is a mesolect, and SM represents an acrolect or a high form of Malay. Nowadays, there is a radio programme called *GEGAR* with the slogan *nombor satu di Pantai Timur* 'number one on the East Coast', which is broadcast in vernacular varieties targeting East Coast Peninsular listeners, but it is typically mixed with colloquial SM. There are also some famous *syair* and *sajak* (forms of Malay poetry) in ITM in Hulu Terengganu, but it is probable that they were originally written in SM and later translated into ITM. Traces of formal usage of SM can be seen in the morphosyntactic structure of these poems, which deviates from the everyday usage of ITM and may seem unnatural to many speakers.

Given their exclusive use in informal settings, NEPMs lack a standardised orthography and are typically not written. For literate younger generations, SM is the preferred written language, although informal communication such as texting and social media may contain unsystematic forms of spelling that reflect local pronunciation. For older generations who did not receive formal education in SM but had religious education, literacy is

largely restricted to reading the Qur'an. Their texting conventions tend to reflect the pronunciation of the vernacular variety, but they are not always used systematically.

It is worth noting that speakers of NEPMs have different ways of referring to their local vernacular, and they are well aware of the distinction between the three varieties. In Kelantan, KM is commonly known as /lɔɣaʔ klatɛ/ 'the Kelantan dialect' (commonly written as ⟨loghat klate⟩), and there is a popular phrase ⟨kecek klate⟩ /kɛcɛʔ klatɛ/ 'to speak Kelantanese', which is also well known outside Kelantan as it reflects the peculiarities of both word usage and pronunciation of KM (cf. SM *cakap kəlantan* 'to speak Kelantanese'). In coastal Terengganu, CTM is referred to as ⟨loghat tganong⟩ /lɔɣaʔ tɡanoŋ/ 'the Terengganu dialect'. In village settings, speakers also use the phrase /cakaʔ kapoŋ/ (cf. SM *cakap kampuŋ*, 'to speak the village variety') to refer to speaking CTM. SM, on the other hand, is referred to as /bahasə suɣaʔ/ (cf. SM *bahasa surat*, 'letter language'). In Hulu Terengganu, speakers use the word /uləʊ/ 'inland, upstream' (or its cognates with variable pronunciation, cf. SM *hulu*) to refer to ITM. They also refer to SM as /bahasɛ suɣaʔ/ 'letter language', and CTM is considered the "city variety", as in the phrase /cakaʔ bandɔ/ 'to speak the city variety' (cf. SM *cakap bandar*). The reported mutual intelligibility is that KM and CTM speakers may be able to understand each other, and ITM speakers can understand both CTM and KM, but neither CTM nor KM speakers understand ITM without sufficient exposure.

Linguistically, the distinctions among NEPMs are primarily marked in the phonological systems (see Chapters 2 and 4), but the three varieties also share many common sound patterns, as shown in Table 1.1. In all three varieties, only three consonants /ʔ, ŋ, h/ are allowed in word-final position, which reflect the merger of earlier final stops to /ʔ/, final nasals to /ŋ/ (with further nasal deletion following *a in KM), and PM *s and *h to /h/. The morphosyntactic features of NEPMs also exhibit more similarities than differences in various aspects (see Chapters 5 and 6).

Table 1.1: A comparison of some words in NEPMs

KM	CTM	ITM	SM	PM	Gloss
Merger of final stops to /ʔ/					
*saya**ʔ***	*saya**ʔ***	*saya**ʔ***	*saya**p***	*saya**p**	'wing'
*laŋi**ʔ***	*laŋi**ʔ***	*laŋi**ʔ***	*laŋi**t***	*laŋi**t**	'sky'
*tase**ʔ***	*tase**ʔ***	*tasɛi**ʔ***	*tasi**k***	*tasi**k**	'lake'
Merger of final nasals to /ŋ/					
tanɛ	*tana**ŋ***	*tana**ŋ***	*tana**m***	*tana**m**	'to plant'
*kiɣi**ŋ***	*kiɣi**ŋ***	*kiɣi**ŋ***	*kiri**m***	*kiri**m**	'to send'
*buɣo**ŋ***	*buɣo**ŋ***	*buɣəʊ**ŋ***	*buru**ŋ***	*buru**ŋ**	'bird'
Merger of *s and *h to /h/					
*ata**h***	*ata**h***	*ata**h***	*ata**s***	*ata**s**	'top'
*pəcɔ**h***	*pəcɔ**h***	*pəcɔ**h***	*pəca**h***	*pəca**h**	'to break'

Based on these shared characteristics, a "Northeastern Peninsular Malay dialect subgroup" has been proposed, which, according to Tadmor (1995: 13–14), includes the varieties spoken in Kelantan, Terengganu, Pahang and southeastern Thailand. Collins (1989: 253–254) previously presented a similar version of this subgrouping, suggesting that KM, CTM, ITM and Pahang Malay must have formed a single dialect group at some point in the past. In *Glottolog 4.7* (Hammarström et al. 2023), "Northeastern Peninsular Malay" is considered a single branch consisting of three subbranches: Kedah-Perak Malay, Kelantan-Patani Malay and Urak Lawoi'. "Terengganu Malay" is classified as a member of the Kelantan-Patani branch, alongside Kelantan, Pahang and Patani-Nonthaburi Malay. The classification of this "Northeastern Peninsular Malay" group is illustrated in Figure 1.5.

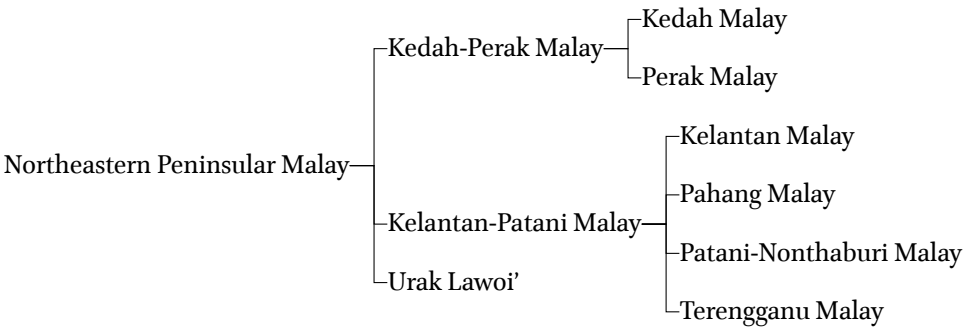

Figure 1.5: Subgrouping of Northeastern Peninsular Malay on *Glottolog 4.7*

However, the basis of these subgrouping proposals is unclear, and the proposed subgroups are not defined by exclusively shared innovations.[7] With a more detailed examination of the historical phonology of NEPMs, I will show that the three varieties do not, in fact, form a lower-level subgroup within Malayic (see Chapter 7). The similarities in their synchronic sound patterns likely have resulted from later diffusion rather than being inherited from an immediate common ancestor.

In the present study, the term "Northeastern Peninsular Malayic" (NEPM) will be used to collectively refer to the varieties spoken in Kelantan and Terengganu for the sake of convenience, but it should be noted that the term is not intended to define a genealogical relationship, but only serves to characterise the geographical area where these varieties are spoken.

1.5 Previous research

Studies on NEPMs so far have covered a wide range of topics with varying degrees of quality. This section presents an overview of the essential literature,

[7] Ajid (2008) also proposed a Patani-Kelantan-Terengganu subgroup, suggesting that Patani Malay and KM have a closer relationship against Terengganu Malay. Unfortunately the methodology used in this study was flawed, and no concrete evidence was presented to support the proposed subgroup. The result of KM and Patani Malay sharing a closer relationship is also hardly surprising, as these two varieties could well be considered the same variety, as mentioned earlier. Moreover, even if all the varieties shown in Figure 1.5 form a tightly-knit subgroup, a more accurate term would be "Northern Peninsular Malay(ic)", as neither Kedah-Perak nor Urak Lawoi' is spoken in the northeast region of the peninsula.

which can be divided into two phases, with the 1960s serving as a dividing line. Early studies primarily consisted of unsystematic observations on the structural peculiarities of NEPMs, while more systematic linguistic research began in the 1960s.

1.5.1 Early studies

The earliest mention of Malayic varieties spoken in Kelantan and Terengganu probably occurs in Munshi Abdullah's *Kisah pelayaran Abdullah ke Kelantan* [The story of Abdullah's voyage to Kelantan], which recounts the writer's experiences on a voyage from Singapore to Kelantan via Pahang and Terengganu in 1838. On the Malay spoken in Kelantan, Abdullah (1949: 44) (translated by A. E. Coope) wrote: "they speak Malay, but their pronunciation is very ugly; they lisp as Tamils do when speaking Malay. Often they leave out a final 't' and add final 'g' and change 'a' to 'o'. But they do not make these changes when writing." In Terengganu,

> "though the people speak Malay, their Malay differs from that of other Malays and sounds strange to the ear; their accent is like that of Kedah Malays. They have a trick of adding a 'g' at the end of words; thus where we say "tuan", they say "tuang"; for "jangan', they say "jangang" and for "bulan", "bulang"; and they say "Alloh" instead of "Allah". This trick however extends only to their speech; they write as we do." (Abdullah 1949: 20–21)

In another early handbook on Kelantan, Graham (1908: 34) also noted that "the Kelantan dialect is a fearsome-sounding jargon in the ear of the Malay of other parts, full of strange clippings and contortions, and sprinkled with words of local manufacture of a Siamese origin, unknown in any other parts of Malaya". These observations were mostly anecdotal, but even nowadays, Malay speakers in Malaysia generally believe that the dialects spoken in Kelantan and Terengganu are unintelligible to speakers from other parts of the country, except perhaps Pahang.

Scholarly studies on Peninsular Malayic varieties began to appear in the early twentieth century. Among the initial contributions to the studies on NEPMs were Sturrock's *Some notes on the Kelantan dialect, and some comparisons with the dialects of Perak and Central Pahang* (1912), Pepys' *A Kelantan glossary* (1916) and McKerron's *A Trengganu vocabulary* (1931), all

authored by officers of the Malayan Civil Service during the colonial period. These studies aimed to document words and expressions specific to these states, along with observations on obvious differences in pronunciation. Lloyd (1921) is another notable contribution, containing transcriptions of native songs and chants recorded in the states of Patani, Kelantan and Kedah ("Lower Siam" in the author's words) sung in the local Malay dialect.[8] The paper offers the first phonetic transcriptions of Kelantan-Patani Malay using International Phonetic Alphabets with great precision. Additionally, the observations made from the comparison between Kelantan-Patani Malay and SM, such as Kelantan-Patani Malay corresponding to SM with changes "occurring chiefly at the end of words, or, at times, of syllables" (Lloyd 1921: 37), are particularly insightful by the standards of their time.

The most important sources from this period are Brown (1927) on KM and Brown (1935) on CTM (referred to as "Terengganu Malay"), later recompiled together with Brown (1921) on Perak Malay and reprinted as Brown (1956). These books comprise of dialogues preceded by brief introductions and general remarks on the usage of personal pronouns, local words and expressions, as well as the pronunciation of these dialects. Brown's works contain informative and mostly accurate observations, and they can further be appreciated in their value in comparative dialectal study. For instance, it was already noted that KM and CTM share many similarities, while ITM reveals striking dissimilarity (Brown 1935: 1, 1956: 124). Unfortunately, the data are transcribed in a confusing and inconsistent orthography that does not correspond to actual pronunciations; instead, the orthography represents how words would have been spelled in SM, which reflects the "true" forms or "a representation of the Malay words in a familiar guise" (Brown 1927: 14) – a doctrine that was also practised in most previous studies on Malay dialects. For examples, KM [tɔʔsɛ] 'not want to' is written as ⟨tak sir⟩ and [ɣɔyaʔ] 'to tell' is written as ⟨ruwiyat⟩ (Brown 1927: 6); the latter is also written as ⟨royat⟩ and ⟨riwayat⟩ elsewhere. While the "standardised" spelling may inform us of the origins of local words, e.g., KM [ɣɔyaʔ] 'to tell' corres-

[8] Though not stated explicitly, it was implied that a single dialect was spoken in these northern states, for which some phonological notes were provided. However, upon examination of the transcriptions and phonological notes, it becomes evident that the recorded dialect was some form of Kelantan-Patani Malay. It also appears that the song recorded in Kedah was in the same dialect, which led to the misconception that northern Malay dialects were homogeneous.

ponds to SM *riwayat* and ultimately comes from Arabic *riwāya*, it disguises the peculiarities of these Malay dialects and makes it difficult for readers with no prior knowledge of these dialects to understand, as they have to remember all the rules while reading the words and dialogues.

On the whole, earlier studies tended to take SM as the point of departure, with which dialects were compared. This approach was based on the unspoken assumption that dialects are secondary to the written language, i.e., they are Malay pronounced in a deviant or an improper way. As a result, the goal was often to identify the "true" forms of local words and expressions, which could be disguised by the use of SM-based orthography (also see comments in Teeuw 1961: 43). As Teeuw (1957: 295) points out in his review of Brown (1956), "it would be important to consider also the Malay dialects in their own rights, and not apriori to make them secondary to and dependent upon the written standard". In spite of these criticisms, earlier works such as Brown's compilations of dialogues remain important sources for understanding the history of NEPMs.

1.5.2 Linguistic studies from 1960s

Modern linguistic research on NEPMs (and Peninsular Malay varieties in general) began in the 1960s. With the exception of works by James T. Collins, almost all studies were carried out by local Malaysian scholars, most of whom are natives of Kelantan or Terengganu.

Nik Safiah's MA thesis (1965) on KM phonology, along with two subsequent articles (1966, 1967) in *Dewan Bahasa*, is among the first studies that treat KM as an independent linguistic entity and offer a systematic analysis of its sound system. While the overall quality is high, the analysis is not always consistent. For instance, in Nik Safiah (1965), she lists nineteen phonemic consonants (/p, b, t, d, c, ɟ, k, g, ʔ, m, n, ɲ, ŋ, s, ɣ, h, l, w, j/) and provides illustrations for contrastive nasal vowels (although it is not specified whether they are phonemic). In contrast, in her later works (1966, 1967), she charts twenty phonemic consonants with the addition of /z/, and no mention is made of nasal vowels. Abdul Hamid (1971) is a BA thesis on the phoneme inventory of KM, and the book titled *Sintaksis dialek Kelantan* [Syntax of the Kelantan dialect] (1994) by the same author is to date the most comprehensive grammar sketch of KM. A summary of the KM phonological history based on Abdul Hamid's data can be found in Adelaar (2005c: 210–212).

However, Abdul Hamid's transcription is quite inconsistent throughout his book, and his analysis of the KM morphological system is particularly debatable. Additionally, Hashim's MA thesis (1974) on KM morphemes and Ajid's (1985) work on the phonology and lexicon of KM (represented by the variety spoken in Pasir Mas) are also noteworthy.

Compared to KM, CTM has received less scholarly attention. While not dedicated to studying CTM, Collins (1980) provides a comparison of Ambon Malay with CTM based on the dialogues compiled by Brown (1935), highlighting the isolating structure of CTM. Abdul Hamid (1990) offers an overview of CTM phonology and some aspects of its morphosyntax. Other BA theses that have been cited in the literature include Othman Omar (1983) on CTM phonology and Kamsiah Salleh (1990) on CTM morphology, but unfortunately I do not have access to them.

ITM has been studied even less extensively but with considerable depth, thanks to Collins' book (1983a), which remains one of the most influential works on a Peninsular Malay dialect. Collins identifies some distinct features of ITM, including the retention of historical high vowels in penultimate syllables and diphthongisation of high vowels in final syllables. It is also noted that ITM has a small inventory of affixes, and passive constructions in ITM are formed with a pre-verbal anaphoric marker *ŋə* or *ɲi*. Also importantly, Collins argues that ITM should not be taken as a subdialect of the ill-defined "Terengganu Malay", but rather a distinct variety on its own. Other published works on related topics include Collins & Naseh Hassan (1981) and Collins (1983b), as well as two unpublished BA theses by Abdullah Junus (1977) and Naseh Hasan (1981). Many of Collins' earlier observations on ITM will be further elaborated and advanced in this dissertation.

In the past two decades, a significant amount of research has been conducted on NEPMs, with a primary focus on phonology and the application of theoretical considerations. Dialectology studies are typically conducted in the context of geolinguistics (see Mackey 1988), aiming to map the differentiation and boundaries of dialects (e.g., Nor Hashimah, Wan Athirah & Khairul 2021; Nor Hashimah, Wan Athirah & Harishon 2021). Mohd Tarmizi (2018a) is an important study focusing on the Malay(ic) varieties spoken in the inland/upstream area on the east coast of the Malay Peninsula and their history, which is particularly relevant to the present dissertation. He hypothesises that the inland area preserves older forms of Malay, as indicated by certain archaic features, and proposes that the spread of Malay varieties

originated from inland regions and then expanded towards the coast. However, I have reservations about the phonological analysis in this work (see below). Moreover, the evidence presented for the spread of Malayic varieties from the inland to the coast is thin (see general discussion in §9.3). Mohd Hilmi et al. (2016, 2018) conducted excellent work on the acoustic aspects of word-initial geminates in KM, following earlier studies on initial geminates in Patani Malay (Abramson 1986, 1987, 2003). The phonological aspects of initial geminates in KM are discussed in §2.2.1. Noraien Mansor et al. (2013) wrote a short monograph on general features of CTM, but it does not offer much advancement compared to previous summaries such as Ismail (1973) and Asmah (1985).

Overall, it is regrettable that, with few exceptions, the morphosyntactic aspects of NEPMs have been generally overlooked, and basic descriptive studies are still lacking. It is also unfortunate that the quality of existing phonological studies is often disappointing, for two reasons.

First, a major issue with many existing phonological studies is the lack of systematic phonemic analysis and differentiation between phonetic realisations and phonological representations. Very often only the phonetic forms are transcribed, and no further attempts are made towards a phonological analysis. For instance, vowel length is sometimes marked in "phonological studies" (e.g., Mohd Tarmizi 2018a, b), yet there is no justification for considering it a distinct feature. Siti Nadiah's thesis (2020) on ITM monophthongs essentially takes all phonetically distinct vowels as phonemes, disregarding some clear allophonic alternations.

Second, a more common and serious problem arises from the lack of distinction between diachronic changes and synchronic derivations. SM or PM approximations have often been taken for granted as the underlying forms from which the phonetic realisations of NEPMs are derived using a set of convoluted rules. Adi Yasran (2005; 2010) analyses the consonant and vowel inventories of KM and formulates the derivation of KM surface forms within the framework of Optimality Theory. However, his analysis lacks justification for the underlying forms of the words being considered. For example, KM [ayɛ] 'chicken' is taken as having derived from the underlying form /ayam/ (cf. SM *ayam*) with the application of phonological rules including vowel raising and nasal deletion. Yet there is no reason to assume that the underlying form of KM [ayɛ] is /ayam/, and a more straightforward analysis would be to simply take /ayɛ/ as the underlying form. Similarly,

Sharifah Raihan (2018) discusses the realisation of consonant clusters consisting of a nasal + a voiceless obstruent in KM and other Malay dialects. She takes KM [gatoŋ] 'to hang' as the surface realisation of the underlying form /gantoŋ/, whereby /n/ in /-nt-/ is deleted in root-internal position. However, the foundation of such an analysis is also problematic: /gantoŋ/ is merely the earlier form from which KM [gatoŋ] developed *diachronically*, which happens to have survived in SM. In the synchronic sound system of KM, there is no evidence that nasal-voiceless obstruent clusters exist at all, even at the underlying level. Similar problems can be found in other publications on NEPMs (as well as other Peninsular Malay varieties), such as Zaharani (2006), Zaharani et al. (2011), Adi Yasran (2011, 2012), Nur Adibah & Sharifah Raihan (2017) and Mohd Tarmizi (2018a, c) and Nor Hashimah, Wan Athirah & Harishon (2021).

The manifestation of this issue culminates in the compilation of glossaries such as *Glosari dialek Terengganu* [A glossary of the Terengganu dialect] (1997). For each CTM word, a standardised spelling that resembles SM orthography is given, which mirrored earlier British linguists' practise of documenting Malay dialects. For instance, [igaʔ] 'to catch' and [tɛpɛʔ] 'to stick' are written as ⟨igat⟩ and ⟨tempek⟩ respectively, which seems to suggest that word-final /t/ and medial consonant sequence /-mp-/ are phonemic. This forced system is fortunately abandoned in *Glosari dialek Kelantan* [A glossary of the Kelantan dialect] (2016), another glossary in the same series that was published about two decades later. Here we can find [sɔʔmɔ] 'always' and [blɛ-blɛ] 'whilst' written as ⟨sokmo⟩ and ⟨ble-ble⟩ (instead of the potentially standardised spelling ⟨belan-belan⟩ which can be found in *Kamus Dewan* [The institute dictionary], Sheikh Othman 2007).

The issues discussed above highlight the need for more descriptive studies of NEPMs. Only then can we gain a better understanding of these languages, including their synchronic systems and diachronic development.

1.6 Present study

The present study has two goals. The first goal is to provide a description of NEPMs by gathering and analysing new data, with a focus on the phonology and morphology of NEPMs, while also providing a concise description of their syntactic structure. Based on synchronic descriptive facts, the second

goal is to establish the phonological and morphological changes that have taken place in NEPMs.

The data for this study were collected during two field trips to Kelantan and Terengganu, conducted from July to November 2018 and August to October 2022. These field trips were a part of the visiting studies at the University of Malaya, at the Academy of Malay Studies in 2018 and at the Faculty of Languages and Linguistics in 2022. A third field trip was originally planned for 2020 but was unfortunately cancelled due to the Covid-19 pandemic.

This section elaborates on the methodology in the present study, describing the fieldwork locations and native speaker consultants in §1.6.1 and explaining data collection and processing in §1.6.2. A summary of transcription conventions follows in §1.6.3.

1.6.1 Fieldwork locations and native speaker consultants

As NEPMs are spoken across an extensive area, the selection of field sites had several considerations. Firstly, the state capitals Kota Bharu and Kuala Terengganu were excluded as they serve as the first entry points for immigrants to these states, and thus are more likely to have received more external influences. For similar reasons, villages were preferred over local towns. Secondly, field sites should be preferably chosen to align with typical Malay settlement patterns. In Kelantan, the river systems play a crucial role in shaping the settlement pattern, while in Terengganu, both the river systems and coasts are important factors (see §1.4.1.1). Finally, feasibility and practicality were also taken into account, with preference given to locations where it was easier to establish relations through my contact network. With these considerations in mind, the following locations were chosen as primary field sites, as shown in Table 1.2 and Figure 1.6.

Table 1.2: Overview of field sites

Variety	Primary field site	Coordinates	District
KM	Kampung Kusial Bharu	5°45'N, 102°08'E	Tanah Merah
CTM	Kampung Gong Sentul	5°20'N, 103°06'E	Kuala Nerus
ITM	Kampung Dusun	5°04'N, 102°56'E	Hulu Terengganu

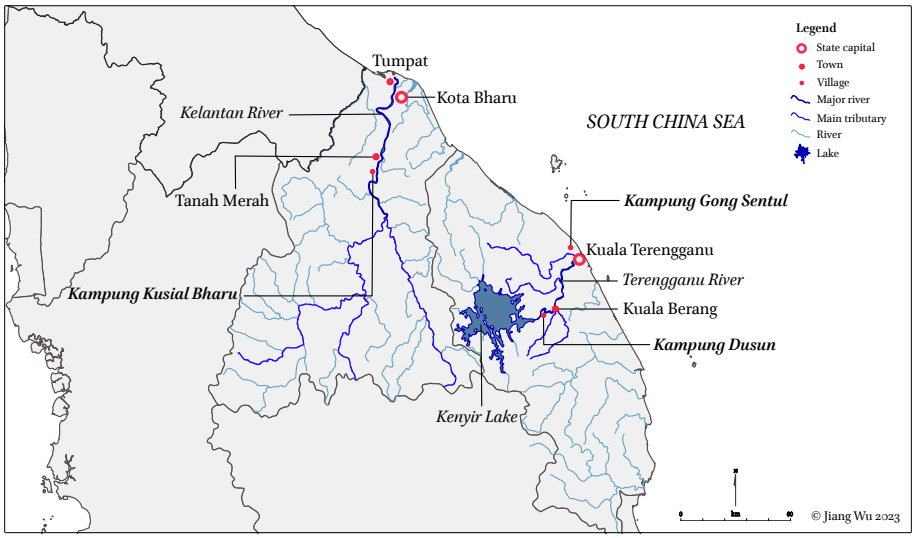

Figure 1.6: Locations of field sites

Data on KM were primarily collected in Kampung Kusial Bharu (*kampung* is Malay for 'village') in the district of Tanah Merah, which is situated in the Kelantan River basin, some fifty kilometres south of the state capital Kota Bharu. The village of Kusial Bharu is on the west bank of the Kelantan River. In addition, several visits were made to Kampung Palekbang in Tumpat.

CTM data were collected in Kampung Gong Sentul in the district of Kuala Nerus, which is located close to the confluence of the Nerus River and the Terengganu River. The town of Kuala Nerus is situated along the coastline of Terengganu, between the city of Kuala Terengganu and the state's airport, approximately seven kilometres from the city centre. Formerly a part of the Kuala Terengganu district, Kuala Nerus is now densely populated with many small villages in close proximity to each other, Kampung Gong Sentul being one of them.

ITM data were collected in several villages across the district of Hulu Terengganu given the vast intra-dialectal variation of ITM (see Collins 1983a). The primary field site is Kampung Dusun, and other villages visited include Kampung Tanjung Baru, Kampung Payang Kayu and Kampung Pasir Nering, where comparative lexical data were collected (see Figure 1.7).

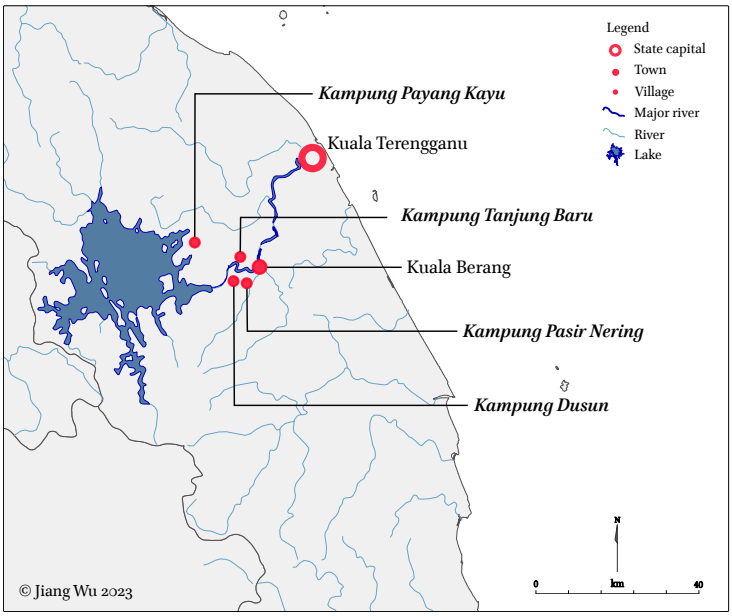

Figure 1.7: Locations of field sites in Hulu Terengganu

In comparison to Kuala Nerus, Hulu Terengganu is more sparsely populated, and the distribution of villages in this region clearly follows the riverine system. Kampung Dusun is a village on the bank of the Terengganu River in its upper reach, approximately ten kilometres from the town of Kuala Berang and fifty kilometres from the capital Kuala Terengganu. Hosting the oldest primary school and one of the oldest mosques in Hulu Terengganu, the village has apparently been of great importance.

In selecting the consultants, the rule of thumb was that non-mobile, older, rural females were preferred. This differs from the NORMs principle proposed by Chambers & Trudgill (1998: 29) for dialectology studies, which prioritises non-mobile, older, rural *males*. The preference for females over males was based on the consideration that males in traditional Malay villages tend to be more mobile than females, and thus are more likely to have received more external influence on their language. Many older men have worked in other states before returning to their hometown, leaving their families in the villages. Furthermore, men generally have higher social status and more interaction with outsiders. Younger generations also tend to move to bigger cities for study or work. Despite the preference for choosing older

consultants, data from some younger speakers were also collected, typically during elicitation sessions, which proved to be easier with younger speakers.

1.6.2 Data collection and processing

The data collection process followed the guidelines outlined by the European General Data Protection Regulation (GDPR, available at https://gdpr-info.eu/), whereby the consultants' consent was obtained before recording their speech. The linguistic data collected can be broadly classified into six types, as listed in Table 1.3. The table also presents an overview of the amount of raw data collected for each variety. Each type of data is briefly described below.

Table 1.3: Overview of raw data

Data type	KM	CTM	ITM
Word lists	≈4 hours	≈3,5 hours	≈5 hours
Narratives	≈20 minutes	≈45 minutes	≈70 minutes
Elicitation	≈6 hours	≈3 hours	≈6,5 hours
Conversations	≈2,5 hours	≈1 hour	≈6 hours
Discussions and interviews	≈1,5 hours	≈45 minutes	≈3 hours
Unrecorded field notes	N/A	N/A	N/A

The first type is word lists, which consist of a basic word list of 260 items, a modified version of the Swadesh 200 word list, and a supplementary list of 309 items. Both lists were recorded for all three varieties, and the basic word lists collected for each variety can be found in Appendix A. Additional word lists focusing on more specific phonological phenomena or words in particular categories were recorded as needed. These lists served as the basis on which phonological analyses were conducted.

The second type consists of narratives. Three standard stories were recorded with the aid of picture books and video clips: *The frog story* (Mayer 1969), a modified version of *The chicken thief story* (Rodriguez 2009) and *The pear story* (Erbaugh 2001, available at http://pearstories.org/). Some of these stories were recorded with multiple consultants. It is acknowledged that narratives obtained with visual stimuli may be less natural than free

narratives (see, e.g., Klamer & Moro 2020). Attempts were also made to record free-style storytelling, but they were unfortunately unsuccessful with most consultants, which could be attributed to two reasons. First, there is not a strong tradition of storytelling in the field sites I visited, and perhaps not in contemporary Malay villages on the peninsula in general. Second, the intermediate language used for data collection was SM, which is closely related to the vernaculars being investigated. The consultants often switched to or mixed their vernaculars with SM when telling free-form stories, which for them was not a natural setting.[9]

The third type of data is from elicitation sessions, which involved various tasks ranging from sentence translations, making sentences using the given words, to elicitation with video stimuli such as the Surrey clips (Fedden et al. 2010) and the *Give*-events clips recorded by Moro & Fricke (2020).

The fourth type of data comes from naturalistic conversations in the vernaculars between two or more interlocutors. Unlike the first three types of data, which were recorded in both video and audio formats (.mp4 and .wav respectively), conversations were only recorded using the audio recorder. Given the spontaneous nature of this type of recording, no videos were recorded so as to minimise potential influences of an artificial interview setup on the natural flow of the conversations.

The fifth type of data includes interviews of the speakers' sociolinguistic background and other types of discussions, such as those on the usage of certain words and constructions. These were typically recorded in SM.

Lastly, during the fieldwork, some observations and notes were made on the spot and written down in notebooks, but not recorded. These belong to the category of unrecorded field notes.

The duration of fieldwork in each site varies for various reasons, which results in differences in the amount of data collected for each NEPM variety. I spent approximately five weeks in Kelantan, four weeks in Coastal Terengganu, and seven weeks in Hulu Terengganu. The largest amount of data has therefore been collected for ITM, whereas the least amount is available for CTM (see Table 1.3).

[9] The close relatedness between the intermediate language and the target language posed a major challenge in data collection. Consultants were sometimes invited to listen to their own speech and point out parts that might have been influenced by SM. Other challenges and difficulties of data collection and processing have been explained in Klamer et al. (2021: 489–491).

Recorded data were transcribed by myself or with the assistance of consultants or student assistants. Word lists were transcribed in Microsoft Excel, some elicitation sessions were transcribed in Microsoft Word, while other recordings were transcribed using the linguistic annotation tool ELAN (The Language Archive, Max Planck Institute for Psycholinguistics, available at https://archive.mpi.nl/tla/elan). The Excel transcriptions (.xlsx) were exported as database files for Standard Format Lexicon (.db), while annotated ELAN files (.eaf) were exported as FLEXTEXT files. Both types of files were then imported into the software FieldWorks Language Explorer (FLEx, SIL International, available at https://software.sil.org/fieldworks/), which provides a corpus-building platform for interlinear glossing of texts and morphological analysis while expanding a lexical database. For each NEPM variety, a corpus was built in FLEx. Translations were done either in ELAN during annotation or in FLEx. Finally, the glossed and translated transcriptions were reimported into ELAN and linked with the recordings. Due to time limitations and the general difficulty in transcribing naturalistic conversations, as well as the varying quality of recordings, not all conversations have been transcribed. Appendix B provides three transcribed sample texts (one for each variety) from different types of recordings, which serve to illustrate the morphosyntactic structures of NEPMs.

In handling the recordings, file naming follows the convention outlined in (1). A file name starts with the language abbreviation, followed by the date of recording, the type of data and its number. The following abbreviations are used for each type of data: "wl" for word lists, "n" for narratives, "e" for elicitation, "cv" for conversations, "d" for discussions and interviews, and "fn" for unrecorded field notes. Additional information such as the content of the recording is sometimes provided, followed by the format of files if recordings were made in multiple formats ("a" for audio and "v" for video). For example, a file with the name "ITM_180907_n01_frogstory_a" shows that it is an audio recording of a frog story (which is a narrative, and the first narrative recorded on that day), recorded on 7th September 2018 in ITM.

(1) LANGUAGE ABBREVIATION_DATE OF RECORDING_TYPE OF DATA AND NUMBER_ADDITIONAL INFORMATION_FORMAT OF FILES

A metadata sheet was created to document the list of sessions, recordings and native speaker consultants. All types of raw data and processed data, as well as the metadata sheet, have been archived in the *Corpora of Kelantan*

Malay, Coastal Terengganu Malay and Inland Terengganu Malay (Wu 2023), available at https://doi.org/10.34894/HWUVLM.

1.6.3 Transcription conventions and citation codes

Linguistic examples in the present study are transcribed using the International Phonetic Alphabet (IPA) with slight modifications. For easier transcription and to align with the tradition in the field, I opted to use /y/ instead of the IPA symbol /j/ for the palatal glide, and /c, j/ instead of IPA /c͡ç, ɟ͡ʝ/ to represent palatal affricates. For example, KM /ayɛ/ 'chicken' and /jaɣi/ 'finger' should be read as /ajɛ/ and /ɟ͡ʝaɣi/ respectively in standard IPA. During initial data processing, words were transcribed in broad IPA following their pronunciations at the phonetic level. After conducting phonological analyses, transcriptions were rendered in phonemic forms. As each NEPM variety has a different phonological system, no attempt has been made to create an orthography for each variety. Linguistic examples are primarily cited in their phonemic transcriptions to facilitate cross-dialectal comparisons, which are either enclosed in forward slash brackets "//" (as practised in the phonology chapters in order to differentiate phonemic forms from phonetic forms, which are given in square brackets "[]"), or simply in *italics* (as practised in other chapters).

When citing example sentences, I provide a slightly shortened name of the corresponding recording as the citation code, which follows the free translation. For instance, the example with the citation code "KM_180812_n01_12" indicates that it is a KM example from line 12 of a narrative recorded on 12th August 2018. The corresponding recording can be found in the corpora by searching for the file name starting with "KM_180812_n01". Morpheme-by-morpheme interlinear glosses are provided following the Leipzig Glossing Rules, with my own additions where necessary. An exception has been made in the transcription of reduplicated forms, where I use a hyphen "-" to link the two morphemes instead of the prescribed tilde "~", which is reserved for indicating free variation between two (or more) forms. A list of standard abbreviations for the glossing of grammatical categories, along with additional abbreviations and a symbol usage guide, can be found in the front matter on pages xvii–xx.

1.7 Structure of this dissertation

The present dissertation is structured into two main parts that align with the research goals.

Part one provides a synchronic description of NEPMs, comprising five chapters. Chapters 2 to 4 focus on the phonology of KM, CTM and ITM respectively, as this is the area where these varieties exhibit the most pronounced differences. Each chapter follows a parallel structure, starting with an examination of the segment inventory, and then building up towards syllable and word structure, as well as phonotactic constraints, before concluding with a brief discussion of stress patterns. Chapters 5 and 6 describe the morphology and syntax of NEPMs. Given the similarities across the morphosyntax of NEPMs, a comparative approach is taken in these two chapters.

Part two delves into the historical development of NEPMs. It consists of two chapters (Chapters 7 and 8), which explore the historical phonology and historical morphology respectively. Building on the synchronic analysis in Part one, these chapters draw comparisons between NEPMs and the preexisting reconstructions in the common ancestral language (in the present case PM) and establish sound changes and morphological changes that have taken place. In these chapters, I show that NEPMs have undergone remarkable sound changes and significant morphological reduction. I also discuss the potential factors contributing to these changes in the history of NEPMs, briefly examining the role of language contact. By analysing the chronological order of sound changes, I argue that NEPMs cannot form a lower-level subgroup within Malayic, despite superficial similarities in their sound patterns. Furthermore, there is no clear evidence of substratal influences, and the morphological reduction was primarily driven by phonological changes.

The main findings of this dissertation are synthesised and summarised in Chapter 9. In addition, I propose a hypothesis regarding the migration patterns of the speakers, combining data from linguistics and historical records. It is evident that ITM stands out as the most divergent and conservative variety among NEPMs, suggesting that its speakers may have a longer history when compared to the coastal population. Speakers of KM and CTM, on the other hand, likely represent a more recent migration, possibly occurring post-Malaccan times. Lastly, I discuss some limitations of the present study and suggest directions for future research.

CHAPTER 2

Phonology of Kelantan Malay

2.1 Introduction

This chapter describes the phonology of KM, covering an examination of the segment inventory in §2.2, syllable structure in §2.3 and word structure in §2.4. Phonotactic constraints on permitted consonant clusters, consonant sequences and vowel sequences are presented in §2.5 to §2.7. The stress pattern is discussed in §2.8, and a summary is provided in §2.9.

2.2 Segment inventory

2.2.1 Consonant system

2.2.1.1 Consonant inventory

Table 2.1 displays the consonant inventory of KM, which consists of twenty native consonant phonemes including seven stops, four nasals, three fricatives, two affricates, two liquids and two glides. Two borrowed consonants with marginal phonemic status are also included in the table, enclosed in parentheses.

Table 2.1: Consonant inventory of KM

		Labial	Dental	Alveolar	Palatal	Velar	Glottal
Stops	voiceless	p	t			k	ʔ
	voiced	b		d		g	
Nasals		m		n	ɲ	ŋ	
Fricatives	voiceless			s		x	h
	voiced			(z)			
Affricates	voiceless				c		
	voiced				j		
Liquids				l, (ɾ)		ɣ	
Glides		w			y		

(c = IPA /c͡ç/, j = IPA /ɟ͡ʝ/, y = IPA /j/)

Among the stops, the voiceless coronal stop is articulated as dental [t̪], whereas its voiced counterpart is articulated as alveolar [d]. This asymmetry in the place of articulation of /t/ and /d/, sometimes labelled as dental discrepancy, is also present in SM and many other Austronesian languages (Henderson 1965: 420–421; Adelaar 1983; Donohue 2009). The palatal obstruents /c/ and /j/, while articulated as affricates, share similar phonological properties with the three sets of non-glottal stops. Therefore, for the purposes of this description, the term "stop" (in the phonemic sense) includes these two affricates. Non-glottal stops have similar distributions, and they exhibit similar patterns in possible combinations with other segments in consonant clusters and sequences (see §2.4.3, §2.5 and §2.6). The phoneme /ɣ/, which is phonetically a voiced velar fricative, is treated as a liquid for two reasons. First, language internal evidence suggests that /ɣ/ patterns with the lateral liquid /l/ in the formation of consonant clusters (see §2.5). Second, /ɣ/ often corresponds to an alveolar tap /ɾ/ (or a trill /r/) in other Malayic varieties (Adelaar 1992: 8).

While most loanwords have been adapted to KM phonology, some retained an unadapted pronunciation. Table 2.2 provides examples of words that contain the borrowed consonants /z/ and /ɾ/ (sometimes realised as a trill [r]).

Table 2.2: Examples of KM words with borrowed consonants

Phoneme	Example	Gloss	Origin
/z/	/**z**ina/	'adultery'	< Arabic *zinā'* [10]
	/u**z**o/	'weak'	< Arabic *ʿudhr*
/ɾ/	/aɾnaʔ/	'rabbit'	< Arabic *arnab*
	/aɾti/	'meaning'	< Sanskrit *artha*

These words ultimately originate from Arabic or Sanskrit. However, most loanwords from these languages (also later Portuguese) have been fully adapted to KM phonology, reflecting regular sound changes because of their long history. Words with unadapted foreign sounds thus likely entered KM more recently, presumably via SM *zina*, *uzur*, *arnab* and *ərti~arti* with the same meanings. The alveolar tap /ɾ/ is also found in English loanwords such as /ɔɾɛŋ/ 'orange', /gɾuʔ/ 'group' and /tɾa/ 'to try'.

The voiceless velar fricative /x/ also occurs in unadapted Arabic loanwords such as /axe/ 'end' < *ākhir*. However, most instances of /x/ do not reflect an unadapted pronunciation but rather result from the regular sound change /xx/ < *kɣ, e.g., /**xx**ɛtɔ/ 'car' < $^{+}$**kɣ**ɛtɔ < SM ***kərɛta*** < Portuguese *carreta* and /**xx**usi/ 'chair' < $^{+}$**kɣ**usi < SM ***kərusi*** < Arabic *kursī*. Therefore, I take /x/ as a native phoneme. See more discussion in §7.5.2.3.

2.2.1.2 Contrasts between consonants

Table 2.3 lists minimal or near-minimal pairs demonstrating the contrasts between similar consonants in KM. When no (near-)minimal pairs are found, the closest pair with contrasting segments is given. Contrasts in word-initial, -medial and -final positions are distinguished on account of the restricted distribution of some consonants. For instance, the velar nasal /ŋ/, the glottal fricative /h/ and the glides almost never occur in initial position. The glottal stop is not phonemic word-initially, and /x/ only appears in the geminated form /xx/. Word-finally, only /ʔ, ŋ, h/ are allowed. See more on consonant distributions in §2.4.3.1.

[10] Unless otherwise noted, the origins of loanwords are cited from Jones (2007).

Table 2.3: Contrasts between consonants in KM

Contrast	Pair	Gloss
Word-initially		
/p/ - /b/	/**p**uloh/	'ten'
	/**b**uloh/	'bamboo'
/b/ - /m/	/**b**ukɔ/	'to open'
	/**m**ukɔ/	'face'
/t/ - /d/	/**t**əbu/	'sugarcane'
	/**d**əbu/	'dust'
/d/ - /n/	/**d**ai/	'forehead'
	/**n**aiʔ/	'to climb, to go up; to ride'
/c/ - /j/	/**c**aɣi/	'to search'
	/**j**aɣi/	'finger'
/j/ - /ɲ/	/**j**awɔ/	'Java'
	/**ɲ**awɔ/	'soul'
/k/ - /g/	/**k**ali/	'times'
	/**g**ali/	'to dig'
/g/ - /ɣ/	/**g**ayɔ/	'to persuade'
	/**ɣ**ayɔ/	'Eid al-Fitr'
/m/ - /n/ - /ɲ/	/**m**amɔh/	'to chew'
	/**n**amɔ/	'name'
	/**ɲ**amoʔ/	'mosquito'
/n/ - /s/	/**n**iyaʔ/	'aim'
	/**s**iyaʔ/	'finished'
/l/ - /n/	/**l**amɔ/	'long (time)'
	/**n**amɔ/	'name'
/l/ - /ɣ/	/**l**ata/	'floor'
	/**ɣ**ata/	'necklace'
Word-medially		
/p/ - /b/	/lɛ**p**a/	'to throw'
	/lɛ**b**a/	'wide'
/b/ - /m/	/tə**b**u/	'sugarcane'
	/tə**m**u/	'to meet'
/t/ - /d/	/u**t**ɛ/	'forest; debt'
	/u**d**ɛ/	'shrimp'

Contrast	Pair	Gloss
/d/ - /n/	/kə**d**a/	'shop'
	/kə**n**a/	'to know'
/c/ - /j/	/a**c**a/	'pickles'
	/a**j**a/	'to teach'
/j/ - /ɲ/	/sə**j**aʔ/	'to toast'
	/sə**ɲ**aʔ/	'quiet'
/k/ - /g/ - /ŋ/	/i**k**aʔ/	'to tie up'
	/i**g**aʔ/	'to catch'
	/i**ŋ**aʔ/	'to think'
/g/ - /ɣ/	/la**g**i/	'again; more; still; yet'
	/la**ɣ**i/	'to run'
/m/ - /n/ - /ŋ/	/ta**m**ɛ/	'park'
	/ta**n**ɛ/	'to plant'
	/ta**ŋ**ɛ/	'hand'
/m/ - /ɲ/	/bu**m**i/	'earth'
	/bu**ɲ**i/	'sound'
/ɲ/ - /ŋ/	/a**ɲ**iŋ/	'dog'
	/a**ŋ**iŋ/	'wind'
/l/ - /n/	/bu**l**oh/	'bamboo'
	/bu**n**oh/	'to kill'
/l/ - /ɣ/	/ja**l**ɛ/	'road'
	/ja**ɣ**ɛ/	'seldom'
/s/ - /h/	/a**s**aʔ/	'smoke'
	/a**h**aʔ/	'Sunday'
/y/ - /w/	/a**y**ɛ/	'chicken'
	/a**w**ɛ/	'cloud'
Word-finally		
/ʔ/ - Ø	/susu**ʔ**/	'to hide'
	/susu/	'milk'
/ŋ/ - Ø	/pəti**ŋ**/	'important'
	/pəti/	'box'
/h/ - Ø	/sɔɣɔ**h**/	'drawer'
	/sɔɣɔ/	'voice'
/ʔ/ - /ŋ/	/lao**ʔ**/	'sea'
	/lao**ŋ**/	'to call'

Contrast	Pair	Gloss
/ʔ/ - /h/	/bɔwɔ**ʔ**/	'to bring'
	/bɔwɔ**h**/	'bottom; below'
/ŋ/ - /h/	/tujo**ŋ**/	'to jump down'
	/tujo**h**/	'seven'

A special type of contrasts is found between singleton and geminate consonants in word-initial position, as demonstrated by (near-)minimal pairs in Table 2.4.

Table 2.4: Contrasts between singletons and geminates in KM

Contrast	Pair	Gloss
/p/ - /pp/	/**p**alɔ/	'nutmeg'
	/**pp**alɔ/	'head'
/b/ - /bb/	/**b**iniŋ/	'wife'
	/**bb**iniŋ/	'to marry (a wife)'
/t/ - /tt/	/**t**upaʔ/	'compact; Tumpat (toponym)'
	/**tt**upaʔ/	'ketupat (k.o. rice cake)'
/d/ - /dd/	/**d**əɣah/	'loud; fast'
	/**dd**əɣah/	'to raise (voice)'
/c/ - /cc/	/**c**aɣɔ/	'method'
	/**cc**aɣɔ/	'to talk'
/j/ - /jj/	/**j**alɛ/	'road'
	/**jj**alɛ/	'to walk'
/k/ - /kk/	/**k**ulaʔ/	'mushroom; mould'
	/**kk**ulaʔ/	'mouldy'
/g/ - /gg/	/**g**aji/	'wage'
	/**gg**aji/	'saw (n.)'
/m/ - /mm/	/**m**aɣi/	'to come'
	/**mm**aɣi/	'cupboard'
/n/ - /nn/	/**n**akɔ/	'jackfruit'
	/**nn**aɣɔ/	'tower; Narathiwat (toponym)'
/ɲ/ - /ɲɲ/	/**ɲ**awɔ/	'soul, life'
	/**ɲɲ**awɔ/	'to breathe'

Contrast	Pair	Gloss
/s/ - /ss/	/**s**iyɛ/	'daytime'
	/**ss**iyɛ/	'to pity'
/l/ - /ll/	/**l**umaʔ/	'crushed'
	/**ll**umaʔ/	'to crush'
/ɣ/ - /ɣɣ/	/**ɣ**ɛhɛ/	'jaw'
	/**ɣɣ**ɛhɛ/	'molar tooth'

Word-initial geminates are cross-linguistically rare (Thurgood 1993; Muller 2001; Kraehenmann 2011). In KM, however, all consonants except for the glottals /h, ʔ/ and the glides /w, y/ can appear as geminates, and they only occur word-initially. The exclusive occurrence of geminates in word-initial position has a diachronic explanation, as they typically originate from the reduction of antepenultimate vowels in trisyllabic words and subsequent assimilation of initial clusters (compare KM /**pp**alɔ/ 'head' and /**ɣɣ**ɛhɛ/ 'molar tooth' with SM cognates ***kəpala*** and ***gəraham***). A more detailed analysis of the origin of geminates is presented in §7.5.2. At the phonetic level, the distinction between singleton-geminate pairs in KM is predominantly reflected in the length of initial consonants, that is, geminate consonants have longer closure duration (Mohd Hilmi et al. 2016). Mohd Hilmi et al. (2018) further show a number of non-durational acoustic parameters where geminates and singletons differ: geminates are associated with shorter post-consonantal vowel duration, greater amplitude and higher fundamental frequency in the early part of the following vowel. At the phonological level, geminates are analysed as a subtype of consonant clusters instead of a separate series of phonemes, as supported by the variation attested between some geminates and non-geminate clusters; see more discussion on syllable structure in §2.3, and on consonant clusters in §2.5.

Note that not all geminates can be clearly contrasted with a corresponding singleton. For instance, the velar nasal occurs in the geminated form /ŋŋ-/, as in /ŋŋale/ 'to flow' and /ŋŋuwaʔ/ 'to yawn', whereas a singleton /ŋ/ rarely occurs initially (except in one instance /ŋaŋɔ/ 'to open wide'). Despite the lack of near-minimal pairs contrasting /ŋ/ and /ŋŋ/, the analysis of a geminate /ŋŋ-/ is on the ground that it has a duration comparable to other geminate nasals. Diachronically, it also originates from the assimilation of earlier clusters consisting of two segments, just as other geminates (com-

pare KM /ŋŋale/ 'to flow' with SM ***məŋalir***). For similar reasons, the status of /xx/ is maintained despite the lack of pairs contrasting /xx/ and /x/.

Many pairs contrasting an initial singleton and an initial geminate are semantically related. This is because some geminates can be analysed as morphologically complex, and initial gemination can be considered a morphophonological process with various functions; see §5.3.2.

2.2.1.3 Phonetic realisations of consonants

Some consonant phonemes in KM have variable realisations at the phonetic level, as summarised in Table 2.5 and explained below.

Table 2.5: Phonetic realisations of some consonant phonemes in KM

Phoneme	Realisation	Environment	Example		Gloss
/h/	[ç]~[h]	i__#	/nipi**h**/	[nipi**ç**]~[nipi**h**]	'thin'
			/kudi**h**/	[kudi**ç**]~[kudi**h**]	'scabies'
	[h]	elsewhere	/kabu**h**/	[kabu**h**]	'fog'
			/ata**h**/	[ata**h**]	'top'
/x/	[xx]~[kk]	#__	/**xx**usi/	[**xx**usi]~[**kk**usi]	'chair'
			/**xx**ɛtɔ/	[**xx**ɛtɔ]~[**kk**ɛtɔ]	'car'

First, the glottal fricative /h/ is often realised as a palatal fricative [ç] in word-final position after a high front vowel /i/, resulting from the coarticulation with the labial compression of that vowel (cf. Japanese /hi/ → [çi]). /nipih/ 'thin' can be articulated as [nipiç]~[nipih], and /kudih/ 'scabies' as [kudiç]~[kudih]. Elsewhere /h/ is realised as [h].

Second, the voiceless velar fricative /x/, which only occurs in the geminated form /xx/, has a phonetic realisation varying from [x] to [k] in word-initial position. For instance, /xxusi/ 'chair' is pronounced as [xxusi]~[kkusi], and /xxɛtɔ/ 'car' as [xxɛtɔ]~[kkɛtɔ], depending on the speaker. Given that phonemic /kk/ also occurs initially, the contrast between /xx/ and /kk/ in this position is sometimes neutralised. However, it does not mean that the two phonemes are in free variation: phonemic /xx/ can be pronounced as [kk], but conversely, phonemic /kk/ is never realised as [xx]. The phonemic status of /x/ is therefore maintained.

2.2.2 Vowel system

2.2.2.1 Vowel inventory

KM has twelve phonemic vowels, as presented in Table 2.6.

Table 2.6: Vowel inventory of KM

	Front	Central	Back
High	i		u, ũ
Mid-high	e		o
Mid-low	ɛ, ɛ̃	ə	ɔ, ɔ̃
Low		a, ã	

There are eight oral vowels and four nasal vowels, and a four-way distinction is made as regards vowel height. The high back vowel /u/, mid-low vowels /ɛ, ɔ/, and the low vowel /a/ have nasal counterparts, which yields a vowel inventory that is larger than many other Malayic varieties (cf. Adelaar 2005c; McDonnell et al. in print). Nasal vowels have limited presence and a restricted distribution, and their functional load seems to be light. Nevertheless, contrasts between a nasal vowel and its oral counterpart can be found in (near-)minimal pairs, see Table 2.8 below. /ĩ, ẽ, õ/ are also included as phonemes in earlier descriptions (e.g., Abdul Hamid 1994), in which the authors argue that these nasal vowels originated from the loss of final nasals and regressive nasalisation, as in /kucĩ/ 'cat' and /ɣacõ/ 'poison' (cf. SM *kuciŋ* and *racun*, also see a summary in Adelaar 2005c). /ẽ/ is transcribed ambiguously in Abdul Hamid's description. Based on my data, however, final nasals are retained in /kuciŋ/ 'cat' and /ɣacoŋ/ 'poison', and no phonemic /ĩ, ẽ, õ/ are attested.

A distinction needs to be made between nasalised vowels and true nasal vowels. Vowel nasality at the phonetic level is common in KM, as nasal consonants spread the feature of nasality rightwards, affecting adjacent vowels. For instance, /makɛ/ 'to eat' and /namɔ/ 'name' are realised as [mãkɛ] and [nãmɔ̃] respectively. Nasality can also spread over glides /w, y/ and glottals /h, ʔ/, affecting more than one vowel: /ɲɲawɔ/ 'to breathe' is realised as [ɲɲãwɔ̃], and /maha/ 'expensive' as [mãhã]. In these cases, the vowels carrying nasality are considered nasalised vowels, which are allophonic

realisations of oral vowels following nasals.[11] However, there are also cases where vowels show nasality independent of nasal consonants, e.g., [ɛ̃] in [kəcɛ̃ʔ] 'small' and [ũ] in [busũʔ] 'smelly'. These vowels are taken as true, phonemic nasal vowels. In a few examples such as [jmãʔãʔ]~[mmãʔãʔ] 'Friday' and [mãʔãh] 'sorry' (< Arabic *jumʿa* and *muʿāf*, cf. SM *jumaat* and *maaf*), it is ambiguous whether the final-syllable [ã] is phonemically nasal or not. On the one hand, final-syllable [ã] in these examples may be seen as resulting from nasal spreading, but on the other hand, Arabic loanwords with similar shapes often evidently have nasal vowels, as in /saʔãʔ/ 'second' and /tɔʔãʔ/ 'devotion' (< *sāʿa* and *ṭāʿa*, cf. SM *saat* and *taat*).[12] I consider the second observation more important and opt for the transcription of /jmaʔãʔ/~/mmaʔãʔ/ 'Friday' and /maʔãh/ 'sorry'.

KM does not have phonemic diphthongs. Vowel sequences constitute two syllables. For instance, /bau/ 'shoulder; smell' is disyllabic, articulated as [ba.u], and /ae/ 'water' is articulated as [a.e]. More discussion on vowel sequences is provided in §2.7.

2.2.2.2 Contrasts between vowels

The phonemic status of KM vowels is demonstrated by minimal or near-minimal pairs presented in Table 2.7 and Table 2.8. Contrasts found in penultimate and final syllables are distinguished in view of the canonical disyllabic word structure and the restricted distribution of some vowels at the word level, see §2.4.3.2.

[11] Similar progressive vowel nasalisation and nasal spreading is reported in SM (Farid 1976: 70) and in Salako (Adelaar 2005b: 23).

[12] It appears that in the borrowing of Arabic loanwords, the original voiced pharyngeal fricative /ʕ/ (transliterated as ⟨ʿ⟩) typically has an effect on nasalising the following vowel. See more discussion in §7.4.3.

Table 2.7: Contrasts between oral vowels in KM

Contrast	Pair	Gloss
Penultimate syllables		
/a/ - /i/	/b**a**tɛ/	'stem'
	/b**i**tɛ/	'star'
/a/ - /u/ - /ɛ/	/m**a**ɣɔh/	'angry'
	/m**u**ɣɔh/	'cheap'
	/m**ɛ**ɣɔh/	'red'
/a/ - /ɔ/	/k**a**li/	'times'
	/k**ɔ**li/	'wok'
/a/ - /ə/	/p**a**ɣɛ/	'machete'
	/p**ə**ɣɛ/	'war'
/i/ - /u/	/t**i**kɛ/	'to stab'
	/t**u**kɛ/	'expert'
/i/ - /ɛ/	/b**i**sɔ/	'poisonous'
	/b**ɛ**sɔ/	'usual'
/i/ - /ɔ/	/l**i**paʔ/	'to fold'
	/l**ɔ**paʔ/	'to jump'
/i/ - /ə/	/p**i**sɛ/	'banana'
	/p**ə**sɛ/	'to order'
/u/ - /ɔ/	/b**u**tɔ/	'blind'
	/b**ɔ**tɔ/	'bottle'
/u/ - /ə/	/m**u**tɔh/	'to vomit'
	/m**ə**tɔh/	'raw'
/ɛ/ - /ɔ/	/b**ɛ**wɔʔ/	'monitor lizard'
	/b**ɔ**wɔʔ/	'to bring'
/ɛ/ - /ə/	/m**ɛ**nɛ/	'toy, game'
	/m**ə**nɛ/	'to win'
/ɔ/ - /ə/	/s**ɔ**ɣɔh/	'drawer'
	/s**ə**ɣɔh/	'to surrender'
Final syllables		
/a/ - /i/ - /u/ - /o/	/at**a**/	'to send'
	/at**i**/	'liver'
	/at**u**/	'ghost'
	/at**o**/	'to arrange'

Contrast	Pair	Gloss
/a/ - /ɔ/	/sap**a**/	'until'
	/sap**ɔ**/	'who'
/a/ - /ɛ/	/lap**a**/	'hungry'
	/lap**ɛ**/	'eight'
/a/ - /e/	/al**a**ʔ/	'tool'
	/al**e**ʔ/	'direction'
/i/ - /ɛ/	/pad**i**/	'paddy'
	/pad**ɛ**/	'field'
/i/ - /ɔ/	/mat**i**/	'to die'
	/mat**ɔ**/	'eye'
/i/ - /e/	/gal**i**/	'to dig'
	/gal**e**/	'loose'
/u/ - /ɔ/	/buk**u**/	'book'
	/buk**ɔ**/	'to open'
/u/ - /e/	/pas**u**/	'flower pot'
	/pas**e**/	'sand'
/u/ - /ɛ/	/baɣ**u**/	'new; have just'
	/baɣ**ɛ**/	'stuff'
/ɛ/ - /ɔ/	/paɣ**ɛ**/	'machete'
	/paɣ**ɔ**/	'rack'
/ɛ/ - /o/	/say**ɛ**/	'to love'
	/say**o**/	'vegetable'
/ɛ/ - /e/	/il**ɛ**/	'lost'
	/il**e**/	'downstream'
/ɔ/ - /e/	/but**ɔ**/	'blind'
	/but**e**/	'grain; CLF'
/ɔ/ - /o/	/mas**ɔ**ʔ/	'ripe'
	/mas**o**ʔ/	'to enter'
/e/ - /o/	/bul**e**h/	'can; to get'
	/bul**o**h/	'bamboo'

Contrasts between oral and nasal vowels are displayed in Table 2.8. Such contrasts are only found in final syllables.

Table 2.8: Contrasts between oral and nasal vowels in KM

Contrast	Pair	Gloss
/a/ - /ã/	/buwa**a**ʔ/	'to do; to make'
	/puwã**a**ʔ/	'to brush (hair)'
/ɛ/ - /ɛ̃/	/kɛcɛʔ/	'to speak'
	/kəcɛ̃ʔ/	'small'
/ɔ/ - /ɔ̃/	/ɛsɔʔ/	'tomorrow'
	/ɛsɔ̃ʔ/	'to scoot over'
/u/ - /ũ/	/bus**u**ʔ/	'ant hill'
	/bus**ũ**ʔ/	'smelly'

2.3 Syllable structure

Syllables in KM have a template of (C)(C)V(C), with an optional onset and an optional coda. A nucleus V is always a monophthong without length distinction. An onset can be a consonant cluster consisting of two consonants, whereas a coda can only have one segment. Table 2.9 offers an overview of permitted syllable types.

Table 2.9: Overview of syllable types in KM

Onset	Nucleus	Coda
	V	
	V	C
C	V	
C	V	C
CC	V	
CC	V	C

The distribution of single consonants at the syllable level is summarised in Table 2.10. A plus sign marks a position where a phoneme is attested, and a minus sign marks a position where a phoneme is not attested. A plus

sign between parentheses indicates that the segment has restricted occurrences in that position. Unusual distributions of segments in identifiable loanwords are not taken into account.

Table 2.10: Distribution of consonants in KM at the syllable level

Position	p	b	t	d	c	j	k	g	ʔ	m	n	ɲ	ŋ	s	x	h	l	ɣ	w	y
onset	+	+	+	+	+	+	+	+	-	+	+	+	+	+	-	+	+	+	+	+
coda	-	-	-	-	-	-	-	-	+	(+)	(+)	(+)	+	-	-	+	-	-	-	-

(+ : attested, - : not attested, (+) : attested but restricted)

A single segment in onset position can be any consonant except for the glottal stop /ʔ/ and the velar fricative /x/, whereas in coda position, only the nasals and the glottals /ʔ, h/ are permitted.[13] The distribution of /x/ is highly restricted; as mentioned earlier, /x/ only appears in the geminated form /xx/ in word-initial onset position. The bilabial, alveolar and palatal nasals /m, n, ɲ/ may occur as a coda, but only when followed by a homorganic voiced obstruent, that is, they only appear in homorganic nasal + voiced obstruent sequences /-m.b-, -n.d-, -ɲ.j-/ in word-medial position. The velar nasal /ŋ/, on the other hand, appears in the consonant sequence /-ŋ.g-/ and in word-final position (see §2.4.3.1).[14]

Complex onsets CC are consonant clusters in the sense that the two consonants belong to the same syllable. They cannot be analysed as consonant sequences that fall into two syllables as they typically occur word-initially, and there is no evidence for syllabic consonants. CC clusters can contain either two identical segments C_xC_x or two dissimilar segments C_xC_y, which I call "geminate clusters" and "non-geminate clusters" respectively. Phonetically, geminate clusters are realised as single long units; phonologically, they are taken as a type of clusters on account of some variation attested between a form with a geminate cluster and a form with a non-geminate cluster,

[13] An initial glottal stop at the phonetic level is always present in vowel-initial words, but there is no contrast between initial [ʔ] and Ø at the phonemic level. A phonemic glottal stop also occurs word-medially as the onset of the final syllable, but it is exclusively attested in loanwords, e.g., /jmaʔãʔ/~/mmaʔãʔ/ 'Friday' and /maʔãh/ 'sorry', as cited previously.

[14] Nasal codas may be represented by an archiphoneme /N/. In word-medial position, the realisation of /N/ shares the same place feature of the following obstruent, and in word-final position, it is realised as a velar nasal [ŋ].

e.g., /tliŋɔ/~/lliŋɔ/ 'ear', /kmaɣiŋ/~/mmaɣiŋ/ 'yesterday' and /jnɛlɔ/~/nnɛlɔ/ 'window'. Attested complex onsets CC are presented in §2.5.

The distribution of oral vowels at the syllable level is summarised in Table 2.11. There are only few constraints.

Table 2.11: Distribution of oral vowels in KM at the syllable level

Syllable type	a	i	u	ɛ	ɔ	e	o	ə
V	+	+	+	+	+	+	+	-
VC	+	+	+	+	+	+	+	-
CV	+	+	+	+	+	+	+	+
CVC	+	+	+	+	+	+	+	+
CCV	+	+	+	+	+	-	-	+
CCVC	+	+	+	+	+	-	-	+

(+ : attested, - : not attested)

First, the schwa /ə/ is not allowed in onsetless syllables (V or VC); in other words, it does not occur initially in a syllable. All other oral vowels can be preceded by an onset and/or followed by a coda. Second, the mid-high vowels /e, o/ are not attested in syllables with a complex onset. This is because in words with a canonical disyllabic shape, /e, o/ only appear in final syllables, whereas syllables with a complex onset only occur in penultimate position (see §2.4.3.2).

The distribution of nasal vowels cannot be summarised in full details given the limited number of instances, but based on what can be observed, nasal vowels typically occur in syllables with a CVC shape where the coda C is a glottal stop /ʔ/. Some examples have been presented in Table 2.8. Some loanwords, function words, ideophones and interjections display a deviant pattern, where a nasal vowel appears in open syllables, as in the following examples: /dɔʔɔ̃/ 'to pray' (< Arabic *duʿāʾ*, cf. SM *doa*), /hɔ̃/ 'AFF', /cã/ 'sound of water ejection', and /wã/, which is an interjection used when one notices something smelly. /ɛ̃/ in /matɔ kaɛ̃/ 'fish hook' is another exception.

2.4 Word structure

In describing word structure, it is necessary to make the distinction between simple words and complex words. A morphologically simple word consists of one morpheme, whereas a complex word is made up of more than one morpheme. Complex words can be affixed forms, which I refer to as "derivatives", compounds or reduplicated forms. For a more detailed description of the morphological system, see Chapter 5.

In this section, I first describe the word shapes of simple words and derivatives separately in §2.4.1 and §2.4.2. As will be shown, despite having different morphological structures, simple words and derivatives share a close similarity in their word shapes with a strong preference for disyllabicity. In §2.4.3, I summarise the phoneme distributions at the word level, which hold for both simple words and derivatives.

2.4.1 Structure of simple words

Simple words in KM are typically disyllabic, but they can also contain one syllable or more than two syllables.

Monosyllabic simple words appear in a (C)(C)V(C) template, with minimally a vowel as the nucleus. Some examples are presented in Table 2.12.

Table 2.12: Monosyllabic simple words in KM

Syllable type	Example	Gloss
V	/a/	'INTERJ'
VC	/eh/	'INTERJ'
CV	/ɲɔ/	'coconut'
CVC	/moŋ/	'2SG'
CCV	/nnɛ/	'six'
CCVC	/mmah/	'gold'

While some monosyllabic content words can be found, monosyllables are more often function words, e.g., the personal pronoun /moŋ/ '2SG', prepositions such as /kɔ/ 'to; AGT' and /di/ 'LOC', and demonstratives /ni/ 'DEM.PROX'

and /tɯ/ 'DEM.DIST'. Monosyllables with an initial vowel are exclusively interjections. Also note that monosyllables with a CCV or CCVC shape typically have a geminate cluster in onset position, except in cases like /mbuh/ 'to blow', /mboŋ/ 'dew' and /ŋgɛ/ 'hornbill', which have an initial cluster consisting of a nasal and a homorganic stop.[15]

The majority of KM simple words are disyllabic with a (C)(C)V(C).(C)V(C) template, which is built upon two of the six possible syllable types presented in Table 2.9. One restriction on such combinations of two syllables is that syllables with a complex onset can only be word-initial. All other logical combinations of two syllables except for ×V.VC and ×CCV.VC are attested and illustrated by the examples in Table 2.13.

Table 2.13: Disyllabic simple words in KM

Syllable type	Example	Gloss
V.V	/a.e/	'water'
V.CV	/a.ti/	'liver'
V.CVC	/i.doʔ/	'to live'
VC.CV	/am.bɔ/	'1SG (polite)'
VC.CVC	/am.biʔ/	'to take'
CV.V	/ba.u/	'shoulder; smell'
CV.VC	/ta.oŋ/	'year'

[15] The analysis of geminate onset clusters in monosyllabic words is not beyond dispute. A few important observations may be laid out here. First, pairs contrasting a singleton C and a geminate CC are not attested in monosyllabic words. Second, monosyllabic content words are always pronounced with initial geminates in careful speech and in isolation (typically in word list elicitation), e.g., [ɲɲɔ] 'coconut', [mmɔʔ] 'mother', [ppaʔ] 'four' and [nnɛ] 'six'. A preliminary examination suggests that the length of initial consonants in these words and the proportion of the consonant duration in the whole syllable are comparable to a typical geminate found in disyllabic words. Third, in natural and connected speech, some monosyllabic content words are consistently pronounced with a geminate, e.g., [ppaʔ] 'four' in [duwɔ puloh ppaʔ] 'twenty four' and [mmah] 'gold' in [pase mmah] 'Pasir Mas (toponym)', whereas others are typically pronounced without a geminate, e.g., [ɲɔ] 'coconut' in [wɔhɲɔ] 'coconut (fruit)'. Based on these observations, I consider that only words consistently pronounced with a geminate have a underlying CCV or CCVC structure. This analysis is also supported diachronically, as monosyllables with a phonemic geminate originate from earlier disyllables with the loss of schwa in the penultimate syllable. Compare KM /ppaʔ/ 'four', /nnɛ/ 'six' and /mmah/ 'gold' with SM cognates *əmpat*, *ənam* and *əmas* (also see §7.5.1).

Syllable type	Example	Gloss
CV.CV	/ka.ki/	'leg; foot'
CV.CVC	/da.ɣɔh/	'blood'
CVC.CV	/paɲ.jɛ/	'long'
CVC.CVC	/puŋ.goŋ/	'buttocks'
CCV.V	/pɣa.u/	'canoe'
CCV.CV	/nna.tɛ/	'animal'
CCV.CVC	/ssə.kiŋ/	'poor'
CCVC.CV	/blaɲ.jɔ/	'expense; bride price'
CCVC.CVC	/ccam.bɔh/	'bean sprouts'

Both geminate clusters and non-geminate clusters can be found as the initial CC onsets in disyllabic simple words. The template (C)(C)V(C).(C)V(C) also shows that a word-medial -C.C- consonant sequence is formed when the coda C of the penultimate syllable precedes the onset C of the final syllable. I use the term "consonant sequences" to refer to such strings of two adjacent consonants that are heterosyllabic, in contrast to "consonant clusters". Attested consonant sequences are discussed in §2.6.

Simple words with more than two syllables are rare, and most of them can be identified as loanwords. Some examples are given in Table 2.14.

Table 2.14: Simple words with more than two syllables in KM

Syllable type	Example	Gloss	Origin
V.CV.CV	/u.ta.ɣɔ/	'north'	< Sanskrit
V.CV.CVC	/ɔ.ga.niʔ/	'organic'	< English
CV.CV.CV	/ba.si.ka/	'bicycle'	< English
CVC.CV.CV	/pus.ta.kɔ/	'library'	< Sanskrit
CV.CV.CV.CVC	/ta.li.bi.siŋ/	'television'	< English

/utaɣɔ/ 'north' and /pustakɔ/ 'library' have their ultimate origin in Sanskrit, but they are likely to have been borrowed directly from SM *utara* and *pustaka*. /ɔganiʔ/ 'organic' and /basika/ 'bicycle' are loanwords from English, possibly via SM *organik* and *basikal*. The quadrisyllabic word /talibisiŋ/ 'television' is also borrowed from English. Some other trisyllabic simple words,

such as /bə.ka.li/ 'maybe' and /ha.li.yɔ/ 'ginger', do not have an obvious traceable foreign origin. However, irregular phonological patterns such as the unexpected initial /h/ in /haliyɔ/ 'ginger' (see §2.4.3.1) suggest that they too were borrowed.

2.4.2 Structure of derivatives

Derivatives are words that have an affix attached to a base form. As regards their word shapes, derivatives in KM have at least two syllables, and a preference for disyllabicity is also attested. Derivatives with more than two syllables are uncommon, and many of them can be arguably taken as loanwords.

Table 2.15 lists some examples of disyllabic derivatives in KM.

Table 2.15: Disyllabic derivatives in KM

Syllable type	Example	Morphological structure	Gloss	
CCV.V	/bɣa.e/	*bɣ-ae*	(INTR-water)	'watery'
CCV.VC	/ssa.iŋ/	*s-saiŋ*	(INTR-friend)	'to befriend'
CCV.CV	/bla.ɣi/	*b-laɣi*	(MID-run)	'to run'
CCV.CVC	/pɣa.bih/	*pɣ-abih*	(CAUS-finished)	'to finish'
CCVC.CV	/mman.di/	*m-mandi*	(CAUS-bath)	'to bath s.o.'
CCVC.CVC	/nnam.bɔh/	*NN$_1$-<t>ambɔh*[16]	(IPFV-add)	'adding'

As will be discussed in §5.3.1, KM only has a small number of affixes, all of which are prefixes, each consisting of two consonants. Given a base form with an initial vowel or a single consonant, prefixation does not add an extra syllable, but generates a disyllabic form with an initial CC cluster (see the morphophonological alternations of prefixes in §5.3.1.1). Importantly, all word shapes attested in disyllabic derivatives are also found in simple words, and the same maximal CCVC.CVC shape is shared.

[16] *NN-* stands for a geminate nasal prefix that undergoes morphophonological alternations of nasal assimilation and substitution depending on the base-initial segment, see §5.3.1.1. The angle brackets <> indicate the base-initial segment that is deleted during this morphophonological process.

Bases with initial CC clusters typically do not undergo prefixation, except in a few examples such as /bə.tɣa.bo/ *bə-tɣabo* (MID-scattered) 'cluttered', /bə.sɣa.buʔ/ *bə-sɣabuʔ* (MID-upset) 'upset' and /bə.glabɔh/ (MID-anxious) 'anxious', whereby prefixation results in trisyllabic derivatives. The bases in all these examples have non-geminate clusters, more specifically comprised of an obstruent followed by a liquid. Bases with initial geminate clusters are never prefixed.

Some trisyllabic words may be analysed as having a nominalising suffix *-ɛ*, as in the following examples:

(1)	/u.ku.mɛ/	*ukum-ɛ*	(penalise-NMLS)	'penalty'
	/ba.la.sɛ/	*balas-ɛ*	(reply-NMLS)	'reply'
	/ha.ɾa.pɛ/	*haɾap-ɛ*	(hope-NMLS)	'hope'
	/ki.sa.ɣɛ/	*kisar-ɛ*	(grind-NMLS)	'grinder'
	/kan.da.ɣɛ/	*kandar-ɛ*	(carry-NMLS)	'shoulder pole'

Nevertheless, the nominalising suffix *-ɛ* is not productive, and the bases ×*ukum-*, ×*balas-*, ×*harap-*, ×*kisar* and ×*kandar* are not synchronically attested (cf. *ukoŋ*, *balah*, *aɣaʔ*, *kisa* and *kanda*, which are attested verbs with the same meaning). Some of these forms also display unexpected sound patterns. For instance, /haɾapɛ/ 'hope' has an unexpected initial /h/ and a tap /ɾ/ (see §2.4.3.1). This raises questions about whether these trisyllabic derivatives are native at all, or possible borrowings from another Malayic variety.

Similarly, some quadrisyllabic words may be analysed as complex, formed with the circumfix *pə- -ɛ*, as shown in (2).

(2)	/pə.ja.lɛ.nɛ/	*pə-jalɛ(n)-ɛ*	(NMLS-road-NMLS)	'journey'
	/pə.ɣu.mɔ.hɛ/	*pə-ɣumɔh-ɛ*	(NMLS-house-NMLS)	'household'
	/pə.jum.pə.ʔɛ/	*pə-jumpə(ʔ)-ɛ*	(NMLS-meet-NMLS)	'meeting'

Here too, there is evidence suggesting that these forms may not be inherited or synchronically derived. First, the circumfixing process is apparently not regular; for instance, /n/ in /jalɛ/ 'road' → /pəjalɛnɛ/ 'journey' is unexplained. Second, some words reveal unexpected sound patterns, such as the consonant sequence /-m.p-/ in /pəjumpəʔɛ/ 'meeting' (see §2.6). I consider words like /pəjumpəʔɛ/ as (nonce) borrowings, possibly from SM *pərjumpaan* (often pronounced as [pəjumpəʔan]) with the adaptation of *-an* to /-ɛ/.

The foregoing examination reveals that although the morphological structures of simple words and derivatives are fundamentally different,

there is a significant similarity in terms of their word shapes. As far as native words are concerned, both types of words exhibit a canonical disyllabic structure with a maximal CCVC.CVC shape. While some monosyllabic simple words and trisyllabic derivatives are allowed, the strong preference for disyllabicity is evident. It is also noteworthy that derivational processes typically do not form words exceeding the maximal disyllabic CCVC.CVC template. In fact, simple words and derivatives do not only share a disyllabic word shape; phonotactic constraints on the distribution of segments at the word level, permitted clusters and sequences, also apply to both simple words and derivatives. This uniformity finds its origin in diachronic processes, as the synchronic constraints are the manifestations of the same phonological changes that have affected both word types. See more in §7.5.2.

2.4.3 Phoneme distribution at the word level

The following phoneme distributions are summarised based on the canonical disyllabic template, with further consideration of monosyllables when relevant.

2.4.3.1 Distribution of consonants

Table 2.16 displays the distribution of single consonants at the word level. Distributional patterns of consonants in loanwords are excluded in the generalisation.

Table 2.16: Distribution of consonants in KM at the word level

Position	p	b	t	d	c	j	k	g	ʔ	m	n	ɲ	ŋ	s	x	h	l	ɣ	w	y
initial	+	+	+	+	+	+	+	+	-	+	+	+	(+)	+	-	(+)	+	+	(+)	(+)
medial	+	+	+	+	+	+	+	+	-	+	+	+	+	+	-	+	+	+	+	+
final	-	-	-	-	-	-	-	-	+	-	-	-	+	-	-	+	-	-	-	-

(+ : attested, - : not attested, (+) : attested but uncommon)

In word-initial position, the glottal stop is not phonemic. The velar fricative /x/ only appears in the geminated form /xx/; the single segment /x/ is also

not allowed in any other position in a word. The presence of /ŋ, h, w, y/ is uncommon. Initial /ŋ, w, y/ are mostly found in shortened variants of some disyllables, such as /ŋa/~/dəŋa/ 'with' (also /ŋɛ/~/dəŋɛ/ 'with'), /wi/~/buwi/ 'to give', /wɔh/~/buwɔh/ 'fruit', /waʔ/~/buwaʔ/ 'to do' and /ya/~/ɣiya/ 'ringgit (currency unit)'.[17] Initial /h/ is found in monosyllabic function words such as /hɔ̃/ 'AFF' and /hɔʔ/ 'REL'. In disyllabic words, initial /ŋ/ is attested in one instance /ŋaŋɔ/ 'to open wide', and /h/ occurs in few examples such as /hoŋga/ 'to run', /hambaʔ/ 'to chase' and /hagɔ/ 'to throw'.[18]

Word-medially, the glottal stop /ʔ/ and the velar fricative /x/ are not allowed. A word-medial /h/ is usually found between alike vowels, e.g., /dɛhɛ/ 'branch', /pɔhɔ/ 'thigh', /pɔhoŋ/ 'tree' and /jahaʔ/ 'bad'. Some exceptions include /pahiʔ/ 'bitter' and /bɣəhi/~/bɣahi/ 'to like (very much); be addicted to'.

In word-final position, only three segments /ʔ, ŋ, h/ are permitted.

2.4.3.2 Distribution of vowels

The distribution of oral vowels at the word level is summarised in Table 2.17.

Table 2.17: Distribution of oral vowels in KM at the word level

Position	a	i	u	ɛ	ɔ	e	o	ə
penultimate	+	+	+	+	+	-	-	+
final	+	+	+	+	+	+	+	-

(+ : attested, - : not attested)

[17] The cognates of /buwɔh/ 'fruit' and /buwaʔ/ 'to do' and /ɣiya/ 'ringgit' in SM are *buah*, *buat* and *rial*, which are analysed as having vowel sequences /-u.a-/ or /-i.a-/. An intervocalic glide breaking up the sequence can sometimes be heard, but glides following a corresponding high vowel (i.e., /w/ following /u/ and /y/ following /i/) are considered non-phonemic (Adelaar 1992: 11). In KM, I treat intervocalic glides following high vowels as phonemic, supported by evidence from the preservation of glides in word-initial position in the shortened forms of disyllabic words, as demonstrated by the examples here. See more in §7.3.5.1.

[18] /hoŋga/ 'to run' appears to be a loanword with the unexpected occurrence of /o/ in the penultimate syllable (see §2.4.3.2). The word itself does not seem native Malayic, but the source is unknown. /hambaʔ/ 'to chase' and /hagɔ/ 'to throw' may be native, but the occurrence of initial /h/ would be unexplained.

The mid-high vowels /e, o/ are not allowed in penultimate syllables. The schwa /ə/ only occurs in non-final syllables and syllables with an onset (§2.3), which means that it does not occur word-initially. Furthermore, among vowels that do appear in final syllables, only /i, e, o/ are allowed before final /ŋ/. Word-final ˣ/-aŋ/, ˣ/-ɛŋ/, ˣ/-ɔŋ/ and ˣ/-uŋ/ are not attested.

Nasal vowels are often attested in monosyllabic words or final syllables in disyllabic words. Some examples are given in Table 2.18.

Table 2.18: Words with nasal vowels in KM

Syllable type	Example	Gloss
monosyllabic	/hɔ̃/	'AFF'
	/sɔ̃ʔ/	'very attractive'
disyllabic	/ttu.wãʔ/	'wart'
	/kə.cɛ̃ʔ/	'small'
	/ɛ.sɔ̃ʔ/	'to scoot over'
	/bu.sũʔ/	'smelly'

2.5 Consonant clusters

Table 2.19 displays attested word-initial consonant clusters C_1C_2 in KM. C_1 is indicated on the left on the vertical line, and C_2 on the top on the horizontal line. Clusters enclosed in parentheses are uncommon, which may either appear as variants of other clusters or be restricted to loanwords.

With the exception of the glottals /h, ʔ/ and the palatal glide /y/, all consonants can occur in a consonant cluster, whether in a geminate cluster or a non-geminate cluster. Within non-geminate clusters, six types of clusters can be categorised based on their components: 1) obstruent + liquid; 2) obstruent + obstruent; 3) obstruent + nasal; 4) nasal + obstruent; 5) nasal + liquid; and 6) obstruent + glide. Each type is explained below, and examples are provided in (3) through (9).

Table 2.19: Attested consonant clusters in KM

C_1 \ C_2	p	b	t	d	c	j	k	g	m	n	ɲ	ŋ	s	x	l	ɣ	w
p	pp	-	-	pd	-	-	-	-	-	pn	-	-	-	-	pl	pɣ	-
b	-	bb	-	-	-	-	-	-	-	-	-	-	bs	-	bl	bɣ	-
t	-	tb	tt	-	-	tj	tk	tg	tm	-	-	-	-	-	tl	tɣ	(tw)
d	-	-	-	dd	-	-	-	-	-	-	-	-	-	-	(dl)	dɣ	-
c	cp	-	-	-	cc	-	-	-	cm	-	-	-	-	-	cl	cɣ	(cw)
j	-	-	-	-	-	jj	-	-	jm	jn	-	-	-	-	jl	jɣ	jw
k	-	kb	-	kd	-	-	kk	-	km	-	-	-	-	-	kl	-	(kw)
g	-	-	-	-	-	-	-	gg	gm	-	-	-	-	-	gl	(gɣ)	gw
m	-	mb	-	-	-	-	-	-	mm	-	-	-	-	-	ml	-	-
n	-	-	-	nd	-	-	-	-	-	nn	-	-	-	-	-	-	-
ɲ	-	-	-	-	-	-	-	-	-	-	ɲɲ	-	-	-	-	-	-
ŋ	-	-	-	-	-	-	-	ŋg	-	-	-	ŋŋ	-	-	-	-	-
s	sp	sb	st	sd	sc	sj	sk	sg	sm	sn	-	-	ss	-	sl	sɣ	-
x	-	-	-	-	-	-	-	-	-	-	-	-	-	xx	-	-	-
l	-	-	-	-	-	-	-	-	-	-	-	-	-	-	ll	-	-
ɣ	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	ɣɣ	-

(- : not attested, ( ) : attested but uncommon)

First, all segments except for /h, ʔ, w, y/ can appear in geminate clusters, as exemplified in (3).

(3) Geminate clusters

/pp-/	/ppalɔ/	'head'
/bb-/	/bbisiʔ/	'to whisper'
/tt-/	/ttinɔ/	'female'
/dd-/	/ddɛɣɔ/	'flag'
/cc-/	/ccayɔ/	'to believe'
/jj-/	/jjabaʔ/	'office'
/kk-/	/kkatɔʔ/	'frog'
/gg-/	/ggɛsɛʔ/	'matches'
/mm-/	/mmaɣi/	'cupboard'
/nn-/	/nnatɛ/	'animal'

/ɲɲ-/	/ɲɲakeʔ/	'disease'
/ŋŋ-/	/ŋŋaji/	'to learn'
/ss-/	/ssəkiŋ/	'poor'
/xx-/	/xxɛtɔ/	'car'
/ll-/	/llabɔ/	'spider'
/ɣɣ-/	/ɣɣamɔ/	'butterfly'

Geminate clusters are also found in a number of monosyllabic words, such as /nnɛ/ 'six' and /mmah/ 'gold'. Some geminate clusters are morphologically complex. For instance, /ddaɣɔh/ 'to bleed' and /lluma ʔ/ 'to crush' are derived from the corresponding bases /daɣɔh/ 'blood' and /luma ʔ/ 'crushed', where the initial geminated segment stands for an intransitive verbal marker and a causative marker respectively. Geminate nasals often represent the nasal prefix *NN*$_1$- 'IPFV' or *NN*$_2$- 'NMLS', e.g., /mmaŋge/ *mm-<p>aŋge* (IPFV-call) 'calling' and /ɲɲapuh/ *ɲɲ-<s>apuh* (NMLS-sweep) 'broom'. The morphological aspects of geminate clusters are examined in more detail in §5.3.

The most common non-geminate clusters have an obstruent followed by a liquid, as shown in (4).

(4) Obstruent + liquid

/pl-/	/plaka/	'thunder'
/pɣ-/	/pɣaŋa/	'behaviour'
/bl-/	/blakɛ/	'back'
/bɣ-/	/bɣani/	'brave'
/tl-/	/tliŋɔ/	'ear'
/tɣ-/	/tɣəbɛ/	'to fly'
/dl-/	/dlimɔ/	'pomegranate'
/dɣ-/	/dɣakɔ/	'insubordinate'
/cl-/	/clakɔ/	'misfortune'
/cɣ-/	/cɣəmiŋ/	'mirror'
/jl-/	/jluwɔʔ/	'to vomit'
/jɣ-/	/jɣəneh/	'clear'
/kl-/	/kladi/	'taro'
/gl-/	/gləga/	'floor'
/gɣ-/	/gɣətɔʔ/	'bridge'
/sl-/	/slaseh/	'basil'
/sɣ-/	/sɣatoh/	'one hundred'

Among these clusters, /dl/ is only attested in one instance, namely /dlimɔ/ 'pomegranate', which is more commonly pronounced as /jlimɔ/ by older speakers. /gɣ-/ is attested in a few words but it always varies with /ɣɣ-/, as seen in /gɣətɔʔ/~/ɣɣətɔʔ/ 'bridge', /gɣɛhɛ/~/ɣɣɛhɛ/ 'molar tooth' and /gɣɛjɔ/ ~/ɣɣɛjɔ/ 'church'. It can thus be concluded that in obstruent + liquid clusters, if C_1 is a stop, C_1 and the following liquid C_2 typically have different places of articulation. The absence of ×/kɣ-/ cluster can be explained by this constraint, as the articulatory similarity between the velar /k/ and /ɣ/ led to the assimilation of earlier +kɣ- > /xx-/ (see §7.5.2.3).

Examples in (5) illustrate the third type of clusters, which consist of two obstruents.

(5) Obstruent + obstruent

/pd-/	/pdiyɔ/	'who'
/bs-/	/bsusoŋ/	'to pile up'
/tb-/	/tbaka/	'tobacco; to be burnt'
/tj-/	/tjatoh/	'to fall (unintentionally)'
/tk-/	/tkəjuʔ/	'to be startled'
/tg-/	/tgəlɛ/	'to sink'
/cp-/	/cpədɔʔ/	'cempedak'
/kb-/	/kbumi/	'to bury'
/kd-/	/kdiyɛ/	'later'
/sp-/	/spuloh/	'ten'
/sb-/	/sbəlah/	'eleven'
/st-/	/ubi stɛlɔ/	'sweet potato'
/sd-/	/sdiɣi/	'alone'
/sc-/	/scawɛ/	'a cup'
/sj-/	/sjatɔ/	'weapon'
/sk-/	/skali/	'most; together'
/sg-/	/sgəlah/	'a glass'

Among the clusters presented in (5), /t/ and /s/ are most frequently attested as C_1, often representing segmentable morphemes. For instance, /tjatoh/ 'to fall (unintentionally)' and /tkəjuʔ/ 'to be startled' have a prefix *t-* 'NVOL' attached to the bases *jatoh* 'to fall' and *kəjuʔ* 'to startle'. In /spuloh/ 'ten' and /sbəlah/ 'eleven', a clitic *s=* 'one' is attached to the bases *puloh* 'ten' and *bəlah* 'teens'. When two stops form a cluster, it is usually the case that a voiceless stop precedes a voiced stop. In careful speech, a schwa can sometimes be

heard between two stops to ease pronunciation, as in [tgəlɛ]~[təgəlɛ] 'to sink' and [tjatoh]~[təjatoh] 'to fall (unintentionally)'. It is therefore debatable whether these words should be analysed as disyllabic with an epenthetic schwa or trisyllabic with a phonemic antepenultimate schwa (with occasional schwa deletion). The reasons for treating /tg-/ and /tj-/ as initial clusters in disyllabic words are twofold. Synchronically, this analysis fits into the overall word structure of KM, and diachronically, these stop + stop clusters share a common history with other clusters that do not show schwa epenthesis; see §7.5.2.3.

The fourth type of clusters has an obstruent followed by a nasal, as shown in (6).

(6) Obstruent + nasal

/pn-/	/pnamɔ/	'full (moon)'
/tm-/	/tmagɔ/	'copper'
/cm-/	/cmuɣu/	'jealous'
/jm-/	/jmaʔãʔ/	'Friday'
/jn-/	/jnɛlɔ/	'window'
/km-/	/kmaɣiŋ/	'yesterday'
/gm-/	/gmalɔ/	'herder'
/sm-/	/smayɛ/	'to pray'
/sn-/	/snaɣa/	'list'

Some of these clusters display variation with geminate clusters, e.g., /jmaʔãʔ/ ~/mmaʔãʔ/ 'Friday', /jnɛlɔ/~/nnɛlɔ/ 'window' and /kmaɣiŋ/~/mmaɣiŋ/ 'yesterday'. Unlike the free variation between /gɣ-/~/ɣɣ-/, however, the variation between obstruent + nasal clusters and geminate clusters appears to be lexically based. For instance, /jnɛlɔ/ 'window' has the variant /nnɛlɔ/, but /jnamɔ/ 'brand' does not exhibit this variation.

The fifth type of clusters, in which a nasal is followed by an obstruent, is attested in a few monosyllabic words listed in (7).

(7) Nasal + obstruent

/mb-/	/mbuh/	'to blow'
/mb-/	/mboŋ/	'dew'
/nd-/	/poʔ ndɛ/	'thunder'
/ŋg-/	/ŋgɛ/	'hornbill'

Another uncommon type of cluster consists of a nasal /m/ followed by a liquid /l/, as illustrated by examples in (8).

(8) Nasal + liquid

/ml-/	/mlayu/	'Malay'
/ml-/	/mlamboŋ/	'bumping'

The last type of clusters has an obstruent followed by a glide /w/, as shown in (9).

(9) Obstruent + glide

/tw-/	/twala/	'towel'	< Portuguese *toalha*	cf. SM *tuala*
/cw-/	/cwacɔ/	'climate'	< Sanskrit *svaccha*	cf. SM *cuaca*
/jw-/	/jwalɛ/	'sale'		
/kw-/	/kwasɔ/	'power'	< *kə-* + Sanskrit *vaśa*	cf. SM *kuasa*
/gw-/	/gwanɔ/	'how'		

This type of clusters is typically found in loanwords, as indicated above,[19] but it is included in the discussion because at least one item is native, namely /gwanɔ/ 'how', which is a contracted form of the compound /lagumanɔ/ (method-which) 'how'.

While the patterns of attested consonant clusters might not seem systematic at first sight, a closer look reveals that the Sonority Sequencing Principle (SSP) is the main constraint outlining permissible clusters. The SSP requires the sonority in a syllable to rise or show plateau from the onset towards the nucleus (Kenstowicz 1994: 254; Blevins 1995: 210). Geminate clusters may be considered as exemplifying sonority plateau, and in most types of non-geminate clusters, namely obstruent + liquid, voiceless stop + voiced stop, obstruent + nasal, nasal + liquid, and obstruent + glide, the requirement of a rising sonority towards the nucleus vowel is met. Exceptions include clusters of /s/ + a stop, which is not uncommon violation of the SSP, and nasal + obstruent clusters such as /mb-/ and /ŋg-/, which are only found in monosyllables. The overall pattern is robust.

[19] If inherited, the expected forms should reflect the contraction of /-u.a-/ to /ɔ/ (compare KM /pɔsɔ/ 'fasting', /kɔli/ 'wok' and /bɔyɔ/ 'crocodile' with SM cognates ***puasa***, ***kuali*** and ***buaya***, see more in §7.5.2.1). Final *-a* in SM *tuala* 'towel' should also correspond to /-ɔ/ in KM with the expected form ˣ/tɔlɔ/.

2.6 Consonant sequences

Word-medial consonant sequences are most commonly homorganic sequences of a nasal + a voiced obstruent, namely /-m.b-, -n.d-, -ɲ.j-, -ŋ.g-/, as illustrated in Table 2.20. Sequences of a nasal + a voiceless obstruent do not occur.

Table 2.20: Homorganic nasal + voiced obstruent sequences in KM

Sequence	Example	Gloss
/-m.b-/	/kam.biŋ/	'goat'
	/am.biʔ/	'to take'
/-n.d-/	/din.diŋ/	'wall'
	/man.di/	'to bathe'
/-ɲ.j-/	/piɲ.jɛ/	'to borrow'
	/paɲ.jɛ/	'long'
/-ŋ.g-/	/piŋ.gɛ/	'waist'
	/tuŋ.gu/	'to wait'

The main reason to analyse these nasal + obstruent combinations as heterosyllabic sequences is that they only occur word-medially, which stands in sharp contrast with consonant clusters, which typically occur word-initially. Furthermore, while nasal + obstruent combinations such as /mb-/ and /ŋg-/ are also attested as word-initial clusters (shown in example (7) above), they are of a rather different nature, as they are only attested in monosyllables with an idiosyncratic history (see §7.5.1).

Another type of word-medial sequences contains a glottal stop /ʔ/ followed by another segment, as illustrated by the examples in Table 2.21. These sequences are often found in (historically) contracted forms or loanwords, as indicated in the table, but the exact origin of medial /-ʔ-/ is not always clear. /mɔʔtɛ/ 'rambutan' and /bɛʔki/ 'to repair' are historically suffixed forms, which can be compared with their SM cognates *rambut-an* (hair-NMLS) and *baik-i* (good-APPL), but the source of /-ʔ-/ is obscure. The origin of /sɔʔmɔ/ 'always' is suggested following Brown (1956: 48), but it also does not explain the occurrence of /-ʔ-/ in this case. /-ʔ.ŋ-/ in /ɣɛʔŋɛ/ 'light (weight)' is unexplained (cf. SM *riŋan*).

Table 2.21: Examples of /-ʔ.C-/ sequences in KM

Sequence	Example	Gloss	Origin
/-ʔ.p-/	/baʔ.pɔ/	'why'	< *buwaʔ* 'do' + †*apa* 'what'
/-ʔ.t-/	/mɔʔ.tɛ/	'rambutan'	
/-ʔ.d-/	/taʔ.dɔʔ/	'not exist'	< †*taʔ* 'NEG' + *ada*(*ʔ*) 'EXIST'
/-ʔ.c-/	/mɔʔ.ciʔ/	'auntie'	< *mɔʔ* 'mother' + *ciʔ* 'sister'
/-ʔ.k-/	/bɛʔ.ki/	'to repair'	
/-ʔ.m-/	/sɔʔ.mɔ/	'always'	< *s*= 'same' + *ɔmɔ* 'age'
/-ʔ.n-/	/saʔ.ni/	'just now'	< *saʔãʔ* 'second' + *ni* 'DEM.PROX'
/-ʔ.ŋ-/	/ɣɛʔ.ŋɛ/	'light (weight)'	
/-ʔ.s-/	/tɔʔ.sɛ/	'not want to'	< *tɔʔ* 'NEG' + *sɛ* 'desire'
/-ʔ.l-/	/tɔʔ.leh/	'cannot'	< *tɔʔ* 'NEG' + *buleh* 'can'
/-ʔ.w-/	/bɛʔ.wɔh/	'feast'	< English *big work*

† indicates earlier forms which are no longer attested.

2.7 Vowel sequences

Two vowels can appear adjacent to each other, forming a vowel sequence. All recorded vowel sequences in KM have the low vowel /a/ as the first component, as presented in Table 2.22. Among the vowels permitted in penultimate syllables, the schwa /ə/ and the mid-low vowels /ɛ, ɔ/ are never directly followed by another vowel. Intervocalic glides following high vowels are taken as phonemic rather than epenthetic (see f.n. 17 in §2.4.3.1), thus excluding possible sequences of ˣ/-i.V-/ and ˣ/-u.V-/.

Table 2.22: Vowel sequences in KM

Sequence	Example	Gloss
/a.i/	/sa.iŋ/	'friend'
/a.u/	/ba.u/	'shoulder; smell'
/a.e/	/a.e/	'water'
/a.o/	/ta.oŋ/	'year'
/a.ɛ̃/	/matɔ ka.ɛ̃/	'fish hook'

Both vowels in vowel sequences are full vowels, each occupying the nucleus of one syllable. The analysis of treating them as vowel sequences rather than diphthongs also fits into the overall phonotactic patterns of KM.

2.8 Stress

In the last part of this phonological description, some aspects of the stress pattern in KM are discussed in a cautious way. A preliminary examination suggests a correlation between stress and a higher pitch, but this observation remains tentative as the acoustic details of stress are not entirely clear. The marking of stress in the following examples is therefore somewhat impressionistic rather than being based on a detailed acoustic study.

At the phonological level, KM has no phonemic stress. There are no pairs of otherwise identical words that differ in their stress pattern. Some tendencies for the placement of stress in disyllabic words are summarised below.

First, words with two open syllables (i.e., words with a CVCV shape) generally have stress on the penultimate syllable. When the penultimate syllable has a schwa, stress falls on the final syllable instead. Some examples are given in (10).

(10)	/tanɛ/	[ˈtanɛ̃]	'to plant'
	/tido/	[ˈtido]	'to sleep'
	/pula/	[ˈpula]	'island'
	/lɛpa/	[ˈlɛpa]	'to throw'
	/bɔyɔ/	[ˈbɔyɔ]	'crocodile'
	/dəbu/	[dəˈbu]	'dust'
	/bəsa/	[bəˈsa]	'big'

There are nevertheless a number of counterexamples that do not comply with the general pattern: /maɣi/ 'to come' and /pase/ 'sand' are most commonly pronounced with stress of the final syllable, i.e., [maˈɣi] and [paˈse], whereas /pəgɛ/ [ˈpəgɛ] 'to hold' and /ɣəgɔ/ [ˈɣəgɔ] 'price' have stress on the penultimate syllable even though the penultimate vowel is a schwa.

Second, in words with a closed syllable ending in a coda /ʔ/, the closed syllable carries stress. For instance, words with a CVCVC shape and a final /ʔ/ have ultimate stress, as in the following examples:

(11)	/kilaʔ/	[kiˈlaʔ]	'lightning'
	/muloʔ/	[mũˈloʔ]	'mouth'
	/kɔpɛʔ/	[kɔˈpɛʔ]	'breast'
	/miɲɔʔ/	[mĩˈɲɔ̃ʔ]	'oil'

Syllables with a coda /ʔ/ also carry stress if they appear in the penultimate position: /mɔʔtɛ/ 'rambutan' and /tɔʔsɛ/ 'not want to' have penultimate stress, i.e., [ˈmɔ̃ʔtɛ] and [ˈtɔʔsɛ].

There does not seem to be a consistent stress pattern in CVCVC words with a final consonant other than /ʔ/; both penultimate stress and ultimate stress can be found:

(12)	*Penultimate stress*		
	/mamɔh/	[ˈmãmɔ̃h]	'to chew'
	/kabuh/	[ˈkabuh]	'fog'
	/buɣoŋ/	[ˈbuɣoŋ]	'bird'
	/kapoŋ/	[ˈkapoŋ]	'village'
	Ultimate stress		
	/tujoh/	[tuˈjoh]	'seven'
	/jatoh/	[jaˈtoh]	'to fall'
	/kuciŋ/	[kuˈciŋ]	'cat'
	/dagiŋ/	[daˈgiŋ]	'meat'

Most importantly, in words with initial geminates, stress always falls on the initial syllable, regardless of whether there is a final consonant or if the penultimate vowel is a schwa. A penultimate schwa following an initial geminate can also be stressed. Examples demonstrating the cooccurrence of initial geminates and initial stress are presented in (13).

(13)	/ttinɔ/	[ˈttinɔ̃]	'female'
	/ppalɔ/	[ˈppalɔ]	'head'
	/mmaɣi/	[ˈmmãɣi]	'cupboard'
	/llabɔ/	[ˈllabɔ]	'spider'
	/ggɛsɛʔ/	[ˈggɛsɛʔ]	'matches'
	/nnusuʔ/	[ˈnnũsuʔ]	'to hide'
	/ssəmɔ/	[ˈssəmɔ̃]	'all'
	/nnətɛ/	[ˈnnə̃tɛ]	'to lie down (face up)'
	/ssəjeʔ/	[ˈssəjeʔ]	'mosque'
	/jjəɣiʔ/	[ˈjjəɣiʔ]	'to call out; to cry'

Other word-initial complex onsets do not always cooccur with initial stress, and stress can be found on either syllable, as shown in (14). In fact, different speakers may place stress in different positions when pronouncing the same word. Both [ˈjɣənẽh] 'clear' and [jɣəˈnẽh] are heard, and [blaˈkɛ] 'back' is also sometimes pronounced as [ˈblakɛ].

(14) *Penultimate stress*

/bɣənɛ/	[ˈbɣənɛ̃]	'to swim'
/tliŋɔ/	[ˈtliŋɔ̃]	'ear'
/tganoŋ/	[ˈtganõŋ]	'rainbow; Terengganu'
/jɣəneh/	[ˈjɣənẽh]	'clear'
Ultimate stress		
/blakɛ/	[blaˈkɛ]	'seven'
/bɣəti/	[bɣəˈti]	'to stop'
/bɣəsiŋ/	[bɣəˈsiŋ]	'to sneeze'
/tɣiŋaʔ/	[tɣiˈŋãʔ]	'to miss (someone)'

The correlation between geminates and stress has been observed in Patani Malay (Yupho 1989; Hajek & Goedemans 2003). Yupho (1989) concluded that primary stress normally falls on the final syllable in Patani Malay but shifts to the first syllable if the onset is a geminate consonant. This correlation is often used as an argument for the heaviness of geminates (see Hayes 1989; Davis 1994; Topintzi 2008; Topintzi & Davis 2017). In KM, however, words do not have ultimate stress by default. While a correlation between word-initial geminates and initial stress is also observed, stress does not "shift" to the first syllable when the initial onset is a geminate.

To sum up, stress in KM is not distinctive. There are some tendencies for stress assignment, but not without variation. It seems that the default pattern is to stress the penultimate syllable unless it contains a schwa, and a closed syllable with a glottal stop /ʔ/ attracts stress. However, due to numerous exceptions, these patterns cannot be firmly established as rules. A more robust pattern is that initial geminates always cooccur with initial stress. The tendencies for stress assignment at the word level might also be overridden at the phrasal level. More phonetic evidence is needed to understand the exact acoustic cues of stress and how stress interacts with other aspects of the phonology.

2.9 Summary

In this chapter, I have described the segment inventory of KM, how segments combine at the syllable level, and how syllables combine to form words. I presented the distribution of phonemes at both the syllable and word levels, and I examined the phonotactic constraints governing permissible clusters and sequences. Additionally, I cautiously discussed the stress pattern in KM.

At the segment level, KM has a total of twenty native phonemic consonants, including nine stops, four nasals, three fricatives, two liquids and two glides. The vowel inventory is fairly large, consisting of twelve phonemic vowels. In addition to /a, i, u, ə/, there are two sets of mid-vowels /e, o/ and /ɛ, ɔ/, along with four phonemic nasal vowels /ã, ɛ̃, ɔ̃, ũ/. There are no phonemic diphthongs.

At the syllable level, KM has a basic template of (C)(C)V(C). Any single consonant except for /ʔ, x/ can occur as an onset, but only nasals and glottals are allowed as codas. Complex onsets CC are permitted, in which the two segments can be either identical or different. Complex onsets with two identical segments manifest as geminates at the phonetic level, but at the phonological level, they are analysed as a subtype of clusters. KM boasts a large inventory of geminate clusters; all segments except for the glottals /ʔ, h/ and the glide /y/ can appear in geminated forms. Attested non-geminate clusters generally follow the SSP. A schwa cannot occur in an onsetless syllable, and nasal vowels are commonly followed by a glottal stop.

At the word level, KM exhibits a canonical disyllabic structure with a (C)(C)V(C).(C)V(C) template. Syllables with a complex onset typically appear word-initially, and word-medial consonant sequences are most commonly homorganic sequences of a nasal + a voiced obstruent. Two vowels occurring adjacent to each other are V.V sequences, which always have /a/ as the first segment. I also showed that despite having different morphological structures, simple words and derivatives share a similar word shape with the same phonotactic constraints on the distribution of phonemes. In initial position, a glottal stop /ʔ/ is not phonemic, and the velar fricative /x/ only occurs in the geminated form /xx/. Initial /ŋ, h, w, y/ are also rare. In medial position, /ʔ/ and /x/ are not allowed in native words. In final position, only three consonants /ʔ, ŋ, h/ are permitted. Mid-high vowels /e, o/

and nasals vowels do not occur in penultimate syllables, whereas the schwa is not permitted in final syllables.

Stress is not phonemic in KM. There is considerable variation as regards stress placement, and further study is required to reveal the acoustic properties of stress. Overall, stress often falls on the penultimate syllable in a disyllabic word, and it shifts to the final syllable when the penultimate syllable has a schwa. Notably, syllables with a coda /ʔ/ tend to attract stress, and initial geminates always cooccur with initial stress.

CHAPTER 3

Phonology of Coastal Terengganu Malay

3.1 Introduction

This chapter presents the phonology of CTM. Its structure follows that of the previous chapter, starting with the presentation of the segment inventory in §3.2, followed by discussions on syllable structure and word structure in §3.3 and §3.4. Phonotactic constraints on permitted consonant clusters, consonant sequences and vowel sequences are examined in §3.5 to §3.7. The stress pattern is discussed in §3.8. Finally, §3.9 summarises the key points in this chapter.

Throughout this chapter, comparisons with KM will be drawn when relevant. Overall, the phonologies of these two varieties exhibit numerous resemblances, particularly in terms of the segment inventory, syllable shapes and word shapes. A notable distinction lies in the distributional constraints on vowels, and some differences can be observed with regard to permitted consonant clusters. These distinctions will be highlighted in this chapter.

3.2 Segment inventory

3.2.1 Consonant system

3.2.1.1 Consonant inventory

The consonant inventory of CTM is displayed in Table 3.1. The twenty native consonant phonemes are identical to those in KM: there are nine stops, four nasals, three fricatives, two liquids and two glides.

Table 3.1: Consonant inventory of CTM

		Labial	Dental	Alveolar	Palatal	Velar	Glottal
Stops	voiceless	p	t			k	ʔ
	voiced	b		d		g	
Nasals		m		n	ɲ	ŋ	
Fricatives	voiceless			s		x	h
	voiced			(z)			
Affricates	voiceless				c		
	voiced				j		
Liquids				l		ɣ	
Glides		w			y		

(c = IPA /c͡ç/, j = IPA /ɟ͡ʝ/, y = IPA /j/)

The asymmetry in the places of articulation of the coronal stops /t/ and /d/ is also attested in CTM: the voiceless stop is realised as dental [t̪], whereas the voiced stop is realised as alveolar [d]. The phoneme /ɣ/ is also treated as a liquid (see §2.2.1.1).

The voiced alveolar fricative /z/, which has a foreign origin and marginal phonemic status, is included in the table. It is found in loanwords such as /zina/ 'adultery' < Arabic *zināʾ* and /zamaŋ/ 'epoch' < Arabic *zamān*, sometimes in variation with the voiced palatal affricate /j/, e.g., [zamaŋ]~[jamaŋ] 'epoch'. The alveolar tap /ɾ/, which was taken as a borrowed phoneme in KM, is not present in CTM. A foreign rhotic sound is always adapted to become the velar liquid /ɣ/, as seen in /aɣənaʔ/ 'rabbit' < Arabic *arnab*, as well as English loanwords such as /stəɣeʔ/ 'straight' and /təɣa/ 'to try'.

3.2.1.2 Contrasts between consonants

Table 3.2 presents minimal or near-minimal pairs contrasting similar consonants in CTM. A closest pair with contrasting segments is given when no minimal pairs can be found. Contrasts found in different positions in a word are differentiated. See more on consonant distributions in §3.4.3.1.

Table 3.2: Contrasts between consonants in CTM

Contrast	Pair	Gloss
Word-initially		
/p/ - /b/	/**p**uloh/	'ten'
	/**b**uloh/	'bamboo'
/b/ - /m/	/**b**ukə/	'to open'
	/**m**ukə/	'face'
/t/ - /d/	/**t**aoŋ/	'year'
	/**d**aoŋ/	'leaf'
/d/ - /n/	/**d**adə/	'chest'
	/**n**ada/	'tone'
/c/ - /j/	/**c**uɣaŋ/	'steep'
	/**j**uɣaŋ/	'canyon'
/j/ - /ɲ/	/**j**awə/	'Java'
	/**ɲ**awə/	'soul'
/k/ - /g/	/**k**ali/	'times'
	/**g**ali/	'to dig'
/g/ - /ɣ/	/**g**ayə/	'style'
	/**ɣ**ayə/	'Eid al-Fitr'
/m/ - /n/	/**m**asiŋ/	'salty'
	/**n**asiʔ/	'rice'
/n/ - /s/	/**n**apɔʔ/	'to see'
	/**s**apɔh/	'garbage'
/l/ - /n/	/**l**amə/	'long (time)'
	/**n**amə/	'name'
/l/ - /ɣ/	/**l**ata/	'floor'
	/**ɣ**ata/	'necklace'

Contrast	Pair	Gloss
Word-medially		
/p/ - /b/	/lɛ**p**ɔ/	'to throw'
	/lɛ**b**ɔ/	'wide'
/b/ - /m/	/cu**b**ə/	'to try'
	/cu**m**ə/	'only'
/t/ - /d/	/u**t**aŋ/	'forest; debt'
	/u**d**aŋ/	'shrimp'
/d/ - /n/	/kə**d**a/	'shop'
	/kə**n**a/	'to know'
/c/ - /j/	/ka**c**aŋ/	'bean'
	/ta**j**aŋ/	'sharp'
/j/ - /ɲ/	/sə**j**aʔ/	'to toast'
	/sə**ɲ**aʔ/	'quiet'
/k/ - /g/ - /ŋ/	/i**k**aʔ/	'to tie up'
	/i**g**aʔ/	'to catch'
	/i**ŋ**aʔ/	'to think'
/g/ - /ɣ/ - /l/	/pə**g**aŋ/	'to hold'
	/pə**ɣ**aŋ/	'war'
	/pə**l**aŋ/	'mango'
/m/ - /n/ - /ŋ/	/ta**m**aŋ/	'park'
	/ta**n**aŋ/	'to plant'
	/ta**ŋ**aŋ/	'hand'
/m/ - /ɲ/	/bu**m**i/	'earth'
	/bu**ɲ**i/	'sound'
/l/ - /n/	/pə**l**oh/	'sweat'
	/pə**n**oh/	'full'
/s/ - /h/	/a**s**aʔ/	'smoke'
	/a**h**aʔ/	'Sunday'
/y/ - /w/	/a**y**aŋ/	'chicken'
	/a**w**aŋ/	'cloud'
Word-finally		
/ʔ/ - Ø	/kɔtɔ**ʔ**/	'box'
	/kɔtɔ/	'dirty'
/ŋ/ - Ø	/kaca**ŋ**/	'bean'
	/kaca/	'to disturb'

Contrast	Pair	Gloss
/h/ - Ø	/subo**h**/	'dawn'
	/subo/	'fertile'
/ʔ/ - /ŋ/	/ido**ʔ**/	'to live'
	/ido**ŋ**/	'nose'
/ʔ/ - /h/	/ata**ʔ**/	'roof'
	/ata**h**/	'top'
/ŋ/ - /h/	/jato**ŋ**/	'heart'
	/jato**h**/	'to fall'

Geminates are also found in CTM, and just as in KM, they only occur word-initially. Minimal or near-minimal pairs contrasting an initial singleton and an initial geminate are presented in Table 3.3.

Table 3.3: Contrasts between singletons and geminates in CTM

Contrast	Pair	Gloss
/p/ - /pp/	/**p**alə/	'nutmeg'
	/**pp**alə/	'head'
/b/ - /bb/	/**b**ini/	'wife'
	/**bb**ini/	'to marry (a wife)'
/t/ - /tt/	/**t**ido/	'to sleep'
	/**tt**ido/	'to fall asleep (unintentionally)'
/d/ - /dd/	/**d**aɣɔh/	'blood'
	/**dd**aɣɔh/	'to bleed'
/c/ - /cc/	/**c**aɣə/	'method'
	/**cc**aɣə/	'to talk'
/j/ - /jj/	/**j**aɲji/	'vow'
	/**jj**aɲji/	'to promise'
/k/ - /kk/	/**k**aɣoŋ/	'sack'
	/**kk**aɣoŋ/	'skink (k.o. lizard)'
/g/ - /gg/	/**g**aji/	'wage'
	/**gg**aji/	'saw (n.)'
/m/ - /mm/	/**m**atə/	'eye'
	/**mm**atə/	'gem'

Contrast	Pair	Gloss
/n/ - /nn/	/**n**anoŋ/	'rainbow'
	/**nn**ənoŋ/	'to stare'
/ɲ/ - /ɲɲ/	/**ɲ**awə/	'soul'
	/**ɲɲ**awə/	'to breath'
/s/ - /ss/	/**s**iyaŋ/	'daytime'
	/**ss**iyaŋ/	'to pity'
/l/ - /ll/	/**l**aki/	'husband'
	/**ll**aki/	'male'
/ɣ/ - /ɣɣ/	/**ɣ**ahaŋ/	'jaw'
	/**ɣɣ**ahaŋ/	'molar tooth'
/w/ - /ww/	/**w**aʔpə/	'why'
	/**ww**apə/	'how many, how much'

Among the twenty native phonemic consonants in CTM, only three segments /ʔ, h, y/ cannot appear in a geminated form. In addition to the fifteen segments that can be contrasted in singleton-geminate pairs as presented above, /x/ appears exclusively in the geminated form /xx/ in word-initial position, as in /xxusi/ 'chair' and /xxəkɔh/ 'to bite'. /ŋ/ also typically occurs geminated, as in /ŋŋaku/ 'to confess' and /ŋŋajɔ/ 'to teach'. The duration of geminates /ŋŋ/ and /xx/ is comparable to that of other geminates, and historically they also originate from the assimilation of two segments in a cluster (/ŋŋ-/ < $^{+}$mŋ- and /xx-/ < $^{+}$kɣ-, see §7.5.2.3). Note that the geminate bilabial glide /ww/ is attested in CTM, but not in KM.

At the phonological level, geminates are analysed as consonant clusters with two identical segments, on account of the variation attested between a form with a geminate cluster and a form with a non-geminate cluster, e.g., /smilaŋ/~/mmilaŋ/ 'nine', /tgəlaŋ/~/ggəlaŋ/ 'to sink' and /bɣəkah/~/wwəkah/ 'to bundle'. Attested consonant clusters are discussed in more detail in §3.5.

3.2.1.3 Phonetic realisations of consonants

Two consonant phonemes, namely the glottal fricative /h/ and the velar fricative /x/, have variable phonetic realisations, which are summarised in Table 3.4 and explained below.

Table 3.4: Phonetic realisations of some consonant phonemes in CTM

Phoneme	Realisation	Environment	Example		Gloss
/h/	[ç]~[h]	i__#	/piti**h**/	[piti**ç**]~[piti**h**]	'money'
			/kudi**h**/	[kudi**ç**]~[kudi**h**]	'scabies'
	[h]	elsewhere	/kabu**h**/	[kabu**h**]	'fog'
			/ata**h**/	[ata**h**]	'top'
/x/	[x]~[k]	#__	/**xx**usi/	[**xx**usi]~[**kk**usi]	'chair'
			/**xx**ijə/	[**xx**ijə]~[**kk**ijə]	'work'

First, when following a high front vowel /i/, word-final /h/ is often realised as a palatal fricative [ç]. /pitih/ 'money' has a variable pronunciation of [pitiç] ~[pitih], and /kudih/ 'scabies' is realised as [kudiç]~[kudih].

Second, a word-initial velar fricative /x/ is realised as [k] by some speakers. Since /x/ only occurs in the geminated form /xx/, the observable variation is between [xx-] and [kk-]. Examples include /xxusi/ [xxusi] ~[kkusi] 'chair', /xxɛtə/ [xxɛtə]~[kkɛtə] 'car' and /xxəpoʔ/ [xxəpoʔ]~[kkəpoʔ] 'k.o. cracker'.

Other consonant phonemes do not show evident variation in their phonetic realisations.

3.2.2 Vowel system

3.2.2.1 Vowel inventory

The vowel inventory of CTM is presented in Table 3.5. There are thirteen phonemic vowels, comprising eight oral vowels and five nasal vowels. The oral vowels have two sets of mid vowels: mid-high /e, o/ and mid-low /ɛ, ɔ/. Compared to KM, CTM features an additional phonemic nasal vowel /ĩ/. The distinction between allophonic nasalised vowels and genuine nasal vowels, as previously explained for KM in §2.2.2.1, also holds true for CTM.

Table 3.5: Vowel inventory of CTM

	Front	Central	Back
High	i, ĩ		u, ũ
Mid-high	e		o
Mid-low	ɛ, ɛ̃	ə	ɔ, ɔ̃
Low		a, ã	

There are no phonemic diphthongs in CTM. Vowel sequences are disyllabic, as in examples like /na.iʔ/ 'to climb' and /ta.u/ 'to know'. Attested vowel sequences in CTM are presented in §3.7. However, it's important to note that diphthongs at the phonetic level can be found as possible realisations of the mid-high vowels /e, o/. For instance, /puteh/ 'white' has a phonetic realisation varying from [puteh] to [puteḭh], and /muloʔ/ can be realised as [muloʔ]~[muloʊ̯ʔ]. See more detail in §3.2.2.3.

3.2.2.2 Contrasts between vowels

Minimal or near-minimal pairs demonstrating the phonemic status of CTM vowels are given in Table 3.6 and Table 3.7. Contrasts found in penultimate syllables and final syllables are distinguished, as words in CTM have a canonical disyllabic structure with distributional constraints on certain vowels at the word level; see §3.4.3.2.

Table 3.6: Contrasts between oral vowels in CTM

Contrast	Pair	Gloss
Penultimate syllables		
/a/ - /i/ - /u/	/b**a**taŋ/	'stem'
	/b**i**taŋ/	'star'
	/b**u**taŋ/	'button'
/a/ - /u/ - /ɛ/	/m**a**ɣɔh/	'angry'
	/m**u**ɣɔh/	'cheap'
	/m**ɛ**ɣɔh/	'red'
/a/ - /ɔ/	/**a**ɣaʔ/	'to hope'
	/**ɔ**ɣaʔ/	'vein'

Contrast	Pair	Gloss
/a/ - /ə/	/p**a**ɣaŋ/	'machete'
	/p**ə**ɣaŋ/	'war'
/i/ - /ɛ/	/b**i**sə/	'poisonous'
	/b**ɛ**sə/	'usual'
/i/ - /ɔ/	/l**i**paʔ/	'to fold'
	/l**ɔ**paʔ/	'to jump'
/i/ - /ə/	/b**i**lə/	'when'
	/b**ə**lə/	'to breed'
/u/ - /ɔ/	/b**u**dɔʔ/	'kid'
	/b**ɔ**dɔh/	'stupid'
/u/ - /ə/	/m**u**tɔh/	'to vomit'
	/m**ə**tɔh/	'raw'
/ɛ/ - /ɔ/	/t**ɛ**mbɔʔ/	'to shoot'
	/t**ɔ**mbɔʔ/	'spear'
/ɛ/ - /ə/	/m**ɛ**naŋ/	'toy, game'
	/m**ə**naŋ/	'to win'
/ɔ/ - /ə/	/t**ɔ**pi/	'cap, hat'
	/t**ə**pi/	'edge'
Final syllables		
/a/ - /i/	/sik**a**ʔ/	'comb'
	/sik**i**ʔ/	'(a) little'
/a/ - /u/	/tə**b**a/	'thick'
	/tə**b**u/	'sugarcane'
/a/ - /ɔ/	/pan**a**h/	'hot'
	/pan**ɔ**h/	'bow'
/a/ - /e/	/sən**a**ŋ/	'easy'
	/sən**e**ŋ/	'Monday'
/a/ - /o/	/tak**a**ʔ/	'to catch'
	/tak**o**ʔ/	'afraid'
/a/ - /ə/	/lim**a**/	'citrus'
	/lim**ə**/	'five'
/i/ - /u/ - /ɔ/	/at**i**/	'liver'
	/at**u**/	'ghost'
	/at**ɔ**/	'to send'

Contrast	Pair	Gloss
/i/ - /e/	/gal**i**/	'to dig'
	/gal**e**/	'loose'
/i/ - /o/	/nip**i**h/	'thin'
	/nip**o**h/	'sugar palm'
/i/ - /ə/	/mat**i**/	'to die'
	/mat**ə**/	'eye'
/u/ - /e/ - /ɔ/	/pas**u**/	'flower pot'
	/pas**e**/	'sand'
	/pas**ɔ**/	'market'
/u/ - /o/	/al**u**/	'pestle'
	/al**o**/	'groove'
/u/ - /ə/	/kay**u**/	'wood'
	/kay**ə**/	'rich'
/ɔ/ - /o/	/bak**ɔ**/	'to burn'
	/bak**o**/	'basket'
/ɔ/ - /ə/	/kɔt**ɔ**/	'dirty'
	/kɔt**ə**/	'city'
/ɛ/ - /e/	/lɛmb**ɛ**ʔ/	'mattress'
	/ləmb**e**ʔ/	'soft'
/e/ - /o/	/bul**e**h/	'can; to get'
	/bul**o**h/	'bamboo'
/e/ - /ə/	/kat**e**/	'bed'
	/kat**ə**/	'to say'
/o/ - /ə/	/jal**o**/	'streak'
	/jal**ə**/	'fishnet'

Table 3.7 presents contrasts between oral and nasal vowels. No clear contrast between /ɛ̃/ and /ɛ/ has been found, as /ɛ̃/ only occurs in one item /mɛcɛ̃ʔ/ 'match' in the corpus. The phonemic status of /ɛ̃/ in /mɛcɛ̃ʔ/ 'match' is nevertheless supported by its occurrence independent of the nasal environment.

Table 3.7: Contrasts between oral and nasal vowels in CTM

Contrast	Pair	Gloss
/a/ - /ã/	/ləpah/	'after; then'
	/ləpãŋ/	'to slap'
/i/ - /ĩ/	/kəciŋ/	'to pee'
	/kəcĩʔ/	'small'
/u/ - /ũ/	/busuʔ/	'ant hill'
	/busũʔ/	'smelly'
/ɔ/ - /ɔ̃/	/ɛsɔʔ/	'tomorrow'
	/kɛsɔ̃ʔ/	'to scoot over'

3.2.2.3 Phonetic realisations of vowels

As mentioned earlier, the mid-high vowels /e, o/ in CTM may be realised as diphthongs [eɪ̯] and [oʊ̯], but this tendency of diphthongisation is only observed in certain environments, conditioned by the presence and the (lack of) nasality of the syllable onset, as well as the presence and the quality of the coda. Also note that mid-high vowels only occur in final syllables (§3.4.3.2).

Diphthongisation of mid-high vowels is observed when the vowel is preceded by a non-nasal or zero onset and followed by a coda /ʔ/ or /h/. For example, /kuleʔ/ 'skin' is realised as [kuleʔ]~[kuleɪ̯ʔ], and /buloh/ 'bamboo' is realised as [buloh]~[buloʊ̯h]. When /e, o/ are preceded by a nasal onset, they are nasalised and do not show the tendency of diphthongisation, regardless of the coda. /bəneh/ 'seed' is consistently realised as [bənẽh], and /pənoh/ 'full' as [pənõh]. In other environments, i.e., when /e, o/ are preceded by a non-nasal onset and followed by a coda /ŋ/ or Ø, they are realised as plain monophthongs. A summary of the phonetic realisations of mid-high vowels in CTM is provided in Table 3.8. For a similar tendency of the diphthongisation of high vowels in ITM, see §4.2.2.3.3.

Table 3.8: Phonetic realisations of mid-high vowels /e, o/ in CTM

Phoneme	Realisation	Environment	Example		Gloss
/e/	[e]~[eɪ̯]	C[-nasal], Ø__ʔ, h#	/kule**ʔ**/	[kule**ʔ**] ~[kule**ɪ̯ʔ**]	'skin'
			/pute**h**/	[pute**h**] ~[pute**ɪ̯h**]	'white'
			/buwe**h**/	[buwe**h**] ~[buwe**ɪ̯h**]	'foam'
	[ẽ]	C[+nasal]__	/bəne**h**/	[bəne**ẽh**]	'seed'
			/səne**ŋ**/	[səne**ẽŋ**]	'Monday'
	[e]	elsewhere	/bibe/	[bibe]	'lip'
			/mmaɣe**ŋ**/	[mmaɣe**ŋ**]	'yesterday'
/o/	[o]~[oʊ̯]	C[-nasal], Ø__ʔ, h#	/tako**ʔ**/	[tako**ʔ**] ~[tako**ʊ̯ʔ**]	'be afraid'
			/bulo**h**/	[bulo**h**] ~[bulo**ʊ̯h**]	'bamboo'
			/lao**ʔ**/	[lao**ʔ**] ~[lao**ʊ̯ʔ**]	'sea'
	[õ]	C[+nasal]__	/pəno**h**/	[pəno**õh**]	'full'
			/jəmo/	[jəmo**õ**]	'to dry'
	[o]	elsewhere	/dapo/	[dapo]	'kitchen'
			/kapo**ŋ**/	[kapo**ŋ**]	'village'

3.3 Syllable structure

Syllables in CTM have a template of (C)(C)V(C). Only a nucleus V is obligatory; any C is optional. A consonant cluster CC is allowed in onset position, whereas a coda, if present, is always simple. Possible syllable types in CTM are summarised in Table 3.9.

Table 3.9: Overview of syllable types in CTM

Onset	Nucleus	Coda
	V	
	V	C
C	V	
C	V	C
CC	V	
CC	V	C

Table 3.10 displays the distribution of single consonants at the syllable level in the native lexicon.

Table 3.10: Distribution of consonants in CTM at the syllable level

Position	p	b	t	d	c	j	k	g	ʔ	m	n	ɲ	ŋ	s	x	h	l	ɣ	w	y
onset	+	+	+	+	+	+	+	+	-	+	+	+	+	+	-	+	+	+	+	+
coda	-	-	-	-	-	-	-	-	+	(+)	(+)	(+)	+	-	-	+	-	-	-	-

(+ : attested, - : not attested, (+) : attested but restricted)

All consonants except for the glottal stop /ʔ/ and the velar fricative /x/ can occur as onsets. /x/ only occurs in the geminated form /xx/ (in word-initial position). In coda position, only the nasals and the glottals /ʔ, h/ are permitted. When occurring in coda position, /m, n, ɲ/ always precede a corresponding homorganic voiced stop, with which they form word-medial consonant sequences /-m.b-, -n.d-, -ɲ.j-/.

Complex onsets CC can be either geminate clusters C_xC_x or non-geminate clusters C_xC_y. The attested combinations of segments in consonant clusters are presented in §3.5.

Table 3.11 summarises the distribution of oral vowels in CTM at the syllable level. The three basic vowels /a, i, u/ and the mid-low vowels /ɛ, ɔ/ have the broadest range of occurrences, as they are allowed in all types of syllables. The schwa /ə/ is not allowed in onsetless syllables V(C). The mid-high vowels /e, o/ do not occur in syllables with a complex onset, as they are only permitted in final syllables (see §3.4.3.2), whereas CCV(C) syllables only occur in non-final position.

Table 3.11: Distribution of oral vowels in CTM at the syllable level

Syllable type	a	i	u	ɛ	ɔ	e	o	ə
V	+	+	+	+	+	+	+	-
VC	+	+	+	+	+	+	+	-
CV	+	+	+	+	+	+	+	+
CVC	+	+	+	+	+	+	+	+
CCV	+	+	+	+	+	-	-	+
CCVC	+	+	+	+	+	-	-	+

(+ : attested, - : not attested)

Nasal vowels are often followed by a coda /ʔ/, as seen in examples like /kəcĩʔ/ 'small', /ɛsɔ̃ʔ/ 'to scoot over' and /busũʔ/ 'smelly'. However, there are a few counterexamples, such as /ĩ/ in /matə kaĩ/ 'fish hook', /ã/ in /ləpãŋ/ 'to hit' and /dɔʔã/ 'to pray' (< Arabic *duʿāʾ*, cf. SM *doa*), as well as /ɔ̃/ in /hɔ̃/ 'AFF'.

3.4 Word structure

CTM words have a fairly rigid disyllabic shape with a (C)(C)V(C).(C)V(C) template. Some morphologically simple monosyllabic words are attested, and words with more than two syllables do exist but are rare, most of which can be identified as loanwords. Just as in KM, simple words and derivatives in CTM share similar word shapes with the same set of phonotactic constraints. To demonstrate this similarity, I will again examine their structures separately. Phoneme distributions at the word level are examined in §3.4.3.

3.4.1 Structure of simple words

Simple words in CTM can be monosyllabic, disyllabic or trisyllabic. Disyllabic words are by far the most common, and trisyllabic words are almost exclusively borrowed.

Table 3.12 presents possible syllable types in monosyllables, which follow a template of (C)(C)V(C). Onsetless syllables V or VC are only attested in interjections like /a/ and /eh/ (see §6.2.13). Monosyllables with a ˣCCV shape are not attested.

Table 3.12: Monosyllabic simple words in CTM

Syllable type	Example	Gloss
V	/a/	'INTERJ'
VC	/eh/	'INTERJ'
CV	/ɲɔ/	'coconut'
CVC	/laʔ/	'to wipe'
CCVC	/ppaʔ/	'four'

The disyllabic template (C)(C)V(C).(C)V(C) applies to most simple words in CTM. Possible shapes of disyllabic words are presented in Table 3.13.

Table 3.13: Disyllabic simple words in CTM

Syllable type	Example	Gloss
V.V	/a.e/	'water'
V.CV	/a.ɣi/	'day'
V.CVC	/i.taŋ/	'black'
VC.CV	/am.bə/	'1SG (polite)'
VC.CVC	/aɲ.jiŋ/	'dog'
CV.V	/da.i/	'forehead'
CV.VC	/la.oʔ/	'sea'
CV.CV	/gi.gi/	'tooth'
CV.CVC	/pu.teh/	'white'
CVC.CV	/man.di/	'to bathe'
CVC.CVC	/tan.doʔ/	'horn'
CCV.V	/pɣa.u/	'canoe'
CCV.CV	/ppa.lə/	'head'
CCV.CVC	/mmi.laŋ/	'nine'
CCVC.CV	/blaɲ.jə/	'expense; bride price'
CCVC.CVC	/ccam.bɔh/	'bean sprouts'

In combinations of two syllables, syllables with a complex onset only occur initially, and words with a ×V.VC or ×CCV.VC shape are not attested. Both geminate and non-geminate clusters can be found in initial position in disyl-

labic simple words. Consonant sequences of the type -C.C- can occur word-medially, and they are discussed in §3.6.

Examples of trisyllabic simple words are given in Table 3.14.

Table 3.14: Trisyllabic simple words in CTM

Syllable type	Example	Gloss	Origin
V.CV.CV	/u.ta.ɣə/	'north'	< Sanskrit
V.CV.CVC	/a.ɣə.naʔ/	'rabbit'	< Arabic
CV.CV.CV	/ba.si.ka/	'bicycle'	< English
CV.CV.CVC	/sə.ŋə.leŋ/	'deliberately'	
CVC.CV.CV	/təŋ.ga.ɣə/	'southeast'	< Malayāḷam
CVC.CV.CVC	/ləm.pɔ.yaŋ/	'ginger plant'	

A foreign origin can be identified for most of these words, as indicated in the table. /utaɣə/ 'north' and /aɣənaʔ/ 'rabbit' are ultimately borrowed from Sanskrit and Arabic, presumably having entered CTM via SM (cf. SM *utara* and *arnab*). In addition to /utaɣə/ 'north', /təŋgaɣə/ 'southeast' is another borrowed term for cardinal directions, and it might have derived from Malayāḷam *ten* 'south' + *kara* 'shore' (Hoogervorst 2015: 77). While it is not surprising that both terms for 'north' and 'southeast' are ultimately borrowed (see Adelaar 1997: 58–61), /təŋgaɣə/ is likely also a direct borrowing from SM *təŋgara* (cf. KM /tgaɣɔ/, which reflects regular sound patterns). The word for /ləmpɔyaŋ/ 'ginger plant' seems Malayic (cf. PM *lAmpuyaŋ 'ginger plant', Adelaar 1992: 58), but the unexpected /-m.p-/ sequence also points to a foreign origin (see §3.6).[20] The trisyllabic structure in a few other simple words, such as /səŋəleŋ/ 'deliberately' and /gədəbuʔ/ 'thump' (an ideophone, the sound of something falling into water), remains unexplained.

3.4.2 Structure of derivatives

CTM shares the same inventory of affixes as KM (see §5.3.1), and similarly, most derivatives are disyllabic prefixed forms. Derivatives with more than two syllables are uncommon, most of which are arguably borrowed.

[20] Note that the word for 'ginger (plant)' in KM /haliyɔ/ is also borrowed. Also in ITM, the words for 'ginger plant' /ləmpuyaŋ/ and 'ginger' /haliyɛ/ may not be native, see Table 4.16 in §4.4.1.

Table 3.15 offers an overview of the syllable types found in disyllabic derivatives in CTM. As evident from the table, these derivatives all have an initial CC cluster. Importantly, all syllable types found in disyllabic derivatives can also be found in simple words, following the same maximal CCVC. CVC template.

Table 3.15: Disyllabic derivatives in CTM

Syllable type	Example	Morphological structure	Gloss	
CCV.V	/bɣa.e/	*bɣ-ae*	(INTR-water)	'watery'
CCV.VC	/ssa.iŋ/	*s-saiŋ*	(INTR-friend)	'to befriend'
CCV.CV	/pɣa.co/	*pɣ-aco*	(CAUS-crushed)	'to crush'
CCV.CVC	/ɲɲi.paŋ/	*NN$_1$-<s>ipaŋ*	(IPFV-clean.up)	'cleaning up'
CCVC.CV	/ssan.dɔ/	*s-sandɔ*	(MID-lean)	'to lean on'
CCVC.CVC	/ŋŋɛn.doŋ/	*NN$_1$-<k>ɛndoŋ*	(IPFV-carry)	'carrying'

Generally speaking, bases with an initial CC cluster cannot be prefixed, and trisyllabic prefixed form are limited to exceptions like /bə.tɣa.bo/ *bə-tɣabo* (MID-scattered) 'cluttered' (see similar exceptions in KM in §2.4.2). Some other trisyllabic derivatives may be analysed as having a nominalising suffix *-aŋ* (corresponding to KM *-ɛ*), as in the following examples:

(1)	/u.ku.maŋ/	*ukum-aŋ*	(penalise-NMLS)	'penalty'
	/ma.ka.naŋ/	*makan-aŋ*	(eat-NMLS)	'food'
	/ba.la.saŋ/	*balas-aŋ*	(reply-NMLS)	'reply (n.)'
	/han.tɔ.ɣaŋ/	*hantɔɣ-aŋ*	(send-NMLS)	'bride price'

Similar to KM, the suffixing process in CTM does not operate at the synchronic level. Verbal bases cannot be systematically and productively suffixed with *-aŋ* to derive nouns. For instance, *balah* 'to reply' cannot be suffixed with *-aŋ* to form *balas-aŋ* (note the mismatch of *-h* and *-s-*). The unexpected sound patterns in certain words, such as /h-/ in /hantɔɣaŋ/ 'bride price' (see §3.4.3.1), also indicate a foreign origin.

Similar generalisations can be made about derivatives with more than three syllables. There are only a few of them in my corpus, as listed in (2), all containing a speculative circumfix *pə- -aŋ*. This circumfix, which historically

formed nouns denoting abstract entities (a cognate of SM *pər- -an* and KM *pə- -ɛ*), is also not productive synchronically.

(2)	/pə.ja.la.naŋ/	*pə-jalan-aŋ*	(NMLS-road-NMLS)	'journey'
	/pə.ni.kɔ.haŋ/	*pə-nikɔh-aŋ*	(NMLS-marry-NMLS)	'marriage'
	/pə.tɔ.lo.ŋaŋ/	*pə-tɔloŋ-aŋ*	(NMLS-help-NMLS)	'help'

Overall, native derivatives in CTM have a canonical disyllabic shape, and there is uniformity between the shapes of simple words and derivatives. CTM parallels KM in this regard, and the structural resemblance between the two types of words stems from the same historical explanation (see 7.5.2).

3.4.3 Phoneme distribution at the word level

Phoneme distributions at the word level in CTM are generalised based on the disyllabic template, with additional considerations of some monosyllabic words. The distributional patterns and constraints summarised in the following sections apply to both simple words and derivatives.

3.4.3.1 Distribution of consonants

The general distributional constraints on single consonants in CTM are similar to those in KM, as summarised in Table 3.16.

Table 3.16: Distribution of consonants in CTM at the word level

Position	p	b	t	d	c	j	k	g	ʔ	m	n	ɲ	ŋ	s	x	h	l	ɣ	w	y
initial	+	+	+	+	+	+	+	+	-	+	+	+	(+)	+	-	(+)	+	+	(+)	(+)
medial	+	+	+	+	+	+	+	+	-	+	+	+	+	+	-	+	+	+	+	+
final	-	-	-	-	-	-	-	-	+	-	-	-	+	-	-	+	-	-	-	-

(+ : attested, - : not attested, (+) : attested but uncommon)

In word-initial position, the glottal stop /ʔ/ is not phonemic, and the velar fricative /x/ only appears in the geminated form /xx/. /h/ is generally not allowed word-initially; it occurs only in monosyllabic function words, e.g., /hɔ̃/ 'AFF' and /hɔʔ/ 'REL'. Initial /ŋ, w, y/ are mostly found in shortened variations of some disyllabic words, e.g., /ŋaŋ/~/dəŋaŋ/ 'with', /wɔh/~/buwɔh/

'fruit', /waʔ/~/buwaʔ/ 'to do' and /ya/ ~/ɣiya/ 'ringgit (currency unit)'. There are nevertheless some exceptions, such as /ŋaŋə/ 'to open wide', /waʔpə/ 'why' (which contains the shortened form /waʔ/ 'to do') and /yə/ '3'.

Word-medially, all segments are permitted except for /ʔ/ and /x/, but a word-medial /h/ is typically found between identical vowels, as in /bahaŋ/ 'to hit' and /pəhə/ 'thigh'. There is nevertheless a trend in younger speakers' speech to reintroduce word-medial [h] in other words that historically had *-h-, presumably under more recent influence of SM. For instance, /dai/ 'forehead' and /taiʔ/ 'excrement' are usually pronounced as [dai] and [taiʔ] by older speakers, but [dahi] and [tahiʔ] can be heard among younger speakers (cf. SM *dahi* and *tahi*).

Word-finally, only /ʔ, ŋ, h/ are permitted.

3.4.3.2 Distribution of vowels

The distribution of oral vowels at the word level is summarised in Table 3.17.

Table 3.17: Distribution of oral vowels in CTM at the word level

Position	a	i	u	ɛ	ɔ	e	o	ə
penultimate	+	+	+	+	+	-	-	+
final	+	+	+	+	+	+	+	+

(+ : attested, - : not attested)

As in KM, the mid-high vowels /e, o/ are not allowed in penultimate syllables. The schwa /ə/ does not occur word-initially since it is not present in syllables with a V(C) shape (§3.3), but it has a wider distribution in CTM as it is also allowed in final open syllables, as in /matə/ 'eye' and /limə/ 'five'. The mid-low front vowel /ɛ/ has a limited occurrence in final syllables, always cooccurring with another mid-low vowel in penultimate syllables and showing vowel height harmony, as seen in words like /gɛgɛ/ 'noisy', /lɛmbɛʔ/ 'mattress' and /mɔlɛʔ/ 'good'.

Nasal vowels only appear in final syllables. Many CTM words with a nasal vowel also have cognates in KM that contain a nasal vowel, e.g., /kəcĩʔ/ 'small', /matə kaĩ/ 'fish hook', /ɛsɔ̃ʔ/ 'to scoot over' and /busũʔ/ 'smelly' (cf. KM /kəcɛ̃ʔ/, /matɔ kaɛ̃/, /ɛsɔ̃ʔ/ and /busũʔ/).

3.5 Consonant clusters

Table 3.18 presents the consonant clusters C_1C_2 attested in CTM, with C_1 indicated on the vertical line and C_2 on the horizontal line. Only clusters found in native words are considered here, and clusters in parentheses are rare.

Table 3.18: Attested consonant clusters in CTM

C_1 \ C_2	p	b	t	d	c	j	k	g	m	n	ɲ	ŋ	s	x	l	ɣ	w
p	pp	-	-	-	-	-	-	-	-	-	-	-	-	-	pl	pɣ	-
b	-	bb	-	-	-	-	-	-	-	-	-	-	-	-	bl	bɣ	-
t	-	tb	tt	-	-	-	-	tg	tm	tn	-	-	-	-	tl	tɣ	-
d	-	-	-	dd	-	-	-	-	-	-	-	-	-	-	(dl)	dɣ	-
c	-	-	-	-	cc	-	-	-	cm	-	-	-	-	-	cl	cɣ	-
j	-	-	-	-	-	jj	-	-	jm	-	-	-	-	-	jl	jɣ	-
k	-	kb	-	kd	-	-	kk	-	-	-	-	-	-	-	kl	-	-
g	-	-	-	-	-	-	-	gg	gm	-	-	-	-	-	gl	-	-
m	-	mb	-	-	-	-	-	-	mm	-	-	-	-	-	ml	-	-
n	-	-	-	-	-	-	-	-	-	nn	-	-	-	-	-	-	-
ɲ	-	-	-	-	-	-	-	-	-	-	ɲɲ	-	-	-	-	-	-
ŋ	-	-	-	-	-	-	-	-	-	-	-	ŋŋ	-	-	-	-	-
s	sp	sb	st	sd	sc	sj	sk	sg	sm	sn	-	-	ss	-	sl	sɣ	-
x	-	-	-	-	-	-	-	-	-	-	-	-	-	xx	-	-	
l	-	-	-	-	-	-	-	-	-	-	-	-	-	-	ll	-	-
ɣ	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	ɣɣ	-
w	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	ww

(- : not attested, ( ) : attested but uncommon)

All consonants except for the glottals /h, ʔ/ and the glide /y/ can occur in a consonant cluster. These clusters typically occur word-initially, except in examples like /bətɣabo/ 'cluttered' as previously noted. Five segments, namely /n, ŋ, ɲ, x, w/, only occur in a geminate cluster. Non-geminate clusters can be classified into five possible segment combinations: 1) obstruent + liquid; 2) obstruent + obstruent; 3) obstruent + nasal; 4) nasal + obstruent; and 5)

nasal + liquid. Examples for each type of clusters are given below.

First, all segments except for /h, ʔ, y/ can appear in a geminate cluster, as shown in (3).

(3) Geminate clusters

/pp-/	/ppalə/	'head'
/bb-/	/bbalɔh/	'to fight'
/tt-/	/ttaɣi/	'sun'
/dd-/	/ddiɣi/	'to stand'
/cc-/	/ccayə/	'to believe'
/jj-/	/jjabaʔ/	'office'
/kk-/	/kkatɔʔ/	'frog'
/gg-/	/gganə/	'how'
/mm-/	/mmaɣeŋ/	'yesterday'
/nn-/	/nnataŋ/	'animal'
/ɲɲ-/	/ɲɲawə/	'to breathe'
/ŋŋ-/	/ŋŋale/	'to flow'
/ss-/	/ssəmə/	'all'
/xx-/	/xxusi/	'chair'
/ll-/	/lliŋə/	'ear'
/ɣɣ-/	/ɣɣamə/	'butterfly'
/ww-/	/wwapə/	'how many, how much'

Geminate clusters may be morphologically complex, as illustrated by the following examples: /ddiɣi/ *d-diɣi* (INTR-self) 'to stand' and /ɲɲawə/ *ɲ-ɲawə* (INTR-soul) 'to breathe'. For a comprehensive analysis of the morphological aspects of initial geminates, see §5.3.2. Geminate clusters are also attested in some monosyllabic words, as in /ppaʔ/ 'four', /nnaŋ/ 'six' and /mmah/ 'gold'.

Second, the most common type of non-geminate clusters has an obstruent followed by a liquid /l/ or /ɣ/, as shown in (4).

(4) Obstruent + liquid

/pl-/	/plitə/	'lamp'
/pɣ-/	/pɣau/	'canoe'
/bl-/	/blakaŋ/	'back'
/bɣ-/	/bɣuwaŋ/	'bear'
/tɣ-/	/tɣəbaŋ/	'to fly'
/tl-/	/tlaŋgɔ/	'to crash (accidentally)'

/dl-/	/dlimə/	'pomegranate'
/dɣ-/	/dɣakə/	'insubordinate'
/cl-/	/clakə/	'misfortune'
/cɣ-/	/cɣəmiŋ/	'mirror'
/jl-/	/jluwɔʔ/	'to vomit'
/jɣ-/	/jɣəneh/	'clear'
/kl-/	/kladi/	'taro'
/gl-/	/gləgɔʔ/	'to boil'
/sl-/	/sluwɔ/	'trousers'
/sɣ-/	/sɣatoh/	'one hundred'

The stop and the liquid in such combinations typically have different places of articulation, with one notable exception being /dl-/ in /dlimə/ 'pomegranate'. Clusters like ˣ/kɣ-/ and ˣ/gɣ-/ are not attested, which were historically assimilated to become /xx-/ and /ɣɣ-/ respectively. Compare CTM /xxɛtə/ 'car' with SM ***kərɛta***, and CTM /ɣɣahaŋ/ 'molar tooth' with SM ***gəraham***. Note that /t/ is dental and /l/ is alveolar, but the adjacency of their places of articulation also often led to the assimilation of ⁺tl- > /ll-/, as in /lliŋə/ 'ear' and /lluŋjuʔ/ 'index (finger)' (cf. SM ***təliŋa*** and ***təluŋjuk***). /tl-/ clusters are nevertheless attested in derivatives like *t-laŋgɔ* (NVOL-crash) 'to crash (accidentally)'.

The third type of clusters has two obstruents, as in (5).

(5) Obstruent + obstruent

/tb-/	/tbaka/	'tobacco'
/tg-/	/tgəlaŋ/	'to sink'
/kb-/	/kbumi/	'to bury'
/kd-/	/kdiyaŋ/	'then, later'
/sp-/	/spuloh/	'ten'
/sb-/	/sbəlah/	'eleven'
/st-/	/ubi stɛlə/	'sweet potato'
/sd-/	/sdaɣə/	'sibling'
/sc-/	/scawaŋ/	'a cup'
/sj-/	/sjabe/	'a pouch'
/sk-/	/skali/	'most'
/sg-/	/sgəlah/	'a glass'

In this type of clusters, /s/ is the most frequently attested C_1 , which is often a clitic *s=* meaning 'a; one; same', as in *s=puloh* (one-ten) 'ten', *s=bəlah* (one=teens) 'eleven', *s=cawaŋ* (one=cup) 'a cup', *s=jabe* (one=pouch) 'a pouch' and *s=gəlah* (one=glass) 'a glass'. In clusters with two stops, it is always the case that a voiceless stop precedes a voiced stop. As in KM, an epenthetic schwa may be heard between two stops, as in /tgəlaŋ/ 'to sink', which may be pronounced as [təgəlaŋ].

The fourth type of clusters, namely obstruent + nasal, is attested in the following examples:

(6) Obstruent + nasal

/tm-/	/tmagə/	'copper'
/tn-/	/tnagə/	'energy'
/cm-/	/cmuɣu/	'jealous'
/jm-/	/jmalaŋ/	'earth gnome'
/gm-/	/gmalə/	'herder'
/sm-/	/smilaŋ/	'nine'
/sn-/	/snaɣa/	'list'

The last two types of clusters are uncommon. The combination of a nasal followed by an obstruent is only found in /mb-/, as in /mbuh/ 'to blow' and /mboŋ/ 'dew'. ×/nd-/ or ×/ŋg-/ is not attested (compare CTM /buɣoŋ kləkeŋ/ 'hornbill' with KM /ŋgɛ/). The combination of a nasal + a liquid is only found in /ml-/, as in /mlayu/ 'Malay' and /mlati/ 'jasmin'.

Another type of clusters has an obstruent followed by a glide /w/ or /y/, as illustrated in (7).

(7) Obstruent + glide

/tw-/	/twala/	'towel'	< Portuguese *toalha*	cf. SM *tuala*
/cw-/	/cwacə/	'climate'	< Sanskrit *svaccha*	cf. SM *cuaca*
/kw-/	/kwali/	'wok'	< Tamil *kuvaḷai*	cf. SM *kuali*
/sw-/	/swaɣə/	'voice'	< Sanskrit *svara*	cf. SM *suara*
/by-/	/byasə/	'usual'	< Sanskrit *abhyāsa*	cf. SM *biasa*

All these words can be identified as loanwords, as indicated above. More importantly, these words not only have a non-Malayic origin, but also do not follow the regular sound changes that most other Sanskrit or Portuguese loanwords have undergone. SM cognates of these words have *-u.a-* or *-i.a-*

sequences, and CTM cognates of words with similar shapes regularly reflect the contraction of $^{+}$-u.a- > /ɔ/ and $^{+}$-i.a- > /ɛ/, as in /pɔsə/ 'fasting', /bɔyə/ 'crocodile' and /bɛwɔʔ/ 'monitor lizard', cf. SM ***pua**sa*, ***bua**ya* and ***bia**wak* (see §7.5.2.1).[21] The deviated sound pattern reflected in the words listed in (7) therefore suggests more recent borrowing. For this reason, clusters of obstruent + glide are not included in Table 3.18.

Overall, the attested consonant clusters in CTM follow the SSP with few exceptions. The common types of non-geminate clusters, namely obstruent + liquid, voiceless stop + voiced stop and obstruent + nasal, comply with the SSP, and /s/ + stop clusters are also cross-linguistically common.

3.6 Consonant sequences

Consonant sequences most commonly consist of a nasal and a homorganic voiced stop, i.e., /-m.b-, -n.d-, -ɲ.j-, -ŋ.g-/, as illustrated in Table 3.19. As in KM, sequences of a nasal + a voiceless stop are also not allowed in CTM.

Table 3.19: Homorganic nasal + voiced stop sequences in CTM

Sequence	Example	Gloss
/-m.b-/	/ɣam.boʔ/	'hair'
	/ləm.bu/	'cattle'
/-n.d-/	/tan.doʔ/	'horn'
	/pɛn.dɛʔ/	'short'
/-ɲ.j-/	/aɲ.jiŋ/	'dog'
	/paɲ.jaŋ/	'long'
/-ŋ.g-/	/tiŋ.gi/	'tall'
	/puŋ.goŋ/	'buttock'

Sequences of a glottal stop /ʔ/ followed by another segment are also attested, as shown in examples in Table 3.20.

[21] Although /pɔsə/ 'fasting' is ultimately derived from Sanskrit *upavāsa*, it is considered native because it reflects regular sound changes.

Table 3.20: /-ʔ.C-/ sequences in CTM

Sequence	Example	Gloss	Origin
/-ʔ.p-/	/waʔ.pə/	'why'	< *buwaʔ* 'do' + †*apa* 'what'
/-ʔ.t-/	/mɔʔ.taŋ/	'rambutan'	
/-ʔ.d-/	/taʔ.di/	'just now'	
/-ʔ.c-/	/mɔʔ.ciʔ/	'auntie'	< *mɔʔ* 'mother' + *ciʔ* 'sister'
/-ʔ.k-/	/bɛʔ.ki/	'to repair'	
/-ʔ.g-/	/kɛʔ.gi/	'later'	
/-ʔ.m-/	/sɔʔ.mɔ/	'always'	< *s*= 'one; same' + *ɔmɔ* 'age'
/-ʔ.n-/	/maʔ.nə/	'meaning'	< Arabic *maʿnā*

† indicates earlier forms which are no longer attested.

These sequences often appear in (historically) contracted forms or loanwords, as indicated in the table. /mɔʔtaŋ/ 'rambutan' and /bɛʔki/ 'to repair' are historically suffixed, which can be compared with KM /mɔʔtɛ/ and /bɛʔki/, SM *rambut-an* 'hair-NMLS' and *baik-i* 'good-APPL'. /-ʔ.d-/ in /taʔdi/ 'just now' and /-ʔ.g-/ in /kɛʔgi/ 'later' are unexplained.

3.7 Vowel sequences

Vowel sequences attested in CTM are presented in Table 3.21.

Table 3.21: Vowel sequences in CTM

Sequence	Example	Gloss
/a.i/	/na.iʔ/	'to climb'
/a.u/	/ba.u/	'shoulder; smell'
/a.e/	/a.e/	'water'
/a.o/	/ta.oŋ/	'year'
/a.ĩ/	/matə ka.ĩ/	'fish hook'

The first vowel in vowel sequences is always the low vowel /a/. Among the vowels permitted in penultimate syllables, the schwa and the mid-low vowels /ɛ, ɔ/ cannot be directly followed by another vowel in final

syllables. Potential vowel sequences with a high vowel as the first component, i.e., ˣ/-i.V-/ and ˣ/-u.V-/, are analysed as /-iyV-/ and /-uwV-/ with a phonemic intervocalic glide, as in examples like /piyə/ 'who', /tiyuʔ/ 'to blow', /duwə/ 'two' and /luwa/ 'outside'. This is evidenced by the alternation between /ya/~/ɣiya/ 'ringgit (currency unit)', /wɔh/~/buwɔh/ 'fruit' and /waʔ/~/buwaʔ/ 'to do', which preserves the glides /y/ and /w/ in word-initial position in the shortened forms. There is one vowel sequence with a nasal vowel, namely /a.ĩ/ in /matə kaĩ/ 'fish hook', a cognate of KM /matɔ kaɛ̃/.

3.8 Stress

Stress is not phonemic in CTM. The acoustic aspects of stress in CTM have not been extensively studied, and the following summaries on the stress placement should be taken as tendencies rather than rules.

Most disyllabic words in CTM have stress on the final syllable, regardless of whether the final syllable is open or closed. Some examples are given in (8).

(8)	/suŋa/	[suˈŋã]	'river'
	/mati/	[mãˈti]	'to die'
	/pase/	[paˈse]	'sand'
	/kɔtɔ/	[kɔˈtɔ]	'dirty'
	/taseʔ/	[taˈseʔ]	'lake'
	/gigiʔ/	[giˈgiʔ]	'to bite'
	/bitaŋ/	[biˈtaŋ]	'star'
	/dagiŋ/	[daˈgiŋ]	'meat'
	/mutɔh/	[mũˈtɔh]	'to vomit'
	/bunoh/	[buˈnõh]	'to kill'

There are nevertheless some counterexamples to this general tendency, as demonstrated by the data in (9). The exact reason for the differences in stress placements is not clear, and it is also likely that stress in either syllable can be accepted by the speakers.

(9)	/lɛbɔ/	[ˈlɛbɔ]	'wide'
	/duɣi/	[ˈduɣi]	'thorn'
	/kilaʔ/	[ˈkilaʔ]	'lightning'

/muloʔ/	[ˈmũloʔ]	'month'
/taŋaŋ/	[ˈtaŋãŋ]	'hand'
/tulaŋ/	[ˈtulaŋ]	'bone'
/juɣuh/	[ˈjuɣuh]	'good'
/baloh/	[ˈbaloh]	'storage house (for paddy)'

Stress on the final syllable is particularly prominent in words with a penultimate schwa, such as /təba/ [təˈba] 'thick', /pəɣoʔ/ [pəˈɣoʔ] 'belly', /tənoŋ/ [təˈnõŋ] 'to weave' and /bəkɔh/ [bəˈkɔh] 'to swell'. If the final syllable has a schwa, stress is usually placed on the penultimate syllable instead, as in (10).

(10)	/dadə/	[ˈdadə]	'chest'
	/matə/	[ˈmãtə]	'eye'
	/buŋə/	[ˈbuŋə̃]	'flower'
	/bɔyə/	[ˈbɔyə]	'crocodile'

A few CVCV words, such as /bələ/ 'to breed' and /ɣəgə/ 'price', have a schwa as the nucleus in both syllables. These words also have stress on the final syllable, i.e., [bəˈlə] and [ɣəˈgə].

Another noteworthy tendency in the stress placement is that words with initial geminates have stress on the initial syllable, even if the penultimate syllable has a schwa. This is shown by the following examples:

(11)	/ttaɣi/	[ˈttaɣi]	'sun'
	/ppəɣaŋ/	[ˈppəɣaŋ]	'to be in a war'
	/nnataŋ/	[ˈnnãtaŋ]	'animal'
	/mmaɣeŋ/	[ˈmmãɣeŋ]	'yesterday'
	/xxəkɔh/	[ˈxxəkɔh]	'to bite'
	/ɣɣima/	[ˈɣɣimã]	'tiger'

In comparison, words with initial non-geminate clusters do not always have initial stress. For instance, /blakaŋ/ 'back' is pronounced as [ˈblakaŋ] with initial stress, but /smilaŋ/ 'nine' and /tgəlaŋ/ 'to sink' often have stress on the final syllable, i.e., [smiˈlaŋ] and [tgəˈlaŋ].

On the whole, stress is not phonemic in CTM. It seems that CTM has a tendency to place stress on the final syllable in disyllabic words, but there is room for variation. When the final syllable has a schwa while the penultimate syllable has a full vowel, stress is often placed on the penultimate syllable. Words with initial geminates have stress on the first syllable.

3.9 Summary

This chapter has provided a description of the phonology of CTM, covering its segment inventory, syllable structure and word structure. Relevant comparisons with KM were made to highlight similarities and differences between the two varieties.

At the segment level, CTM has twenty phonemic consonants and thirteen phonemic vowels, including five nasal vowels /ã, ĩ, ũ, ɛ̃, ɔ̃/. There are no phonemic diphthongs, but the mid-high vowels /o, e/ have a tendency to be diphthongised to [eɪ̯] and [oʊ̯] in certain environments.

At the syllable level, CTM syllables follow a template of (C)(C)V(C). All consonants except for /ʔ/ and /x/ can appear as a single onset, but only nasals and glottals /ʔ, h/ are allowed as a coda. Complex onsets CC can be either geminate clusters with two identical segments or non-geminate clusters with two different segments. All consonants excluding /ʔ, h, y/ can appear in a geminate cluster, and the combinations of segments in non-geminate clusters generally comply with the SSP. The velar fricative /x/ only appears as a geminate cluster /xx/. As for vowels, a schwa cannot occur in an onsetless syllable. Nasal vowels also have a limited distribution, and they are commonly followed by a glottal stop.

At the word level, CTM words have a canonical disyllabic structure with a (C)(C)V(C).(C)V(C) template, which is shared by both simple words and derivatives. Complex onsets typically occur word-initially, and consonant sequences can be formed word-medially, with the most common type being a nasal followed by a homorganic voiced stop. Two adjacent vowels form a vowel sequence, in which the first segment is always /a/. Distributional constraints on segments at the word level are fairly complex. Word-initially, a glottal stop is not phonemic, and the glottal fricative /h/ and the glides /w, y/ also have limited presence. As mentioned above, /x/ only occurs in the geminated form /xx/, and /ŋ/ also typically occurs as geminate /ŋŋ/ in word-initial position. Word-medially, /ʔ/ and /x/ are not allowed in inherited vocabulary. In word-final position, only /ʔ, ŋ, h/ are permitted. As far as vowels are concerned, the mid-high vowels /e, o/ are not allowed in penultimate position, and the mid-low front vowel /ɛ/ has a limited occurrence in final syllables, reflecting vowel height harmony. Nasal vowels are restricted to final syllables.

Stress is not phonemic in CTM. In general, there is a tendency to stress the final syllable in disyllabic words, but there are instances where stress placement varies. For instance, if the final syllable contains a schwa or the penultimate syllable has an initial geminate, stress falls on the penultimate syllable.

The phonology of CTM exhibits several striking similarities when compared to that of KM. Both varieties share the same consonant inventory and the basic templates for syllables and words. Moreover, the phonotactic constraints on the distribution of consonants, permissible consonant clusters and sequences, as well as vowel sequences, are largely analogous. For instance, in both varieties /ŋ, h, w, y/ do not commonly occur in word-initial position, and only three segments /ʔ, ŋ, h/ are permitted word-finally. Nevertheless, there are some noteworthy differences between the two varieties. In the consonant system, CTM has a geminate glide /ww/ which is not attested in KM. In the vowel system, while both varieties have a similar vowel inventory, different distributional constraints apply. For example, while a schwa is not allowed in KM in final syllables, it can appear word-finally in CTM.

CHAPTER 4

Phonology of Inland Terengganu Malay

4.1 Introduction

This chapter describes the phonology of ITM, primarily based on the variety spoken in Kampung Dusun. Given the considerable variation across different subvarieties of ITM (see Collins 1983a), the specific subvariety described here is referred to as "Dusun". The description begins by examining the segment inventory of Dusun in §4.2, followed by discussions on syllable structure in §4.3 and word structure in §4.4. Permitted consonant clusters, consonant sequences and vowel sequences are presented from §4.5 to §4.7. §4.8 discusses the stress pattern in Dusun. §4.9 summarises this chapter.

When comparing the phonology of Dusun to that of KM and CTM, it becomes evident that there are both shared similarities and notable differences. At the segment level, the consonant inventory of Dusun is identical to that of KM and CTM with twenty native phonemes, but the vowel system of Dusun presents distinct features. Moreover, differences can be observed in the phonotactic constraints on the distribution of segments, as will be discussed in detail throughout this chapter. Some important comparative findings regarding the phonologies of the NEPMs are summarised in §4.10.

4.2 Segment inventory

4.2.1 Consonant system

4.2.1.1 Consonant inventory

The consonant inventory of Dusun is presented in Table 4.1. There are twenty native consonants, consisting of nine stops, four nasals, three fricatives, two liquids and two glides.

Table 4.1: Consonant inventory of Dusun

		Labial	Dental	Alveolar	Palatal	Velar	Glottal
Stops	voiceless	p	t			k	ʔ
	voiced	b		d		g	
Nasals		m		n	ɲ	ŋ	
Fricatives	voiceless			s		x	h
	voiced			(z)			
Affricates	voiceless				c		
	voiced				j		
Liquids				l		ɣ	
Glides		w			y		

(c = IPA /c͡ç/, j = IPA /ɟ͡ʝ/, y = IPA /j/)

Asymmetry in the phonetic values of the dental-alveolar stops is also displayed in Dusun, with the voiceless stop articulated as dental [t̪] and its voiced counterpart as alveolar [d]. /ɣ/ is also treated as a liquid rather than a fricative.

The voiced alveolar fricative /z/ is taken as a marginal phoneme, as it is found in loanwords such as /zamaŋ/ 'epoch' < Arabic *zamān* (which also alternates with [jamãŋ]) and /bizɛ/ 'different' < Sanskrit *bheda* (presumably via SM *bɛza*). It cannot be taken as a free variation of /j/, as the realisation of /j/ does not show variation in other words.

4.2.1.2 Contrasts between consonants

Phonemic contrasts between similar consonants in Dusun are demonstrated by (near-)minimal pairs presented in Table 4.2. Contrasts found in

word-initial, -medial and -final positions are distinguished. The distribution of consonants in Dusun is discussed in §4.4.3.1.

Table 4.2: Contrasts between consonants in Dusun

Contrast	Pair	Gloss
Word-initially		
/p/ - /b/	/**p**adaŋ/	'suitable'
	/**b**adaŋ/	'body'
/b/ - /m/	/**b**ukɛ/	'open'
	/**m**ukɛ/	'face'
/t/ - /d/	/**t**əbəʊ/	'sugarcane'
	/**d**əbəʊ/	'dust'
/d/ - /n/	/**d**ai/	'forehead'
	/**n**aiʔ/	'to climb, to go up; to ride'
/c/ - /j/	/**c**aɣɛi/	'to search'
	/**j**aɣɛi/	'finger'
/j/ - /ɲ/	/**j**awɛ/	'Jawa'
	/**ɲ**awɛ/	'soul'
/k/ - /g/	/**k**alɛi/	'times'
	/**g**alɛi/	'to dig'
/k/ - /x/	/**k**usɛi/	'to share'
	/**x**usɛi/	'chair'
/g/ - /ɣ/ - /l/	/**g**ata/	'itchy'
	/**ɣ**ata/	'necklace'
	/**l**ata/	'floor'
/m/ - /n/	/**m**ikɛ/	'3PL'
	/**n**akɛ/	'jackfruit'
/l/ - /n/	/**l**amɛ/	'long (time)'
	/**n**amɛ/	'name'
Word-medially		
/p/ - /b/ - /m/	/li**p**ɔ/	'to throw'
	/li**b**ɔ/	'wide'
	/li**m**ɔ/	'lemon, citrus'
/t/ - /d/	/u**t**ɔŋ/	'debt'
	/u**d**ɔŋ/	'shrimp'

Contrast	Pair	Gloss
/d/ - /n/	/kə**d**a/	'shop'
	/kə**n**a/	'to know'
/c/ - /j/	/ka**c**ɔŋ/	'bean'
	/ka**j**ɔŋ/	'Kajang (toponym)'
/j/ - /ɲ/	/sə**j**aʔ/	'to evaporate'
	/sə**ɲ**aʔ/	'quiet'
/k/ - /g/- /ŋ/	/i**k**aʔ/	'to tie up'
	/i**g**aʔ/	'to catch'
	/i**ŋ**aʔ/	'to think'
/g/ - /ɣ/	/pə**g**ɔŋ/	'to hold'
	/pə**ɣ**ɔŋ/	'war'
/m/ - /n/	/tə**m**uŋ/	'to encounter'
	/tə**n**uŋ/	'to weave'
/n/ - /ŋ/	/ta**n**aŋ/	'to plant'
	/ta**ŋ**aŋ/	'hand'
/n/ - /ɲ/	/ku**n**iŋ/	'yellow'
	/ku**ɲ**iʔ/	'turmeric'
/l/ - /n/	/pə**l**əʊh/	'sweat'
	/pə**n**əʊh/	'full'
/l/ - /ɣ/	/bə**l**ah/	'ten'
	/bə**ɣ**ah/	'husked rice'
/s/ - /h/	/a**s**aʔ/	'smoke'
	/a**h**aʔ/	'Sunday'
/y/ - /w/	/la**y**ɔ/	'to sail'
	/la**w**ɔ/	'beautiful'
Word-finally		
/ʔ/ - Ø	/iku**ʔ**/	'to follow'
	/iku/	'tail; CLF'
/ŋ/ - Ø	/panda**ŋ**/	'pandan'
	/panda/	'clever'
/h/ - Ø	/ɣama**h**/	'to squeeze'
	/ɣama/	'many'
/ʔ/ - /h/	/cucəʊ**ʔ**/	'to skewer'
	/cucəʊ**h**/	'to light up'

Contrast	Pair	Gloss
/ʔ/ - /ŋ/	/tulɔ**ʔ**/	'to push'
	/tulɔ**ŋ**/	'bone'
/h/ - /ŋ/	/buka**h**/	'to bundle'
	/buka**ŋ**/	'NEG'

Initial geminates are also attested in Dusun. Table 4.3 lists (near-) minimal pairs contrasting a singleton and a geminate in word-initial position.

Table 4.3: Contrasts between singletons and geminates in Dusun

Contrast	Pair	Gloss
/p/ - /pp/	/**p**alɛ/	'nutmeg'
	/**pp**alɛ/	'head'
/b/ - /bb/	/**b**au/	'smell'
	/**bb**au/	'smelly'
/t/ - /tt/	/**t**upaʔ/	'compact'
	/**tt**upaʔ/	'k.o. rice cake'
/d/ - /dd/	/**d**iɣɛi/	'self'
	/**dd**iɣɛi/	'to stand'
/c/ - /cc/	/**c**abɔŋ/	'branch'
	/**cc**abɔŋ/	'branched'
/j/ - /jj/	/**j**alaŋ/	'road'
	/**jj**alaŋ/	'to walk'
/k/ - /kk/	/**k**alɛi/	'times'
	/**kk**alɛi/	'most'
/g/ - /gg/	/**g**ajɛi/	'wage'
	/**gg**ajɛi/	'saw (n.)'
/m/ - /mm/	/**m**asɔʔ/	'ripe'
	/**mm**asɔʔ/	'to cook'
/n/ - /nn/	/**n**akɛ/	'jackfruit'
	/**nn**aɣɛ/	'tower'
/ɲ/ - /ɲɲ/	/**ɲ**awɛ/	'soul'
	/**ɲɲ**awɛ/	'to breathe'
/s/ - /ss/	/**s**alɔh/	'mistake'
	/**ss**alɔh/	'to make mistake'

Contrast	Pair	Gloss
/l/ - /ll/	/laŋiʔ/	'sky'
	/llaŋiʔ/	'palate'
/ɣ/ - /ɣɣ/	/ɣahɔŋ/	'jaw'
	/ɣɣahaŋ/	'molar tooth'

Among the twenty phonemic consonants in Dusun, the glottals /ʔ, h/ and the glides /w, y/ do not occur as geminates. As in KM and CTM, no clear contrast between /ŋ/ and /ŋŋ/ is found in Dusun, as the velar nasal /ŋ/ mostly occurs in a geminated form /ŋŋ/, e.g., /ŋŋali/ 'to flow' and /ŋŋakəʊ/ 'to confess'. There is also no clear contrast between initial /x/ and /xx/, which appear to have a complementary distribution. In careful speech, /xx-/ is attested before /a, ə/ whereas /x-/ is attested before /i, u/, but there is more variation in natural speech production. As will be discussed in more detail in §7.5.2.3, /xx-/ originates from the assimilation of earlier ⁺kɣ- clusters, which was then reduced to /x-/ before the high vowels /i, u/, as in the following examples: Sanskrit *karaṇa* > SM *kərani* > /xxaniŋ/ 'clerk', Portuguese *carreta* > SM *kərɛta* > ⁺xxɛta > /xitɛ/ 'car', and Arabic *kursiyy* > SM *kərusi* > ⁺xxusi > /xusɛi/ 'chair' (cf. CTM /xxaniŋ/, /xxɛtə/ and /xxusi/). The velar liquid /ɣ/ and its geminate counterpart /ɣɣ/ have been reduced in a similar way. Before the high vowels /i, u/, ⁺ɣɣ- was reduced to /ɣ-/, and ⁺ɣ- was reduced to Ø. Compare Dusun /ɣudɛi/ 'drill' with SM *gərudi* (presumably via ⁺ɣɣudi), /iŋaŋ/ 'light' with SM *riŋan*, and /usɛ/ 'deer' with SM *rusa*.

At the phonological level, geminates in Dusun are analysed as a subtype of consonant clusters. Variations between forms with geminate clusters and forms with a non-geminate clusters are also attested in Dusun, as in examples like /tmakɔ/~/mmakɔ/ 'tobacco' and /smayɔŋ/~/mmayɔŋ/ 'to pray'.

4.2.1.3 Phonetic realisations of consonants

Similar to KM and CTM, the glottal fricative /h/ and the velar fricative /x/ exhibit variable phonetic realisations in Dusun, which are summarised in Table 4.4. Other consonants are realised fairly consistently.

Table 4.4: Phonetic realisations of some consonant phonemes in Dusun

Phoneme	Realisation	Environment	Example		Gloss
/h/	[ç]~[h]	i__#	/nipi**h**/	[nĩpi**ç**]~[nĩpi**h**]	'thin'
			/piti**h**/	[piti**ç**]~[piti**h**]	'money'
	[h]	elsewhere	/ata**h**/	[ata**h**]	'top'
			/kabu**h**/	[kabu**h**]	'fog'
			/putɛi**h**/	[putæe̯**h**]	'white'
			/jatəʊ**h**/	[jatɐo̯**h**]	'to fall'
/x/	[x]~[k]	#__	/**x**usi/	[**x**usi]~[**k**usi]	'chair'
			/**xx**ətah/	[**xx**ətah]~[**kk**ətah]	'paper'

Word-final /h/ following a high front vowel /i/ is often realised as a voiceless palatal fricative [ç]. It is noteworthy that final-syllable /i/ in Dusun has a tendency to be diphthongised in closed syllables (see more detail in §4.2.2.3); accordingly, /-ih/ can be realised as [-iç]~[-ɪi̯ç] following non-nasal onsets and as [-ĩç]~[-ɛ̃ĩ̯ç] following nasal onsets. For example, /pitih/ is realised as [pitiç]~[pitɪi̯ç] 'money' and /manih/ as [mãnĩç]~[mãnɛ̃ĩ̯ç] 'sweet'.

The velar fricative /x/ and its geminated form /xx/ tend to merge with the velar stop /k/ or /kk/ in word-initial position, which is primarily observed in the speech of younger speakers. Examples include /xusɛi/ [xusɛi]~ [kusɛi] 'chair', /xxəjɛ/ [xxəjɛ]~[kkəjɛ] 'work' and /xxətah/ [xxə-tah]~[kkətah] 'paper'.

4.2.2 Vowel system

4.2.2.1 Vowel inventory

The vowel inventory of Dusun consists of ten monophthongs and two diphthongs. The ten monophthongs include six oral vowels and four nasal vowels, as presented in Table 4.5.

Table 4.5: Inventory of monophthongs in Dusun

	Front	Central	Back
High	i, ĩ		u, ũ
Mid	ɛ	ə	ɔ, ɔ̃
Low		a, ã	

In addition to the three basic vowels /a, i, u/ and a schwa, there is only one set of mid vowels /ɛ, ɔ/ in Dusun, as opposed to KM and CTM which contrast mid-low vowels with mid-high vowels. The four nasal vowels /ã, ĩ, ũ, ɔ̃/ have limited presence, and no clear minimal pairs contrasting a nasal vowel and a corresponding oral vowel have been identified. Yet, these nasal vowels are considered phonemic for two reasons: first, they occur independent of nasal contexts, as in /duʔã/ 'to pray', /kəcĩʔ/ 'small', /busũʔ/ 'rotten' and /hɔ̃/ 'AFF', which means that they are not allophonic variations of oral vowels. Second, they show consistent pronunciation without variation.

In contrast to phonemic nasal vowels, nasalised vowels are prevalent in Dusun due to the process of progressive nasal spreading. Nasality can also spread over glides and glottals, affecting more than one vowel (see §2.2.2.1 for a similar phenomenon explained for KM). For instance, /naiʔ/ 'to climb' and /ŋŋuwaʔ/ 'to yawn' are articulated as [nãĩʔ] and [ŋŋũwãʔ] with both vowels carrying nasality. Additionally, the presence of nasal consonants and vowel nasalisation can affect the height of some vowels, as will be discussed in §4.2.2.3.

Dusun has two phonemic diphthongs /ɛi/ and /əʊ/. There is a larger number of phonetically distinctive diphthongs, but based on their complementary distributions, some are analysed as possible realisations of the high vowels /i, u/, and some as allophones of the same diphthong phoneme. The current analysis is motivated in §4.2.2.3.

4.2.2.2 Contrasts between vowels

The phonemic status of Dusun vowels is demonstrated by minimal or near-minimal pairs presented in Table 4.6. Contrasts found in penultimate and final syllables are distinguished. Phonotactic constraints on vowel distributions are discussed in §4.4.3.2.

Table 4.6: Contrasts between oral vowels in Dusun

Contrast	Pair	Gloss
Penultimate syllables		
/a/ - /i/ - /u/	/m**a**ɣɔh/	'angry'
	/m**i**ɣɔh/	'red'
	/m**u**ɣɔh/	'cheap'
/a/ - /ə/	/p**a**ta/	'beach'
	/p**ə**ta/	'bitter bean'
/i/ - /ə/	/b**i**lɛ/	'when'
	/b**ə**lɛ/	'to breed'
/u/ - /ə/	/b**u**nɔŋ/	'to swim'
	/b**ə**nɔŋ/	'thread'
Final syllables		
/a/ - /i/	/mas**a**ŋ/	'sour'
	/mas**i**ŋ/	'salty'
/a/ - /u/ - /ɛi/	/tup**a**/	'squirrel'
	/tup**u**/	'dull'
	/tup**ɛi**/	'cap, hat'
/a/ - /ɛ/	/pət**a**/	'bitter bean'
	/pət**ɛ**/	'map'
/a/ - /ɔ/	/pak**a**/	'to wear; to use'
	/pak**ɔ**/	'expert'
/i/ - /u/ - /ɔ/	/uk**i**/	'to carve'
	/uk**u**/	'to measure'
	/uk**ɔ**/	'to scratch'
/i/ - /ɛ/	/kat**i**/	'bed'
	/kat**ɛ**/	'to say'
/i/ - /ɔ/ - /əʊ/	/pas**i**/	'sand'
	/pas**ɔ**/	'market'
	/pas**əʊ**/	'flower pot'
/u/ - /ɛ/	/but**u**/	'bottle'
	/but**ɛ**/	'blind'
/ɛ/ - /ɔ/	/lim**ɛ**/	'five'
	/lim**ɔ**/	'lemon, citrus'
/ɛi/ - /i/	/tap**ɛi**/	'but'
	/tap**i**/	'to come forward'

Contrast	Pair	Gloss
/ɛi/ - /ɛ/	/mat**ɛi**/	'to die'
	/matɛ/	'eye'
/ɛi/ - /ɔ/	/pag**ɛi**/	'morning'
	/pagɔ/	'fence'
/əʊ/ - /a/	/təb**əʊ**/	'sugarcane'
	/təb**a**/	'thick'
/əʊ/ - /u/	/dəb**əʊ**/	'dust'
	/dəb**u**/	'splash'
/əʊ/ - /ɛ/	/kay**əʊ**/	'wood'
	/kay**ɛ**/	'rich'
/ɛi/ - /əʊ/	/at**ɛi**/	'liver'
	/at**əʊ**/	'ghost'

4.2.2.3 Phonetic realisations of vowels

The vowels in Dusun exhibit complex variations in their phonetic realisations. Both allophonic and free variations exist. This section provides a more detailed description of these vowel variations.

4.2.2.3.1 The low vowel

The low vowel /a/ has a fairly consistent realisation as [a], except when following a nasal consonant, in which case it is nasalised as [ã]. For instance, /m**a**k**a**n/ 'to eat' and /p**a**n**a**h/ 'warm' are realised as [m**ã**k**a**n] and [p**a**n**ã**h] respectively.

4.2.2.3.2 The mid vowels

The mid front unrounded vowel /ɛ/ is usually realised as [ɛ]. When following a nasal consonant, it is nasalised and raised. The result of raising can be as high as [ĩ] in careful speech, although in connected speech it is often [ẽ]. The allophonic variations of /ɛ/ are illustrated in Table 4.7.

Table 4.7: Phonetic realisations of /ɛ/ in Dusun

Phoneme	Realisation	Environment	Example		Gloss
/ɛ/	[ĩ]~[ẽ]	C[+nasal]__	/manɛ/	[mãnĩ]~[mãnẽ]	'where'
			/limɛ/	[limĩ]~[limẽ]	'five'
			/buŋɛ/	[buŋĩ]~[buŋẽ]	'flower'
			/taɲɛ/	[taɲĩ]~[taɲẽ]	'to ask'
	[ɛ]	C[-nasal]__	/matɛ/	[mãtɛ]	'eye'
			/usɛ/	[usɛ]	'deer'
			/gulɛ/	[gulɛ]	'sugar'
			/diyɛ/	[diyɛ]	'3SG'

The mid back round vowel /ɔ/ is lowered to [ɑ] when preceding a velar nasal /ŋ/; in essence, /ɔŋ/ is phonetically realised as [ɑŋ]. A phonemic contrast between /aŋ/ and /ɔŋ/ is thus in fact a contrast between [aŋ] and [ɑŋ] at the phonetic level, as can be seen from the following minimal pairs:

(1)	a.	/utaŋ/	[utaŋ]	'forest; rattan'
		/utɔŋ/	[utɑŋ]	'debt'
	b.	/piŋgaŋ/	[piŋgaŋ]	'plate'
		/piŋgɔŋ/	[piŋgɑŋ]	'waist'
	c.	/padaŋ/	[padaŋ]	'to fit'
		/padɔŋ/	[padɑŋ]	'field'

These two "a"-like sounds have been reported in some subvarieties of ITM, transcribed as [æ] vs. [a] (Collins 1983a: 32–37). In Dusun, the phonetic values of these two sounds are [a] vs. [ɑ]. The low back vowel [ɑ] is taken as an allophone of /ɔ/ before a velar nasal, and a phonetic realisation of ˣ[ɔŋ] is never found. When following a nasal onset, /ɔ/ is also nasalised, and /ɔŋ/ is realised as [ɑ̃ŋ]. In other contexts, /ɔ/ is realised as [ɔ]. The allophonic variations of /ɔ/ are summarised in Table 4.8.

Table 4.8: Phonetic realisations of /ɔ/ in Dusun

Phoneme	Realisation	Environment	Example		Gloss
/ɔ/	[ɑ]	__ŋ#	/utɔŋ/	[utɑŋ]	'debt'
			/blakɔŋ/	[blakɑŋ]	'back'
			/musɔŋ/	[mũsɑŋ]	'civet cat'
	[ã]	C[+nasal]__ŋ#	/mimɔŋ/	[mĩmãŋ]	'indeed'
			/bunɔŋ/	[bunãŋ]	'to swim'
			/pinɔŋ/	[pinãŋ]	'areca nut'
	[ɔ̃]	C[+nasal]__Ø, ʔ, h#	/dəŋɔ/	[dəŋɔ̃]	'to hear'
			/anɔʔ/	[anɔ̃ʔ]	'child'
			/timɔh/	[timɔ̃h]	'tin'
	[ɔ]	elsewhere	/bakɔ/	[bakɔ]	'to burn'
			/kapɔʔ/	[kapɔʔ]	'axe'
			/daɣɔh/	[daɣɔh]	'blood'

4.2.2.3.3 The high vowels

The high vowels /i, u/ exhibit various realisations in different environments, conditioned by a number of factors including their position in a word, the presence of an onset and coda, as well as the nasality of the onset.

In non-final syllables, /i, u/ are consistently realised as [i, u] unless following a nasal onset, which nasalises the subsequent vowel. For example, /iku/ 'tail' and /kubɔ/ 'buffalo' are realised as [iku] and [kubɔ], but /nnikɔh/ 'to marry' and /mudɛ/ 'young' are realised as [nnĩkɔh] and [mũdɛ].

In final syllables, there is more variation in the realisations of the high vowels, which tend to be diphthongised. However, this tendency only occurs in final closed syllables with an onset. For instance, the final-syllable /i/ in /kuliʔ/ 'skin' and /u/ in /muluʔ/ 'month' can be realised as [i]~[ɪi̯] and [u]~[ʊu̯] respectively, varying across speakers. Therefore, /kuliʔ/ 'skin' can be phonetically realised as [kuliʔ]~[kulɪi̯ʔ], and /muluʔ/ 'month' as [mũluʔ]~ [mũlʊu̯ʔ]. When the onset of the final syllable is a nasal, as in /kamiŋ/ '1PL.EXCL' or /bunuh/ 'to kill', diphthongisation of high vowels results in [ɛ̃ĩ̯] and [ɔ̃ʊ̯] respectively, exhibiting the centralisation effect of

nasal onsets, as compared to [ɪ̯] and [ʊ̯] (Beddor 1983; Beddor et al. 1986). /kamiŋ/ '1PL.EXCL' is often pronounced as [kamɛ̃ɪ̯ŋ], and /bunuh/ 'to kill' as [bunə̃ʊ̯h], but [kamĩŋ] and [bunũh] are also heard. When the final syllable is onsetless, as in /kaiŋ/ 'cloth' and /jauh/ 'far', the high vowels are never diphthongised. Table 4.9 illustrates the tendency of diphthongisation of /i/ and /u/ in final closed syllables.

Table 4.9: Diphthongisation of high vowels in final closed syllables in Dusun

Phoneme	Realisation	Environment	Example		Gloss
/i/	[i]~[ɪ̯]	C[-nasal]__C#	/kuliʔ/	[kuliʔ]~[kulɪ̯iʔ]	'skin'
			/ssəkiŋ/	[ssəkiŋ]~[ssəkɪ̯iŋ]	'poor'
			/nipih/	[nĩpiç]~[nĩpɪ̯iç]	'thin'
	[ĩ]~[ɛ̃ɪ̯]	C[+nasal]__C#	/tumiʔ/	[tumĩʔ]~[tumɛ̃ɪ̯ʔ]	'heel'
			/kamiŋ/	[kamĩŋ]~[kamɛ̃ɪ̯ŋ]	'1PL.EXCL'
			/bənih/	[bənĩh]~[bənɛ̃ɪ̯h]	'seed'
	[i]	Ø__C#	/jaiʔ/	[jaiʔ]	'to sew'
			/kaiŋ/	[kaiŋ]	'cloth'
/u/	[u]~[ʊ̯]	C[-nasal]__C#	/cukuʔ/	[cukuʔ]~[cukʊ̯uʔ]	'enough'
			/kəbuŋ/	[kəbuŋ]~[kəbʊ̯uŋ]	'farm'
			/tikuh/	[tikuh]~[tikʊ̯uh]	'mouse'
	[ũ]~[ə̃ʊ̯]	C[+nasal]__C#	/səmuʔ/	[səmũʔ]~[səmə̃ʊ̯ʔ]	'ant'
			/tənuŋ/	[tənũŋ]~[tənə̃ʊ̯ŋ]	'to weave'
			/bunuh/	[bunũh]~[bunə̃ʊ̯h]	'to kill'
	[u]	Ø__C#	/tauŋ/	[tauŋ]	'year'
			/jauh/	[jauh]	'far'

Diphthongisation also fails to occur in final open syllables. High vowels in this position are always realised as monophthongs, e.g., /kati/ [kati] 'bed' and /sayu/ [sayu] 'vegetable'. They are nasalised when following a nasal onset, e.g., /bani/ [banĩ] 'buttress root', /cumi/ [cumĩ] 'beautiful', /umu/ [umũ] 'age' and /timu/ [timũ] 'east'.

4.2.2.3.4 The diphthongs

The two phonemic diphthongs /ɛi/ and /əʊ/ in Dusun only occur in final syllables (see §4.4.3.2). They display allophonic variations, conditioned by the coda of the syllable.

In open syllables, /ɛi/ and /əʊ/ are realised as [ɛi̯] and [əʊ̯], e.g., /apɛi/ [apɛi̯] 'fire' and /buləʊ/ [buləʊ̯] 'feather'. In closed syllables, they are lowered to [æɛ̯] and [ɐɔ̯] respectively, as in /tasɛiʔ/ [tasæɛ̯ʔ] 'lake' and /tujəʊh/ [tujɐɔ̯h] 'seven'. No minimal pairs contrasting [ɛi̯] with [ɐɔ̯] or [əʊ̯] with [ɐɔ̯] are found, as these two sets of diphthongs show clear complementary distributions. The following minimal pairs in (2) thus do not only contrast final open syllables with closed syllables, but also have different nuclei in final syllables.

(2)	a.	/kucɛi/	[kucɛi̯]	'lock'
		/kucɛiŋ/	[kucæɛ̯ŋ]	'cat'
	b.	/buwɛi/	[buwɛi̯]	'to give'
		/buwɛih/	[buwæɛ̯h]	'foam'
	c.	/kaləʊ/	[kaləʊ̯]	'if'
		/kaləʊŋ/	[kalɐɔ̯ŋ]	'to throw'
	d.	/buləʊ/	[buləʊ̯]	'feather'
		/buləʊh/	[bulɐɔ̯h]	'bamboo'

Diphthongs in Dusun do not usually occur after nasal onsets except in a few special cases, e.g., the deictics /nɛiŋ/ 'DEM.PROX' and /sinɛiŋ/ 'here', as well as the personal pronoun /məʊŋ/ '2SG'.[22] The phonetic realisations of diphthongs in these words are lowered and nasalised, i.e., [næ̃ɛ̯ŋ], [sinæ̃ɛ̯ŋ] and [mɐ̃ɔ̯ŋ] respectively.

Table 4.10 summarises the phonetic realisations of diphthongs in Dusun.

[22] In these cases, the diphthongs are always followed by a velar nasal /ŋ/. Diphthongs never occur in an open syllable with a nasal onset, see §4.3.

Table 4.10: Phonetic realisations of diphthongs in Dusun

Pho-neme	Real-isation	Environment	Example		Gloss
/ɛi/	[ɛi̯]	__#	/atɛi/	[atɛi̯]	'liver'
			/mmipɛi/	[mmĩpɛi̯]	'to dream'
	[æɛ̯]	__C#	/tasɛiʔ/	[tasæɛ̯ʔ]	'lake'
			/kucɛiŋ/	[kucæɛ̯ŋ]	'cat'
			/putɛih/	[putæɛ̯h]	'white'
	[æ̃ɛ̯̃]	C[+nasal]__C#	/nɛiŋ/	[næ̃ɛ̯̃ŋ]	'this'
			/sinɛiŋ/	[sinæ̃ɛ̯̃ŋ]	'here'
/əʊ/	[əʊ̯]	__#	/batəʊ/	[batəʊ̯]	'stone'
			/abəʊ/	[abəʊ̯]	'ash'
	[ɐɔ̯]	__C#	/tandəʊʔ/	[tandɐɔ̯ʔ]	'horn'
			/idəʊŋ/	[idɐɔ̯ŋ]	'nose'
			/tujəʊh/	[tujɐɔ̯h]	'seven'
	[ɐ̃ɔ̯̃]	C[+nasal]__C#	/məʊŋ/	[mɐ̃ɔ̯̃ŋ]	'2SG'

4.2.2.3.5 The nasal vowels

Nasal vowels in Dusun do not show variation in their phonetic realisations, and they are consistently pronounced with nasality. Some examples of words with a nasal vowel are given in (3).

(3) /duʔã/ 'to pray'
/suʔã/ 'to question'
/sikĩʔ/ '(a) little'
/kəcĩʔ/ 'small'
/matɛ kaĩ/ 'fish hook'
/busũʔ/ 'smelly'
/clabũʔ/ 'plop' (an ideophone)
/hɔ̃/ 'AFF'
/kisɔ̃ʔ/ 'to scoot over'

4.3 Syllable structure

Syllables in Dusun have a template of (C)(C)V(C). Only a nucleus V is obligatory, and diphthongs take the slot of one V. In onset position, a consonant cluster CC is allowed, whereas a coda can only be a simple C. Possible syllable types in Dusun are presented in Table 4.11.

Table 4.11: Overview of syllable types in Dusun

Onset	Nucleus	Coda
	V	
	V	C
C	V	
C	V	C
CC	V	
CC	V	C

Table 4.12 summarises the distribution of consonants in Dusun at the syllable level.

Table 4.12: Distribution of consonants in Dusun at the syllable level

Position	p	b	t	d	c	j	k	g	ʔ	m	n	ɲ	ŋ	s	x	h	l	ɣ	w	y
onset	+	+	+	+	+	+	+	+	(+)	+	+	+	+	+	+	+	+	+	+	+
coda	-	-	-	-	-	-	-	-	+	(+)	(+)	(+)	+	(+)	-	+	-	-	-	-

(+ : attested, - : not attested, (+) : attested but restricted)

Distributional constraints on single consonants in Dusun are similar to those in KM and CTM, but there are some important differences. In onset position, while a glottal stop is never phonemic in the native lexicon of KM and CTM, there are two instances of an onset glottal stop in Dusun, as in /ma.ʔa.mɛ/ 'butterfly' and /ma.ta.ʔa.ɣɛi/ 'sun' (which are apparently native, reflecting an earlier full reduplicated form or a compound, cf. SM *rama-rama* and *mata-hari*).[23] In coda position, only nasals and the glottals

[23] Alternatively, the underlying forms of these words may be analysed as /ma.a.mɛ/

/ʔ, h/ are permitted. Similar to KM and CTM, when /m, n, ɲ/ occur as a coda in Dusun, they only appear in word-medial position, and they are always followed by a homorganic voiced stop to form a consonant sequence /-m.b-, -n.d-, -ɲ.j-/. Importantly, the voiceless alveolar sibilant /s/ is also attested as a coda in Dusun, occurring in a penultimate syllable and immediately followed by another consonant, as seen in words like /məs.tɛ/~/mis.tɛ/ 'mangosteen', /is.na/ 'Monday' and /mis.kiŋ/ 'poor'. There are two reasons to analyse /s/ as the coda of the penultimate syllable instead of part of a complex onset of the final syllable. First, a complex onset is never found in word-medial position in other words; treating ˣst-, ˣsn- and ˣsk- as onset clusters would alter the basic disyllabic template in Dusun (see §4.4). Second, /mis.kiŋ/ 'poor' has a variant /ssə.kiŋ/ 'poor', which suggests that the sibilant belongs to the first syllable. It is worth mentioning that two of the three examples provided are ultimately loanwords: /isna/ 'Monday' < Arabic *ithnain* and /miskiŋ/ 'poor' < Arabic *miskīn* (cf. SM *isnin* and *miskin*). The origin of /məs.tɛ/~/mis.tɛ/ 'mangosteen' is not clear.[24] Therefore, the presence of a coda /s/ in the penultimate syllable might be a foreign feature, but no conclusion can be drawn yet, hence it was indicated by (+) in Table 4.12. Word-medial consonant sequences -C.C- are discussed further in §4.6.

Combinations of consonants in CC clusters in onset position will be discussed in §4.5. Just as in KM and CTM, CC clusters in Dusun can consist of either two identical segments C_xC_x or two dissimilar segments C_xC_y.

The distributional patterns of vowels at the syllable level are fairly complex, as summarised in Table 4.13.

---

and /ma.ta.a.ɣɛi/, whereby the glottal stop between two [a] at the phonetic level derives from a rule of glottal stop epenthesis between two identical vowels, e.g., /maamɛ/ → [mãʔãmĩ]~[mãʔãmẽ] (for the raising of final /ɛ/ following a nasal onset, see §4.2.2.3.2). However, this analysis is also not optimal as no other words in Dusun exhibit glottal stop epenthesis. These words thus represent idiosyncratic cases regardless of the phonemic status of the glottal stop.

[24] It is not listed as a loanword in Jones (2007), and it has cognates in KM (/ssətɔ/) and in CTM (/ssətə/), which would reflect an earlier form ⁺mista.

Table 4.13: Distribution of oral vowels in Dusun at the syllable level

Syllable type	a	i	u	ɛ	ɔ	ə	ɛi	əʊ
V	+	+	+	+	+	-	-	-
VC	+	+	+	-	+	-	-	-
CV	+	+	+	+	+	+	+	+
CVC	+	+	+	-	+	+	+	+
CCV	+	+	+	-	-	+	-	-
CCVC	+	+	+	-	-	+	-	-

(+ : attested, - : not attested)

Only /a, i, u/ can occur with all types of syllables. The mid vowels /ɛ, ɔ/ and the diphthongs /ɛi, əʊ/ do not occur in syllables with a complex onset, as they typically appear in final syllables at the word level (see §4.4.3.2), whereas CCV(C) syllables are restricted to non-final positions. Furthermore, the mid front vowel /ɛ/ only occurs in open syllables (C)V, and neither the schwa /ə/ nor the diphthongs /ɛi, əʊ/ are permitted in onsetless syllables V(C).[25] The distribution of diphthongs is also conditioned by the nasality of the onset. As mentioned earlier, diphthongs do not usually occur following a nasal onset except in a few function words, e.g., /nɛiŋ/ 'DEM.PROX', /sinɛiŋ/ 'here' and /məʊŋ/ '2SG', where they are surrounded by two nasals.

Nasal vowels in Dusun are also often followed by a glottal stop /ʔ/, as in /kəcĩʔ/ 'small', /sikĩʔ/ '(a) little' and /busũʔ/ 'smelly'. Some exceptions include /matɛ kaĩ/ 'fish hook' and /duʔã/ 'to pray' < Arabic *duʿāʾ*.

[25] Words like /buwɛi/ 'to give', /buwɛih/ 'foam' and /iyəʊ/ 'shark' could potentially be analysed as /bu.ɛi/, /bu.ɛih/ and /i.əʊ/ respectively, placing the diphthongs /ɛi, əʊ/ in onsetless syllables. However, the phonemic status of intervocalic glides following a high vowel is argued for based on independent evidence, namely the alternation between forms like /wɛi/~/buwɛi/ 'to give', see §4.4.3.1. The incompatibility of diphthongs with onsetless syllables can be explained from a diachronic perspective, as these diphthongs reflect earlier high vowels, and the diphthongisation process only took place in syllables with an onset, see §7.4.4.2.

4.4 Word Structure

The majority of words in Dusun, whether morphologically simple or derivative, have a canonical disyllabic shape with a (C)(C)V(C).(C)V(C) template. A small number of morphologically simple monosyllables are also found in the corpus. As in KM and CTM, words with more than two syllables in Dusun are typically borrowed.

The phonological structure of simple words and derivatives in Dusun are examined in §4.4.1 and §4.4.2 respectively. Phoneme distributions at the word level are presented in 4.4.3.

4.4.1 Structure of simple words

Simple words in Dusun can be mono-, di-, tri- and quadri-syllabic.

Table 4.14 presents possible syllable types in monosyllables, which have a (C)(C)V(C) shape. Monosyllables with a ˣCCV shape are not attested, and V(C) syllable types are only found in interjections. While a handful of content words can be monosyllabic, monosyllables are more commonly function words, such as the personal pronoun /məʊŋ/ '2SG', prepositions /ŋaɲ/ 'with' and /də/ 'LOC', as well as the negator /dɔʔ/ 'NEG'.

Table 4.14: Monosyllabic simple words in Dusun

Syllable type	Example	Gloss
V	/a/	'INTERJ'
VC	/ɛh/	'INTERJ'
CV	/gɛi/	'to go'
CVC	/nɔʔ/	'to want'
CCVC	/mbuh/	'to blow'

The majority of simple words in Dusun consist of two syllables and follow a (C)(C)V(C).(C)V(C) template. Any C is optional, but syllables with a complex onset CC can occur word-initially. All other possible combinations of two syllables, except for ˣV.VC and ˣCCV.VC, are attested and presented in Table 4.15.

Table 4.15: Disyllabic simple words in Dusun

Syllable type	Example	Gloss
V.V	/a.i/	'water'
V.CV	/i.ku/	'tail; CLF'
V.CVC	/i.gaʔ/	'to catch'
VC.CVC	/am.bɛiʔ/	'to take'
CV.V	/da.i/	'forehead'
CV.VC	/ta.uŋ/	'year'
CV.CV	/pa.si/	'sand'
CV.CVC	/da.ɣɔh/	'blood'
CVC.CV	/jaŋ.ga/	'awkward'
CVC.CVC	/pin.dɛiʔ/	'short'
CCV.V	/pɣa.u/	'canoe'
CCV.CV	/ŋŋa.jɛ/	'deliberately'
CCV.CVC	/nna.tɔŋ/	'animal'
CCVC.CV	/blaɲ.jɛ/	'expense; bride price'
CCVC.CVC	/jɣəm.bəʊŋ/	'to clash'

Simple words with more than two syllables are less common in Dusun. As previously noted, there is one idiosyncratic trisyllabic native word, namely /maʔamɛ/ 'butterfly'. Other trisyllabic or quadrisyllabic words are typically loanwords, as shown in Table 4.16.

Table 4.16: Simple words with more than two syllables in Dusun

Syllable type	Example	Gloss	Origin
V.CV.CV	/u.ta.ɣɛ/	'north'	< Sanskrit
V.CV.CVC	/a.ɣə.naʔ/	'rabbit'	< Arabic
CV.CV.CV	/ba.si.ka/	'bicycle'	< English
CV.CV.CVC	/ma.jə.ɣiŋ/	'margarine'	< English
CVC.CV.CVC	/ləm.pu.yaŋ/	'ginger plant'	
CCV.CV.VC	/stu.ka.iŋ/	'stocking'	< English
CV.CV.CV.CVC	/ta.lɛ.bi.sɛiŋ/	'television'	< English

/utaɣɛ/ 'north' and /aɣənaʔ/ 'rabbit' are ultimately borrowed from Sanskrit and Arabic, presumably having entered ITM via SM (cf. SM *utara* and *arnab*). Some clear English loanwords are also trisyllabic or quadrisyllabic, e.g., /basika/ 'bicycle' (cf. SM *basikal*), /majəɣiŋ/ 'margarine', /stukaiŋ/ 'stocking' and /talɛ-bisɛiŋ/ 'television' (also /tibi/). /ləmpuyaŋ/ 'ginger plant' is also likely borrowed, as suggested by the unexpected /-m.p-/ sequence (see §4.6), but the source remains unclear.

4.4.2 Structure of derivatives

Dusun has a smaller inventory of affixes compared to KM and CTM (see §5.3.1). Derivatives are typically disyllabic and prefixed, as shown in Table 4.17. All prefixed disyllables have an initial CC cluster, and they share the maximal template CCVC.CVC with simple words.

Table 4.17: Disyllabic derivatives in Dusun

Syllable type	Example	Morphological structure	Gloss	
CCV.V	/bɣa.i/	*bɣ-ai*	(INTR-water)	'watery'
CCV.VC	/ssa.iŋ/	*s-saiŋ*	(INTR-friend)	'to befriend'
CCV.CV	/tba.kɔ/	*t-bakɔ*	(NVOL-burn)	'to be burnt'
CCV.CVC	/jja.laŋ/	*j-jalaŋ*	(INTR-road)	'to walk'
CCVC.CV	/ttiŋ.ga/	*t-tiŋga*	(NVOL-leave)	'to be left behind'
CCVC.CVC	/ŋŋam.baʔ/	*NN$_1$-ambaʔ*	(IPFV-chase)	'chasing'

As in KM and CTM, there are some examples of prefixed trisyllables with *bə-* 'INTR; MID' in Dusun, all of which have bases with initial non-geminate clusters, typically consisting of an obstruent and a liquid, as shown in (4).

(4)

/bə.tɣa.bu/	*bə-tɣabu*	(MID-scattered)	'cluttered'
/bə.gli.sɔh/	*bə-glisɔh*	(MID-anxious)	'anxious'
/bə.cla.bũʔ/	*bə-clabũʔ*	(INTR-plop)	'to make a plop sound'
/bə.sli.muʔ/	*bə-slimuʔ*	(INTR-blanket)	'to cover (oneself) with a blanket'

Other trisyllabic derivatives in Dusun are often historically suffixed with a nominaliser *-aŋ* (cf. KM *-ɛ* and CTM *-aŋ*), as seen in (5). However, there are

only a few instances with this suffix, and the suffixing process is not active at the synchronic level. For example, *manis-aŋ* 'palm sugar' cannot be synchronically derived from *manih* 'sweet'. Following Collins (1983a: 57), I consider words with the *-aŋ* suffix as loanwords.

(5)	/u.ku.maŋ/	*ukum-aŋ*	(penalise-NMLS)	'penalty'
	/ma.ka.naŋ/	*makan-aŋ*	(eat-NMLS)	'food'
	/ma.ni.saŋ/	*manis-aŋ*	(sweet-NMLS)	'palm sugar'

A similar conclusion can be drawn for derivatives with more than three syllables, which seem to be historically circumfixed with *pə- -aŋ* (cf. KM *pə- -ɛ*, CTM *pə- -aŋ* and SM *pər- -an*), as in (6). These words also cannot be synchronically derived, as the hypothetical bases ˣmɔhɔn, ˣkapuŋ and ˣjaɲji are not attested.[26]

(6)	/pə.mɔ.hɔ.naŋ/	*pə-mɔhɔn-aŋ*	(NMLS-apply-NMLS)	'application'
	/pə.ka.pu.ŋan/	*pə-kapuŋ-aŋ*	(NMLS-village-NMLS)	'settlement'
	/pə.jaɲ.ji.aŋ/	*pə-jaɲji-aŋ*	(NMLS-agree-NMLS)	'agreement'

I conclude that as far as native words are considered, both simple words and derivatives in Dusun are canonically disyllabic. Additionally, a small number of morphologically simple monosyllables and trisyllables, as well as some prefixed trisyllables, are also attested.

4.4.3 Phoneme distribution at the word level

The following distributional patterns are summarised primarily based on the canonical disyllabic template, and the phonotactic constraints on the segment distributions apply to both simple words and derivatives.

4.4.3.1 Distribution of consonants

Distributional patterns of consonants at the word level in Dusun are comparable to those in KM and CTM, as displayed in Table 4.18.

[26] The native word for 'to apply' is /mitɔʔ/, and the attested forms for 'village' and 'to promise' are /kapəʊŋ/ and /jaɲjɛi/ with diphthongs.

Table 4.18: Distribution of consonants in Dusun at the word level

Position	p	b	t	d	c	j	k	g	ʔ	m	n	ɲ	ŋ	s	x	h	l	ɣ	w	y
initial	+	+	+	+	+	+	+	+	-	+	+	+	(+)	+	+	(+)	+	+	(+)	(+)
medial	+	+	+	+	+	+	+	+	-	+	+	+	+	+	-	+	+	+	+	+
final	-	-	-	-	-	-	-	-	+	-	-	-	+	-	-	+	-	-	-	-

(+ : attested, - : not attested, (+) : attested but uncommon)

Word-initially, a glottal stop is not phonemic. Word-initial /ŋ, h/ are typically found in function words, e.g., /ŋə/ 'ANAPH', /ŋaŋ/~/dəŋaŋ/ 'with; and' and /ŋaʔ/~/saŋaʔ/ 'very', /hɔʔ/~/haʔ/ 'REL' and /hɔ̃/ 'AFF'. There are only a few content words with initial /ŋ/ or /h/, i.e., /ŋaŋɛ/ 'to open wide', /hagɛ/ 'price', /hatɛ/ 'property' and /humbaŋ/ 'to throw'.[27] Initial glides /w, y/ are mostly restricted to shortened variants of some disyllables, as in /wɛi/~/buwɛi/ 'to give', /wɔh/~/buwɔh/ 'fruit' and /ya/~/iya/ 'ringgit (currency)', but they also occur in one disyllable /wayaʔ/ 'to tell'. Initial /x/ is attested but only when it is followed by a high vowel /i/ or /u/, e.g., /xitɛ/ 'car' and /xusɛi/ 'chair'; elsewhere it occurs geminated.

Word-medially, /ʔ/ and /x/ are not permitted. Unlike in KM and CTM, where a word-medial /-h-/ is only allowed between alike vowels, /-h-/ in Dusun is also found between two dissimilar vowels, e.g., /pahuŋ/ 'tree', /pahɛ/ 'thigh' and /lahi/~/lihi/ 'neck'.

Word-finally, only /ʔ, ŋ, h/ are allowed.

4.4.3.2 Distribution of vowels

The distribution of oral vowels in Dusun at the word level is presented in Table 4.19.

[27] /hagɛ/ 'price' and /hatɛ/ 'property' have variants /ɣəgɛ/ and /ɣətɛ/, which seem to be the more conservative pronunciations. /hagɛ/ and /hatɛ/ are probably the pronunciations under the influence of SM cognates *harga* and *harta*, ultimately from Sanskrit *argha* and *artha*. Initial /h/ in /humbaŋ/ 'to throw' is unexplained.

Table 4.19: Distribution of oral vowels in Dusun at the word level

Position	a	i	u	ɛ	ɔ	ə	ɛi	əʊ
penultimate	+	+	+	(-)	(-)	+	-	-
final	+	+	+	+	+	-	+	+

(+ : attested, - : not attested, (+): attested but uncommon)

In penultimate syllables, only /a, i, u, ə/ are generally allowed. However, some contracted forms and loanwords exhibit a deviant pattern. For instance, a mid vowel /ɔ/ is found in words like /mɔʔciʔ/ 'auntie' < *mɔʔ* 'mother' + *ciʔ* 'sister', and /wɔʔki/~/waʔki/ 'representative (n.)' < Arabic *wakīl*. A mid vowel /ɛ/ is attested in /lɛnɛiŋ/ 'now', derived from *kala* 'time' (< Sanskrit *kāla*) + *nɛiŋ* 'DEM.PROX'. Furthermore, while the schwa /ə/ is permitted in penultimate syllables, it does not occur word-initially. In final syllables, all oral vowels expect for the schwa /ə/ are allowed.

Nasal vowels are only attested in final syllables.

4.5 Consonant clusters

Table 4.20 presents attested consonant clusters C_1C_2 in Dusun, with C_1 indicated on the left on the vertical line, and C_2 on the top on the horizontal line. Clusters enclosed in parentheses are infrequent, either representing variants of other clusters or limited to loanwords.

C_1 and C_2 can be either identical or different, and all consonants except for the glottals /h, ʔ/ can occur in a consonant cluster. Non-geminate clusters in Dusun can be further classified into six groups based on their components: 1) obstruent + liquid; 2) obstruent + obstruent; 3) obstruent + nasal; 4) nasal + obstruent; 5) nasal + liquid; and 6) obstruent + glide. Examples for each type of clusters are given from (7) to (13).

Table 4.20: Attested consonant clusters in Dusun

C_1 \ C_2	p	b	t	d	c	j	k	g	m	n	ɲ	ŋ	s	x	l	ɣ	w	y
p	pp	-	-	pd	pc	-	-	pg	-	-	-	-	-	-	pl	pɣ	(pw)	
b	-	bb	bt	bd	-	bj	-	bg	-	-	-	-	-	-	bl	bɣ	bw	(by)
t	tp	tb	tt	-	-	tj	tk	tg	tm	tn	-	tŋ	ts	-	tl	tɣ	(tw)	-
d	-	-	-	dd	-	-	-	-	-	-	-	-	-	-	(dl)	dɣ	dw	-
c	cp	-	-	-	cc	-	ck	-	cm	-	-	-	-	-	cl	cɣ	(cw)	-
j	-	-	-	-	-	jj	-	-	jm	jn	-	-	-	-	jl	jɣ	-	-
k	-	-	-	kd	-	-	kk	-	-	kn	-	-	-	-	kl	(kɣ)	(kw)	-
g	-	-	-	-	-	-	-	gg	-	-	-	-	-	-	gl	(gɣ)	gw	-
m	-	mb	-	-	-	-	-	-	mm	-	-	-	-	-	ml	mɣ	-	-
n	-	-	-	-	-	-	-	-	-	nn	-	-	-	-	-	-	-	-
ɲ	-	-	-	-	-	-	-	-	-	-	ɲɲ	-	-	-	-	-	-	-
ŋ	-	-	-	-	-	-	-	-	-	-	-	ŋŋ	-	-	-	-	-	-
s	sp	sb	st	sd	sc	sj	sk	sg	sm	sn	-	-	ss	-	sl	sɣ	(sw)	-
x	-	-	-	-	-	-	-	-	-	-	-	-	-	xx	-	-	-	-
l	-	-	-	-	-	-	-	-	-	-	-	-	-	-	ll	-	-	-
ɣ	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	ɣɣ	-	-

(- : not attested, ( ) : attested but uncommon)

First, all segments except for the glottals and glides are attested in a geminate cluster, as exemplified in (7).

(7) Geminate clusters

/pp-/	/ppalɛ/	'head'
/bb-/	/bbalɔh/	'to fight'
/tt-/	/ttuwaʔ/	'warts'
/dd-/	/ddiɣɛi/	'to stand'
/cc-/	/ccayɛ/	'to believe'
/jj-/	/jjaɲjɛi/	'to promise'
/kk-/	/kkatɔʔ/	'frog'
/gg-/	/ggajɛi/	'saw (n.)'
/mm-/	/mmipɛi/	'to dream'
/nn-/	/nnatɔŋ/	'animal'

/ɲɲ-/	/ɲɲapəʊ/	'broom'
/ŋŋ-/	/ŋŋali/	'to flow'
/ss-/	/ssəjiʔ/	'mosque'
/xx-/	/xxətah/	'paper'
/ll-/	/llakɛi/	'male'
/ɣɣ-/	/ɣɣəbəʊʔ/	'cupboard'

Second, the most common non-geminate clusters consist of an obstruent followed by a liquid, as in the following examples:

(8) Obstruent + liquid

/pl-/	/plitɛ/	'lamp'
/pɣ-/	/pɣaŋa/	'behaviour'
/bl-/	/blakɛ/	'all'
/bɣ-/	/bɣətɛi/	'to stop'
/tl-/	/tliŋɛ/	'ear'
/tɣ-/	/tɣimɛ/	'to receive'
/dl-/	/dlimɛ/	'pomegranate'
/dɣ-/	/dɣakɛ/	'insubordinate'
/cl-/	/clabũʔ/	'plop'
/cɣ-/	/cɣəmiŋ/	'mirror'
/jl-/	/jlagɛ/	'soot'
/jɣ-/	/jɣənih/	'clear'
/kl-/	/kladɛi/	'taro'
/kɣ-/	/kɣaniŋ/	'clerk'
/gl-/	/glabɛ/	'spider'
/gɣ-/	/gɣəbɔŋ/	'(hair) spread out'
/sl-/	/sluwɔ/	'trousers'
/sɣ-/	/sɣatuh/	'one hundred'

As in KM and CTM, there is a tendency in Dusun to avoid a stop and a liquid in such combinations sharing the same place of articulation, but there are some exceptions. /dl-/ is attested in one instance /dlimɛ/ 'pomegranate'. /kɣ-/ and /gɣ-/ are recorded in younger speakers' speech, e.g., /kɣaniŋ/ 'clerk' and /gɣəbɔŋ/ '(hair) spread out', but these words are pronounced with /xx-/ and /ɣɣ-/ among older speakers, i.e., /xxaniŋ/ and /ɣɣəbɔŋ/. I take the pronunciation of /kɣ-/ and /gɣ-/ as a result of SM's influence (cf. SM *kərani* and *gərbaŋ*), hence marked by parentheses in Table 4.20.

The third type of clusters consists of an obstruent followed by another obstruent, which are much more common in Dusun than in KM and CTM, as illustrated in (9).

(9) Obstruent + obstruent

/pd-/	/pdiyɛ/	'who'
/pc-/	/pcumɛ/	'free'
/pɡ-/	/pɡuwaŋ/	'lawyer'
/bt-/	/btəmuŋ/	'to meet'
/bd-/	/bdiɣɛ/	'flag'
/bj-/	/bjəɣəʊʔ/	'soaking wet'
/bk-/	/bkabuh/	'foggy'
/bɡ-/	/bɡəɣɔʔ/	'to depart'
/tp-/	/tpakɛ/	'be forced to'
/tb-/	/tbuwaŋ/	'hornet'
/tj-/	/tjatəʊh/	'to fall (unintentionally)'
/tk-/	/tkəjuʔ/	'to be startled'
/tɡ-/	/tɡaɣɛ/	'southeast'
/ts-/	/tsakuʔ/	'to be hung'
/cp-/	/cpədɔʔ/	'cempedak'
/ck-/	/ckalɛiʔ/	'to smash'
/kd-/	/kdiyaŋ/	'later'
/sp-/	/spuləʊh/	'ten'
/sb-/	/sbəlah/	'eleven'
/st-/	/stəŋɔh/	'half'
/sd-/	/sdiɣɛi/	'on one's own'
/sc-/	/scawaŋ/	'a cup'
/sj-/	/sjabi/	'a pouch'
/sk-/	/skulɔh/	'school'
/sɡ-/	/sɡəlah/	'a glass'

Three segments are most commonly attested as C_1 in this type of clusters, namely /b, t, s/, which often represent the prefix *b-* 'INTR; MID', *t-* 'NVOL' or the clitic *s=* 'a; one; same'. Stop + stop clusters usually consist of a voiceless stop and a voiced stop, or two voiceless stops, but there are some exceptions. The voiced stop /b-/ is sometimes followed by a voiceless stop, as in /btəmuŋ/ *b-təmuŋ* (MID-meet) 'to meet' and /bkabuh/ *b-kabuh* (INTR-fog) 'foggy', but these forms typically alternate with forms with a geminate

cluster, i.e., /ttəmuŋ/ and /kkabuh/. For the allomorphic realisations of prefixes, see §5.3.1.1.

The fourth type of clusters consists of an obstruent followed by a nasal, as listed (10).

(10) Obstruent + nasal

/tm-/	/tmakɔ/	'tobacco'
/tn-/	/tnagɛ/	'energy'
/tŋ-/	/tŋəlaŋ/	'to sink'
/cm-/	/cmuɣəʊ/	'jealous'
/jm-/	/jmalɔŋ/	'earth gnome'
/jn-/	/jnamɛ/	'brand'
/kn-/	/knuɣɛi/	'feast'
/sm-/	/smayɔŋ/	'to pray'
/sn-/	/snaɣa/	'list'

Some of these examples exhibit variation between a non-geminate cluster and a geminate cluster, as in /tmabɔ/~/mmakɔ/ 'tobacco' and /smayɔŋ/~/mmayɔŋ/ 'to pray'.

The fifth type of clusters, in which a nasal is followed by an obstruent, is only found as /mb-/, attested in two monosyllabic words in (11). ×/nd-/ or ×/ŋg-/ is not attested.

(11) Nasal + obstruent

/mb-/	/mbuh/	'to blow'
/mb-/	/mbuŋ/	'dew'

The combinations of a nasal and a liquid are found in /ml-/ and /mɣ-/, as illustrated by examples in (12).

(12) Nasal + liquid

/ml-/	/mlayəʊ/	'Malay'
/ml-/	/mlaɣɛ/	'traveller'
/mɣ-/	/mɣatɛ/	'prevalent'

The last type of clusters consists of an obstruent followed by a glide /w/ or /y/, as demonstrated in (13).

(13) Obstruent + glide

/pw-/	/pwasɛ/	'to fast'	< Sanskrit *upavāsa*	cf. SM *puasa*
/bw-/	/bwayɛ/	'crocodile'		
/tw-/	/twala/	'towel'	< Portuguese *toalha*	cf. SM *tuala*
/dw-/	/dwanɛ/	'where'		
/cw-/	/cwacɛ/	'climate'	< Sanskrit *svaccha*	cf. SM *cuaca*
/kw-/	/kwalɛi/	'wok'	< Tamil *kuvaḷai*	cf. SM *kuali*
/gw-/	/gwanɛ/	'how'		
/sw-/	/swaɣɛ/	'voice'	< Sanskrit *svara*	cf. SM *suara*
/by-/	/byasɛ/	'usual'	< Sanskrit *abhyāsa*	cf. SM *biasa*

This type of clusters is often found in loanwords, as indicated above, but it is also attested in at least three native words, namely /bwayɛ/ 'crocodile', /dwanɛ/ 'where' and /gwanɛ/ 'how' (cf. SM *buaya*, *di mana* and *bagai-mana*). This suggests that the pattern of a cluster comprising an obstruent followed by a glide is generally permissible in Dusun; these clusters are thus included in Table 4.20.

To summarise, Dusun attests a broader range of consonant clusters when compared to KM and CTM. Although the general patterns of clustering are similar, and the SSP can still be seen as playing a central role in governing permissible clusters, it appears to be applied with slightly more flexibility in Dusun. Notable exceptions include clusters consisting of a voiced stop and a voiceless stop, as found in derivatives such as /btəmuŋ/ *b-təmuŋ* (MID-meet) 'to meet' and /bkabuh/ *b-kabuh* (INTR-fog) 'foggy'. These clusters are nevertheless subject to regressive assimilation to become geminates, which also highlights the effect of the SSP.

4.6 Consonant sequences

The most common consonant sequences in Dusun are homorganic nasal + voiced stop sequences, namely /-m.b-, -n.d-, -ɲ.j-, -ŋ.g-/, as demonstrated by the examples in Table 4.21. Nasal + voiceless stop sequences are not allowed in native words.

Table 4.21: Homorganic nasal + voiced stop sequences in Dusun

Sequence	Example	Gloss
/-m.b-/	/tim.bu/	'to float'
	/am.bɛiʔ/	'to take'
/-n.d-/	/pan.da/	'clever'
	/mən.dɛ/	'what'
/-ɲ.j-/	/aɲ.jɛiŋ/	'dog'
	/blaɲ.jɛ/	'expense; bride price'
/-ŋ.g-/	/tiŋ.gɛi/	'tall'
	/tuŋ.gəʊ/	'to wait'

Dusun also permits sequences with a fricative /s/ followed by another consonant, as in /-s.t-/ in /məs.tɛ/~/mis.tɛ/ 'mangosteen', /-s.k-/ in /mis.kiŋ/ 'poor' and /-s.n-/ in /is.na/ 'Monday'. Such sequences are not attested in either KM or CTM, and as mentioned earlier in §4.3, they may have been borrowed.

Another type of sequences in Dusun consists of a glottal stop directly followed by another consonant, as listed in Table 4.22.

Table 4.22: /-ʔ.C-/ sequences in Dusun

Sequence	Example	Gloss	Origin
/-ʔ.p-/	/baʔ.pɛ/	'why'	< *buwaʔ* 'do' + †*apa* 'what'
/-ʔ.t-/	/saʔ.təʊ/	'Saturday'	< Arabic *sabt*
/-ʔ.d-/	/taʔ.dɛi/	'just now'	
/-ʔ.c-/	/mɔʔ.ciʔ/	'auntie'	< *mɔʔ* 'mother' + *ciʔ* 'sister'
/-ʔ.k-/	/biʔ.kɛi/	'to repair'	
/-ʔ.g-/	/kiʔ.gɛi/	'later'	
/-ʔ.m-/	/suʔ.mu/	'always'	< *s=* 'same' + *umu* 'age'
/-ʔ.n-/	/saʔ.nɛiŋ/	'just now'	< *saʔãʔ* 'second' + *nɛiŋ* 'DEM.PROX'
/-ʔ.s-/	/piʔ.sɛ/	'examination'	< Sanskrit *parīkṣā*

† indicates earlier forms which are no longer attested.

Several origins can be identified for words with this type of sequences: (historically) contracted forms, loanwords or historically suffixed forms.

The origins for contracted forms and loanwords are indicated in Table 4.22. /biʔkɛi/ 'to repair' is a historically suffixed form (cf. SM *baik-i* 'good-APPL'; /-ɛi/ in Dusun /biʔkɛi/ corresponds to SM *-i* 'APPL'. /taʔdɛi/ 'just now' and /kiʔgɛi/ 'later' have unexplained /-ʔ.d-/ and /-ʔ.g-/ sequences.

4.7 Vowel sequences

Table 4.23 presents the vowel sequences found in Dusun.

Table 4.23: Vowel sequences in Dusun

Sequence	Example	Gloss
/a.i/	/a.i/	'water'
/a.u/	/ta.u/	'to know'
/a.ĩ/	/matɛ ka.ĩ/	'fish hook'

In total there are only three recorded vowel sequences, all of which begin with the low vowel /a/. Among the vowels that are generally allowed in penultimate syllables (/a, i, u, ə/), the schwa /ə/ cannot be directly followed by another vowel. The high vowel /i/ and /u/ also do not occur as the first component in a vowel sequence, as potential sequences like ˣ/iV/ and ˣ/uV/ are analysed as having a corresponding phonemic glide, i.e., /iyV/ and /uwV/, as in /liyaʔ/ 'to see', /diyɛ/ '3SG', /buwaʔ/ 'to do' and /buwɛih/ 'foam'. This analysis is supported by alternations observed in words such as /waʔ/~/buwaʔ/ 'to do', /wɔh/~/buwɔh/ 'fruit' and /ya/~/iya/ 'ringgit (currency)', where the shortened monosyllabic forms preserve an initial glide. Other logical combinations of /a/ + another vowel, such as ˣ/aɛ/ and ˣ/aɔ/, are also not attested.

4.8 Stress

Stress is not phonemic in Dusun. In disyllabic words, stress is fairly consistently placed on the final syllable, whether that syllable is open or closed. This pattern is shown in the following examples:

(14) *Final open syllables*

/pata/	[paˈta]	'beach'
/matɛ/	[mãˈtɛ]	'eye'
/layɔ/	[laˈyɔ]	'to sail'
/kati/	[kaˈti]	'bed'
/kiɣɛi/	[kiˈɣɛi̯]	'left (side)'
/batəʊ/	[baˈtəʊ̯]	'stone'
Final closed syllables		
/sakiʔ/	[saˈkiʔ]~[saˈkɪi̯ʔ]	'sick'
/muluʔ/	[mũˈluʔ]~[mũˈləʊ̯ʔ]	'mouth'
/taɣɛiʔ/	[taˈɣæɛ̯ʔ]	'to pull'
/katəʊʔ/	[kaˈtɐɔ̯ʔ]	'to hit'
/awaŋ/	[aˈwaŋ]	'cloud'
/bitɔŋ/	[biˈtɑŋ]	'star'
/musɛiŋ/	[mũˈsæɛ̯ŋ]	'season'
/idəʊŋ/	[iˈdɐɔ̯ŋ]	'nose'
/tikuh/	[tiˈkuh]~[tiˈkəʊ̯h]	'rat'
/miɣɔh/	[mĩˈɣɔh]	'red'
/putɛih/	[puˈtæɛ̯h]	'white'
/tujəʊh/	[tuˈjɐɔ̯h]	'seven'

This pattern of stress placement aligns with the observation that more vowels are permitted in final syllables and that more phonemic contrasts are found in this position, suggesting a higher prominence of final syllables. The exclusive occurrence of diphthongs in final syllables is also likely connected to final-syllable stress. Diachronically, diphthongs in Dusun developed from earlier high vowels (/ɛi/ < *i and /əʊ/ < *u, see more in §7.4.4), and it is plausible that only stressed high vowels were diphthongised. Synchronically, the interaction between diphthongisation and stress further affects other high vowels in final closed syllables, as illustrated by examples such as /sakiʔ/ [saˈkiʔ]~[saˈkɪi̯ʔ] 'sick' and /muluʔ/ [muˈluʔ]~[muˈləʊ̯ʔ] 'mouth'.[28]

Exceptions to the general pattern of stressing final syllables can be found in words with initial geminates, where the stress is often placed on the initial syllable. The following words are commonly pronounced with initial stress:

[28] Kerinci, another Malayic variety spoken in Sumatra, has diphthongs in final syllables with similar origins, which also cooccur with ultimate stress (Steinhauer & Usman 1978; Steinhauer 2002).

(15)	/llakɛi/	[ˈllakɛi̯]	'male'
	/ŋŋali/	[ˈŋŋãli]	'to flow'
	/nnatɔŋ/	[ˈnnãtɑŋ]	'animal'
	/jjalaŋ/	[ˈjjalaŋ]	'to walk'
	/ŋŋuaʔ/	[ˈŋŋũwãʔ]	'to yawn'

However, the association between initial geminates and initial stress is not always consistent, as some other words with initial geminates may be pronounced with stress on the final syllables, e.g., /mmayɛiŋ/ [mmãˈɣæɛ̯ŋ] 'yesterday' and /ɲɲusuʔ/ [ɲɲũˈsuʔ]~[ɲɲũˈsəʊ̯ʔ] 'to hide'.

To sum up, stress is not phonemic in Dusun, and the final syllable of disyllabic words is typically stressed. Words with initial geminates may have stress on the initial syllable, although this pattern is not always consistent.

4.9 Summary

This chapter has described the phonology of ITM spoken in Kampung Dusun and compared it with the phonologies of KM and CTM.

Dusun has a consonant inventory consisting of twenty native phonemic consonants and a vowel inventory of ten phonemic monophthongs, which include six oral vowels /a, i, u, ɛ, ɔ, ə/ and four nasal vowels /ã, ĩ, ũ, ɔ̃/. It also has two phonemic diphthongs /ɛi/ and /əʊ/. The phonetic realisations of oral vowels are particularly complex, with conditioning environments that involve the presence and nasality of the onset consonant, as well as the presence of the coda consonant. Following a nasal consonant, all oral vowels are nasalised phonetically. In addition, the mid front vowel /ɛ/ is raised to [ĩ]~[ẽ] when following nasals, and the mid back vowel /ɔ/ is lowered to [ɑ] when preceding a velar nasal /ŋ/. The high vowels /i, u/ have a tendency to be diphthongised in final closed syllables, and the diphthongs /ɛi, əʊ/ are lowered to [æɛ̯, ɐɔ̯] respectively in closed syllables. The nasal vowels, on the other hand, are consistently realised with nasality, although their presence is limited.

The basic syllable structure in Dusun is (C)(C)V(C). All segments can be an onset (with the glottal stop /ʔ/ having a restricted occurrence), whereas in coda position, only the glottals /ʔ, h/, the nasals /m, n, ɲ, ŋ/ and the fricative /s/ are permitted. A complex onset cluster CC can consist of either

two identical segments or dissimilar ones, and variation is sometimes observed between geminate and non-geminate clusters. All segments except for the glottals /ʔ, h/ can occur in a consonant cluster, but the glides /w, y/ only occur in non-geminate clusters. Attested non-geminate clusters generally comply with the SSP, but some exceptions exist, typically resulting from the process of prefixation. Regarding vowels, a nucleus V can be either a monophthong or a diphthong. Not all vowels can cooccur with an onset and/or a coda, as there are many restrictions on their distribution at the syllable level. The schwa and the diphthongs do not occur in onsetless syllables, and the mid front vowel /ɛ/ only appears in open syllables. Following a complex onset CC, only /a, i, u, ə/ are allowed. Nasal vowels often occur in closed syllables with a coda /ʔ/.

Words in Dusun typically follow a disyllabic structure with a (C)(C)V(C).(C)V(C) pattern, which applies to both morphologically simple words and derivatives. Additionally, a small number of monosyllabic simple words and trisyllabic derivatives are also attested. In word-initial position, /ʔ/ is not phonemic, and /ŋ, h, w, y/ also have limited presence. /ʔ, x/ do not occur in word-medial intervocalic position, and only /ʔ, h, ŋ/ are permitted in word-final position. Consonant clusters typically occur word-initially, except in a few instances of trisyllabic prefixed forms where they occur in penultimate onset position. In disyllabic words, only /a, i, u, ə/ are generally allowed in penultimate syllables, whereas the schwa is not permitted in final syllables. Most word-medial sequences are homorganic nasal + voiced stop sequences, but Dusun also allows some /-ʔ.C-/ sequences, as well as /-s.C-/ sequences in a few examples. In vowel sequences formed by two adjacent vowels, the first one can only be /a/.

Dusun does not have phonemic stress. Word stress is largely predictable, typically falling on the final syllable in disyllabic words. Ultimate stress can also be associated with the large inventory of vowels permitted in final syllables and the tendency for high vowels to diphthongise in this position.

4.10 Comparison of NEPM phonologies

So far, I have described the phonology of all three varieties of NEPMs, and it is evident that there are many strikingly similar features characterising NEPMs as a whole. As regards the consonant systems, the inventories are

identical across all three varieties, and consonants also have nearly identical distributions, revealing a clear left-oriented consonantal asymmetry in syllable structure and word structure (Ogloblin 2018: 330). There are many restrictions on the distribution of consonants at the right edge of a syllable and a word: only the glottals /ʔ, h/ and the nasals /m, n, ɲ, ŋ/ are permitted in coda position (and possibly also /s/ in Dusun), among which only /ʔ, ŋ, h/ are allowed in word-final position. In contrast, the majority of consonants are permitted at the beginning of a word. Consonant clusters also attested in word-initial position, and they include geminates, which are cross-linguistically quite rare. NEPMs also share many similarities in syllable and word shapes. The basic syllable template is (C)(C)V(C), and words are generally disyllabic with a (C)(C)V(C).(C)V(C) shape. Similar generalisations can also be drawn for the patterns of permitted consonant clusters and consonant sequences in all three varieties, albeit with minor differences.

The main phonological differences among these varieties are primarily found in their vowel systems, with ITM (Dusun) displaying several peculiarities in particular. Unlike KM and CTM, Dusun lacks a contrast between mid-low and mid-high vowels, and the mid-vowels /ɛ, ɔ/ are generally not allowed in penultimate syllables. Furthermore, Dusun has two phonemic diphthongs, which contrast with all vowels except for /ə/ in final syllables. Each variety also has its own phonotactic constraints on vowel distribution. For example, a schwa is not allowed in final syllables in KM and Dusun, but it is permitted in final open syllables in CTM. Nonetheless, all three varieties share the feature of having phonemic nasal vowels, with words containing nasal vowels often being cognates.

Many phonological similarities observed across NEPMs at the synchronic level can be attributed to their shared historical development. A detailed examination of the evolution of their phonological systems will be provided in Chapter 7.

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 lluma*ʔ (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 lluma*ʔ '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/ → [ˌjjɛ] '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 tuwɛ 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 *ɲɲ-<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_I- '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
aŋjiŋ tu təɣuh lagi, diyə ɲɲ-<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~nVʔ* or *tVh~tVʔ* respectively, the V being any non-high monophthong. The factors determining the choice of *V* in *nVh~nVʔ* or *tVh~tVʔ* and the choice of the final consonant remain unclear, but *-h* appears to be more common than *-ʔ*. These weakened demonstratives are considered phonologically deficient, as there are no other words in ITM ending in *-əh*, *-əʔ*, *-ɛh* or *-ɛʔ*.

Table 5.3: Demonstratives in ITM

Full forms	Weakened forms	Gloss
nɛiŋ	*nah~naʔ* *nɛh~nɛʔ* *nɔh~nɔʔ* *nəh~nəʔ*	'DEM.PROX'
təʊ	*tah~taʔ* *tɛh~tɛʔ* *tɔh~tɔʔ* *təh~təʔ*	'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-slimuʔ* (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 *NN₁-* 'IPFV' and *NN₂-* '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

NN₁-akaʔ	(IPFV-lift)	→	*ŋŋ-akaʔ*	'lifting'
NN₁-ɣukah	(IPFV-climb)	→	*ŋŋ-<ɣ>ukah*	'climbing'
NN₁-paŋge	(IPFV-call)	→	*mm-<p>aŋge*	'calling'
NN₁-tanɛ	(IPFV-plant)	→	*nn-<t>anɛ*	'planting'
NN₂-cɛtɔʔ	(NMLS-print)	→	*ɲɲ-<c>ɛtɔʔ*	'printer'
NN₁-kuteʔ	(IPFV-pick)	→	*ŋŋ-<k>uteʔ*	'picking'
NN₂-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 *NN₁-* seems to occur before a base with initial *l* and surface as a singleton *m-*, i.e., ʔ*NN₁-lamboŋ* (IPFV-bump) → *m-lamboŋ* 'bumping'. However, the allomorphic alternation of *NN₁-* → *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 *NN₁-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 *NN₁-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'
likəʊ̈	'twist'	→	*b-likəʊ̈-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-baɣu.*
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

aɲjɔʔ	'to move'	→	*bɣ-aɲjɔʔ*	'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 baɲɔʔ 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 ppuwaŋ 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 *ubaʔ~ɔbaʔ* 'medicine', *pɣ-* derives *pɣ-ubaʔ~pɣ-ɔbaʔ* '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'
ubaʔ	'medicine'	→	*pɣ-ubaʔ*	'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

toʔ bɔmɔh tu=lah ***pɣ-ubaʔ*** *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 *waʔ* '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 *waʔ* '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 *NN₁-* 'IPFV'

The prefix *NN₁-* 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), *NN₁-* 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 *NN₁-* 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 *NN₁-* 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 *NN₁-* is very common, which casts doubt on the analysis of *NN₁-* 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 tu?*
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 *NN₁-* 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 kuwaʔ baũ,*
indeed when 1PL blend DEM.PROX strong smell
kitɔ ***nn-<t>umih*** *poŋ bakeʔ tapi mɛmɛ sədaʔ.*
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* ***ɲɲ-<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 *NN₁-* in my CTM and ITM corpora, I assume *NN₁-* functions in the same way. Some examples are given in (47) and (48).

(47) CTM

a. *masə diyə* ***ɲɲ-<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=kə?*
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 sakeʔ	(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 sakiʔ	(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 *alaʔ nnənoŋ* (tool weave) 'loom', *alaʔ ŋŋuko* (tool measure) 'measuring tools', and ITM *jaɣoŋ ŋŋaiʔ* (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 *pɛsəŋjɛ* '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'	+	-	-	-
NN₁-	'IPFV'	-	+	+	-
NN₂-	'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-pəɣɛ***	'to be at war'
bɩwɔh	'fruit'	→	***b-bɩwɔh***	'to bear fruit'
kɩwɔh	'gravy'	→	***k-kɩwɔh***	'to become gravy'
jalɛ	'road'	→	***j-jalɛ***	'to walk'
saiŋ	'friend'	→	***s-saiŋ***	'to befriend'
ɣayɔ	'Eid al-Fitr'	→	***ɣ-ɣayɔ***	'to celebrate Eid al-Fitr'
CTM				
bɩɲiŋ	'sound'	→	***b-bɩɲiŋ***	'to make sound'
bau	'smell'	→	***b-bau***	'smelly'
diɣi	'self'	→	***d-diɣi***	'to stand'
jalaŋ	'road'	→	***j-jalaŋ***	'to walk'
kawaŋ	'friend'	→	***k-kawaŋ***	'to befriend'
ITM				
biniŋ	'wife'	→	***b-biniŋ***	'to marry (a wife)'
daɣɔh	'blood'	→	***d-daɣɔh***	'to bleed'
cabɔŋ	'branch'	→	***c-cabɔŋ***	'branched'
kəbuŋ	'farm'	→	***k-kəbuŋ***	'to farm'
gɩnɛ	'use'	→	***g-gɩnɛ***	'useful'

(54) KM

pɩwɔʔ 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-jəmo*** *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-susoŋ****.*
much INTERJ book MID-arrange
'There are a lot of books piling up.' (CTM_220927_e02_103)

(60) ITM

budɔʔ təʊ ***s-sandɔ*** *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

bəto	'correct'	→	***b-bəto***	'to correct s.th.'
tido	'to sleep'	→	***t-tido***	'to put s.o. to sleep'
dəɣah	'loud'	→	***d-dəɣah***	'to raise (voice)'
kəɣiŋ	'dry'	→	***k-kəɣiŋ***	'to dry s.th.'
gadiʔ	'thick'	→	***g-gadiʔ***	'to thicken'
siyaʔ	'finished'	→	***s-siyaʔ***	'to finish'
*luma*ʔ	'crushed'	→	***l-luma*ʔ**	'to crush'

CTM

bəsɔ	'big'	→	***b-bəsɔ***	'to enlarge'
təɣaŋ	'clear'	→	***t-təɣaŋ***	'to clarify'
*kuwa*ʔ	'strong'	→	***k-kuwa*ʔ**	'to strengthen'
*maso*ʔ	'to enter'	→	***m-maso*ʔ**	'to stuff'
mandi	'to bathe'	→	***m-mandi***	'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

pagi	'morning'	→	***p-pagi***	'(in the) early morning'
jaŋɔʔ	'pretty'	→	***j-jaŋɔʔ***	'very pretty'
pəkaʔ	'thick'	→	***p-pəkaʔ***	'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 *daɣɔh* → *d-daɣɔh* (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-lumaʔ*** *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-ɣayɔ.***
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) <u>Gemination as a non-volitional marker</u>

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 sədaʔ 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 ×*katɔʔ* and ×*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.

CHAPTER 6

Word classes and basic syntax

6.1 Introduction

This chapter presents an overview of word classes and basic syntactic structures in NEPMs.

Word classes are categories of words that share similar morphosyntactic properties. Given the limited productive morphology in NEPMs, there are not sufficient criteria to distinguish word classes on morphological grounds. Word classes are therefore primarily determined based on their syntactic behaviour. The chapter begins by describing two open word classes: nouns in §6.2.1 and verbs in §6.2.2. NEPMs do not have a separate adjectival class that can be clearly distinguished from verbs; instead, words expressing property concepts are formally treated as a subclass of verbs called stative verbs. Closed word classes are introduced in the subsequent sections: adverbs in §6.2.3, pronouns in §6.2.4, demonstratives and deictics in §6.2.5, quantifiers and numerals in §6.2.6, classifiers in §6.2.7, interrogative words in §6.2.8, negators in §6.2.9, prepositions in §6.2.10 and conjunctions in §6.2.11. Discourse particles and interjections are discussed in §6.2.12 and §6.2.13.

The structure of noun phrases is outlined in §6.3. Verb phrases are discussed in the basic clausal syntax in §6.4. §6.5 summarises this chapter.

6.2 Word classes

6.2.1 Nouns

Nouns are words that function as the heads of noun phrases (NPs), which serve as arguments within clauses. Semantically, nouns typically refer to persons, places, things and abstract concepts. Syntactically, nouns may be identified by their collocation with modifiers such as demonstratives, possessive pronouns, quantifiers, numerals and classifiers. In the KM example in (1), the two nouns *jiɣɛ* 'neighbour' and *ɣumɔh* 'house' are modified by the (possessive) pronoun *diyɔ* '3' and the demonstrative *tu* 'DEM.DIST' respectively. The two NPs in which they occur are indicated by square brackets. The structure of NPs is described in more detail in §6.3.

(1) KM
[***jiɣɛ*** *diyɔ*]$_{NP}$ *baka* [***ɣumɔh*** *tu*]$_{NP}$.
neighbour 3 burn house DEM.DIST
'His neighbour burnt the house.' (KM_180827_e02_24)

(2) and (3) present similar examples from CTM and ITM. The nouns *ɔɣaŋ* 'person' and *bula* 'ball' are modified by the quantifier *ɣama* 'many' and the numeral-classifier combination *s=buti* 'one=CLF' respectively.

(2) CTM
[*ɣama* ***ɔɣaŋ***]$_{NP}$ *mandi s=suŋa tu.*
many person bathe LOC=river DEM.DIST
'Many people bathe in the river.' (CTM_181029_e02_60)

(3) ITM
[*s=buti* ***bula***]$_{NP}$ *jatəʊh daɣɛi mijɛ.*
one=CLF ball fall from table
'A ball fell from the table.' (ITM_180909_e02_39)

Other syntactic properties of nouns include: 1) they can occur in prepositional phrases following the head (e.g., ITM *daɣɛi* ***mijɛ*** 'from table' in (3) above); 2) they can function as modifiers of nouns (e.g., KM *tukɛ* ***kəboŋ*** 'farmer' in (4)); 3) they are negated with nominal negators, namely KM *bukɛ* and CTM/ITM *bukaŋ*, as illustrated in (5) (more accurately non-verbal

negators, see in §6.2.9). Example (4) also demonstrates that in addition to functioning as arguments, nouns and NPs can serve as predicates without any copulas. A nominal predicate is further illustrated in (6).

(4) KM
ayɔh aku [***tukɛ*** *kəboŋ*]$_{NP}$.
father 1SG craftsman farm
'My father is a farmer.' (KM_180830_e01_14)

(5) ITM
bukaŋ *kkatɔʔ haʔ tubiʔ, tikuh haʔ tubiʔ dayɛi lubɔŋ.*
NEG frog REL come.out rat REL come.out from hole
'It was not a frog that came out from the hole, but a rat.'
(ITM_180907_n01_12)

(6) CTM
budɔʔ ni [*anɔʔ mɔʔciʔ*]$_{NP}$*=kəʔ*
kid DEM.PROX child auntie=Q
'Is this kid auntie's child?' (CTM_181029_e02_17)

Like in most other western Austronesian languages, nouns in NEPMs are not morphologically marked for gender, case or number. Except for nominal compounds and reduplicated forms, the great majority of nouns are morphologically simple.

There is no special morphology that is characteristic of nouns. The only nominalising prefix *NN*$_2$- is formally identical to the verbal aspect marker *NN*$_1$-, hence not indicative of the nominal status of a word. There are also no prefixes that can only be applied to nominal bases. For instance, the verbal prefix *bɣ*- 'INTR; MID' does not only occur on nouns as an intransitive verbal marker, but also on verbs as a middle (voice) marker (see §5.3.1.2). The word class of nouns is therefore defined only on the basis of their semantics and syntactic properties.

6.2.2 Verbs

Verbs are words that denote actions, states or properties. Syntactically, verbs function as the heads of predicates, often without any morphology. This can be seen in the following examples where the verbs are marked in bold:

(7) KM
diyɔ poŋ ***akaʔ*** *s=bako,* ***lətɔʔ*** *atah basika diyɔ.*
3 then lift one=basket put top bike 3
'He lifted a basket (of pears) and put it on his bike.'
(KM_180814_n01_23)

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

(9) ITM
diyɛ ***baŋuŋ, caɣɛi*** *dalaŋ kasuʔ,* ***caɣɛi*** *dalaŋ bujəʊŋ,*
3SG get.up search inside shoe, search inside bottle
bukɛ *pitəʊ-malɛiŋ,* ***bukɛ*** *pitəʊ-malɛiŋ* ***paŋgi*** *kkatɔʔ ŋə.*
open door-thief, open door-thief call frog ANAPH
'He got up, searched inside the shoes, searched inside the bottle, opened the windows, opened the windows and called the frog.'
(ITM_180907_n01_6)

Verbs in NEPMs can be morphologically complex with one of the verb-forming prefixes (*bɣ-* 'INTR; MID', *tɣ-* 'NVOL', *pɣ-* 'CAUS; FCT' and *NN₁-* 'IPFV'). Words with one of these prefixes are easily recognisable as verbs.

Other syntactic properties of verbs include: 1) they are negated with the verbal negators, namely KM *tɔʔ* and CTM/ITM *dɔʔ*;[38] 2) they may cooccur with aspectual and modal markers such as *doʔ* 'PROG'. As can be seen in the KM example in (10), the verb *ita-ita* 'to peep (repetitively)' follows the progressive marker *doʔ*, and *gaba* 'to notice' is negated with *tɔʔ*.

(10) KM
diyɔ doʔ ***ita-ita,*** *tɛŋɔʔ tuwɛ* *diyɔ tɔʔ* ***gaba,*** *diyɔ poŋ akaʔ.*
3 PROG RDP-peep look owner 3 NEG notice, 3 then lift
'He was peeping; seeing that the owner didn't notice, he just lifted (a basket).' (KM_180814_n01_22)

[38] Occasionally verbal predicates can also be negated with KM *bukɛ* or CTM/ITM *bukaŋ*, see §6.2.9.

Verbs can be classified into intransitive, transitive and ditransitive verbs depending on how many arguments they take. For example, KM *ita-ita* 'RDP-peep' in (10) is an intransitive verb following the subject argument *diyɔ* '3'; ITM *bukɛ* 'open' in (9) is a transitive verb following the subject *diyɛ* '3SG' and preceding the object *pitəʊ-malɛiŋ* 'window'. Example (11a) illustrates an ditransitive verb *tuŋjəʊʔ* 'to show' in ITM, which takes three arguments: the subject *luʔmaŋ* 'Lukman (a person name)', the indirect object *mɔʔ* 'mother' and the direct object *gambɔ* 'picture'. In ditransitive clauses, the indirect object generally follows the verb immediately and precedes the direct object. However, more commonly, the recipient/beneficiary is demoted to the oblique role introduced in a prepositional phrase, as shown in (11b).

(11) ITM

a. *luʔmaŋ* ***tuŋjəʊʔ*** *mɔʔ* *gambɔ* *dalaŋ* *talipuŋ.*
Lukman show mother picture inside phone
'Lukman shows his mother the pictures on his phone.'
(ITM_220923_e01_19)

b. *luʔmaŋ* ***tuŋjəʊʔ*** *gambɔ* *dalaŋ* *talipuŋ* [*kə mɔʔ*]$_{PP}$.
Lukman show picture inside phone to mother
'Lukman shows the pictures on his phone to his mother.'
(ITM_220923_e01_18)

NEPMs do not have a distinct class of adjectives. Words expressing qualities or attributes (i.e., semantic adjectives, see Dryer 2007) function grammatically in a verb-like manner, as they can act as intransitive predicates, and their morphosyntactic properties in this slot are similar to those of intransitive verbs. They are also negated with verbal negators KM *tɔʔ* or CTM/ITM *dɔʔ*. Consider the following examples:

(12) KM

a. *sɔɣɔ* *moŋ* ***kɔhɔ.***
voice 2SG soft
'Your voice is soft.' (KM_180825_e01_24)

b. *sɔɣɔ* *moŋ* ***tɔʔ*** ***dəɣah.***
voice 2SG NEG loud
'Your voice is not loud.' (KM_180827_e01_11)

While some differences between semantic adjectives and prototypical verbs can be identified, whether they can be used as criteria to recognise adjectives as a distinct word class or a subclass of verbs is sometimes arbitrary (see the "subclass problem" in Haspelmath 2001; also see Schachter & Shopen 2007: 19). In the case of NEPMs, I consider that the differences are not robust enough to establish a separate class of adjectives. Semantics adjectives are subsumed as a subclass of verbs called "stative verbs", as opposed to "dynamic verbs" denoting actions or events. Two parameters for potential grammatical variation between stative verbs and dynamic verbs are discussed below.

The first parameter concerns the possibilities of serving as attributive modifiers for nouns within NPs. Stative verbs can directly modify a head noun, whereas dynamic verbs generally need to be placed in a relative clause when modifying a noun. In (13a), *batu* 'stone' is directly modified by *bəsɔ* 'big', whereas in (13b), a relativiser *hɔʔ* is used so that *budɔʔ* 'kid' is modified by *cuɣi* 'to steal' in the NP.

(13) CTM

a. *batu* ***bəsɔ*** *tu*
stone big DEM.DIST
'the big stone' (CTM_181023_n02_32)

b. *budɔʔ hɔʔ* ***cuɣi***
kid REL steal
'the kid who steals' (CTM_181025_n02_53)

There are nevertheless a few instances of dynamic verbs directly serving as attributive modifiers in NPs, as shown in (14) and (15). In this regard, stative verbs and dynamic verbs cannot be clearly differentiated based on their potential differences as modifiers within an NP.

(14) KM
pɔleh ləpah dɔh [*ɔɣɛ* ***c-cuɣi*** *tu*]$_{NP}$ *di supəmakeʔ.*
police release already person IPFV-steal DEM.DIST LOC supermarket
'The police has released the guy stealing at the supermarket.'
(KM_180827_e02_34)

(15) ITM

[*buɣəʊŋ-atəʊ* ***kluwɔ*** *tɔh*]$_{NP}$ *tɣəjuŋ* ...
bird-ghost come.out DEM.DIST jump ...
'The owl that came out plunged ...' (ITM_180907_n02_21.1)

The second parameter involves the possibility of being gradable and comparable. Comparative or superlative degrees in NEPMs are expressed with adverbs meaning 'more' or 'most'. The following examples demonstrate that KM/CTM *lagi* (and its reduced form *agi*) or ITM *lagɛi* 'more' either precedes or follows the stative verbs being modified, and KM *skali*, CTM *skali~kkali* or ITM *skalɛi~kkalɛi* 'most' follows the stative verbs (see §6.4.1.2 for more discussion on comparative and superlative constructions).

(16) KM

bəsa ***lagi***	(big more)	'bigger'
mudɔ ***lagi***	(young more)	'younger'
lagi *səda*ʔ	(more delicious)	'more delicious'
*səda*ʔ ***skali***	(delicious most)	'most delicious'

(17) CTM

bəsɔ ***agi***	(big more)	'bigger'
lɛbɔ ***lagi***	(wide more)	'wider'
lagi *bəɣa*ʔ	(more heavy)	'heavier'
tiŋgi ***kkali***	(tall most)	'tallest'

(18) ITM

*kəcĩ*ʔ ***lagɛi***	(small more)	'smaller'
jauh ***lagɛi***	(far more)	'further'
lagɛi *payɔh*	(more difficult)	'more difficult'
bəsɔ ***skalɛi***	(big most)	'biggest'

While dynamic verbs are not gradable and cannot appear in comparative constructions, they can also be modified by (*l*)*agi* or *lagɛi* which has the meaning of 'again'. The constructions of a dynamic verb + (*l*)*agi* or *lagɛi*, as shown in (19) to (21), are parallel to the constructions of a stative verb + (*l*)*agi* or *lagɛi* presented above. This observation further suggests that semantic adjectives are better treated as a subclass of verbs, and the exact meaning of the adverbial modifier depends on the semantics of the verbs being modified.

(19) KM
diyɔ poŋ ɣukah spulɔʔ nɔʔ kuteʔ ***lagi.***
3 then climb again want pick again
'He climbed back up the tree and wanted to pick (pears) again.'
(KM_180814_n01_9)

(20) CTM
diyə ɲŋ-<s>alɔʔ ***agi.***
3 IPFV-bark again
'It keeps on barking.' (CTM_181023_n02_24)

(21) ITM
dɔʔ, guɣɛiŋ ***lagɛi!***
NEG fry again
'No, keep frying!' (ITM_180917_cv01_101)

In sum, there are not sufficient morphosyntactic criteria to recognise a separate adjectival class.

6.2.3 Adverbs

Adverbs constitute a heterogeneous class of words that serve to modify various non-nominal constituents including verbs, other adverbs, clauses or sentences. They express concepts such as degree, frequency, locative or temporal settings. Some common adverbs are listed in Table 6.1.

Table 6.1: Adverbs in NEPMs

Category	Meaning	KM	CTM	ITM
Degree	'very (much)'	*suŋgoh, saŋaʔ*	*saŋaʔ*	*suŋgəʊh, saŋaʔ*
	'more'	*lagi~agi*	*lagi~agi*	*lagɛi*
	'most'	*skali*	*skali~kkali*	*skalɛi~kkalɛi*
	'(not) at all'	*lasoŋ*	*lasoŋ*	*lasəʊŋ*
	'a lot'	*baɲɔʔ*	*baɲɔʔ*	*baɲɔʔ*
	'(a) little'	*sikiʔ*	*sikiʔ*	*sikĩʔ~ikĩʔ*

Category	Meaning	KM	CTM	ITM
Frequency	'often'	*kəɣɛʔ, acaʔ*	*kəɣɛʔ, ɣajiŋ*	*ɣajiŋ*
	'always'	*sɔʔmɔ*	*sɔʔmɔ*	*suʔmu*
	'sometimes'	*kadɛ*	*skali-kalə*	*kadɔŋ-kadɔŋ*
	'usually'	*bɛsɔ, skalɔ*	*bɛsɔ*	*byasɛ*
	'seldom'	*jaɣɛ*	*jaɣaŋ*	*jaɣɔŋ*
Locative	'here'	*sini~siniŋ*	*sini~siniŋ*	*sinɛiŋ, dinɛiŋ*
	'there'	*situ*	*situ*	*sitəʊ, ditəʊ*
	'there; yonder'	*noŋ, dənoŋ*	*noŋ, dinoŋ~dənoŋ*	*nuŋ, sinuŋ, dinuŋ*
Temporal	'now'	*lɔni*	*ləniŋ*	*lɛnɛiŋ*
	'just now'	*taʔdi, saʔni*	*taʔdi~aʔdi, saʔdi*	*taʔdɛi, saʔnɛiŋ*
	'later'	*kdiyɛ*	*kdiyaŋ, kɛʔgi*	*kdiyaŋ, kiʔgɛi*
	'recently'	*baɣu ni~niŋ*	*baɣu ni~niŋ*	*baɣəʊ nɛiŋ*
	'today'	*aɣi ni~niŋ*	*aɣi ni~niŋ*	*aɣɛi nɛiŋ*
	'tomorrow'	*ɛsɔʔ*	*ɛsɔʔ*	*isəʊʔ*
	'the day after tomorrow'	*lusɔ*	*lusə*	*lusɛ*
	'yesterday'	*mmaɣiŋ*	*mmaɣeŋ*	*mmaɣɛiŋ*
	'already'	*dɔh*	*dɔh*	*dɔh*
Evidential	'only, just'	*cumɔ, sajɔ*	*sajə*	*sajɛ*
	'indeed; really'	*mɛmɛ*	*mɛmaŋ*	*mimɔŋ*
	'also'	*jugɔʔ, pulɔʔ*	*jugɔʔ, pulɔʔ*	*jugɛ~ugɛ, pulɔʔ*
	'probably'	*bəkali, koʔ*	*koʔ*	*kuʔ*
	'also; even; then'	*poŋ*	*poŋ*	*puŋ*
Interrogative	'when'	*bilɔ*	*bilə*	*bilɛ*
	'where'	*manɔ, mmanɔ*	*manə, mmanə*	*manɛ, dwanɛ*
	'whereto'	(*kwanɔ*)[39]	*kwanə*	*kwanɛ*
	'how'	*gganɔ, gwanɔ, lagumanɔ*	*gganə*	*gwanɛ*
	'why'	*baʔpɔ*	*baʔpə~waʔpə*	*baʔpɛ~waʔpɛ*

[39] KM *kwanɔ* 'whereto' could be elicited, but it is probably obsolete. In naturalistic data, only *manɔ* 'where' is used, e.g., *gi manɔ* 'go where', cf. ITM *gɛi kwanɛ*.

NEPMs lack distinctive manner adverbs. The meaning equivalents of manner adverbs are expressed by stative verbs without formal changes. This is illustrated in (22) to (24). As all subcategories of adverbs have limited members, adverbs are regarded as a closed word class.

(22) KM
diyɔ mmasɔʔ ***mɔlɛʔ.***
3 cook good
'It cooks well.' (KM_180820_cv03_23)

(23) CTM
pah tu, aŋiŋ utaɣə poŋ tiyuʔ ***kuwaʔ*** *sapa=lah ...*
after DEM.DIST wind north also blow strong until=SFP ...
'Then the north wind blows hard until ...' (CTM_220928_n01_05)

(24) ITM
mɔʔ saʔnɛiŋ baɣəʊ tutuʔ ***ɣapaʔ.***
mother just.now have.just close tight
'I've just closed it tightly.' (ITM_180917_cv01_98)

Adverbs denoting frequency or degree usually occur immediately adjacent to the verbs they modify, as shown in (25). Other types of adverbs, such as the temporal adverb *baɣəʊ nɛiŋ* 'recently' in (26), have more flexible positions, as they can have scope over the entire clause.

(25) KM
hɔʔ pɛseŋ diyɔ mmasɔʔ ***sɔʔmɔ.***
REL kind 3 cook always
'The kind that she always cooks.' (KM_180820_cv03_154)

(26) ITM
baɣəʊ nɛiŋ, *kamiŋ gɛi tah pəkaŋ təh ...*
have.just DEM.PROX 1PL.EXCL go DEM.DIST Pekan DEM.DIST ...
'Recently we went to Pekan.' (ITM_180926_cv02_52)

Since adverbs form a "catch-all" category, many adverbs presented above can be classified into other word classes. Locative setting adverbs and interrogative adverbs will be discussed in more detail in §6.2.5 and §6.2.8.

6.2.4 Pronouns

Pronouns are free forms that function to fill the position of a noun or an NP in a clause (Payne 1997: 43). This section focuses on personal pronouns and relative pronouns. Demonstrative pronouns and interrogative pronouns are discussed in §6.2.5 and §6.2.8 respectively.

6.2.4.1 Personal pronouns

Personal pronouns in NEPMs distinguish singular and plural numbers, and three persons. For some pronominal forms, however, the number or person distinction is neutralised. The full paradigms of personal pronouns are presented in Table 6.2, Table 6.3 and Table 6.4, followed by notes on the use of pronouns in each variety.

Table 6.2: Personal pronouns in KM

	SG	PL
1	*aku*, *kawɛ*, *ambɔ*, *kitɔ*	*kitɔ*
2	*moŋ*, *dɛmɔ*	*dɛmɔ*, *moŋ ssəmɔ*
3	*diyɔ*, *iyɔ*	*dɛmɔ*, *diyɔ*, *iyɔ*

In KM, *aku* is the most frequent singular form among first-person pronouns. In addition, *kawɛ* and *ambɔ* (cf. SM *kawan* 'friend' and *hamba* 'slave') can be used by speakers to refer to themselves for honorific purposes. The 1PL pronoun *kitɔ* does not distinguish clusivity. While the consultants also reported that *kitɔ* can function as a singular pronoun, no example illustrating this usage has been found in the corpus.

Regarding second-person pronouns, *moŋ* is the neutral term used among friends, acquaintances and people from the same generation. Kinship terms and titles, such as *mɔʔ* 'mother', *umi* 'mother' and *pɔʔciʔ* 'uncle', are often employed as forms of address for showing respect (these terms can also be used as first-person pronouns to refer to the speakers themselves). There is no distinct form of a 2PL pronoun. To overtly address several listeners, one can use the 2SG pronoun *moŋ* combined with a modifier indicating plurality, such as *ssəmɔ* 'all'. It also also been reported that the 3PL pronoun *dɛmɔ* can function as 2SG and 2PL pronouns (in the latter case

it may combine with words like *sɛʔ* 'group', *puwɔʔ* 'group' or *ɣama* 'many', i.e., *sɛʔ dɛmɔ*, *puwɔʔ dɛmɔ* or *dɛmɔ ɣama*, also see Abdul Hamid 1994: 86), but this usage is yet to be attested.

Diyɔ and *dɛmɔ* are the common 3SG pronoun and 3PL pronoun respectively. *Diyɔ* or *i̇yɔ* can also refer to plural referents if the context allows, as in (27). The exact difference between *diyɔ* and *i̇yɔ* is unclear. Both can refer to animate or inanimate referents, but in general *i̇yɔ* does not occur often.

(27) KM
*ɔɣɛ tuwɔ deh, **diyɔ** payɔh.*
person old SFP 3 difficult
'Old people are difficult.' (KM_180820_cv03_152)

Table 6.3: Personal pronouns in CTM

	SG	PL
1	*aku*, *ambə*, *sayə*	*kitə*
2	*moŋ*	*moŋ* (*ssəmə*)
3	*yə*, *diyə*	*yə*

The pronominal system in CTM is similar to that in KM. In addition to the commonly used 1SG pronoun *aku*, *ambə* and *sayə* can be used in polite address. The 1PL pronoun *kitə* does not distinguish clusivity. The main differences between CTM and KM pronouns are in the third-persons: *yə* is the more common form, and there is no equivalent to KM *dɛmɔ* in CTM. To overtly express or emphasise the plurality of third-person referents, *sɛʔ-sɛʔ yə* (RDP-group 3) can be used.

Table 6.4: Personal pronouns in ITM

	SG	PL
1	*akəʊ*	*kitɛ* (INCL), *kamiŋ* (EXCL)
2	*məʊŋ*	*mikɛ*
3	*diyɛ*	*dimɛ* (HUM), *diyɛ* (NHUM)

The pronominal system in ITM exhibits more differences when compared with that in KM and CTM. There is only one pronoun for each person in singular forms, namely *akəʊ* '1SG', *məʊŋ* '2SG' and *diyɛ* '3SG'. 1PL pronouns have a distinction of clusivity: *kitɛ* includes the hearers, whereas *kamiŋ* excludes the hearers. Furthermore, ITM has a dedicated 3PL pronoun *dimɛ*, which is restricted to human referents.[40] For 3PL non-human referents, *diyɛ* is used.

Unlike many Malayic languages, NEPMs do not have a set of clitic forms for personal pronouns. In possessive constructions, all pronouns appear in full forms following the possessum, e.g., KM *ɣumɔh aku* (house 1SG) 'my house', *sɔɣɔ moŋ* (voice 2SG) 'your voice'; CTM *buku aku* (book 1SG) 'my book', *basika diyə* (bike 3) 'his bike'; ITM *ayɔh məʊŋ* (father 2SG) 'your father', *saiŋ dimɛ* (friend 3PL) 'their friend'.

In addition to the personal pronouns discussed above, KM has another pronominal form *ɲɔ*, and ITM has *ŋə*. They are analysed as anaphoric pronouns, referring to an antecedent that was mentioned earlier in the discourse, either a situation or a person/object. In (28), *ɲɔ* refers to the scenario described in the first two clauses, which 'would get cursed'. In (29), *ŋə* refers to *kəɣɛ* 'monkey' mentioned in the first clause, and it replaces the head noun in the NP *kəɣɛ s=iku* (monkey one=CLF) 'a monkey'.

(28) KM

ɔɣɛ təpɔh nɔʔ makɛ ɣayɔ ptamɔ,
person book want eat Eid.al-Fitr first

ɣayɔ kduwɔ tɔʔleh makɛ lagi meh,
Eid.al-Fitr second cannot eat yet SFP

ɲɔ *supɔh k=ɔɣɛ=lah.*
ANAPH curse AGT=person=SFP

'People make reservations to eat for the first Eid al-Fitr, but can't even get them on the second Eid al-Fitr day; it would get cursed by people.' (KM_180820_cv03_141–142)

[40] ITM *dimɛ* and KM *dɛmɔ* are cognates, which reflect an earlier form +dima. It might be a contraction of +dia with another morpheme. Another cognate of this form is found in Perak Malay as *dema*, which according to Brown (1921: x) can be decomposed to *dia-ma*, derived from the 3SG pronoun *dia* with the suffixation of a plural marker *-ma* on personal pronouns.

(29) ITM
kəɣɛ nɔʔ maɣɛi tapi buwɔh ambutaŋ gamɔʔ a,
monkey want come take fruit rambutan guess INTERJ
*mmaɣɛiŋ adɛ napɔʔ **ŋə** s=iku.*
yesterday EXIST look ANAPH one=CLF
'It seems like monkeys are coming to pick rambutan, yesterday I did see one.' (ITM_180917_cv01_107)

6.2.4.2 Relative pronouns

Relative pronouns are pronouns that mark relative clauses, which typically function as noun modifiers within NPs. Relative pronouns in NEPMs are KM *hɔʔ*, CTM *hɔʔ* and ITM *hɔʔ~haʔ*. Examples illustrating their usage are presented in (30) to (32).

(30) KM
tapi gaʔ [*diyɔ* [***hɔʔ** masɔ aɣoŋ*]$_{REL}$]$_{NP}$ [***hɔʔ** pətiŋ*]$_{REL/NP}$.
but EMPH 3 REL when steam REL important
'But the steaming process is important.' (Lit. 'But the one when steaming is the one that is important.') (KM_180820_cv03_130)

(31) CTM
yə nɔʔ laɣi daɣipadə [*buɣoŋ-atu* [***hɔʔ** kəjɔ yə tu*]$_{REL}$]$_{NP}$.
3 want run from bird-ghost REL chase 3 DEM.DIST
'He wanted to run away from the owl that was chasing him.' (CTM_181023_n02_32)

(32) ITM
a. [***haʔ** tɛh*]$_{REL/NP}$ *jadɛi ləkaʔ tagɛih, uŋgəʊh.*
REL DEM.DIST become stick addictive very
'That becomes addictive, very much.' (ITM_180930_cv01_65)
b. [*anɔʔ diyɛ* [***hɔʔ** ppuwaŋ*]$_{REL}$]$_{NP}$, [***hɔʔ** tuwɛ təʊ*]$_{REL/NP}$ *adɛ*
child 3SG REL female REL old DEM.DIST EXIST
taʔdɛi.
just.now
'Her eldest daughter was there just now.' (ITM_180926_cv02_4)

Relative pronouns often conjoin an antecedent head noun and serve to add more information about the referent. When the head noun is unspecified, not expressed or elided, relative clauses may be headless, which can often be translated as 'the one that is', as illustrated by *hɔʔ pətiŋ* 'what is important' in (30), *haʔ tɛh* 'that one' in (32a), and *hɔʔ tɯwɛ təʊ* 'the old one' in (32b). There are also some relative clauses where the relative pronoun is omitted, as in (33).

(33) ITM
[*uɣɔŋ* Ø [***ɲn-<s>andɔ*** *təpɛi dindɛiŋ*]$_{REL}$ *təʊ*]$_{NP}$ *makaŋ pisɔŋ.*
person Ø IPFV-lean side wall DEM.DIST eat banana
'The person leaning on the wall is eating banana.'
(ITM_180921_e03_11)

6.2.5 Demonstratives and deictics

Demonstratives are reference indicators whose meanings are deictic as they vary depending on space and time. NEPMs make a two-way distinction between proximal and distal demonstratives, indicating the relative distance between the referents and the deictic centre, which is roughly the location of the speaker/hearer. Table 6.5 lists the demonstratives in NEPMs.

Table 6.5: Demonstratives in NEPMs

	KM	CTM	ITM
PROX	*ni~niŋ*	*ni~niŋ*	*nɛiŋ*
DIST	*tu*	*tu*	*təʊ*

The variation between KM/CTM *ni* and *niŋ* is phonological rather than functional (see §7.3.2). Syntactically, demonstratives can be either adnominal, which combine with other elements in an NP and occur at the right periphery, as in (34a), or pronominal, which can substitute an NP on their own, as in (34b).

(34) KM
a. [*duɣiyɛ hɔʔ bəsa ŋa manih* ***tu***]$_{NP}$
durian REL big and sweet DEM.DIST
'that big and sweet durian' (KM_180825_e01_13)

b. [***ni***]$_{NP}$ [*ula*]$_{NP}$.
DEM.PROX snake
'This is a snake.' (KM_180816_e02_25)

As mentioned in §5.2.3, ITM also has a series of reduced demonstratives in the form of *nVh~nVʔ* 'DEM.PROX' and *tVh~tVʔ* 'DEM.DIST', with V being any non-high vowel. However, it is not clear which factors determine the choice of demonstratives in ITM. As shown in (35), various forms of demonstratives can appear in a single sentence, and the selection of vowel in *nVh~nVʔ* and *tVh~tVʔ* also does not seem to be dependent on the quality of vowels in words surrounding the demonstratives.

(35) ITM
baɣəʊ ***nɛiŋ,*** *kamiŋ gɛi* ***tah*** *pəkaŋ* ***təh,***
have.just DEM.PROX 1PL go DEM.DIST Pekan DEM.PROX
ikuʔ ***nɛh.***
follow DEM.PROX
'Recently we went to Pekan, and we followed this (route).'
(ITM_180926_cv02_52–53)

In addition to functioning like demonstratives within NPs, *nVh~nVʔ* and *tVh~tVʔ* can also behave as particles serving the pragmatic function of calling attention. This usage is illustrated in (36) to (39). In these cases, *nVh~nVʔ* or *tVh~tVʔ* can often be translated as 'look', 'like this' or 'like that'.

(36) ITM
taɣɛiʔ ujəʊŋ taʔdɛi, ***nɛh,*** *tɔʔ təŋɔh.*
pull end just.now DEM.DIST/PART put middle
'You pull the tips, like this, and put in the middle.'
(ITM_180917_cv01_51-52)

(37) ITM
A. *apɛi təh* *kəcĩʔ!*
fire DEM.DIST small
'Lower the heat!'
B. *kəcĩʔ dɔh.*
small already
'It's already on low.'

A. *eh, kəcĩʔ lagɛi, ɔ* ***tah.***
INTERJ small more INTERJ DEM.DIST/PART
'Even lower, yeah like that.' (ITM_180917_cv01_68-70)

(38) ITM
tiŋuʔ romantiʔ ŋaʔ ***təʊ*** ***tah.***
look romantic(ENG) very DEM.DIST DEM.DIST/PART
'They look very romantic like that.' (ITM_180907_n02_36)

(39) ITM
tɛʔ *anɔʔ adɛ pulɔʔ du=iku nn-<s>usuʔ nuŋ.*
PART child EXIST also two=CLF IPFV-hide there
'Look there are two kittens hiding there.' (ITM_220910_cv01_11)

Related to demonstratives are deictic adverbs. Unlike demonstratives, however, deictic adverbs make a three-way distinction regarding the distance with respect to the deictic centre, as shown in Table 6.6.

Table 6.6: Deictic adverbs in NEPMs

Meaning	KM	CTM	ITM
'here'	*sini~siniŋ*	*sini~siniŋ*	*sinɛiŋ, dinɛiŋ*
'there'	*situ*	*situ*	*sitəʊ, ditəʊ*
'there; yonder'	*noŋ, dənoŋ*	*noŋ, dinoŋ~dənoŋ*	*nuŋ, sinuŋ, dinuŋ*

Historically, deictic adverbs are derived from demonstratives. The first two sets of deictic adverbs meaning 'here' and 'there' are constructed on the basis of corresponding proximal and distal demonstratives (KM/CTM *ni*(*ŋ*) and *tu*, ITM *nɛiŋ* and *təʊ*), combined with *si* or *di*.[41] The third set of deictic adverbs is based on *noŋ* or *nuŋ* (cf. SM *nun* 'yonder', archaic), which can also be combined with *di/də* or *si*.

[41] *Di* was originally a preposition with multiple functions, including indicating locative relationships (Adelaar 2005a). The meaning of *si* (or possibly *s-*, see Adelaar 1992: 127) is unclear. Note that in ITM the locative preposition is *də* instead of *di*, which suggests that the derivation from demonstratives to deictic adverbs is not synchronically active.

6.2.6 Quantifiers and numerals

Quantifiers and numerals are words used to indicate the quantity of the referents. They take the same syntactic slot and share a number of syntactic properties, but numerals differ from non-numeral quantifiers in that they typically require a classifier (see §6.2.7) when quantifying nouns.

Quantifiers and numerals usually occur as modifiers within NPs, preceding the head nouns. Numerals may also follow head nouns, see §6.3 for more detail on the word order property. When the referent is clear from the context and omitted, quantifiers and numerals (more often the combination of a numeral + a classifier) can be NP heads. In (40a), *ɣama* 'many' directly modifies the noun *uɣɔŋ* 'person'. In (40b), *duwɛ* 'two' quantifies the number of *anɔʔ* 'child' with *uɣɔŋ*, the classifier for humans. In (40c), the referent of *tujəʊh lapaŋ iku* (seven eight CLF) can be inferred from the preceding context, and the numeral + classifier combination takes up the full NP. Quantifiers and numerals can also function as quantifier/numeral predicates (QPs), as illustrated in (41), (42) and (43).

(40) ITM

a. [***ɣama*** *uɣɔŋ*]$_{NP}$ *makaŋ siɣɛih, ...*
many person eat betel ...
'Many people eat betel nuts ... ' (ITM_180930_cv01_25.1)

b. *diyɛ adɛ* [***duwɛ*** *uɣɔŋ anɔʔ*]$_{NP}$.
3SG have two CLF child
'He has two children.' (ITM_180923_n01_23)

c. *anɔʔ diyɛ puŋ adɛ jugɛ,* [***tujəʊh lapaŋ*** *iku*]$_{NP}$.
child 3SG also EXIST also seven eight CLF
'There were also his children, seven or eight.'
(ITM_180907_n02_35)

(41) KM
pəɣɔh asɛ-lima=kɛ, diyɔ [***baɲɔʔ***]$_{QP}$.
squeeze sour-citrus=TAG, 3 much
'When you squeeze the lime, you know, there's a lot (of juice).'
(KM_180820_cv03_69)

(42) ITM
kubɔ diyɛ [***ɣama***]$_{QP}$, *tanɔh ai diyɛ* [***baɲɔʔ***]$_{QP}$ ...
buffalo 3 many land water 3 much
'He had a lot of buffaloes and land.' (Lit. '(the amount of) buffalo is many, (the amount of) land is a lot.') (ITM_220920_cv01_34)

(43) ITM
nasiʔ [***tigɛ***]$_{QP}$, *ikaŋ* [***duwɛ***]$_{QP}$, *dauŋ* [***sɛ***]$_{QP}$ ...
rice three fish two leaf one ...
'(When ordering) Three portions of rice, two fishes, one portion of salad ...' (Lit. '(the amount of) rice is three, (the amount of) fish is two, (the amount of) leave is one.') (ITM_220910_fn)

6.2.6.1 Numerals

NEPMs have a decimal numeral system that is typical of Malayic languages. The basic cardinal numerals are listed in Table 6.7.

Among the numerals for 'one', *sɔ*/*sə*/*sɛ* are used in counting and as numeral predicates, as shown in (44) and (43) above. KM/CTM *satu* and ITM *satəʊ* 'one' are only used as the last digit in the formation of higher numbers, e.g., CTM *duwə puloh satu* '21'.

(44) KM
diyɔ kali, diyɔ mmike, diyɔ poŋ bilɛ, ***sɔ*** *duwɔ,* ***sɔ*** *duwɔ.*
3 count 3 think 3 then count one two one two
'He was counting; he pondered and counted, one two, one two.'
(KM_180814_n01_43)

Table 6.7: Basic numerals in NEPMs

Meaning	KM	CTM	ITM
'one'	*sɔ, satu*	*sə, satu*	*sɛ, satəʊ*
'two'	*duɯɔ*	*duɯə*	*duɯɛ*
'three'	*tigɔ*	*tigə*	*tigɛ*
'four'	*ppaʔ*	*ppaʔ*	*ppaʔ*
'five'	*limɔ*	*limə*	*limɛ*
'six'	*nnɛ*	*nnaŋ*	*nnaŋ*
'seven'	*tujoh*	*tujoh*	*tujəʊh*
'eight'	*lapɛ*	*lapaŋ*	*lapaŋ*
'nine'	*smilɛ*	*smilaŋ~mmilaŋ*	*smilaŋ*
'ten'	*s=puloh*	*s=puloh*	*s=puləʊh*
'teens'	*bəlah*	*bəlah*	*bəlah*
'eleven'	*s=bəlah*	*s=bəlah*	*s=bəlah*
'twelve'	*duɯɔ bəlah*	*duɯə bəlah*	*duɯɛ bəlah*
'thirteen'	*tigɔ bəlah*	*tigə bəlah*	*tigɛ bəlah*
'twenty'	*duɯɔ puloh*	*duɯə puloh*	*duɯɛ puləʊh*
'twenty-one'	*duɯɔ puloh satu*	*duɯə puloh satu*	*duɯɛ puləʊh satəʊ*
'thirty'	*tigɔ puloh*	*tigə puloh*	*tigɛ puləʊh*
'hundred'	*ɣatoh*	*ɣatoh*	*ɣatuh*
'thousand'	*ɣibu*	*ɣibu*	*ɣibəʊ*
'million'	*jutɔ*	*jutə*	*jutɛ*

When followed by a noun, a classifier or some other numeral base such as *puloh/puləʊh* 'ten' or *bəlah* 'teens', *sɔ/sə/sɛ* are cliticised to *s=*, as can be seen in (45). Cliticisation does not occur if the quantity of 'one' is emphasised, as illustrated in (46).

(45) KM

budɔʔ tu poŋ ambiʔ buɯɔh pɛ tu bɯɔʔ gi kə saiŋ
kid DEM.DIST then take fruit pear DEM.DIST bring go to friend

diyɔ=lah, bagi ***s=bute s=ɔɣɛ.***
3=SFP give one=CLF one=person

'The boy took the pears and brought them to his friends, giving one pear to each person.' (KM_180814_n01_38)

(46) ITM

a. *adɛ* ***sɛ*** *butu atah lata.*
EXIST one bottle top floor
'There is one bottle on the floor.' (ITM_180909_e01_19)

b. *akəʊ adɛ tujəʊh uɣɔŋ anɔʔ,* ***sɛ*** *ppuwaŋ, nnaŋ llakɛi.*
1SG have seven CLF child one female six male
'I have seven children, one girl and six boys.'
(ITM_180923_n01_4)

The numerals for 'two' (*duwɔ*, *duwə* and *duwɛ*) may be cliticised to *du=* preceding vowel-initial classifiers, as in KM/CTM *du=ɛkɔ* or ITM *du=iku* 'TWO=CLF'.

Numerals from 11–19 are formed with the digit 1–9 followed by *bəlah* 'teens'. Higher numbers with millions, thousands, hundreds or tens are formed based on the template in (47), illustrated by CTM *limə jutə ppaʔ ɣibu tigə ɣatoh duwə puloh satu* '5,004,321'.

(47) DIGIT + *jutɔ/jutə/jutɛ* + DIGIT + *ɣibu/ɣibəʊ* + DIGIT + *ɣatoh/ɣatuh* + DIGIT + *puloh/puləʊh* + DIGIT

Ordinal numerals are formed periphrastically using the word for 'number' (KM/CTM *nɔmbɔ*, ITM *numbɔ*) plus corresponding cardinal numbers. The ordinal numerals meaning 'first' are also often expressed with words meaning 'beginning', 'early' or 'before', e.g., KM *mulɔ-mulɔ* 'RDP-beginning', CTM *hɔʔ mulə* 'the first; the early one' and ITM *hɔʔ duləʊ* 'the one from before' in ITM. *Ptamɔ* 'first' is also attested in KM, presumably a loanword from SM *pərtama*. (48) presents an additional example of an ordinal numeral *numbɔ tigɛ* 'third' in ITM.

(48) ITM
anɔʔ hɔʔ ***numbɔ tigɛ*** *namɛ ɣahiŋ.*
child REL number three name Rahim
'The third child is called Rahim.' (ITM_180923_n01_12)

An alternative strategy of constructing ordinal numerals with the prefix *k-* is occasionally attested, e.g., KM *k-duwɔ* and ITM *k-duwɛ* 'second', but the restricted occurrences of this derivational strategy suggests that it is likely borrowed from SM.

6.2.6.2 Non-numeral quantifiers

Common non-numeral quantifiers are presented in Table 6.8.

Table 6.8: Non-numeral quantifiers in NEPMs

Meaning	KM	CTM	ITM
'many'	*ɣama*	*ɣama*	*ɣama*
'much, many'	*baɲɔʔ*	*baɲɔʔ*	*baɲɔʔ*
'(a) little'	*sikiʔ*	*sikiʔ*	*sikĩʔ~ikĩʔ*
'all'	*ssəmɔ*	*ssəmə*	*smuwɛ*

Among these quantifiers, *ɣama* is restricted to quantifying humans and big animals. *Baɲɔʔ* is used to quantify all other referents, as illustrated in (49). Also compare ITM *ɣama ləmbəʊ* 'many cows' and *baɲɔʔ səmuʔ* 'many ants'.

(49) ITM

a. *uɣɔŋ nɔʔ waʔ* ***baɲɔʔ baŋunaŋ*** *baɣəʊ b=bandɔ təʊ.*
person want make many building new LOC=city DEM.DIST
'They want to build many new buildings in the city.'
(ITM_180919_e01_42)

b. ***baɲɔʔ pitih*** *diyɛ katɛ.*
much money 3 say
'She said he had a lot of money.' (ITM_220910_cv01_285)

When used as quantifiers, KM/CTM *sikiʔ* and ITM *sikĩʔ~ikĩʔ* '(a) little' appear to be restricted to quantifying uncountable nouns, as shown in (50). There are no dedicated words meaning 'a few, some' that are compatible with countable nouns; the semantic equivalents are expressed with constructions like 'two three (four)', as illustrated in (51) and (52).

(50) ITM
lətɔʔ ugɛ ***ikĩʔ susəʊ*** *ah.*
put also little milk INTERJ
'Also add a bit of milk.' (ITM_180917_cv01_31)

(51) KM
*buwaʔ siki*ʔ ***duwɔ tigɔ ppaʔ*** *bute ladɔ,*
make little two three four CLF chilli

*sədaʔ suŋgoh kitɔ makɛ nasi*ʔ*=kɛ?*
delicious very 1PL eat rice=TAG

'I made some chillies, they were very delicious when paired with rice, you know?' (ITM_180920_cv03_190)

(52) ITM
*dudəʊ*ʔ *umɔh pidah* ***duwɛ tigɛ*** *aɣɛi, nɔ*ʔ *gɛi pəkaŋ, jauh jugɛ.*
live house Pidah two three day want go Pekan far also

'(We are going to) stay at Pidah's house for a few days, then go to Pekan, but it's also far.' (ITM_180926_cv02_48)

6.2.7 Classifiers

NEPMs employ a small number of classifiers, which reflect the conceptual categorisation of nouns when being counted. The most salient features involved in the classification of nouns by classifiers are the distinction between human and non-human referents, and the animacy of non-human referents. Other relevant features include the size and shape of the objects. Common classifiers attested in NEPMs are listed in Table 6.9.

Table 6.9: Classifiers in NEPMs

Usage	KM	CTM	ITM	Literal meaning
for humans	*ɔɣɛ*	*ɔɣaŋ*	*uɣɔŋ*	'person'
for non-human animates	*ɛkɔ*	*ɛkɔ*	*iku*	'tail'
for small objects and fruits	*bute*	*bute~ute*	*buti*	'seed'
for big objects	*buwɔh*	*buwɔh*	*buwɔh*	'fruit'
for long or tall objects	*batɛ*	*bataŋ*	*batɔŋ*	'trunk'
for a bunch of bananas	*sikaʔ*	*sikaʔ*	*sikaʔ*	'comb'
for separated parts of fruits or plants	*ulah*	*ulah*	*ulah*	
for flat and thin objects	*la*	*la*	*la*	
for lumpy items	*kətu*	...	–	

Classifiers in the first four rows are most commonly used. As also indicated in the table, most classifiers are nouns in origin, and they are still used as such. Presumably in order to avoid the concatenation of a classifier that is identical to the head noun, *ɔɣɛ/ɔɣaŋ/uɣɔŋ* 'person; human being' are quantified by numerals directly without a classifier (e.g., KM *tigɔ ɔɣɛ* 'three people' instead of ˣ*tigɔ ɔɣɛ ɔɣɛ*). Nouns for measurement (weight, time, frequency, etc.) or currency are also not quantified by classifiers.

Syntactically, classifiers occur with countable nouns, and they only occur when numerals or interrogative quantifiers are present, which they immediately follow. As mentioned in §6.2.6, when the head noun is omitted, it is usually the combination of a numeral + a classifier that stands as the full NP, as illustrated in (40c) and (53a) below. There are nevertheless a few examples of classifiers being left out together with head nouns, as in (53b).

(53) ITM

a. *ŋə buwɛi* [*s*=***buti***]$_{NP}$ *s=uɣɔŋ, buwɔh pɛ təʊ.*
ANAPH give one=CLF one=person fruit pear DEM.DIST
'He gave one pear to each person.' (ITM_180919_n01_38)

b. *bagɛi* [*sɛ*]$_{NP}$ *s=uɣɔŋ*
give one one=person
'(He) gave one (pear) to each person.' (ITM_180927_n02_54)

Younger speakers of CTM also produced constructions like (54) where an expected classifier between the numeral and the head noun is absent, which seems to suggest that classifiers are not strictly obligatory. Overall speaking, however, classifiers are almost always used when countable nouns are quantified, and it is possible that sentences like (53b) and (54) reflect nonstandard usage.

(54) CTM
sə batu ni
one stone DEM.PROX
'a stone' (CTM_181023_n02_23)

In addition to classifiers, NEPMs have measure words that are used to indicate a particular quantity of the referents, such as KM *s=**kilo** satɛ* (one=kilo coconut.milk) 'one kilo of coconut milk', CTM *s=**bako** wɔh pɛ* (one=basket

fruit pear) 'a basket of pears' and ITM *s=**jabi** ai* (one=pouch water) 'a pouch of water'. Unlike classifiers, measure words are not restricted to quantifying countable nouns. They also do not reflect the categorisation of nouns by their inherent properties, but only indicate how the referents are measured.

6.2.8 Interrogatives

Interrogatives are words (or clitics) used to form questions. Two types of interrogatives can be distinguished in NEPMs, which I call "interrogative words" and "interrogative particles" respectively. Interrogative words are used to ask non-polar questions, replacing the constituents being asked. Depending on the grammatical categories of the replaced constituents, interrogative words cut across several word classes including pronouns, determiners, adverbs and quantifiers. Interrogative particles, on the other hand, are optionally used in clause-final position to form polar questions.

Table 6.10 provides a list of interrogative words in NEPMs. The fullest set of examples demonstrating the usage of each interrogative word can be found for ITM, and they are are given first in (55). Selected examples from KM and CTM are given in (56) and (57).

Table 6.10: Interrogative words in NEPMs

Category	Meaning	KM	CTM	ITM
Pronoun	'what'	*ggapɔ*	*pəndə~məndə*	*məndɛ*
	'who'	*sapɔ, pdiyɔ*	*sapə, piyə*	*pdiyɛ~piyɛ*
Determiner	'which'	*manɔ*	*manə*	*manɛ*
Adverb	'when'	*bilɔ*	*bilə*	*bilɛ*
	'where'	*manɔ, mmanɔ*	*manə, mmanə*	*manɛ, dwanɛ*
	'whereto'	(*kwanɔ*)	*kwanə*	*kwanɛ*
	'how'	*gganɔ, gwanɔ, lagumanɔ*	*gganə*	*gwanɛ*
	'why'	*baʔpɔ*	*baʔpə~waʔpə*	*baʔpɛ~waʔpɛ*
Quantifier	'how much, how many'	*bɣapɔ*	*wwapə*	*bɣapɛ*

(55) ITM

a. *mɔʔ caɣɛi* ***mənd*ɛ***
mother search what
'What are you (addressing his mother) looking for?'
(ITM_180917_cv01_71)

b. ***pdiy*ɛ** *tibɔ kə məʊŋ?*
who flip for 2SG
'Who will flip (the dough) for you?' (ITM_180917_cv01_3)

c. *mamaʔ* ***man*ɛ***
Mamat which/where
'Which Mamat?' (or 'Mamat from where?')
(ITM_180930_cv01_10)

d. ***bil*ɛ** *pɔʔciʔ nɔʔ gɛi b=bandɔ?*
when uncle want go LOC=city
'When is uncle going to the city?' (ITM_180919_e01_15)

e. *diyɛ ambɛiʔ* ***dwan*ɛ***
3SG take where
'Where did he get them?' (ITM_180930_cv01_14)

f. *dɔh dimɛ təʊ gɛi* ***kwan*ɛ***
well 3PL DEM.DIST go to.where
'Well, where have they gone?' (ITM_180926_cv01_8)

g. *dɔʔ, məʊŋ waʔ* ***gwan*ɛ** *məʊŋ dɔʔ tibɔ?*
NEG 2SG make how 2SG NEG flip
'No, how are you making it if you don't flip (the dough)?'
(ITM_180917_cv01_5)

h. ***waʔp*ɛ** *paka sluwɔ pindɛiʔ=jə tah?*
why wear pants short=just DEM.DIST
'Why are you only wearing shorts like that?'
(ITM_180917_cv01_42)

i. ***bɣap*ɛ** *kəpɛiŋ?*
how.many piece
'How many pieces?' (ITM_180930_cv01_60)

(56) KM

a. *ŋŋ-aja* ***ggapɔ?***
IPFV-teach what
'What (does she) teach?' (KM_180820_cv03_107)

b. ***sapɔ*** *diyɔ?*
who 3
'Who's that?' (KM_180820_cv03_144)

c. ***manɔ*** *nn-<t>utoʔ bəkah?*
where NMLS-close container
'Where is the lid of the container?' (KM_221025_e02_101)

d. *waʔ* ***lagumanɔ*** *ɔɣɛ kusia ni?*
make how person Kusial DEM.PROX
'How does this person from Kusial make this?'
(KM_180820_cv03_14)

e. *katɔ* ***baʔpɔ*** *diyɔ nɔʔ gi mɔh kaʔ niŋ?*
say why 3 want go house sister Ning
'He said, why does he want to go to Ning's house?'
(KM_180816_cv01_14)

(57) CTM

a. *ɛsɔʔ aɣi* ***məndə?***
tomorrow day what
'What day is it tomorrow?' (CTM_181022_cv01_8.2)

b. *xxəjə* ***mmanə*** *hɔʔ tu ɔ??*
work where REL DEM.DIST INTERJ
'Where does she work?' (CTM_181024_cv02_38)

c. *mɔʔ wi kə* ***piyə*** *baɣu ni baju,*
mother give to who have.just DEM.PROX shirt
mɔʔ kabɔ aʔdi?
mother tell just.now
'What did you just say, who did you give the shirt to recently?'
(CTM_181022_cv01_47)

d. ***gganə*** *məʔciʔ yaŋ nəʔ mitəʔ kuɣaŋ?*
how auntie Yam want request less
'How is it if auntie Yam wants to ask for discount?'
(CTM_181024_cv02_3)

e. *jəh* ***wwapə*** *ah?*
grade how.many INTERJ
'Which grade is she in?' (CTM_220928_cv01_6)

As can be seen from the examples, interrogative pronouns and determiners typically take up the same syntactic slot as the replaced constituent, without being fronted to the clause-initial position. Interrogative adverbs, on the other hand, like other types of adverbs, have more flexible positions. The word for 'what' can also be used as a filler in hesitation, or refer to something that the speaker cannot recall at the moment of utterance, as illustrated in (58).

(58) KM
diyɔ ŋŋ-aja ***ggapɔ,***
3 IPFV-teach whatchamacallit
nahu=kɔ ***ggapɔ*** *eh.*
grammar=Q whatchamacallit INTERJ
'She teaches grammar or whatever.' (KM_180820_cv03_108.1)

Many interrogative words are historically complex forms, derived based on PM *apa 'what' and *mana 'which, where'.[42] Interrogatives words formed based on *apa 'what' include:

(59) KM *ggapɔ* 'what' < *muga apa (thing what)[43]
CTM *pəndə~məndə*, ITM *məndɛ* 'what' < *apa bənda (what thing)
KM *sapɔ*, CTM *sapə* 'who' < *si-apa (PERS-what)
KM *pdiyɔ*, CTM *piyə*, ITM *pdiyɛ~piyɛ* 'who' < *apa dia (what 3SG)
KM *baʔpɔ*, CTM *baʔpə~waʔpə*, ITM *baʔpɛ~waʔpɛ* 'why' < *buat apa (do what)
KM *bɣapɔ*, CTM *wwapə*, ITM *bɣapɛ* 'how much, how many' < *baraʔ apa (INDEF what)

[42] *-a is regularly reflected as *-ɔ*, *-ə* and *-ɛ* in KM, CTM and ITM respectively, see §7.4.3.

Note that no reflexes of *apa 'what' are attested by themselves, and all interrogative words in (59) are morphologically simple at the synchronic level.

The interrogatives words in (60) are formed based on *mana 'which, where', which is also reflected as *manɔ/manə/manɛ* 'which, where'. Some of these interrogative words may be analysed as complex synchronically, e.g., KM *m=manɔ*, CTM *m=manə* (LOC=where) 'in/at/on/from where', and KM *lagu-manɔ* (method-which) 'how'.

(60) KM *mmanɔ*, CTM *mmanə*, ITM *dwanɛ* 'where' < *di mana (LOC where)
KM *gganɔ*, *gwanɔ*, CTM *gganə*, ITM *gwanɛ* 'how' < *bagai mana (kind which)
KM *lagumanɔ* 'how' < *lagu mana (method which)

Unlike interrogative words, interrogative particles only serve to indicate the status of a sentence as a polar question. KM uses the clitic *=kɔ* or *=kə* for this purpose, and CTM and ITM have *=kə*, as illustrated in (61) to (63). The usage of these particles is optional. Example (63) also shows that interrogative particles may be followed by a verbal negator, with which they form a tag meaning 'or not'.

(61) KM
*doʔ mahaʔ=**kɔ**ʔ*
live Maahad=Q
'(Does he) live in Maahad?' (KM_180820_cv03_105)

(62) CTM
*aɣɔʔ moŋ nn-<t>anaŋ padi=**kə**ʔ*
father 2 IPFV-plant paddy=Q
'Is your father a rice farmer?' (CTM_181029_e02_15)

(63) ITM
*nɔh, ɣasɛ manih=**kə** dɔʔ?*
DEM.PROX, feel sweet=Q NEG
'Here, does it taste sweet or not?' (ITM_180917_cv01_62)

[43] Cf. KM *muɣɔ* 'thing, item; seemingly'. The origin of *ggapɔ* 'what' < *muga apa is suggested by Ruslan Uthai (2011: 87).

Additionally, interrogative particles can be used to present alternatives or options, functioning like a conjunction meaning 'or' that connects two or more constituents. This usage is exemplified in (64) and (65). However, rather than suggesting that *kɔ* and *kə* are conjunctions, it may be more appropriate to view this usage as an extension of their primary function as interrogative particles. For instance, in (66), the sentence is also grammatical without the alternative constituent *gganə* 'how', suggesting that *=kə* in this position can be interpreted as a regular interrogative particle.

(64) KM
*ikɛ kəɣiŋ gɔɣeŋ=**kɔ** cicɔh budu=**kɔ** jadi dɔh.*
fish dry fry=Q little budu=Q accomplish already
'Either frying the fish dry or adding a little bit of budu (k.o. sauce), it's done.' (KM_180820_cv03_163)

(65) ITM
*məʊŋ ɣasɛ ikĩʔ sayu nəh, masiŋ=**kə** tawɔ?*
2SG taste little vegetable DEM.PROX salty=Q bland
'You try to taste the vegetable, is it salty or bland?'
(ITM_220915_e03_28)

(66) CTM
*mɛmaŋ ɔdə ŋaŋ ʃaɣikaʔ yə=**kə** gganə?*
indeed order(ENG) with company(SM) 3=Q how
'Did you really order them from the company or what?'
(CTM_181024_cv02_15)

6.2.9 Negators

Negators are words associated with negative polarity, expressing the falsity of an assertion or a proposition. Five major types of negators can be distinguished in NEPMs, as presented in Table 6.11.

Table 6.11: Negators in NEPMs

Category	Meaning	KM	CTM	ITM
Question	'no'	*dɔʔ*	*dɔʔ*	*dɔʔ*
Verbal	'not'	*tɔʔ*	*dɔʔ*	*dɔʔ*
Non-verbal	'no, not'	*bukɛ*	*bukaŋ*	*bukaŋ*
Prohibitive	'don't'	*jaŋɛ*	*jaŋaŋ*	*jaŋaŋ*
Negative existential	'there is not'	*taʔdɔʔ*	*taʔdɔʔ*	*taʔdɔʔ*

The first type of negators is the question negator or negative particle, which is *dɔʔ* in all three varieties. It is the counterpart of the affirmative particle *hɔ̃* 'AFF'; both particles are used in isolation, specifically in response to questions or antecedent assertions by showing (dis)agreement (cf. English 'yes' and 'no').

The second type is the verbal negator KM *tɔʔ* or CTM/ITM *dɔʔ*, which negates a verbal predicate. The usage of question negators and verbal negators is illustrated in (67) and (68). It is worth noting that a formal distinction between these two categories is only made in KM.

(67) KM

a. *iyɔ* ***tɔʔ*** *buboh bawɛ s=ulah ggapɔ?*
3 NEG put garlic one=CLF what
'They don't put a clove of garlic or things like that?'
(KM_180816_cv01_64)

b. ***dɔʔ, tɔʔ*** *buboh bawɛ-puteh,* ***tɔʔ*** *buboh bawɛ-bəsa,* ***tɔʔ*** *buboh.*
NEG NEG put onion-white NEG put onion-big NEG put
'No, they don't put garlic or onion.' (KM_180816_cv01_65)

(68) ITM

a. *diyɛ gambi təʊ buwaʔ kɔya tɛh ikĩʔ.*
3 gambier DEM.DIST CAUS hallucinate DEM.DIST little
'Gambier makes you a bit high.' (ITM_180930_cv01_45)

b. ***dɔʔ,*** *diyɛ* ***dɔʔ*** *kaya.*[44]
NEG 3 NEG hallucinate
'No, it doesn't.' (ITM_180930_cv01_47)

The third category are the non-verbal negators, which are *bukɛ* in KM and *bukaŋ* in CTM and ITM. These negators are used to negate a wider range of constituents, including nominal predicates, contrastive verbal predicates, prepositional predicates or an entire clause. (69) and (70) demonstrate the usage of CTM/ITM *bukaŋ* as a nominal negator and a prepositional negator respectively, and (71) to (74) present examples of contrastive negation, whereby the negated proposition is followed by an affirmative alternative (i.e., 'not X, but Y'). In these cases, the constituents negated by *bukɛ/bukaŋ* are not restricted to nominal predicates as in (73), but also verbal predicates in (71) and (72), as well as a prepositional predicate in (74).

(69) CTM
ikaŋ hɔʔ kitə bəli tu ***bukaŋ*** *ikaŋ suŋa.*
fish REL 1PL buy DEM.DIST NEG fish river
'The fish we bought is not river fish.' (CTM_220927_e02_15)

(70) ITM
ikaŋ haʔ kamiŋ juwa təh ***bukaŋ*** *dayɛi d=ai.*
fish REL 1PL.EXCL sell DEM.DIST NEG from LOC=water
'The fish we sell are not from the river.' (ITM_220922_e01_2)

(71) KM
diyɔ ***bukɛ*** *bɔdɔh, malah=yə.*
3 NEG stupid lazy=just
'He's not stupid, but just lazy.' (KM_180831_e01_15)

(72) CTM
dɔʔ, yə ***bukaŋ*** *nn-<t>anaŋ padi, yə n-nɛgə.*
NEG 3 NEG IPFV-plant paddy 3 IPFV-trade
'No, he doesn't plant paddy, he does business.' (CTM_181029_e02_16)

(73) ITM
upɛ-upɛ ***bukaŋ*** *dahaŋ kayəʊ, tandəʊʔ usɛ.*
seemingly NEG branch wood antler deer
'It seems that it was not a tree branch, but the antler of a deer.'
(ITM_180907_n01_27)

[44] *Kaya* in (68b) and *kɔya* in (68a) are variant forms with the same meaning.

(74) ITM
tupɛi ... ***bukaŋ*** *t=taŋaŋ, diyɛ atah ppalɛ.*
hat ... NEG LOC=hand 3SG top head
'The hat is not on the hand, but on the head.' (ITM_220915_e03_16)

Non-verbal negators can also be used to negate an entire clause. In (75), for example, KM *bukɛ* negates *tɔʔ bəli* 'not buy', which is a verbal clause that has already been negated by *tɔʔ*. (76) exemplifies a rhetorical question in which CTM *bukaŋ* negates the complete clause that follows it.

(75) KM
bukɛ *tɔʔ bəli, diyɔ tɔʔsɛ.*
NEG NEG buy 3 not.want
'It's not that (we) don't buy (it for her); she doesn't want it.'
(KM_180820_cv03_226)

(76) CTM
bukaŋ *mɔʔ pileh dɔh ayi tuʔ*
NEG mother choose already day DEM.DIST
'Didn't you choose it already that day?' (CTM_181029_cv01_49)

The fourth type of negators is the prohibitive negator KM *jaŋɛ* or CTM/ITM *jaŋaŋ*, which occurs in clause-initial position, introducing imperative clauses. This is illustrated in (77).

(77) ITM
jaŋaŋ, jaŋaŋ *təʊ lagɛi, dɔʔ basɔh lagɛi tah.*
PROH PROH DEM.DIST yet NEG wet yet DEM.DIST
'Don't, don't (do it like) that, it's not wet yet.' (ITM_180917_cv01_97)

Finally, existential clauses with KM *adɔ*, CTM *adə* or ITM *adɛ* 'EXIST' (see §6.4) are negated with *taʔdɔʔ*, as illustrated in (78) and (79). *Taʔdɔʔ* also means 'not have', and it is used as the negative counterpart of *adɔ/adə/adɛ* as a verb meaning 'have', as shown in (80).

(78) CTM
wi k=kakɔʔ yə, ***taʔdɔʔ*** *ɔɣaŋ makaŋ ɣumɔh yə.*
give to=sister 3 NEG.EXIST person eat house 3
'Give it to her sister, no one eats them at their place.'
(CTM_181022_cv01_37)

(79) ITM
akəʊ puŋ dɔʔ tau, ***taʔdɔʔ*** *dinɛiŋ.*
1SG also NEG know NEG.EXIST here
'I don't know, it's not here.' (ITM_180919_e01_22)

(80) KM
... ***taʔdɔʔ*** *nnagɔʔ gaʔ, ɔbaʔ baɲɔʔ.*
... NEG.EXIST energy EMPH medicine many
'(She) doesn't have energy, (but she needs to eat) a lot of medicine.'
(KM_180820_cv03_217)

NEPMs do not have a negative aspectual marker meaning 'not yet' (cf. SM *bəlum*). The meaning equivalent is expressed by the verbal negator or the existential negator combined with the adverb meaning 'yet' (KM/ITM *lagi* and ITM *lagɛi*), as exemplified in (81) and (82).

(81) KM
toʔ laki, toʔ laki, ***taʔdɔʔ lagi.***
mister husband mister husband NEG.EXIST yet
'Husband, husband, don't have one yet.' (KM_180812_wl01)

(82) ITM
dɔʔ *baŋuŋ* ***lagɛi.***
NEG rise yet
'It hasn't risen yet.' (ITM_180917_cv01_11)

A few other negative verbs are worth mentioning here:

(83) KM *tɔʔsɛ*, CTM *taʔamboh*, ITM *taʔambəʊh* 'not want' (cf. *nɔʔ* 'want')
KM *taʔpɔ*, CTM *taʔpə*, ITM *taʔpɛ* 'no problem' < *tak apa (NEG what)
KM/CTM *taʔyɔh* 'not necessary' < *tak payah (NEG difficult)
KM *tɔʔleh*, CTM *dɔʔleh* 'cannot' < *tɔʔ/dɔʔ buleh* (NEG can)

6.2.10 Prepositions

Prepositional phrases (PPs) consist of prepositions followed by NPs. These prepositions express various semantic relations between the prepositional

phrase and the rest of the clause, including location, instrument, accompaniment, among others. Table 6.12 lists prepositions in NEPMs. Words enclosed between parentheses are rare, and they may be borrowed from SM.

Table 6.12: Prepositions in NEPMs

Meaning	KM	CTM	ITM
'LOC'	*di*	*di*	*də*
'to; for'	*kɔ~kə*	*kə*	*kə*
'AGT'	*kɔ~kə*	*di*	*də*
'from'	*daɣi,* (*daɣipadɔ*)	*daɣi,* (*daɣipadə*)	*daɣɛi,* (*daɣipadɛ*)
'at; from'	*padɔ*	*padə*	*padɛ*
'inside'	*dalɛ*	*dalaŋ*	*dalaŋ~laŋ*
'(on) top'	*atah*	*atah*	*atah*
'below'	*bawɔh~bɔwɔh*	*bɔwɔh*	*bawɔh*
'(in) front'	*dəpɛ*	*dəpaŋ*	*dəpaŋ*
'behind'	*blakɛ*	*blakaŋ*	*blakɔŋ*
'after'	*ləpah~pah*	*ləpah~pah*	*ləpah~pah*
'before'	*bəloŋ*	*sbəloŋ*	(*sbəluŋ*)
'near'	*dəkaʔ*	*dəkaʔ*	*dəkaʔ~kaʔ*
'with'	*dəŋɛ~dəŋa~ŋɛ~ŋa*	*dəŋaŋ~ŋaŋ~ŋə*	*dəŋaŋ~ŋaŋ~ŋə*
'until'	*sapa*	*sapa*	*sapa*
'about'	*pasa*	*pasa*	*pasa*
'like'	*macɛ, supɔ*	*macaŋ*	*macaŋ*

The general locative prepositions KM/CTM *di* and ITM *də* indicate the location of the following NP. They are glossed as 'LOC' and can be translated as 'at; in; on'. *Di* or *də* can be cliticised as *d=* preceding vowel-initial nouns (§5.2.3); furthermore, location can also be marked by the gemination of the initial consonant of the locative noun (§5.3.2.3).

KM *kɔ~kə* and CTM/ITM *kə* cover a large range of functions including indicating movements towards the following NP and introducing recipients or beneficiaries. For example, KM *kɔ* in (84a) indicates the movement towards *gaoŋ* 'canyon', and in (84b), it introduces *miru* 'Amirul (a person name)' as the recipient. In (84c) and (84d), *kɔ* or *kə* serves a more general grammatical function of introducing an oblique argument, glossed as 'PREP'.

(84) KM

a. *jatoh daɣipadɛ cuɣɛ tu, jatoh* [***kɔ*** *gaoŋ*]$_{PP}$.
fall from steep DEM.DIST fall to canyon
'(They) fall from the steep (cliff), and fell into the canyon.'
(KM_180812_n01_25)

b. *aku wi* [***kɔ*** *miru*]$_{PP}$ *buku tu.*
1SG give to Amirul book DEM.DIST
'I gave Amirul that book.' (KM_180825_e01_31)

c. *diyɔ iŋaʔ* [***kɔ*** *ɣatiŋ kayu*]$_{PP}$.
3 think PREP branch wood
'He thought that they were branches.' (KM_180812_n01_19)

d. *tɔʔ kənɛ lasoŋ* [*kə diyɔ*]$_{PP}$ *tu gaʔ.*
NEG remember at.all PREP 3 DEM.DIST EMPH
'(She) doesn't remember him at all.' (KM_180826_cv01_9.2)

Similar functions are attested for CTM/ITM *kə*, as exemplified in (85) and (86). In (85a) and (86a), *kə* expresses the motion in the direction of the following noun (*bɔwɔh* 'bottom; below' and *tuki* 'Turkey'), and in (85b) and (86b), it introduces a recipient. In (86c), *kə* introduces *məʊŋ* '2SG' as the beneficiary. In (85c) and (86d), *kə* is a general preposition. (86d) also shows that *kə* can be cliticised to *k*= preceding vowel-initial words, which is a process similar to the cliticisation of *di/də* → *d*= (§5.2.3).

(85) CTM

a. *yə ambiʔ mɛjə, yə tɔlɔʔ* [***kə*** *bɔwɔh*]$_{PP}$.
3 take table 3 push to bottom
'She grabbed the table and pushed it down (the stairs).'
(CTM_181023_e01_18)

b. *cəʔ wi buŋə* [***kə*** *yə*]$_{PP}$.
IMP give flower to 3
'Give the flowers to him.' (CTM_220927_e02_43)

c. *yə poŋ kabɔ* [***kə*** *aɲjiŋ tu*]$_{PP}$ *suɣuh səɲaʔ.*
3 also tell PREP dog DEM.DIST command quiet
'He asked the dog to be quiet.' (CTM_181023_n02_46)

(86) ITM

a. *ləpah gɛi* [***kə*** *tuki*]$_{PP}$ *nuŋ, gɛi buwaʔ umɣɔh lələ̞ʊ.*
then go to Turkey there go do pilgrimage immediately
'Then we went to Turkey, and we went on a pilgrimage right away.' (ITM_180923_n01_30)

b. *ipah kiɣiŋ salaŋ* [***kə*** *məʊŋ*]$_{PP}$*, ayɔh.*
Ipah send greeting to 2SG father
'Ipah sent greetings to you, dad.' (ITM_180926_cv01_12.1)

c. *pdiɣɛ tibɔ* [***kə*** *məʊŋ*]$_{PP}$*?*
who spread for 2SG
'Who will flip (the dough) for you?' (ITM_180917_cv01_3)

d. *adɛiʔ taku*ʔ [***kə*** *talɛi*]$_{PP}$*, diɣɛ iŋa*ʔ [***k****=ulɔ*]$_{PP}$*.*
younger.sibling afraid PREP rope 3SG think PREP=snake
'The younger kid was afraid of the rope, (because) (s)he thought it was a snake.' (ITM_220915_e03_53)

Another function of KM *kɔ~kə*, CTM *di* and ITM *də* (or their cliticised forms) is to mark agents in passive constructions (see more in §6.4.1). (87) to (89) present some examples of passive constructions in NEPMs, which have the word order of patient–verb–agent. The agents are introduced in prepositional phrases headed by KM *kɔ* in (87), CTM *di* in (88) and ITM *də* in (89).

(87) KM

anɔʔ aku kənɔ ttɛ [***kɔ*** *ceʔgu*]$_{PP}$ *aɣi ni di skɔlɔh.*
child 1SG ADVS slap AGT teacher day DEM.PROX LOC school
'My child was slapped by the teacher at school today.'
(KM_180827_e01_28)

(88) CTM

*abih ikaŋ hɔ*ʔ *bəli p=pasɔ taʔdi makaŋ* [***di*** *kuciŋ*]$_{PP}$*.*
finished fish REL buy LOC=market just.now eat AGT cat
'The fish that (I) bought at the market was eaten by the cat.'
(CTM_180919_e02_52)

(89) ITM

ləpah təʊ, ɑŋjɛiŋ kənɛ kəjɔ [***də*** *tbuwaŋ tɔh*]$_{PP}$*, ...*
after DEM.DIST dog ADVS chase AGT hornet DEM.DIST ...
'Then the dog was chased by the hornets ...' (ITM_180907_n02_22.1)

A number of prepositions have dual or trial word class memberships. The prepositions indicating spatial relations ('inside', 'top', 'bottom', 'front' and 'behind') are essentially nouns, and their prepositional usage emerges out of the nominal sources diachronically. When used as nouns, these spatial terms follow a general preposition, with which they form a PP. In (90a), KM *di* 'LOC' is cliticised to *d=* preceding *atah* 'top', which is nominal. With the omission of the general prepositions, these spatial terms can introduce PPs directly and function as prepositions themselves, as illustrated in (90b).

(90) KM

a. *... diyɔ doʔ* [***d=atah***]$_{PP}$ *dɔh ...*
... 3 sit LOC=top already ...
'He was already sitting on top (of the tree).' (KM_180814_n01_13)

b. *aɲiŋ doʔ* [***atah*** *ppalɔ diyɔ*]$_{PP}$.
dog sit top head 3
'The dog was sitting on his head.' (KM_180812_n01_28)

A few prepositions can be used as conjunctions, including *ləpah~pah* 'after; then' and *sapa* 'until', and words meaning 'with' can be used as conjunctions meaning 'and' (§6.2.11). Additionally, *sapa* is used as a verb meaning 'to reach; to come true', from which the prepositional and conjunction usage ultimately derives. The usage of *sapa* with various meanings is illustrated by ITM examples in (91). *Sapa* is a verb meaning 'to come true' in (91a), whereas in (91b), it may be analysed as either a verb 'to reach' or a preposition 'until'. A clearer prepositional usage of *sapa* is illustrated in (91c), and in (91d), *sapa* is a conjunction connecting two clauses.

(91) ITM

a. *ajaʔ diyɛ nɔʔ jadɛi duʔtu təʊ*
aspiration 3SG want become doctor(ENG) DEM.DIST
sapa *jugɛ.*
come.true also
'His aspiration of becoming a doctor came true.'
(ITM_180921_e01_39)

b. *diyɛ puŋ dudəʊʔ atah tandəʊʔ usɛ təʊ,*
3SG also sit top antler deer DEM.DIST
bɔwɔʔ laɣɛi d=usɛ ***sapa*** *təbɛiŋ.*
bring run AGT=deer reach/until cliff
'He was sitting on the deer's antler, being carried away by the deer reaching the cliff.' (ITM_180907_n01_22)

c. *ambaʔ-ambaʔ,* [***sapa*** *malaŋ*]$_{PP}$ *pulɔʔ, dɔʔ daŋ jugɛ.*
RDP-chase until night also NEG have.time also
'They kept chasing until the evening, but they still couldn't make it.' (ITM_180927_n01_19)

d. *batɔŋ kayəʊ atah papaŋ,* ***sapa*** *landəʊŋ papaŋ təʊ.*
stem wood top plank until sag plank DEM.DIST
'A piece of wood is on the plank ... until the plank sags.' (ITM_180921_e03_35)

6.2.11 Conjunctions

Conjunctions connect words, phrases and clauses. An overview of conjunctions attested in NEPMs is provided in Table 6.13.

Table 6.13: Conjunctions in NEPMs

Meaning	KM	CTM	ITM
'and'	*dəŋɛ~dəŋa~ŋɛ~ŋa*	*dəŋaŋ~ŋaŋ~ŋə*	*dəŋaŋ~ŋaŋ*
'but'	*tapi*	*tapi*	*kadɔŋ,* (*tapɛi*)
'if'	*kalu*	*kalu*	*kaləʊ*
'because'	*səbaʔ~baʔ*	*səbaʔ~baʔ*	*səbaʔ,* (*xxənɛ*)
'so'	*jadi*	*jadi*	(*jadɛi*)
'after; then'	*ləpah~pah*	*ləpah~pah*	*ləpah~pah*
'before'	*bəloŋ*	*sbəloŋ*	(*sbəluŋ*)
'until'	*sapa*	*sapa*	*sapa*
'when'	*masɔ, bilɔ*	*masə, bilə*	*masɛ*

While conjunctions can be traditionally divided into coordinating and subordinating conjunctions (Schachter & Shopen 2007: 45), there is no systematic grammatical distinction between these two types of constructions in

NEPMs. The distinction can still be made on semantic grounds (see Haspelmath 2007: 46–48); for example, the words for 'and', 'or' and 'but' express coordination, whereas 'if', 'when' and 'although' express subordination.

NEPMs do not make extensive use of coordinating conjunctions for the purpose of linking clauses. Subordinating conjunctions such as *kalu/kaləʊ* 'if' and *masɔ/masə/masɛ* 'when' occur more frequently, as illustrated in (92) and (93).

(92) KM
kalu tu gaʔ, buboh blacɛ sikiʔ,
TOP DEM.DIST EMPH put shrimp.paste little

kalu *tɔh boh blacɛ poŋ taʔpɔ dɔh.*
if NEG put shrimp.paste also no.problem already

'In that case, (you can) also add a bit of shrimp paste, but if not it's also fine.' (KM_180820_cv03_80)

(93) ITM
uɣɔŋ dɔʔ ubəʊh majəɣiŋ, uɣɔŋ ubəʊh majəɣiŋ
person NEG put margarine(ENG) person put margarine(ENG)
kdiyaŋ, ***masɛ*** *udɔh ulɛi.*
later when finish knead

'People don't add margarine, they add margarine later, when they finish kneading.' (ITM_180917_cv01_34)

The coordination of two clauses, on the other hand, is typically realised by juxtaposition. In (94), for instance, six clauses describing a series of events are juxtaposed without any conjunctions.

(94) ITM
diyɛ baŋuŋ, caɣɛi dalaŋ kasuʔ, caɣɛi dalaŋ bujəʊŋ,
3 get.up search inside shoe search inside bottle

bukɛ pitəʊ-malɛiŋ, bukɛ pitəʊ-malɛiŋ paŋgi kkatɔʔ ŋə.
open door-thief open door-thief call frog ANAPH

'He got up, searched inside the shoes, searched inside the bottle, opened the window, opened the window and called the frog. ' (ITM_180907_n01_06)

In fact, KM *dəŋɛ~dəŋa* and CTM/ITM *dəŋaŋ* 'and' (and their shortened forms) are restricted to linking words and phrases (primarily NPs), and they do not link clauses. As mentioned in §6.2.10, the same words also function as prepositions meaning 'with'. When they connect two NPs, it is not always clear whether they are conjunctions 'and' or prepositions 'with'. Such an ambiguity is illustrated in (95).

(95) KM
lalu [*s=ɔɣɛ* *llaki*]$_{NP}$ ***dəŋa*** [*s=ɛkɔ* *kambiŋ*]$_{NP}$.
pass one=person male with/and one=CLF goat
'A man with/and a goat passed by.' (KM_180814_n01_10)

Some conjunctions such as KM *ataupoŋ* 'or' and *walaupoŋ* 'although' only occur a few times in the corpora. In view of their unusual trisyllabic shapes, it is likely that these forms are borrowed or calqued from SM (cf. *ataupun* 'or' and *walaupun* 'although').

6.2.12 Discourse particles

Discourse particles are a group of words that serve to express various pragmatic functions such as topicalisation and emphasis. They typically do not serve syntactic functions, and they are often not directly translatable. Tags are included in this category.

NEPMs exhibit considerable variation with regard to the use of discourse particles, and KM in particular utilises a large number of such particles. Table 6.14 lists discourse particles attested in NEPMs.

Table 6.14: Discourse particles in NEPMs

Meaning	KM	CTM	ITM
'FOC; SFP'	*=lah*	*=lah*	*=lah*
'SFP'	*deh*	*deh*	–
'SFP'	*meh*	...	–
'TOP'	*kalu*	...	–
'TAG'	*=kɛ*	*=kaŋ*	*=kaŋ*
'EMPH'	*gaʔ*	...	*gaʔ*
'IMP'	–	*cəʔ*	*cəʔ*

In all three varieties, the enclitic *=lah* is used as both a focus marker and a sentence-final particle for various purposes including softening the tones, emphasis and reassurance. When used as a focus marker, *=lah* can follow various kinds of constituents which are often (but not necessarily) fronted to clause-initial positions, as illustrated in (96) to (98). KM and CTM also use *iyɔ=lah* (3=FOC, with a variant *iyɔlɔh*) as a conventionalised affirmative expression, which can be translated as 'indeed, right'. A similar expression can be found in CTM, which takes the form of *yə=lah* (3=FOC) or *yə ah* (3 INTERJ).

(96) KM
*tu=**lah*** *cɣitɔ diyɔ.*
DEM.DIST=FOC story 3
'That's the story.' (KM_180812_n01_35)

(97) CTM
*tibə-tibə dalaŋ lubaŋ tu, tubeʔ=**lah** s=ɛkɔ tikuh.*
suddenly inside hole DEM.DIST come.out=FOC one=CLF rat
'Suddenly a rat came out of the hole.' (CTM_181023_n02_22)

(98) ITM
*bulɛih=**lah** mɔʔciʔ ŋaŋ pɔʔciʔ gɛi j-jalaŋ təpaʔ-təpaʔ uɣɔŋ.*
can=FOC auntie and uncle go INTR-road RDP-place person
'We (auntie and uncle) could travel to other places.'
(ITM_180923_n01_28)

Some examples of *=lah* used as a sentence-final particle are given in (99) to (101).

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

(100) CTM
*ɣusə tu poŋ tɛŋɔʔ=jə=**lah**.*
deer DEM.DIST also look=just=SFP
'The deer was just watching.' (CTM_181023_n02_43.1)

(101) ITM

*mulɛ tɔh panda, lɛnɛiŋ dɔʔ tau=**lah**.*
beginning DEM.DIST good.at now NEG know=SFP
'I was good at it before, but now I don't know.'
(ITM_180917_cv01_23)

In addition to *=lah*, KM has other discourse particles like *deh* and *meh* which also occur in sentence-final positions, and they are conveniently referred to as sentence-final particles ('SFP'). Some examples are presented in (102). It appears that *deh* is often used to ask confirmation or seek attention, which may also stand on its own, as shown in (102b). The usage of *meh* needs further investigation.

(102) KM

a. *stai mɛʔ diyɔ tumih **deh**?*
style(ENG) mother 3 sauté SFP
'My mom's style is to sauté, right?' (KM_180820_cv03_6)

b. ***deh**, b-bəlɔh ikɛ=lah supɔ ggitu?*
SFP CAUS-split fish=FOC like like.that
'So you half the fish like that?' (KM_180820_cv03_58)

c. *ɔɣɛ təpɔh nɔʔ makɛ ɣaɣɔ ptamɔ,*
person book want eat Eid.al-Fitr first
*ɣayɔ kduwɔ tɔʔleh makɛ lagi **meh**.*
Eid.al-Fitr second cannot eat yet SFP
'People make reservations to eat for the first Eid al-Fitr, but can't even get them on the second Eid al-Fitr day.
(KM_180820_cv03_141)

KM *kalu* 'TOP' occurs in clause-initial position, introducing an NP as the topic in the discourse. It should be distinguished from *kalu* 'if', which is a conjunction that links two clauses (see §6.2.11). *Kalu* as a topic marker can often be translated as 'as for, regarding', as exemplified in (103).

(103) KM

***kalu** ikɛ poŋ, hɔʔ diyɔ bɛsɔ makɛ=jə=lah.*
TOP fish even REL 3 usual eat=just=SFP
'As for fish, only the kinds that she usually eats.'
(KM_180820_cv03_153.2)

Another two discourse particles attested in KM are the tag marker *=kɛ* 'TAG', which can be translated as 'you know; right', and the emphasis marker *gaʔ*, which follows a number of elements including nouns, verbs and conjunctions like *tapi* 'but'. These two particles are illustrated in (104) and (105).

(104) KM
*diyɔ tu=**kɛ**, mugɔ budɔʔ ttinɔ ɲɔ,*
3 DEM.DIST=TAG seemingly kid female ANAPH
*jaŋɔʔ budɔʔ **gaʔ**, nɔʔ pɛkɔŋ s=macɛ.*
pretty kid EMPH want throw same=sort
'That one, you know, it was a girl, a pretty girl, but he still wants to hit her anyway.' (KM_180816_cv01_30)

(105) KM
*tapi **gaʔ** diyɔ hɔʔ masɔ aɣoŋ hɔʔ pətiŋ.*
but EMPH 3 REL when steam REL important
'But the steaming process is important.' (KM_180820_cv03_130)

CTM/ITM *=kaŋ* 'TAG' is the equivalent of KM *=kɛ*. CTM and ITM have one particle that is not attested in KM, namely the imperative marker *cəʔ* 'IMP'. Examples illustrating its usage are given in (106) and (107).

(106) CTM
***cəʔ** aleh jaʔ kkusi, sakoʔ ah.*
IMP move a.while chair hook INTERJ
'Move the chair a bit; something is hooked.' (CTM_220927_e02_72)

(107) ITM
kaiŋ təʊ basɔh,
cloth DEM.DIST wet
***cəʔ** məʊŋ gɛi ŋŋ-<k>əɣɛiŋ kaiŋ təʊ sikĩʔ lagɛi.*
IMP 2SG go IPFV-dry cloth DEM.DIST little more
'That cloth is wet, try to dry it a bit more.' (ITM_180921_e01_6)

6.2.13 Interjections

Interjections are utterances that express various spontaneous emotions or reactions. They often occur on their own and do not serve syntactic functions. Table 6.15 presents common interjections attested in NEPMs. The

phonetic realisations of these interjections exhibit considerable variation, and the rough transcription provided here is not intended to capture their phonetic details.

Table 6.15: Interjections in NEPMs

Expressions	KM	CTM	ITM
emphasis, attention-seeking	*a~ah~hah*	*a~ah~hah*	*a~ah~hah*
contempt, confirmation-seeking	*e~eh~heh*	*e~eh~heh*	*ɛ~ɛh~hɛh*
confusion	*a~ha*	*a~ha*	*a~ha*
astonishment, disappointment	*ɔlɔh*	...	*ɔlɔh*
understanding, realisation	*ɔ~ɔ̃*	*ɔ~ɔʔ*	*ɔ*
disappointment, sympathy	*ado*	*ado*	*adu*
contempt	...	...	*iʃ~uʃ*

Examples illustrating the usage of interjections are given in (108) to (110), each presenting a short conversation between two interlocutors. In (108), *ado* expresses sympathy. In (109), the first speaker gave the instruction *cakaʔ kitə* '(speak) the local dialect', and the second speaker replied in confusion with *ha*, after which the first speaker repeated *cakaʔ kitə*. In (110), *uʃ* shows contempt.

(108) KM

A. *taʔdɔʔ ɔɣɛ kənɛ, katɔ.*
NEG.EXIST person remember say
'Nobody remembers him, she said.'

B. ***ado eh.***
INTERJ INTERJ
'What a pity!' (KM_180816_cv01_12.2–13)

(109) CTM

A. *cakaʔ kitə.*
speech 1PL
'(Speak) the local dialect.'

B. ***ha****ʔ*
huh
'Huh?'

A. *cakaʔ kitə.*
speech 1PL
'(Speak) the local dialect.' (CTM_181022_cv01_18–20)

(110) ITM

A. *dɔʔ, məʊŋ waʔ gwanɛ məʊŋ dɔʔ tibɔʔ*
NEG 2SG make how 2SG NEG flip
'No, how are you making (the roti canai) if you don't flip (the dough)?'

B. *uɣɔŋ waʔ macaŋ lipɛiŋ.*
person make like pancake
'People make it like a pancake.'

A. ***uʃ.***
pfft
'Pfft.' (ITM_180917_cv01_5–7)

6.3 Noun phrases

Noun phrases (NPs) are larger units headed by nouns, which serve the same grammatical functions as nouns. This section outlines the basic NP structure, discussing the constituents that can occur in NPs, their order and possible variations.

Words from a number of word classes discussed in §6.2 can occur in NPs, including demonstratives, quantifiers and numerals, classifiers and possessive pronouns. These constituents typically have the following order in an NP:

(111) quantifier/numeral – classifier – noun – attributive modifier – possessor – relative clause – demonstrative

NPs featuring all possible constituents are extremely rare (if occurring at all) in naturalistic data. Example (112) from KM demonstrates an NP where the head noun *budɔʔ* 'kid' is quantified by a pronominal numeral + a classifier, and followed by a demonstrative. In (113), *bini* 'wife' is modified by a post-nominal possessive pronoun, a relative clause and a demonstrative. Similarly, in the ITM example in (114), the head noun *kkatɔʔ* 'frog' is followed by a relative clause and a demonstrative.

(112) KM
... [s=ɔɣɛ budɔʔ ni]$_{NP}$,
... one=CLF kid DEM.PROX

diyɔ pɛɣɔ kkatɔʔ dalɛ s=buɯɔh bɔtɔ.
3 breed frog inside one=CLF bottle.
'This boy ... he kept a frog in a bottle.' (KM_180812_n01_2)

(113) KM
mugɔ [bini diyɔ [hɔʔ ptamɔ]$_{REL}$ tu]$_{NP}$,
seemingly wife 3 REL first DEM.PROX

diyɔ ggapɔ, uzo.
3 whatchamacallit sick
'It seems that his first wife was sick.' (KM_180820_cv03_103)

(114) ITM
... caɣɛi lɯwɔ umɔh nɔʔ
... search outside house want

[kkatɔʔ [haʔ ilɔŋ taʔdɛi]$_{REL}$ tah]$_{NP}$.
frog REL disappear just.now DEM.DIST
'(He) searched outside the house to look for the frog that went missing.' (ITM_180907_n02_12.2)

The head nouns of NPs may also be omitted or ellipted, resulting in NPs that only consist of a numeral + a classifier, or headless relative clauses (§6.2.6 and §6.2.4.2).

Deviating from the typical constituent order schematised in (111), numeral + classifier combinations may follow head nouns, as illustrated in (115). It appears that when numeral + classifier combinations are placed post-nominally, more emphasis is given to the referents rather than their quantities, but the meaning difference is rather subtle. The two examples in (116) come from the same elicitation session where the consultant was asked to describe pictures. As can be seen, they have parallel constructions except for the order of the numeral + classifier combination with respect to the noun.

(115) KM

... diyɔ buwi=lah [buwɔh pɛ tigɔ bute]$_{NP}$.
... 3 give=FOC fruit pear three CLF

'He gave away three pears.' (KM_180814_n01_37)

(116) ITM

a. *adɛ [kayəʊ apɛi paʔ batɔŋ]$_{NP}$,*
EXIST wood fire four CLF

tigɛ batɔŋ pandɔʔ, sɛ batɔŋ paɲjɔŋ.
three CLF short one CLF long

'There are four sticks of firewood; three sticks are short, one stick is long.' (ITM_180921_e03_17)

b. *adɛ [tigɛ buti batəʊ]$_{NP}$, duwɛ buti kəcĩʔ, sɛ buti bəsɔ.*
EXIST three CLF stone two CLF small one CLF big

'There are three stones; two are small, one is big.'
(ITM_180921_e03_39)

There is no alienability distinction in possessive constructions. The possessor, either a noun, a pronoun or an NP, follows the head noun which is the possessum. Also as mentioned in 6.2.4.1, there are no dedicated possessive pronouns in clitic forms. Some further examples of possessive constructions are given in (117).

(117) ITM

a. *tandəʊʔ usɛ*
antler deer

'deer's antler'

b. *lakɛi məʊŋ*
husband 2SG

'your husband'

c. *anɔʔ uɣɔŋ sbəlɔh umɔh*
child person side house

'neighbour's child'

At the clausal level, NPs may function as arguments of verbs, nominal predicates or complements of prepositions. See more discussion below in §6.4.

6.4 Basic clause structure

This section outlines the structure of simple clauses, which are grammatical units that minimally consist of a predicate. A clause typically also has a subject; the predicate ascribes properties or states something about the subject. For a brief discussion on clause combination, see §6.2.11.

The most common type of predicates contains a verb phrase (VP) headed by a verb. Verbal clauses as such are discussed in §6.4.1. Non-verbal clauses, such as those containing a nominal predicate or a quantifier/numeral predicate, as well as prepositional clauses and existential clauses, are described in §6.4.2.

6.4.1 Verbal clauses

Verbal clauses can be classified along two primary parameters: transitivity and dynamic vs. stative.

Transitivity relates to the number of arguments a verb takes. Verbal clauses can contain an intransitive verb with one argument, a transitive verb with two arguments, or a ditransitive verb with three arguments (which only has limited occurrences, see below). In declarative main clauses, the basic word order for intransitive clauses is subject–verb, and for transitive clauses, it is subject–verb–object. A further distinction can be made between the single argument of an intransitive verb (S), the most agent-like argument of a transitive verb (A) and the most patient-like argument of a transitive verb (P).

Verbal clauses can also be divided into dynamic and stative verbal clauses depending on whether the main verb is a dynamic or a stative verb. The discussion below is organised along this parameter: dynamic verbal clauses are discussed in §6.4.1.1, and stative verbal clauses are discussed in §6.4.1.2.

6.4.1.1 Dynamic verbal clauses

Example (118) presents two simple dynamic verbal clauses from CTM. (118a) is an intransitive clause with the order of SV, and (118b) is an transitive clause with an AVP order.

(118) CTM

a. [*aku*]$_S$ *mandi.*
1SG bathe
'I'm bathing.' (CTM_181029_e01_24)

b. [*aku*]$_A$ *m-mandi* [*anɔʔ aku*]$_P$.
1SG CAUS-bathe child 1SG
'I'm bathing my child.' (CTM_181029_e01_25)

A ditransitive clause has two object arguments, whereby the indirect object argument, which is often a recipient (R), precedes the direct object argument, typically the theme (T). Some examples are presented in (119) and (120), and earlier in (11a). Only a few ditransitive verbs are attested, such as KM/CTM *buwi~wi*, ITM *buwɛi~wɛi* 'to give', KM *aja*, CTM/ITM *ạjɔ* 'to teach', and KM/CTM *tuŋjoʔ*, ITM *tuŋjəʊʔ* 'to show'. Even for these verbs that have the possibility to take three arguments, the recipients of actions like 'to give' are more commonly introduced in PPs following the object argument, as in (121).

(119) ITM
[*uɣɔŋ təʊ*]$_A$ *maɣɛi,* ***wɛi*** [*tuwaŋ umɔh*]$_R$ [*buŋɛ*]$_T$.
person DEM.DIST come give owner house flower
'That person came and gave the host some flowers.'
(ITM_220918_e01_6)

(120) KM
[*asma*]$_A$ ***aja*** [*nɔdiŋ*]$_R$ [*bahasa klatɛ*]$_T$.
Asma teach Nordin language(SM) Kelantan
'Asma teaches Nordin Kelantanese.' (KM_221025_e02_9)

(121) ITM
adɛ duwɛ uɣɔŋ ppuwaŋ,
EXIST two person female
[*hɔʔ b-diɣɛi təʔ*]$_A$ ***wɛi*** [*ai*]$_T$ [*kə hɔʔ dudəʊʔ*]$_R$.
REL INTR-stand DEM.DIST give water to REL sit
'There are two women, the one who's standing gave water to the one who's sitting.' (ITM_220915_e02_1)

While the basic word order in declaration main clauses is SV or AVP, there is some variation. First, the verb in an intransitive clause may be fronted to pre-subject position for the effect of emphasis, with or without the focus marker *=lah*. (122) illustrates a VS order as such, and examples with pre-subject intransitive verbs marked by *=lah* were given earlier in (97) and (98).

(122) KM
*... **lalu** budɔʔjatɛ s=ɔɣɛ dəŋa basika.*
... pass.by kid male one=CLF with bike
'A boy with a bike passed by.' (KM_180814_n01_14)

Second, the order of PVA or PV is attested in passive constructions, where the grammatical subject is the patient of the action denoted by the verb. The agent is introduced in a post-verbal PP headed by KM *kɔ~kə*, CTM *di*, ITM *də* or their cliticised forms, as mentioned in §6.2.10 and illustrated in (87) to (89).

It is worth noting that two types of passive constructions may be distinguished, and they allow different constituent orders. The first type is the adversative passive, which suggests that an action or an event was unpleasant or undesirable, and the patient is negatively affected (Kroeger 2005: 279). In this type of constructions, an auxiliary verb marking adversativity, namely KM *kənɔ*, CTM *kənə* and ITM *kənɛ*, precedes the main verb (cf. SM *kəna*, discussed in Koh 1990: 167; Chung 2005; Nomoto & Kartini 2012). Adversative passives allow the constituent order PV or PVA, meaning that the agents may be unexpressed. Examples demonstrating these two possibilities are presented in (123) and (124).

(123) KM
a. [*ɣumɔh tu*]$_{P}$ *kənɔ bakɔ* [*kɔ jiɣɛ diyɔ*]$_{A}$.
house DEM.DIST ADVS burn AGT neighbour 3
'That house was burnt by his neighbour.' (KM_180825_e01_41)
b. [*ɣumɔh tu*]$_{P}$ *kənɔ bakɔ.*
house DEM.DIST ADVS burn
'That house was burnt.' (KM_180827_e01_24)

(124) CTM

a. *aŋjiŋ tu poŋ layi*
dog DEM.DIST then run

səbaʔ [*yə*]$_{P}$ *kənə kəjɔ* [*di tbuwaŋ tu*]$_{A}$.
because 3 ADVS chase AGT hornet DEM.DIST

'The dog ran away because it was chased by the hornets.'
(CTM_181023_n02_31.2)

b. [*anɔʔ pɔʔ maŋ*]$_{P}$ *kənə igaʔ.*
child uncle Man ADVS catch

'Uncle Man's child was caught.' (CTM_220927_e02_115)

The second type of passive constructions does not have the adversative marker, and they do not necessarily encode the adversative meaning. Unlike the first type of passives where the passive meaning is primarily expressed by the adversative marker, this type of passive is only marked by the word order of PVA, and the agents must be expressed at the syntactic level, introduced by an agent marker. This is illustrated in (125). Note that there is no voice-marking on verbs in any type of clauses; without *də ayɔh ŋə* 'by her father', *tulɔʔ* 'to push' in (125) could have been interpreted as an action that initiated by the girl, rather than affecting the girl.

(125) ITM

budɔʔ ppuwaŋ d-diɣɛi atah mijɛ,
kid female INTR-self top table

[*ŋə*]$_{P}$ *tulɔʔ* [*də ayɔh ŋə*]$_{A}$.
ANAPH push AGT father ANAPH

'A girl was standing on the table, and she was pushed by his father.'
(ITM_180909_e01_44–45)

More importantly, even if the agent of an action is unspecific, a dummy agent meaning 'person' has to occur. In (126) and (127), for instance, there is no specific agent for the actions 'to curse' and 'to take', and the more natural English translations would be passive sentences without overt agents. In NEPMs, however, the agent 'person' needs to be introduced, as in KM *k=ɔɣɛ* and ITM *d=uɣɔŋ* 'AGT=person'.

(126) KM
[*ɲɔ*]$_{P}$ *supɔh* [***k***=***ɔɣɛ***]$_{A}$=*lah*.
ANAPH curse AGT=person=SFP
'It would get cursed (by people).' (KM_180820_cv03_142)

(127) ITM
pɔʔ samaʔ dɔʔ sədɔ dɔh [*buwɔh pɛ*]$_{P}$ *ambɛiʔ* [***d***=***uɣɔŋ***]$_{A}$.
uncle Samat NEG realise already fruit pear take AGT=person
'Pak Samat didn't realise that his pears were taken (by someone).'
(ITM_180927_n02_30)

In addition to the adversative marker, the main verbs in dynamic verbal clauses can be preceded by auxiliary verbs encoding certain aspectual or modal expressions, as well as causativity. Pre-verbal auxiliary verbs attested in NEPMs are listed in Table 6.16. Some examples illustrating the combination of an auxiliary verb + a main verb are presented in (128) to (130). As also indicated in the table, auxiliary verbs often grammaticalised from lexical verbs (or nouns, as in the case of *təŋɔh* 'middle'), and they are still used as such.

Table 6.16: Auxiliary verbs in NEPMs

Meaning	KM	CTM	ITM	
'PROG'	*dudoʔ~doʔ*	*doʔ*	*dudəʊʔ~duʔ*	< 'to sit; to stay'
'PROG'	*təŋɔh*	*təŋɔh*	*təŋɔh*	< 'middle'
'can'	*buleh*	*buleh*	*bulɛih*	< 'to get'
'CAUS'	*waʔ*	*waʔ*	*waʔ*	< 'to do; to make'
'CAUS'	–	–	*wɛi*	< 'to give'
'ADVS; must'	*kənɔ*	*kənə*	*kənɛ*	< 'to suffer; to affect'

(128) KM
diyɔ ***dudoʔ d-diɣi*** *sambe taŋɛ diyɔ* ***doʔ*** ***pəgɛ*** *pəɣoʔ*.
3 PROG INTR-self whilst hand 3 PROG hold belly
'He is standing there with his hands holding his belly.'
(KM_180816_e02_9)

(129) CTM

yə ambeʔ s=bako wɔh pɛ niŋ,
3 take one=basket fruit pear DEM.DIST

yə ***waʔ nai*** *basika diyə.*
3 CAUS go.up bike 3

'He took a basket of pears and put it on his bike.' (Lit. 'cause to go up his bike') (CTM_181025_n02_24.2)

(130) ITM

nɔʔ ***wɛi cai*** *sikĩʔ.*
want CAUS liquid little

'Want to make it a bit more liquid.' (ITM_180917_cv01_18)

Other pre-verbal elements in a verbal predicate include the verbal negators and other negative verbs (§6.2.9). Adverbial expressions may precede or follow the main verb: adverbs denoting frequency precede the verb, but degree adverbs and locative setting adverbs typically follow the verb. Evidential adverbs may be pre-verbal or post-verbal. The variation displayed by the order of verbs and adverbial expressions is demonstrated by KM examples in (131). *Bɛsɔ* 'usual(ly)' and *mɛmɛ* 'really' precede the verb in their clauses, whereas *=jə* 'just', *sajɔ* 'only' and *suŋgoh* 'very much' follow the verb.

(131) KM

a. *kalu ikɛ poŋ, hɔʔ diyɔ* ***bɛsɔ*** *makɛ=****jə****=lah.*
TOP fish even REL 3 usual eat=just=SFP

'As for fish, only the kinds that she usually eats.'
(KM_180820_cv03_153.2)

b. *diyɔ* ***mɛmɛ*** *pɛkoŋ* ***suŋgoh****.*
3 really throw very.much

'He really throws (things) a lot.' (KM_180816_cv01_35)

c. *tapi diyɔ gunɔ ladɔ-ija* ***sajɔ****, ladɔ-sɔlɔʔ.*
but 3 use chilli-green only chilli-solok

'But they use green chilli only, solok chilli.'
(KM_180820_cv03_52)

Lastly, a verbal predicate may consist of several verbs which are strung together tightly to form a serial verb construction (SVC), as illustrated by KM

capo masoʔ 'mix add' in (132), CTM *blaŋgɔ ɣəbɔh* 'crash fall' in (133) and ITM *bbaɣɛiŋ tlətɔŋ* 'lie lie (on the back)' in (134). Furthermore, since semantic adjectives are taken as stative verbs, which may also function as manner adverbs (§6.2.3), the combination of a dynamic verb + a stative verb may be considered a type of SVCs; see examples in (22) to (24).

(132) KM
hɔ̃ diyɔ ***capo masoʔ*** *budu, ggapɔ deh.*
AFF 3 mix add budu what SFP
'Yeah she mixed it with budu (k.o. sauce) or things like that.'
(KM_180820_cv03_194)

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

(134) ITM
uɣɔŋ ***b-baɣɛiŋ tlətɔŋ*** *atah padɔŋ.*
person MID-lie lie.down top field
'A person is lying on his back on the ground.' (ITM_180921_e03_10)

SVCs like these code a single concurrent event with multiple verbs acting together as a single predicate and sharing the same arguments (Aikhenvald 2006: 1). An example of the juxtaposition of verbs which does not constitute a SVC is given in (135), in which the first verb *luɣəʊh* 'to fall' has the argument *buwɔh ɲɔ* 'coconut', whereas the second verb *tiyuʔ* 'to blow' has a different argument *aŋiŋ* 'wind'. (135) is therefore best viewed as biclausal.

(135) ITM
buwɔh ɲɔ luɣəʊh tiyuʔ d=aŋiŋ.
fruit coconut fall blow AGT=wind
'The coconut fell being blown by the wind.' (ITM_180921_e03_15)

6.4.1.2 Stative verbal clauses

Stative verbal clauses have stative verbs as predicates, which are always intransitive. They follow the basic word order of SV, as seen in examples (12)

and (136). The verb may be fronted for emphasis, often with the expression of exclamation, as demonstrated in (137) to (139).

(136) ITM
uɣɔŋ ŋə ***lawɔ,*** *təpaʔ ŋə* ***mulɛiʔ.***
person ANAPH beautiful place ANAPH good
'The people were beautiful and the places were nice.'
(ITM_180923_n01_37)

(137) KM
... ***daɣaʔ*** *ɔɣɛ kusia ni!*
... stupid person Kusial DEM.PROX
'How stupid is this guy from Kusial!' (KM_180820_cv03_19.2)

(138) CTM
hɔ̃ ***mɔlɛʔ*** *ah ni, kəcĩʔ eh.*
AFF good INTERJ DEM.PROX small INTERJ
'Ah this one is good, it's small.' (CTM_181022_cv01_15)

(139) ITM
... ɔ ***kəɣa-kətəʊŋ=lah*** *utɛi-cana məʊŋ!*
... INTERJ hard-thud=FOC roti.canai 2SG
'Ah your roti canai is hard as rock!' (ITM_180917_cv01_27)

Stative verbal clauses share many properties with dynamic intransitive verbal clauses, such as taking the verbal negator (KM *tɔʔ* and CTM/ITM *dɔʔ*). Many adverbs can also occur in both types of clauses. In addition to KM/CTM (*l*)*agi* and ITM *lagɛi* 'again; more; still; yet' (see §6.2.2), *dɔh* 'already' is also compatible with both dynamic and stative verbal clauses.[45] Examples illustrating the usage of *dɔh* 'already' are given in (140) and (141).

(140) KM
a. *... diyɔ* ***ɣukah*** *pɔkɔʔ* ***dɔh*** *...*
... 3 climb tree already ...
'He is already climbing the tree.' (KM_180814_n01_13)

[45] *Dɔh* appears to have another function as a discourse marker that can be translated to 'well', typically occurring on its own or in clause-initial position. See examples (17) and (19) in Appendix B.3.

b. *diyɔ mɛmɛ* ***lamɔ dɔh.***
3 indeed long already
'He's already been like that for a long time.'
(KM_180816_cv01_16)

(141) ITM

a. *mɔʔ makaŋ* ***dɔh*** *siɣɛih hah?*
mother eat already betel INTERJ
'Did you already eat the betel?' (ITM_180930_cv01_3)

b. *kəcĩʔ* ***dɔh.***
small already
'It's already on low.' (ITM_180917_cv01_69)

However, an important difference between dynamic and stative clauses is the more restricted use of auxiliaries in stative verbal clauses. Stative verbal clauses cannot be modified by aspectual or modal auxiliaries, and among the auxiliary verbs listed in Table 6.16, only the causative markers *waʔ* and *wɛi* can cooccur with stative verbs, as shown in (142) to (144).

(142) KM
hɔ̃ diyɔ ɣamah, mɛmɛ diyɔ ***waʔ pəkaʔ.***
AFF 3 squeeze really 3 CAUS thick
'Yeah she squeezed it and really made it very thick.'
(KM_180820_cv03_125)

(143) CTM
adeʔ aku ***waʔ ilaŋ*** *buku aku.*
younger.sibling 1SG CAUS disappeared book 1SG
'My younger brother/sister lost my book.' (CTM_181029_e02_79)

(144) ITM

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

b. ***wɛi luma?*** *lələʊ* *ladɛ* *təʊ* *sbəluŋ nɔ?* *gunɛ* *tah.*
CAUS crushed immediately chilli DEM.DIST before want use DEM.DIST
'Crush the chilli before you use it.' (ITM_180921_e01_4)

6.4.2 Non-verbal clauses

Non-verbal clauses have predicates that are not headed by a verb. NEPMs have several types of non-verbal clauses, which may contain nominal predicates, quantifier predicates or numeral predicates. Locational clauses and existential clauses may be viewed as special types of non-verbal clauses (or semi-verbal clauses, a term that is employed by Donohue 1999). They could contain a verbal predicate, but they also differ from prototypical verbal predicates in that they are not compatible with any auxiliary verbs.

6.4.2.1 Nominal predicates

The most common type of non-verbal clauses has nominal predicates headed by NPs, as already mentioned in the discussion on nouns and illustrated in (4) and (6). Some additional examples are provided in (145) to (147).

(145) KM
tu=lah [*cɣitɔ diyɔ*]$_{NP}$.
DEM.DIST=FOC story 3
'That's the story.' (KM_180812_n01_35)

(146) CTM
diyə [*anɔ? jiɣaŋ* *dəpaŋ ɣumɔh*]$_{NP}$.
3 child neighbour front house
'He is the child of the opposite door neighbour.'
(CTM_181029_e02_18)

(147) ITM
umu akəʊ [*limɛ puləʊh ppa? tauŋ*]$_{NP}$.
age 1SG five tens four year
'I'm 54 years old.' (Lit. 'My age is 54 years.') (ITM_180923_n01_3)

Like verbal predicates, nominal predicates follow the subject. Nominal clauses like these essentially have two nominal constituents that are juxtaposed without copulas. Auxiliary verbs are not allowed in nominal clauses. Adverbs, on the other hand, may occur in nominal clauses, as exemplified in (148) to (150).

(148) KM
Ø [baɲɔʔ kkaɣɔ]$_{NP}$ ***dɔh****=kɛʔ*
Ø much stuff already=TAG
'That's already a lot of stuff, no?' (KM_180820_cv03_256)

(149) CTM
... pi baʔ ciʔ yaŋ [ɔɣaŋ sini]$_{NP}$ ***dɔh,*** *moŋ katə ggitu.*
... but because auntie Yam person here already 2SG tell like.that

'... but since auntie Yam is from here, you tell her that.'
(CTM_181024_cv02_11.2)

(150) ITM
nɛh ***sajɛ*** *[ɛkspɛrimɛŋ]$_{NP}$.*
DEM.PROX just experiment(ENG)
'It's just an experiment.' (ITM_180917_cv01_78)

6.4.2.2 Quantifier/numeral predicates

A second type of non-verbal clauses has quantifiers or numerals as predicates, as shown in (41) to (43), as well as in (151) and (152).

(151) KM
... taʔdɔʔ nnagɔʔ gaʔ, ɔbaʔ [baɲɔʔ]$_{QP}$.
... NEG.EXIST energy EMPH medicine much
'(She) doesn't have energy, (but she needs to eat) a lot of medicine.'
(Lit. 'medicine is a lot') (KM_180820_cv03_217.2)

(152) ITM
taʔdɛi bawɔʔ tigɛ, nɛh [duwɛ]$_{QP}$=jə dɔh.
just.now bring three DEM.PROX two=just already
'I brought three (baskets) just now, but now this is only two.'
(ITM_180927_n02_60)

Quantifier or numeral predicates also follow the subject. This contrasts with quantifiers occurring in NPs, which precede the head noun (see §6.3); compare *ɔbaʔ baɲɔʔ* '(the amount of) medicine is a lot' in (151) with *baɲɔʔ ɔbaʔ* 'a lot of medicine'.

6.4.2.3 Locational clauses

Locational clauses consist of a subject that identifies the person or thing whose location is being described, followed by a locational predicate that may be verbal or prepositional. These clauses are commonly used to answer questions with the locative interrogative 'where'. In fact, the question 'where is X?' itself may be viewed as a specific type of locational clause, in which the predicate may be fronted, as in (153a) and (154). In (155), however, the locative interrogative remains *in situ*.

(153) KM

a. ***m=manɔ*** *kɔtɔ-baɣu?*
LOC=where Kota-Bharu
'Where is Kota Bharu?' (KM_221025_e02_28)

b. *kɔtɔ-baɣu doʔ klatɛ.*
Kota-Bharu sit Kelantan
'Kota Bharu is in Kelantan.' (KM_221025_e02_30)

(154) CTM
m=manə *ayaŋ ... kawaŋ yə ni, dɔʔ juɲpə caɣi agi*
LOC=where chicken ... friend 3 DEM.PROX NEG meet search yet
eh.
INTERJ
'Where is the chicken ... his friends still cannot find him.'
(CTM_181029_n01_33)

(155) ITM
suɣaʔ məʊŋ tulih mmaɣɛiŋ ***dwanɛ?***
letter 2SG write yesterday where
'Where is the letter you wrote yesterday?' (ITM_180921_e01_14)

(153b) exemplifies a locational clause that answers the question in (153a). It has a verbal predicate that is headed by a positional verb *doʔ* 'to sit'. In KM

and CTM, locational clauses typically require the positional verb *dudoʔ~doʔ* (also *udoʔ* in CTM), as further illustrated in (156) and (157). The positional verb introduces the location in space, which could be expressed by a noun, but more commonly by a locative PP, either headed by a preposition or a cliticised geminated segment which serves the same locative function, as in KM *k=kaki* 'on the foot' in (156c) and CTM *t=tali* 'on the rope' in (157c). Another positional verb 'to hang' is illustrated in (157d), but its usage is restricted to its literal meaning which describes the state of 'being hung'. In comparison, *dudoʔ~udoʔ~doʔ* has apparently been grammaticalised and acquired a more general meaning 'to be located'.

(156) KM

a. *tɔpi* ***doʔ*** *atah ppalɔ.*
cap sit top head
'A cap is on the head.' (KM_180829_e01_6)

b. *lapu* ***dudoʔ*** *di siliŋ.*
lamp sit LOC ceiling(ENG)
'A lamp is on the ceiling.' (KM_180829_e01_14)

c. *kasuʔ* ***dudoʔ*** *k=kaki.*
shoe sit LOC=foot
'A shoe is on the foot.' (KM_180829_e01_22)

(157) CTM

a. *ckəlaʔ* ***dudoʔ*** *di mɛjə.*
chewing.gum(ENG) sit LOC table
'Chewing gum is on the table.' (CTM_181029_e01_59)

b. *sudu* ***doʔ*** *bɔwɔh kaiŋ.*
spoon sit below cloth
'A spoon is underneath a cloth.' (CTM_180829_e01_29)

c. *ɲɲ-<s>akoʔ* ***udoʔ*** *t=tali.*
NMLS-hang sit LOC=rope
'Hangers are on the rope.' (CTM_181029_e01_39)

d. *baju koʔ* ***gatoŋ*** *di ɲɲ-<s>akoʔ baju.*
shirt coat(ENG) hang LOC NMLS-hang shirt
'A coat hangs on the cloth hanger.' (CTM_181029_e01_11)

ITM, on the other hand, does not have a grammaticalised positional verb (examples similar to (157d) are attested). Locational clauses typically have prepositional predicates, as illustrated in (158).

(158) ITM

a. *ukuʔ təh də muluʔ uɣɔŋ.*
cigarette DEM.DIST LOC mouth person
'The cigarette is on the mouth.' (ITM_180921_e02_39)

b. *kasuʔ təh k=kakɛi.*
shoe DEM.DIST LOC=foot
'The shoe is on the foot.' (ITM_180921_e02_21)

Additionally, locational clauses in all three varieties of NEPMs may use an existential predicate headed by the existential verb 'there is', as illustrated in (159) to (161). These locational clauses may be alternatively viewed as existential clauses with a location, see §6.4.2.4 below.

(159) KM
buku tu ***adɔ*** *ataḥ mɛjɔ.*
book DEM.DIST EXIST top table
'The book is on the table.' (KM_221025_e02_35)

(160) CTM
kɔtɔʔ ***adə*** *biɣa dindiŋ.*
box EXIST edge wall
'A box is right next to the wall.' (CTM_181023_e01_50)

(161) ITM
planta təʊ ***adɛ*** *də matɛ pisɔ.*
margarine DEM.DIST EXIST LOC eye knife
'Margarine is on the blade.' (ITM_180921_e02_12)

6.4.2.4 Existential clauses

Existential clauses are introduced by existential verbs meaning 'there is' or 'there isn't', namely KM *adɔ*, CTM *adə* and ITM *adɛ* 'EXIST' and their negative counterpart *taʔdɔʔ* 'NEG.EXIST', which historically reflects the combination

of a pre-verbal negator *tak and the affirmative existential verb *ada (final *ʔ* in *taʔdɔʔ* is unexplained).

Two types of existential clauses may be distinguished, one type with a location and one type without. Existential clauses with a location have an existential predicate following the subject which is the person/object whose existence or absence is asserted (the "existant"). Some examples of affirmative existential predicates were presented in (159) to (161) above, and negative existential predicates are illustrated in (162) and (163).

(162) CTM
buku tu ***taʔdɔʔ*** *atah kkusi, atah mɛjə.*
book DEM.DIST NEG.EXIST top chair top table
'The book is not on the chair, but on the table.'
(CTM_220927_e02_53)

(163) ITM
k-kəjuʔ *pagɛi* *isəʊʔ,*
NVOL-startle morning tomorrow
tiŋuʔ tah *kkatɔʔ* ***taʔdɔʔ*** *dalaŋ bujəʊŋ ŋə.*
look DEM.DIST frog NEG.EXIST inside bottle ANAPH
'The next morning, (the boy) was surprised to see that the frog was not in the bottle.' (ITM_180907_n01_4)

In existential clauses without a location, existential verbs are typically clause-initial, followed by the existant. Some examples are given in (164) to (166). (164b) and (166b) illustrate the idiomatic expression of *taʔdɔʔ ɔɣɛ/uɣɔŋ* (also CTM *taʔdɔʔ ɔɣaŋ*) (NEG.EXIST person) 'there is no person', which can often be translated as 'nobody'.

(164) KM
a. ***adɔ*** *ɔɣɛ* *jatɛ* *tuɣoŋ* *xxɛtɔ puteh ...*
EXIST person male go.down car white ...
'There was a man getting off a white car.' (KM_180816_cv01_2)
b. *...* ***taʔdɔʔ*** *ɔɣɛ* *kənɛ,* *katɔ.*
... NEG.EXIST person remember say
'Nobody remembers him, she said.' (KM_180816_cv01_12.2)

(165) CTM

a. *pəʔciʔ dɔʔ sədɔ* ***adə*** *budɔʔ llaki lalu.*
uncle NEG realise EXIST kid male pass.by
'The uncle didn't realise that there was a boy passing by.'
(CTM_181025_n01_22)

b. *ah* ***taʔdɔʔ*** *aŋiŋ pulɔʔ, kənə gi isi dɔh aŋiŋ ah.*
INTERJ NEG.EXIST wind also must go fill already wind INTERJ
'Ah the tyre is flat again (lit. 'there is no wind (in the tyre)'), need to pump it.' (CTM_220927_e02_140)

(166) ITM

a. ***adɛ*** *s=buwɔh umɔh nɛh ...*
EXIST one=CLF house DEM.PROX ...
'There is this house ...' (ITM_180927_n03_1.1)

b. ***taʔdɔʔ*** *uɣɔŋ təgu ŋə kə diyɛ.*
NEG.EXIST person greet ANAPH to 3SG
'Nobody greeted him.' (ITM_180926_cv02_6)

The existential verbs are also used as lexical verbs meaning 'to have' or 'to not have', as illustrated in (167).

(167) ITM

a. *akəʊ* ***adɛ*** *xitɛ, duwɛ buwɔh xitɛ akəʊ, diyɛ katɛ.*
1SG have car two CLF car 1SG 3SG say
'He said: I have cars, two cars.' (ITM_220922_cv03_35.2)

b. *padɔŋ ɣupuʔ taʔ* ***taʔdɔʔ*** *ataʔ,*
field grass DEM.DIST not.have roof
tapɛi təpaʔ dudəʊʔ taʔ diyɛ ***adɛ*** *ataʔ.*
but place sit DEM.DIST 3SG have roof
'The field doesn't have roof, but the sitting area has roof.'
(ITM_220910_cv01_46)

Another extended usage of *adɛ* as a verum focus marker is attested in ITM, which is used before dynamic verbal predicates to emphasise the expression of truth of a proposition (Lohnstein 2016, citing Höhle 1988; Höhle 1992). In the examples in (168), *adɛ* does not have an existential meaning, and its

presence is not necessary for the grammaticality of the sentences. Instead, it serves to enhance the emphasis on the events being described.

(168) ITM

a. ***adɛ*** *balɛiʔ sənɛiŋ ɣasɛ a.*
VF return just.now feel INTERJ
'Seems he did come back just now.' (ITM_180926_cv01_30)

b. *... mmaɣɛiŋ* ***adɛ*** *napɔʔ ŋə s=iku.*
... yesterday VF see ANAPH one=CLF
'Yesterday I did see one.' (ITM_180917_cv01_107.2)

6.5 Summary

This chapter has provided an overview of the word classes and basic syntactic structures of NEPMs, including discussions of the NP structure and basic clausal syntax.

Due to the limited productive morphology, word classes are primarily defined based on their syntactic properties. NEPMs have two major open word classes: nouns and verbs. Nouns typically function as heads in NPs which in turn function as arguments in clauses, whereas verbs function as predicates. Nouns and verbs also differ in their collocation with negators: nouns are negated with the non-verbal negators (KM *bukɛ* and CTM/ITM *bukaŋ*), whereas verbs are primarily negated with the verbal negators (KM *tɔʔ* and CTM/ITM *dɔʔ*). NEPMs do not have a separate word class of adjectives. Semantic adjectives share many morphosyntactic similarities with intransitive verbs, on the basis of which they are subsumed as a subclass of verbs called "stative verbs", as opposed to "dynamic verbs".

Adverbs are considered constituting a closed word class, and there are no distinct manner adverbs. Instead, the functions of manner verbs are fulfilled by stative verbs. Other closed word classes discussed in this chapter include pronouns, demonstratives, quantifiers and numerals, classifiers, prepositions and conjunctions. Interrogatives, negators, discourse particles and interjections may not constitute a unified word class with clear definitions on syntactic grounds, but they were discussed in individual sections for ease of reference and cross-linguistic comparisons.

In NPs, the typical constituent order is as follows: quantifiers, numerals and classifiers precede the head noun, while other attributive modifiers, including both nouns and verbs, as well as possessors, relative clauses and demonstratives follow the head noun. Unlike many other Malayic languages, NEPMs lack distinct possessive pronouns in clitic forms. Additionally, the pronominal systems of KM and CTM deviate from the typical Malayic pronominal system in that they lack the inclusive–exclusive distinction in the first-person plural pronouns.

Clauses in NEPMs can be divided into verbal or non-verbal. Within verbal clauses, dynamic verbal clauses can be intransitive, transitive or ditransitive, whereas stative verbal clauses are always intransitive. Dynamic ditransitive clauses are nevertheless rare. The typical word orders for intransitive and transitive clauses are SV and AVP respectively, but variations such as VS and PVA can also occur. PV(A) word order is typically found in passive constructions. These are formed either with the auxiliary verb expressing adversativity (KM *kənɔ*, CTM *kənə* and ITM *kənɛ*), which may allow for the omission of the post-verbal agent introduced in a PP, or solely with the PVA word order, in which case the agent must be expressed. Notably, transitive verbs in NEPMs do not have voice-marking morphology. Non-verbal clauses may have NPs, quantifiers or numerals as predicates. In addition, locational clauses and existential clauses may be viewed as semi-verbal clauses. KM and CTM locational clauses typically have a verbal predicate headed by a positional verb, whereas ITM locational clauses more commonly have prepositional predicates. Existential clauses may or may not contain a location; when they do, they have a structure similar to locational clauses.

CHAPTER 7

Phonological history

7.1 Introduction

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

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

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

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

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

7.2 Some notes on the reconstruction of PM

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

retentions from PAn/PMP" (Adelaar 1992: 3).

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

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

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

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

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

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

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

(diphthongs: *-aw, *-ay)

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

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

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

7.3 Reflexes of PM consonants

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

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

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

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

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

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

Table 7.4: Overview of reflexes of PM final consonants

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

Table 7.5: Overview of reflexes of PM consonant sequences

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

7.3.1 PM stops

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

Table 7.6: Reflexes of PM initial and intervocalic stops

PM	KM	CTM	Dusun	SM	Gloss
Initial stops					
*pasir	*pase*	*pase*	*pasi*	*pasir*	'sand'
*taŋan	*taŋɛ*	*taŋaŋ*	*taŋaŋ*	*taŋan*	'hand'
*caciŋ	*caciŋ*	*caciŋ*	*cacɛiŋ*	*caciŋ*	'worm'
*kulit	*kuleʔ*	*kuleʔ*	*kuliʔ*	*kulit*	'skin'
*bakar	*baka*	*bakɔ*	*bakɔ*	*bakar*	'to burn'
*dagiŋ	*dagiŋ*	*dagiŋ*	*dagɛiŋ*	*dagiŋ*	'meat'
*jauh	*jaoh*	*jaoh*	*jauh*	*jauh*	'far'
*gigi	*gigi*	*gigi*	*gigɛi*	*gigi*	'tooth'
Intervocalic stops					
*api	*api*	*api*	*apɛi*	*api*	'fire'
*putih	*puteh*	*puteh*	*putɛih*	*putih*	'white'
*pəcah	*pəcɔh*	*pəcɔh*	*pəcɔh*	*pəcah*	'to break'
*sakit	*sakeʔ*	*sakeʔ*	*sakiʔ*	*sakit*	'sick'
*dəbu	*dəbu*	*dəbu*	*dəbəʊ*	*dəbu*	'dust'
*babah	*bawɔh*	*bawɔh*	*bawɔh*	*bawah*	'below'
*hidup	*idoʔ*	*idoʔ*	*iduʔ*	*hidup*	'to live'
*tajəm	*tajɛ*	*tajaŋ*	*tajaŋ*	*tajam*	'sharp'
*tiga	*tigɔ*	*tigə*	*tigɛ*	*tiga*	'three'

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

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

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

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

PM	KM	CTM	Dusun	SM	Gloss
*sayap	*sayaʔ*	*sayaʔ*	*sayaʔ*	*sayap*	'wing'
*hidup	*idoʔ*	*idoʔ*	*iduʔ*	*hidup*	'to live'
*laŋit	*laŋiʔ*	*laŋiʔ*	*laŋiʔ*	*laŋit*	'sky'
*mulut	*muloʔ*	*muloʔ*	*muluʔ*	*mulut*	'mouth'
*tasik	*taseʔ*	*taseʔ*	*tasɛiʔ*	*tasik*	'lake'
*duduk	*dudoʔ*	*dudoʔ*	*dudəʊʔ*	*duduk*	'to sit'

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

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

Table 7.8: Reflexes of PM *-ʔ

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Iban	KM	CTM	Dusun	SM	Gloss
Iban *-ʔ* : KM *-ʔ* : CTM *-ʔ* : Dusun *-ʔ* : SM *-k*					
adiʔ	*adiʔ*	*adiʔ*	*adɛiʔ*	*adik*	'younger sibling'
akaʔ[a]	*kakɔʔ*	*kakɔʔ*	*kakɔʔ*	*kakak*	'older sister'
baliʔ	*baleʔ*	*baleʔ*	*balɛiʔ*	*bali|k*	'to return'
datuʔ[b]	*toʔ*	*toʔ*	*təʊʔ*	*datuk*	'grandfather'
ən|daʔ	*dɔʔ*	*dɔʔ*	*dɔʔ*	*ti|da|k*	'no, not'
iniʔ	*nɛnɛʔ*	*nɛnɛʔ*	*nɛiʔ*	*nɛnɛk*	'grandmother'
pintaʔ	*mitɔʔ*	*mitɔʔ*	*mitɔʔ*	*pintak*[c]	'to request'
Iban *-ʔ* : KM *-ʔ* : CTM *-ʔ* : Dusun *-ʔ* : SM *Ø*					
asiʔ	*nasiʔ*	*nasiʔ*	*nasiʔ*	*nasi*	'cooked rice'
taiʔ	*taiʔ*	*taiʔ*	*taiʔ*	*tahi*	'excrement'
Iban *-ʔ* : KM *Ø* : CTM *Ø* : Dusun *Ø* : SM *Ø*					
antiʔ	*n|nati*	*n|nati*	–	*nanti*	'to wait'
akuʔ	*ŋŋ|aku*	*ŋŋ|aku*	*ŋŋ|akəʊ*	*məŋ-aku*	'to confess'
bəriʔ	*buwi*	*buwi*	*buwɛi*	*bəri*	'to give'
bukaʔ	*bukɔ*	*bukə*	*bukɛ*	*buka*	'to open'
butaʔ	*butɔ*	*butə*	*butɛ*	*buta*	'blind'
ucuʔ	*cucu*	*cucu*	*cucəʊ*	*cucu*	'grandchild'
daguʔ	*dagu*	*dagu*	*dagəʊ*	*dagu*	'chin'
saʔ	*sɔ*	*sə*	*sɛ*	*satu*	'one'
duriʔ	*duɣi*	*duɣi*	*duɣɛi*	*duri*	'thorn'

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

[a] Iban *akaʔ* 'older sibling'.
[b] Iban *datuʔ* 'nobleman, chief'.
[c] Archaic, also *pinta* and *minta* without *-k*.
[d] Also *isiʔ* 'to fill'.
[e] Final *-ŋ* in NEPMs is a later innovation, see §7.3.2.
[f] Might not be a valid cognate set. Correspondence of Iban/SM *-m-* : NEPM *-ŋ-* is unexplained, and Iban *-aʔ* : KM *-a* : CTM *-ə* : Dusun *ɔ* : SM *a* is also irregular. See more discussions in §7.3.2.

In the first set of correspondences, namely Iban *-ʔ* : NEPM *-ʔ* : SM *-k*, all varieties have a final consonant. This is mostly found in words belonging to certain semantic categories: out of the seven examples listed in Table 7.9, four

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

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

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

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

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

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

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

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

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

7.3.2 PM nasals

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

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

Table 7.10: Reflexes of PM initial and intervocalic nasals

PM	KM	CTM	Dusun	SM	Gloss
Initial nasals					
*mata	*matɔ*	*matə*	*matɛ*	*mata*	'eye'
*naiʔ	*naiʔ*	*naiʔ*	*naiʔ*	*naik*	'to go up'
*ɲaŋa(ʔ)	*ŋaŋɔ*	*ŋaŋə*	*ŋaŋɛ*	*ŋaŋa*	'to open wide'
*ɲamuk	*ɲamoʔ*	*ɲamoʔ*	*ɲamuʔ*	*ɲamuk*	'mosquito'
Intervocalic nasals					
*rumah	*ɣumɔh*	*ɣumɔh*	*umɔh*	*rumah*	'house'
*tanah	*tanɔh*	*tanɔh*	*tanɔh*	*tanah*	'earth, soil'
*taɲaʔ	*taɲɔ*	*taɲə*	*taɲɛ*	*taɲa*	'to ask'
*buŋa(ʔ)	*buŋɔ*	*buŋə*	*buŋɛ*	*buŋa*	'flower'

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

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

Table 7.11: Reflexes of PM final nasals

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

[a] PM *pətəŋ 'dark, obscure'.

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

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

Table 7.12: Final velar nasal accretion

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

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

7.3.3 PM fricatives

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

Table 7.13: Reflexes of PM *s

PM	KM	CTM	Dusun	SM	Gloss
*sayap	*saya*ʔ	*saya*ʔ	*saya*ʔ	*sayap*	'wing'
*səmpit	*səpe*ʔ	*səpe*ʔ	*səpi*ʔ	*səmpit*	'narrow'
*asəp	*asa*ʔ	*asa*ʔ	*asa*ʔ	*asap*	'smoke'
*bəsar	*bəsa*	*bəsɔ*	*bəsɔ*	*bəsar*	'big'
*atas	*ata**h***	*ata**h***	*ata**h***	*atas*	'top'
*nipis	*nipi**h***	*nipi**h***	*nipi**h***	*nipis*	'thin'

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

Table 7.14: Reflexes of PM *h

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

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

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

7.3.4 PM liquids

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

Table 7.15: Reflexes of PM liquids

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

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

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

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

In one example, an intervocalic *r is reflected as *w on the surface, and the schwa preceding *-r- changed to *u* accordingly, as in *bəriʔ 'to give' > KM/CTM *buwi*, Dusun *buwɛi*. The history of this word displays a rather irregular development, as by rule intervocalic *r in disyllables is preserved, even when the penultimate vowel was a schwa, e.g., *bəras 'uncooked rice' > NEPM *bəɣah*, *kəriŋ 'dry' > KM/CTM *kəɣiŋ*, Dusun *kəɣɛiŋ*. Yet, parallel developments of *-ər- > *-u-* can be found in disyllables containing a tautosyllabic ⁺-ər- directly preceding another consonant, i.e., ⁺-ər.C- > -u.C-. Examples include *tə**r**bit 'to emerge' > KM/CTM ***tu**beʔ*, Dusun ***tu**biʔ* and *kə**r**baw 'buffalo' > KM/CTM ***ku**ba*, Dusun ***ku**bɔ*.[55] In all likelihood, the development of *bə**r**iʔ 'to give' > ⁺bə**r**i > KM/CTM ***bu**wi*, Dusun ***bu**wɛi* has a similar trajectory with an intermediate stage ⁺bui, which also reflects *-ər- > ⁺-u-. The sound change presumably began with the weakening of the liquid *r, followed by a merger with the preceding schwa. An epenthetic glide in ⁺bui [buwi] was later reinterpreted as phonemic, followed by diphthongisation of ultimate *i in Dusun, hence KM/CTM *buwi* and Dusun *buwɛi* (see §7.4.4.2). What is unusual in the case of *bəriʔ 'to give' is that the sound change *-ər- > ⁺-u- affected a disyllabic word with an intervocalic *-r-, and it only affected this particular word.

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

7.3.5 PM glides

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

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

7.3.5.1 PM intervocalic glides

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

Table 7.16: Reflexes of PM intervocalic glides

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

According to the reconstructed phonotactic constraints, PM intervocalic glides did not occur following corresponding high vowels. Words such as *buah 'fruit', *dua(ʔ) 'two', *ia '3SG' and *tiup 'to blow' were reconstructed with VV sequences, rather than intervocalic phonemic glides (×buwah, ×duwa(ʔ), ×iya or ×tiyup). There are two reasons to reconstruct these constraints in PM. First, materials from the six varieties on which Adelaar's reconstruction was based invariably have a non-phonemic glide in these environments, or at least analysed as so. For instance, *y* in SM "does not occur adjacent to schwa or *i/e*, nor does *w* occur adjacent to schwa or

u/o, although non-phonemic glides are heard" (Adelaar 1992: 11). In Iban, "semivowels do not occur adjacent to schwa or to a vowel of the same colouring (the [y] heard between *u* and *a/u*, and the [w] heard between *u* and *a/i*, are non-phonemic glides)" (Adelaar 1992: 26). Second, glides in similar environments were not reconstructed as phonemic in higher-order proto languages, and VV sequences were favoured. Consider PMP *buaq 'fruit' > PM *buah (PMP *q > PM *h), PMP *duha 'two' > PM *dua(ʔ) (PMP *h > PM *Ø), PMP *ia '3SG' > PM *ia and PMP *tiup 'to blow' > PM *tiup.[56] Both internal and external evidence therefore seem to give grounds for the analysis of non-phonemic glides following corresponding high vowels in PM.

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

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

Table 7.17: Shortening of disyllables with intervocalic glides

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

7.3.5.2 PM final glides

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

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

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

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

Table 7.18: Reflexes of PM vowel-glide sequences

PM	KM	CTM	Dusun	SM	Gloss
*pula**w**	*pula*	*pula*	*pulɔ*	*pula**w***	'island'
*hija**w**	*ija*	*ija*	*ijɔ*	*hija**w***	'green'
*lanta**y**	*lata*	*lata*	*lata*	*lanta**y***	'floor'
*suŋ**ay**	*suŋa*	*suŋa*	*suŋa*	*suŋ**ay***	'river'

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

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

(4) Loss of final approximants in KM

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

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

(5) Loss of final approximants in CTM

*pula**w**	>	*pula*	'island'
*hija**w**	>	*ija*	'green'
*lanta**y**	>	*lata*	'floor'
*suŋ**ay**	>	*suŋa*	'river'
*ak**ar**	>	*akɔ*	'root'
*lay**ar**	>	*layɔ*	'sail'
*ju**al**	>	*juwa*	'to sell'
*tuŋg**al**	>	*tuŋga*	'single'

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

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

*an**a**k	>	*anɔʔ*	'child'
*miŋ**a**k	>	*miŋɔʔ*	'oil'
*kak**a**ʔ	>	*kakɔʔ*	'older sibling'
*dar**a**h	>	*daɣɔh*	'blood'
*pəc**a**h	>	*pəcɔh*	'to break'

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

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

(7) Loss of final approximants in Dusun

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

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

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

*pula**w**	>	*pulɔ*	'island'
*aka**r**	>	*akɔ*	'root'
*ana**k**	>	*anɔʔ*	'child'
*kaka**ʔ**	>	*kakɔʔ*	'older sibling'
*dara**h**	>	*dayɔh*	'blood'
*binta**ŋ**	>	*bitɔŋ*	'star'

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

*lanta**y**	>	*lata*	'floor'
*jua**l**	>	*juwa*	'to sell'
*saya**p**	>	*sayaʔ*	'wing'
*bəra**t**	>	*bəɣaʔ*	'heavy'
*hita**m**	>	*itaŋ*	'black'
*jala**n**	>	*jalaŋ*	'road'
*ata**s**	>	*atah*	'top'

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

they did with other consonants. The development of ultimate *-a- in final *-aC sequences was only sensitive to the place feature of final consonants, and the "special vocalic properties" of glides played no role.

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

(10)

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

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

It must be conceded that there are Malayic varieties in which earlier *-aw and *-ay were monophthongised while other *VC sequences are preserved, indicating that vowel-glide sequences were grouped based on certain properties. In Jakarta Malay, for instance, *-aw and *-ay have been monophthongised to *o* and *e* respectively, as in *pul**o*** 'island' and *ram**e*** 'crowded' (Wallace 1976: 122, cf. SM *pul**aw*** and *ram**ay***), whereas final *C in most other *VC sequences are preserved (except for *-ah > *ɛ*). One could argue that the assumption of vowel-glide sequences being more likely to monophthongise over time has a general phonetic basis, since [aw] and [ay] are indistinguishable from diphthongs [au] and [ai] at the phonetic level, whereas other VC sequences are not diphthong-like phonetically, hence less prone to monophthongise. However, if the inherent tendency to be monophthongised is generalised at the phonetic level, it would imply that the phenomenon is cross-linguistically universal and probably common in other languages and language families. If that is the case, it also makes little sense to label *-aw and *-ay in Malayic and Austronesian languages alone as "diphthongs". These vowel-glide sequences can be simply treated as VC while having their special properties acknowledged. After all, each consonant phoneme is unique, and it is not surprising that phonological

changes are sensitive to different features, therefore targeting groups of consonants in different ways. Labelling vowel-glide sequences as "diphthongs" is unnecessarily confusing. PM *-w and *-y should be considered as nothing more than phonemic glides in final position.

7.3.6 PM consonant sequences

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

Table 7.19: Reflexes of PM consonant sequences

PM	KM	CTM	Dusun	SM	Gloss
*ru**mp**ut	*ɣu**p**uʔ*	*ɣu**p**uʔ*	*u**p**uʔ*	*ru**mp**ut*	'grass'
*bin**t**aŋ	*bitɛ*	*bitaŋ*	*bitɔŋ*	*bin**t**aŋ*	'star'
*ci**ɲc**in	*ci**c**iŋ*	*ci**c**iŋ*	*ci**c**iŋ*	*ci**ɲc**iŋ*	'ring'
*bə**ŋk**ak	*bə**k**ɔʔ*	*bə**k**ɔʔ*	*bə**k**ɔʔ*	*bə**ŋk**ak*	'to swell'
*la**ŋs**uŋ	*la**s**oŋ*	*la**s**oŋ*	*la**s**əʊŋ*	*la**ŋs**uŋ*	'directly'
*ram**b**ut	*ɣam**b**oʔ*	*ɣam**b**oʔ*	*ɣam**b**uʔ*	*ram**b**ut*	'hair'
*din**d**iŋ	*din**d**iŋ*	*din**d**iŋ*	*din**d**ɛiŋ*	*din**d**iŋ*	'wall'
*pa**ɲj**aŋ	*pa**ɲj**ɛ*	*pa**ɲj**aŋ*	*pa**ɲj**ɔŋ*	*pa**ɲj**aŋ*	'long'
*ta**ŋg**aʔ	*ta**ŋg**ɔ*	*ta**ŋg**ə*	*ta**ŋg**ɛ*	*ta**ŋg**a*	'ladder'

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

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

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

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

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

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

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

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

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

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

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

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

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

KM	CTM	Dusun	SM	Gloss
*k**u**ba*	*k**u**bɔ*	*k**u**bɔ*	*kə**r**baw*	'buffalo'
*t**u**beʔ*	*t**u**beʔ*	*t**u**biʔ*	*tə**r**bit*	'to emerge'
*t**u**joŋ, tɣəjoŋ*	*tɣəjoŋ*	*t**u**juŋ*	*tə**r**jun*	'to jump down'
–	*wwəkah*[a]	*b**u**kah*	*bə**r**kas*	'bundle'
bɣənah	*wwənah*	*b**u**nah*	*bə**r**(ə)nas*	'rice ears'
mməciʔ[b]	*ŋŋəciʔ*	*mm**u**ciʔ*	*pə**r**cik~pə**r**cit*	'to squirt'
bɣəsiŋ	*bɣəsiŋ*	*b**u**siŋ*	*bə**r**sin*	'to sneeze'
tɣəbɛ	*tɣəbaŋ*	*t**u**bɔŋ*	*tə**r**(ə)baŋ*	'to fly'

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

Three more sets of correspondences may be added:

1) KM *bɣənɛ* : CTM *bɣənaŋ* : Dusun *b**u**nɔŋ* : SM *bə**r**ənaŋ* 'to swim'
2) KM *b**u**wi* : CTM *b**u**wi* : Dusun *b**u**wɛi* : SM *bə**r**i* 'to give' < PM *bə**r**iʔ
3) KM *p**u**yuʔ* : CTM *p**u**yuʔ* : Dusun *p**u**yəʊʔ* : SM *pə**r**iuk* 'cooking pot' < PMP *pariuk

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

Now let us consider another piece of evidence from the perspective of syllable structure. The uncertainty in the reconstruction of *-r(ə)C-

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

Table 7.22: Reduction of PM trisyllables

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

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

As mentioned earlier, the correspondence between NEPM *-u-* and SM *-ər(ə)-* is not always regular. In some instances, KM and CTM have *-ɣə-* corresponding to Dusun *-u-* and SM *-ər-*, e.g., KM/CTM ***bɣəsiŋ*** : Dusun ***busiŋ*** : SM ***bərsin*** 'to sneeze', and in a few other cases, it also appears that Dusun fails to reflect +-ər- as *-u-*, e.g., Dusun ***cɣəmiŋ*** : SM ***cər(ə)min*** 'mirror'. I suggest that these NEPM forms with unexpected *-ɣəC-* are the results of more recent borrowings, or simply an adaptation of the pronunciation of SM words. There are several reasons for this suggestion. First, *-ɣəC-* is found in some apparent loanwords, e.g., KM *bɣəseh*, Dusun *bɣəsɛih* 'clean' < SM *bərsih* (native terms

are KM/CTM *cuci* and Dusun *cucɛi*), and NEPM *xxətah* 'paper' < ⁺kɣətah < SM *kərtas* < Arabic *qirṭas* (⁺kɣ- > *xx-* is a regular reciprocal assimilation, see §7.5.2.3). Second, *-ɣəC-* also appears in the sound adaptation of toponyms, e.g., Dusun *pɣəlih* < SM *Pərlis* (name of a northern Malay state). Third, variation between *-uC-* and *-ɣəC-* is sometimes attested, as in KM *tujoŋ~tɣəjoŋ* 'to jump down'. Older speakers tend to prefer *tujoŋ*, while younger speakers often use *tɣəjoŋ*. Some Dusun speakers also occasionally pronounce *tubɔŋ* 'to fly' as [tɣəbɔŋ] and *bunɔŋ* 'to swim' as [bɣənɔŋ] (cf. SM *tərbaŋ* and *bərənaŋ*), but when asked for confirmation, they would correct the pronunciations to *tubɔŋ* and *bunɔŋ*. These observations all indicate that *-ɣəC-* has a shallower history, and *-u-* bears the more authentic pronunciation. In essence, there is a two-layer reflexes of PM *-ərC-. The original outcome is *-uC-*, but the sound change of *-ərC- > *-uC-* must have ceased to operate at a relatively early stage. Later borrowings of SM *-ərC-* have been adapted to *-ɣəC-*, which presumably underwent an intermediate stage of ⁺-əɣəC-, e.g., SM *bərsih* 'clean' > ⁺bəɣəsih > KM *bɣəseh*, Dusun *bɣəsɛih* (deletion of antepenultimate schwas). It is not surprising that Dusun attests more forms with *-uC-* whereas KM and CTM often have *-ərC-*, since Dusun is the most conservative variety among NEPMs with least external influence.

Based on the cognate sets discussed above, I reconstruct:

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

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

[62] *pərcit with final *-t instead of *ˣ-k is reconstructed based on Dusun *-iʔ*. *ˣ-ik would have been reflected as *-ɛiʔ in Dusun, see §7.4.4.2.

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

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

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

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

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

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

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

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

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

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

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

7.4 Reflexes of PM vowels

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

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

Table 7.23: Overview of reflexes of PM vowels

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

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

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

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

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

7.4.1 PM penultimate *a and *ə

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

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

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

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

7.4.2 PM penultimate *i and *u

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

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

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

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

7.4.3 PM ultimate *a and *ə

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

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

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

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

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

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

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

PM	KM	CTM	Dusun	SM	Gloss
Open syllables					
*dada	*dadɔ*	*dadə*	*dadɛ*	*dada*	'chest'
*buŋa(ʔ)	*buŋɔ*	*buŋə*	*buŋɛ*	*buŋa*	'flower'
*limaʔ	*limɔ*	*limə*	*limɛ*	*lima*	'five'
Closed syllables					
*anak	*anɔʔ*	*anɔʔ*	*anɔʔ*	*anak*	'child'
*ləmək	*ləmɔʔ*	*ləmɔʔ*	*ləmɔʔ*	*ləmak*	'grease'
*kakaʔ	*kakɔʔ*	*kakɔʔ*	*kakɔʔ*	*kakak*	'older sister'
*pəcah	*pəcɔh*	*pəcɔh*	*pəcɔh*	*pəcah*	'to break'
*akar	*aka*	*akɔ*	*akɔ*	*akar*	'root'
*ulər	*ula*	*ulɔ*	*ulɔ*	*ular*	'snake'
*pulaw	*pula*	*pula*	*pulɔ*	*pulaw*	'island'
*bintaŋ	*bitɛ*	*bitaŋ*	*bitɔŋ*	*bintaŋ*	'star'
*pətəŋ	*pətɛ*	*pətaŋ*	*pətɔŋ*	*pətaŋ*	'afternoon'
*hayam	*ayɛ*	*ayaŋ*	*ayaŋ*	*ayam*	'chicken'
*hitəm	*itɛ*	*itaŋ*	*itaŋ*	*hitam*	'black'
*jalan	*jalɛ*	*jalaŋ*	*jalaŋ*	*jalan*	'road'
*simpən	*sipɛ*	*sipaŋ*	*sipaŋ*	*simpan*	'to save'
*sayap	*sayaʔ*	*sayaʔ*	*sayaʔ*	*sayap*	'wing'
*asəp	*asaʔ*	*asaʔ*	*asaʔ*	*asap*	'smoke'
*surat	*suɣaʔ*	*suɣaʔ*	*suɣaʔ*	*surat*	'letter'
*bulət	*bulaʔ*	*bulaʔ*	*bulaʔ*	*bulat*	'round'
*bəras	*bəɣah*	*bəɣah*	*bəɣah*	*bəras*	'uncooked rice'
*bələs	*balah*	*balah*	*balah*	*balas*	'to reply'
*jual	*juwa*	*juwa*	*juwa*	*jual*	'to sell'
*təbəl	*təba*	*təba*	*təba*	*təbal*	'thick'
*lantay	*lata*	*lata*	*lata*	*lantay*	'floor'

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

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

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

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

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

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

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

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

7.4.4 PM ultimate *i and *u

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

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

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

7.4.4.1 Lowering in KM and CTM

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

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

PM	KM	CTM	SM	Gloss
Retention in open syllables				
*kaki	*kaki*	*kaki*	*kaki*	'foot; leg'
*api	*api*	*api*	*api*	'fire'
Retention in closed syllables				
*nasiʔ	*nasiʔ*	*nasiʔ*	*nasi*	'cooked rice'
*kirim	*kiɣiŋ*	*kiɣiŋ*	*kirim*	'to send'
*aɲin	*aɲiŋ*	*aɲiŋ*	*aɲin*	'wind'
*caciŋ	*caciŋ*	*caciŋ*	*caciŋ*	'worm'
Regular lowering in closed syllables				
*air	*ae*	*ae*	*air*	'water'
*paŋgil	*paŋge*	*paŋge*	*paŋgil*	'to call'
*pilih	*pileh*	*pileh*	*pilih*	'to choose'
Irregular lowering in closed syllables				
*kulit	*kuleʔ*	*kuleʔ*	*kulit*	'skin'
*tasik	*taseʔ*	*taseʔ*	*tasik*	'lake'
*tulis	*tuleh*	*tuleh*	*tulis*	'to write'
*tumit	*tumiʔ*	*tumiʔ*	*tumit*	'heel'
*(b)isik	*b|bisiʔ*	*b|bisiʔ*	*bisik*	'to whisper'
*nipis	*nipih*	*nipih*	*nipis*	'thin'

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

PM	KM	CTM	SM	Gloss
Retention in open syllables				
*batu	*batu*	*batu*	*batu*	'stone'
*malu	*malu*	*malu*	*malu*	'shamed'
Retention in closed syllables				
*kayuʔ	*kayu*	*kayu*	*kayu*	'wood'
Regular lowering in closed syllables				
*alur	*alo*	*alo*	*alur*	'groove'
*tumpul	*tupo*	*tupo*	*tumpul*	'dull'
*bunuh	*bunoh*	*bunoh*	*bunuh*	'to kill'
*jarum	*jaɣoŋ*	*jaɣoŋ*	*jarum*	'needle'
*daun	*daoŋ*	*daoŋ*	*daun*	'leaf'
*buruŋ	*buɣoŋ*	*buɣoŋ*	*buruŋ*	'bird'
Sporadic lowering in closed syllables				
*hidup	*idoʔ*	*idoʔ*	*hidup*	'to live'
*mulut	*muloʔ*	*muloʔ*	*mulut*	'mouth'
*duduk	*dudoʔ*	*dudoʔ*	*duduk*	'to sit'
*ratus	*ɣatoh*	*ɣatoh*	*ratus*	'hundred'
*t|iup	*tiyuʔ*	*tiyuʔ*	*tiup*	'to blow'
*rumput	*ɣupuʔ*	*ɣupuʔ*	*rumput*	'grass'
*gəmuk	*gəmuʔ*	*gəmuʔ*	*gəmuk*	'fat'
*tikus	*tikuh*	*tikuh*	*tikus*	'rat'

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

Given the irregular nature of the lowering of high vowels before certain

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

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

7.4.4.2 Diphthongisation in ITM

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

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

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

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

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

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

(12)

*hat**i**	>	*at**ɛi***	'liver'
*bə**l**i	>	*bə**l**ɛi*	'to buy'
*dur**i**ʔ	>	*du**ɣ**ɛi*	'thorn'
*tas**i**k	>	*tas**ɛ**iʔ*	'lake'
*put**i**h	>	*put**ɛ**ih*	'white'
*cac**i**ŋ	>	*cac**ɛ**iŋ*	'worm'
*bat**u**	>	*bat**ə**ʊ*	'stone'
*kuk**u**	>	*kuk**ə**ʊ*	'nail'
*dag**u**ʔ	>	*dag**ə**ʊ*	'chin'
*dud**u**k	>	*dud**ə**ʊʔ*	'to sit; to stay'
*tuj**u**h	>	*tuj**ə**ʊh*	'seven'
*hid**u**ŋ	>	*id**ə**ʊŋ*	'nose'

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

(13)	*kulit	>	*kuli*ʔ	'skin'
	*kirim	>	*kiɣiŋ*	'to send'
	*ciɲcin	>	*ciciŋ*	'ring'
	*tulis	>	*tulih*	'to write'
	*pasir	>	*pasi*	'sand'
	*paŋgil	>	*paŋgi*	'to call'
	*hidup	>	*idu*ʔ	'to live'
	*mulut	>	*mulu*ʔ	'mouth'
	*jarum	>	*jaɣuŋ*	'needle'
	*turun	>	*tuɣuŋ*	'to go down'
	*ratus	>	*ɣatuh*	'hundred'
	*ikur	>	*iku*	'tail'
	*tumpul	>	*tupu*	'dull'

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

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

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

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

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

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

(15)	*bini	>	*biniŋ*	'wife'
	⁺manik[69]	>	*maniʔ*	'bead'
	*bənih	>	*bənih*	'seed'
	⁺kuniŋ	>	*kuniŋ*	'yellow'
	*laŋit	>	*laŋiʔ*	'sky'
	*aŋin	>	*aŋiŋ*	'wind'
	*manis	>	*manih*	'sweet'
	*pəɲuʔ	>	*pəɲuŋ*	'turtle'
	*ɲamuk	>	*ɲamuʔ*	'mosquito'
	*bunuh	>	*bunuh*	'to kill'
	⁺tənuŋ	>	*tənuŋ*	'to stare'
	*m|inum	>	*minuŋ*	'to drink'
	*tənun	>	*tənuŋ*	'to weave'

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

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

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

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

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

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

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

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

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

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

(16) *kaki > *kakaɪ* 'foot; leg'
*dagiŋ > *dagaɪŋ* 'meat'
*batu > *bataʊ* 'stone'
*kayuʔ > *kayaʊ* 'wood'
*hiduŋ > *idaʊŋ* 'nose'

(17) *tasik > *tasaʔ* 'lake'
*putih > *putah* 'white'
*duduk > *dudaʔ* 'to sit'
*tujuh > *tujah* 'seven'

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

(18) *kulit > *kuleɪʔ* 'skin'
⁺licin > *liceɪŋ* 'smooth'
*pasir > *paseɪ* 'sand'
*hidup > *idoʊʔ* 'to live'
*mulut > *muloʊʔ* 'mouth'
*jarum > *jaɣoʊŋ* 'needle'
*kabus > *kaboʊh* 'fog'
*ikur > *ikoʊ* 'tail'
*tumpul > *tupoʊ* 'dull'

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

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

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

(20)	*jahit	>	*jaiʔ*	'to sew'
	*main	>	*maiŋ*	'to play'
	*air	>	*ai*	'water'
	*laut	>	*lauʔ*	'sea'
	*daun	>	*dauŋ*	'leaf'

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

(21)	*tiup	>	*tiyoʊʔ*	'to blow'
	*cium	>	*ciyoʊŋ*	'to sniff'

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

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

(22)	*bin**i**	>	*bineɪŋ*	'wife'
	*bən**i**h	>	*bənеɪh*	'seed'
	⁺kun**i**ŋ	>	*kuneɪŋ*	'yellow'
	*laŋ**i**t	>	*laŋeɪʔ*	'sky'
	*aŋ**i**n	>	*aŋeɪŋ*	'wind'
	*man**i**s	>	*maneɪh*	'sweet'
	*ɲam**u**k	>	*ɲamoʊʔ*	'mosquito'
	*bun**u**h	>	*bunoʊh*	'to kill'
	*tən**u**n	>	*tənoʊŋ*	'to weave'
	*m|in**u**m	>	*minoʊŋ*	'to drink'

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

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

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

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

The two diphthongs ⁺aɪ and ⁺aʊ subsequently underwent monophthongisation to *a* with offglide deletion, but only ⁺aɪ and ⁺aʊ before *-k and *-h were affected (see example 16). Simultaneously, VV sequences *-a.i- and *-a.u- before *Ø or back non-approximants were affected by similar changes, resulting in the deletion of ultimate high vowels (see example 19). At this point, the reflexes of ultimate *i and *u in PK should resemble what is presented

in Table 7.33. *i and *u following nasal onsets and/or preceding other codas were still retained as monophthongs.

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

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

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

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

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

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

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

7.4.4.3 Nasalisation

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

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

PM	KM	CTM	Dusun	SM	Gloss
*kəc**i**l/*kəc**i**k	*kəcɛ̃ʔ*	*kəcĩʔ*	*kəcĩʔ*	*kəc**i**l*	'small'
PMP *kaw**i**l	*kaɛ̃*	*kaĩ*	*kaĩ*	*kai**l***	'fishhook'
*bus**u**k	*busũʔ*	*busũʔ*	*busũʔ*	*bus**u**k*	'rotten'

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

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

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

7.5 Syllable reduction

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

7.5.1 Syllable reduction in PM disyllables

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

Table 7.36: Syllable reduction in PM disyllables

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

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

7.5.2 Syllable reduction in PM trisyllables

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

Table 7.37: Syllable reduction in earlier trisyllables

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

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

7.5.2.1 Vowel contraction

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

Table 7.38: Vowel contraction in earlier trisyllables

PM	KM	CTM	Dusun	SM	Gloss
*bi**a**wak	*bɛwɔʔ*	*bɛwɔʔ*	–	*bi**a**wak*	'monitor lizard'
–	*bɛsɔ*	*bɛsə*	*byasɛ*	*bi**a**sa*[a]	'usual'
*ma-**i**rah	*mɛɣɔh*	*mɛɣɔh*	*mi**ɣ**ɔh*	*mɛrah*	'red'
*b**a**ik-i[b]	*bɛʔk|i*	*bɛʔk|i*	*bi**ʔ**k|ɛi*	*b**a**ik-i*	'to repair'
*bu**h**aya	*bɔyɔ*	*bɔyə*	*bwayɛ*	*bu**a**ya*	'crocodile'
–	*sɔɣɔ*	*sɔɣə*	*swaɣɛ*	*su**a**ra*[c]	'voice'
...	*kɔli*	*kuali*	*kwalɛi*	*ku**a**li*	'wok'
–	*pɔsɔ*	*pɔsə*	*pwasɛ*	*pu**a**sa*[d]	'to fast'

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

7.5.2.2 Antepenultimate vowel syncope

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

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

Table 7.39: Antepenultimate vowel syncope in earlier trisyllables

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

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

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

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

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

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

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

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

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

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

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

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

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

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

7.5.2.3 Cluster assimilation

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

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

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

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

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

Table 7.41: Syllable reduction and cluster assimilation in earlier trisyllables

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

[a] Dusun *maʔkaɣəʊŋ* is unexplained.

as noted in §7.5.1.

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

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

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

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

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

In a few other instances, the assimilation of earlier non-geminate clusters appears to be progressive; compare KM *ssəmɔ*, CTM *ssəmə* 'all' with SM *səmua*. KM *ssəmɔ* and CTM *ssəmə* probably had an earlier ⁺sm- cluster, which has been progressively assimilated to *ss-*. Also compare KM *ssəjeʔ*, CTM *mməjeʔ*, Dusun *ssəjiʔ* 'mosque' with SM *maṣjid*, which ultimately comes from Arabic *maṣjid*. The origin of this loanword had a non-native medial *-ṣj-* sequence, which was presumably broken up by a schwa when

borrowed into NEPMs, i.e., +masəjid. The reduction of this trisyllable resulted in an initial +ms- cluster, which was regressively assimilated to *ss-* in KM and Dusun, but progressively to *mm-* in CTM.

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

Table 7.43: Initial geminates in some animal names

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

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

7.5.3 Interim summary

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

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

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

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

7.6 Relative chronology of sound changes

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

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

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

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

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

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

At first glance, it appears that the common changes attested in all three varieties, namely (4)–(7) and (9)–(10), are shared innovations that might be reconstructable to an immediate ancestral language, namely Proto NEPM. However, the chronological order of sound changes shows that changes in final consonants (4)–(7) must have been preceded by changes of ultimate vowels, which are distinct in each variety. There is no single vowel change that affected all three varieties in the exact same way in this position, ex-

cept that (2) is shared by KM and CTM. Even though the sound change (1) ultimate *a, *ə > ⁺a > ɔ is recurring, the environment in which the rounding took place differed across NEPMs (see §7.4.3). In other words, it is not possible that (4)–(7) took place in a common ancestral language; instead, they must have recurred independently or resulted from diffusion, after ultimate vowel changes had already affected NEPMs in distinct ways. The only sound changes that might be reconstructed to an earlier stage are (9) and (10), as well as (2) for KM and CTM, but all these changes are fairly common in Malayic varieties in general, therefore offering little value in subgrouping.

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

7.7 Summary

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

Some general trends and important observations in the phonological history of NEPMs are summarised as follows. In the consonant system, changes were primarily observed in word-final position. Final stops have merged to *-ʔ*, and final nasals have been neutralised to *-ŋ*. Fricatives merged to *-h*, and all approximants were lost. Word-medial consonant sequences comprising a nasal + a homorganic voiceless stop or *s were reduced to their obstruent components. Overall, the development of consonants from PM to NEPMs exhibits a trend of reduction. In contrast, the development of vowels may be viewed as a matter of complexification. All three varieties have acquired more vowels compared to their ancestral language, as a

result of the lowering of earlier high vowels, the raising of low vowels, and the nasalisation of oral vowels in certain environments. While Dusun (and ITM in general) retains a conservative feature of preserving penultimate high vowels (Anderbeck in print), ultimate high vowels underwent diphthongisation, with complex conditions determined by the presence and the nasality of the onset, as well as the presence and the place feature of the coda. At the syllable level, some PM disyllables have been reduced to monosyllables as the result of losing initial *(h)ə, and earlier trisyllables have been reduced to disyllables through vowel contraction or antepenultimate vowel syncope. The reduced disyllables have word-initial consonant clusters, some of which were further affected by cluster assimilation, resulting in geminate clusters.

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

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

CHAPTER 8

Morphological history

8.1 Introduction

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

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

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

8.2 Affixes reconstructed in PM

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

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

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

Table 8.1: Affixes reconstructed in PM

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

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

Needless to say, PM was far from being isolating. In comparison, KM and CTM only have five prefixes, namely *bɣ-* 'INTR; MID', *pɣ-* 'CAUS; FCT', *tɣ-* 'NVOL', *NN₁-* 'IPFV' and *NN₂-* 'NMLS' (§5.3.1). ITM has an even smaller inventory with four prefixes as it lacks the causative/factitive prefix. In addition to prefixation, however, NEPMs utilise the process of initial gemination to realise certain grammatical functions (§5.3.2). The general evolution of the morphology from PM to NEPMs can thus be characterised as a process of reduction with some traits of innovation.

8.3 Morphological retention and innovation

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

(1)	*(mb)Ar-	'INTR'	>	*bɣ-*	'INTR; MID'
	*pAr_1-	'TR'	>	*pɣ-*	'CAUS; FCT'
	*tAr-	'NVOL'	>	*tɣ-*	'NVOL'
	*mAN-	'AGT; INTR'	>	*NN₁-*	'IPFV'
	*pAN-	'NMLS'	>	*NN₂-*	'NMLS'

Except for *mAN- 'AGT; INTR' > *NN₁-* 'IPFV', other prefixes generally retain the original meanings and functions, despite some analytical differences. For instance, I treat NEPM *bɣ-* as a middle (voice) marker when attached to verbal bases (§5.3.1.2), whereas this function was subsumed as part of the intransitive marker for PM *(mb)Ar-. Similarly, PM *pAr_1- was considered a prefix forming transitive verbs, but a distinction was made in NEPM *pɣ-* between a causative marker when prefixed to verbal bases and a factitive marker when prefixed to nominal bases (§5.3.1.4). The reason behind the semantic shift from *mAN- 'AGT; INTR' to *NN₁-* 'IPFV' is unclear, and it may be considered an innovation. However, as pointed out in §5.3.1.5, aspectual functions of cognates to *NN₁-* appear to have a wider distribution in Malayic

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

Formally, the changes reflected in the phonological forms of these prefixes follow regular sound changes. In §7.5, I demonstrated that PM trisyllables underwent syllable reduction and became disyllables in NEPMs, commonly through antepenultimate vowel syncope and subsequent cluster assimilation. Importantly, these sound changes affected both simple words and prefixed derivatives in the same way. Since the canonical shape of PM roots was disyllabic, prefixes typically fell on the antepenultimate syllables, which were the targets of syllable reduction. The phonological evolution of these prefixes retained from PM is described in the following sections. I first consider *(mb)Ar-, *pAr$_1$- and *tAr- in §8.3.1. *mAN- and *pAN- are treated in §8.3.2.

8.3.1 PM *(mb)Ar-, *pAr$_1$- and *tAr-

To illustrate the sound changes reflected in PM *(mb)Ar-, *pAr$_1$- and *tAr-, some examples are given in (2) to (4).

(2) PM > KM

***(mb)Ar**-anak	>	***bɣ**-*anɔʔ*	(INTR-child)	'to give birth'
***(mb)Ar**-layar	>	***b**-*laya*	(INTR-sail)	'to sail'
***(mb)Ar**-lari	>	***b**-*laɣi*	(MID-run)	'to run'
***pAr$_1$**-habis	>	***pɣ**-*abih*	(CAUS-finished)	'to finish'
***pAr$_1$**-hati	>	***pɣ**-*ati*	(FCT-liver)	'to observe'
$^+$**pAr$_1$**-lumat	>	***p**-*lumaʔ*	(CAUS-crushed)	'to crush'
$^+$**tAr**-iŋat	>	***tɣ**-*iŋaʔ*	(NVOL-think)	'to remember'
***tAr**-bakar	>	***t**-*baka*	(NVOL-burn)	'to be burnt'

(3) PM > CTM

***(mb)Ar**-anak	>	***bɣ**-*anɔʔ*	(INTR-child)	'to give birth'
***(mb)Ar**-lari	>	***b**-*laɣi*	(MID-run)	'to run'
***pAr$_1$**-habis	>	***pɣ**-*abih*	(CAUS-finished)	'to finish'
$^+$**pAr$_1$**-hancur	>	***pɣ**-*aco*	(CAUS-crushed)	'to crush'
***pAr$_1$**-hati	>	***pɣ**-*ati*	(FCT-liver)	'to observe'
$^+$**tAr**-iŋat	>	***tɣ**-*iŋaʔ*	(NVOL-think)	'to remember'
***tAr**-bakar	>	***t**-*bakɔ*	(NVOL-burn)	'to be burnt'

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

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

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

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

(5) PM > KM/CTM

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

(6) PM > ITM

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

In other cases, the prefix would first form a non-geminate cluster with the base-initial consonant. This non-geminate cluster, like other non-geminate clusters resulting from the reduction of morphologically simple trisyllables, was subject to further cluster assimilation. As discussed in §7.5.2.3, clusters violating the SSP were typically assimilated regressively to become geminates. For prefixed forms with *(mb)Ar-, *pAr$_1$- or *tAr-, it is expected that *b-*, *p-* and *t-* were assimilated to base-initial segments that were equally or less sonorous. The function of these original prefixes is thus realised by an initial geminated segment on the surface. This development of PM prefixes is illustrated in (7) to (9).

(7) PM > KM/CTM

***(mb)Ar**-jalan	>	***j**-jalɛ*	(INTR-road)	'to walk'
***(mb)Ar**-jəmur	>	***j**-jəmo*	(MID-jəmo)	'to sunbathe'
***pAr$_1$**-kəriŋ	>	***k**-kəɣiŋ*	(CAUS-dry)	'to dry s.th.'
***pAr$_1$**-tidur	>	***t**-tido*	(CAUS-sleep)	'to put s.o. to sleep'
$^+$**tAr**-kəjut	>	***k**-kəjuʔ*	(NVOL-startle)	'to be startled'
$^+$**tAr**-saŋkut	>	***s**-sakoʔ*	(NVOL-hang)	'to be hung'

(8) PM > CTM

***(mb)Ar**-diri	>	***d**-diɣi*	(INTR-self)	'to stand'
***(mb)Ar**-jəmur	>	***j**-jəmo*	(MID-jəmo)	'to sunbathe'
$^+$**pAr$_1$**-kuat	>	***k**-kuwaʔ*	(CAUS-strong)	'to strengthen'
***pAr$_1$**-tidur	>	***t**-tido*	(CAUS-sleep)	'to put s.o. to sleep'
$^+$**tAr**-kəjut	>	***k**-kəjuʔ*	(NVOL-startle)	'to be startled'
$^+$**tAr**-saŋkut	>	***s**-sakoʔ*	(NVOL-hang)	'to be hung'

(9) PM > ITM

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

As a result, following regular sound changes, PM *(mb)Ar-, *pAr$_1$- and *tAr- are expected to have three sets of reflexes under different phonological conditions: they are reflected as *bɣ-*, *pɣ-* and *tɣ-* before bases with initial vowels or *h, as *b-*, *p-* and *t-* before bases with a more sonorous initial consonant,

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

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

$$*(\text{mb})\text{Ar-}, *\text{pAr}_1\text{-}, *\text{tAr-} > \begin{cases} b\gamma\text{-}, p\gamma\text{-}, t\gamma\text{-} & / __*(\text{h})\text{V} \\ b\text{-}, p\text{-}, t\text{-} & / __*\text{C with higher sonority} \\ \text{base-initial C} & / __\text{other } *\text{C} \end{cases}$$

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

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

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

(11) KM

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

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

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

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

8.3.2 PM *mAN- and *pAN-

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

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

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

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

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

(12) PM > KM

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

(13) PM > CTM

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

(14) PM > ITM

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

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

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

The retention of *mAN- and *pAN- is less clear in other environments. As mentioned in §5.3.1.1, *NN₁-* 'IPFV' and *NN₂-* 'NMLS' are generally not attested before bases with an initial voiced obstruent, an initial liquid or an initial nasal, except in a few instances. PM *mAN- appears to be reflected as *m-* before a base with an initial liquid, namely KM *m-lamboŋ* (IPFV-bump) 'bumping' < ⁺mAN-lambuŋ (⁺mAlambuŋ). In another example, an initial *p-* may be seen as the reflex of *pAN-, namely NEPM *p-lawɔʔ* (NMLS-lie) 'liar, the habit of lying' < ⁺pAN-lawak (cf. SM *pəlawak*). In view of these two examples, it is possible to argue that *mAN- and *pAN- are reflected as *m-* and *p-* before bases with an initial liquid. If this holds true, it would necessitate an expansion of the synchronic inventory of NEPM affixes. However, given the limited number of instances, these *m-* and *p-* cannot be explained satisfactorily for now. It is also likely that these forms are nonce borrowings.

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

Consider the anticipated development of *mAN- before bases with an initial voiced obstruent in ITM. Since *N was realised as a homorganic nasal without nasal substitution in this environment, PM forms prefixed with *mAN- should have taken on a $*C_1VC_2.C_3V(C).(C)V(C)$ structure with a $*\text{-}C_2.C_3\text{-}$ cluster consisting of a nasal + a voiced obstruent, e.g., *mAN-bəli 'ACT-buy' → *mAmbəli. Following regular sound changes, *mAmbəli is expected to be reflected as *mməlɛi* in ITM with the deletion of *b in the *-mb- cluster;[78] that is, *mAN- should have been reflected as *NN-* triggering nasal substitution before voiced obstruents. Nevertheless, this pattern is not attested. ITM forms like *b-bəlɛi* 'to go shopping' and *j-juwa* 'to trade' presumably derive via initial gemination, although the precise source of

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

this imperfective marker is not entirely clear.[79]

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

(15) PM > KM/CTM

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

(16) PM > ITM

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

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

(17) PM > KM/CTM

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

(18) PM > ITM

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

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

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

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

8.4 Morphological reduction

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

8.4.1 Fossilisation of PM *-i and *-an

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

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

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

(19) PM > KM

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

(20) PM > CTM

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

(21) PM > ITM

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

8.4.2 Loss of other affixes

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

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

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

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

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

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

*(mb)Ar-	'INTR'	*-i	'APPL'	*kA- -an_1	'NMLS'
*pAr_1-	'TR'	*-aʔ	'SUBJ'	*kA- -an_2	'NMLS'
*tAr-	'NVOL'	*-an_1	'DISTR'	*pAN- -an	'NMLS'
*mAN-	'ACT; INTR'	*-an_2	'NMLS'	*pAr- -an	'NMLS'
*maka-	'TR.CAUS'	*-An	'NMLS'		
*pAN-	'NMLS'				
*pAr_2-	'NMLS'				

As described earlier, NEPMs only have reflexes of *(mb)Ar-, *pAr_1-, *tAr-, *mAN- and *pAN-. A striking pattern can be revealed from a closer examination of the retention and loss of affixes: prefixes tend to be retained, whereas all suffixes and circumfixes are lost, except the few instances of fossilised *-i and *-an.[86] Out of seven prefixes reconstructed to PM, only *maka- 'TR.CAUS' and *pAr_2- 'NMLS' are not inherited. Moreover, *pAr_1- 'TR' is also not inherited in ITM. The divergent histories between prefixes on the one hand and suffixes/circumfixes on the other hand suggest that the morphological reduction was not random. Furthermore, it should be emphasised again that derivatives in NEPMs, either derived synchronically or historically, have a

[86] It is worth noting that the generalisation is at odds with some previous reports. Ras (1970: 439–411) takes note of a productive nominaliser *-ɛ* in KM as a continuation of *-an_2 'NMLS', and a few examples of an applicative suffix *-kɛ* which corresponds to SM *-kan* and reflects a PM preposition *akAn. In my KM corpus, however, no instances of *-kɛ* or *akɛ* are attested. Their equivalent *-kaŋ* or *akaŋ* is also not found in CTM or ITM. However, as Ras himself points out, some of these suffixed formed cited in his study might have derived from the written standard language. Abdul Hamid (1994) dedicates a whole chapter to describing various affixes in KM, including nominalising affixes *pə-*, *pə- -ɛ̃* and *kə- -ɛ̃*, which supposedly reflect PM *pAr_2-/*pAN-, *pAN- -an/*pAr- -an and *kA- -an_2. These affixes are also not attested in my data. My analysis of NEPM morphological history nevertheless accords with Collins' early documentation of ITM morphology (1983: 52–55), which only includes three prefixes, roughly transcribed as *NN-*, *bəɣ-* and *tə-*, and no suffixes.

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

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

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

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

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

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

8.4.3 Contact-induced change?

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

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

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

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

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

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

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

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

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

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

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

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

Table 8.3: NEPM basic lexical items without SM cognates

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

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

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

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

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

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

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

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

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

8.5 Summary

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

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

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

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

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

CHAPTER 9

General discussion and conclusions

9.1 Overview

In this dissertation, I have examined the Malayic varieties spoken in Kelantan and Terengganu (NEPMs) with two goals: first, to provide sketch grammars of NEPMs (presented in Chapters 2 to 6), and second, to investigate NEPMs' historical development (discussed in Chapters 7 and 8). In this concluding chapter, I summarise and synthesise the main findings in §9.2. Furthermore, I explore the implications of the linguistic data for the Malayic migration history in §9.3 by placing the linguistic findings against the non-linguistic background summarised in §1.4.1. In §9.4, I acknowledge some limitations in the present study and point at some directions for future research.

9.2 Summary of main findings

In the phonology, all three varieties of NEPMs share a consonant inventory comprised of twenty native phonemic consonants. These include nine stops /p, b, t, d, c, j, k, g, ʔ/ (among which /t/ is dental [t̪], /c/ and /j/ are phonetic-

ally affricates), four nasals /m, n, ɲ, ŋ/, three fricatives /s, x, h/, two liquids /l, ɣ/ and two glides /w, y/. This consonant inventory is typical of Malayic languages, except for the voiceless velar fricative /x/, which originated from the reciprocal assimilation of the two segments making up earlier $^{+}$kɣ- clusters. As a result, /x/ only occurs word-initially, primarily in the geminated form /xx-/. Geminate consonants, which are analysed as clusters consisting of two identical segments, stand out as a remarkable feature in the consonant systems of NEPMs. Almost all consonants, except for glottals and glides, can appear in their geminated form and are exclusively attested in word-initial position. CTM also has a geminated bilabial glide /ww-/, resulting from the reciprocal assimilation within earlier $^{+}$bɣ- clusters (thus parallel to $^{+}$kɣ- > /xx-/). These geminate clusters are characterised by a longer duration of articulation, and they can often be contrasted with their singleton counterparts in (near-)minimal pairs. They occasionally display variation with non-geminate clusters, from which they arose diachronically. Moreover, geminate clusters can occur in both morphologically simple and complex words. In the latter case, they can often be analysed as having derived from the morphophonological process of initial gemination.

The vowel systems of NEPMs differ more significantly. KM and CTM make a four-way distinction with regard to vowel height, each having eight oral vowels /a, i, u, ə, ɛ, ɔ, e, o/. Additionally, KM has four nasal vowels /ã, ũ, ɛ̃, ɔ̃/, and CTM has five nasal vowels /ã, ĩ, ũ, ɛ̃, ɔ̃/. ITM, as represented by the subvariety spoken in Kampung Dusun, lacks the set of mid-high vowels, thus having only six oral vowels /a, i, u, ə, ɛ, ɔ/. It also features four nasal vowels /ã, ĩ, ũ, ɔ̃/. ITM further differs from KM and CTM in having two phonemic diphthongs /ɛi, əʊ/, and its vowel phonemes generally display more complex allophonic variation, conditioned by the presence and the nasality of the onset, as well as the presence of the coda.

NEPMs share similar syllable and word structures, yet they vary in their phonotactic constraints on phoneme distributions, particularly concerning vowel distribution. The canonical syllable template is (C)(C)V(C). Words typically comprise two syllables following a (C)(C)V(C).(C)V(C) pattern. With the exception of the glottal stop, all consonant phonemes can occupy the onset position of a syllable, among which the velar fricative /x/ only occurs as geminated /xx/ in KM and CTM. In the coda position, only nasals and glottals are allowed. The syllable template further reveals that consonant clusters only occur in the onset position, which may consist of

two identical or two dissimilar segments. Non-geminate clusters usually comply with the SSP. ITM allows the largest range of possible combinations in consonant clusters, but non-geminate clusters that violate the SSP tend to be assimilated regressively to become geminates. This tendency is also observable in some clusters that conform to the SSP.

At the word level, syllables with a complex onset typically occur initially. Exceptions are found in a few trisyllabic derivatives, where non-geminate clusters occur word-medially after a prefix *bə-* 'INTR; MID'. Consonant sequences formed across two syllables commonly consist of a nasal + a homorganic voiced obstruent, but sequences of /-ʔ.C-/ are also attested. In word-final position, only three consonants /ʔ, ŋ, h/ are allowed. NEPMs thus exhibit a strong left-oriented consonantal asymmetry in their syllable and word structures (Ogloblin 2018: 330). As for vowels, ultimate syllables allow for more vowels and more phonemic contrasts, which may be taken as an indicator of their greater prominence. In KM and CTM, the mid-high vowels /e, o/ are only permitted in ultimate syllables. In ITM, both the mid vowels and the diphthongs exclusively occur in this position. Nasal vowels, which carry a relatively low functional load, are also solely attested in ultimate syllables. Another noteworthy generalisation about the word structure in NEPMs is that both morphologically simple words and derivatives share a similar word shape, following the same maximal CCVC.CVC template, and the phonotactic constraints apply to both word types.

In the morphological system, NEPMs have a small inventory of affixes with an overwhelming preference for prefixing. KM and CTM have five prefixes, namely *bɣ-* 'INTR; MID', *tɣ-* 'NVOL', *pɣ-* 'CAUS; FCT', *NN$_1$-* 'IPFV' and *NN$_2$-* 'NMLS'. ITM has one prefix less as it lacks the causative/factitive marker. Overall, NEPMs display a largely isolating-analytic profile. The prefixation processes are further limited by the phonological constraints on permissible word-initial clusters. The full forms of *bɣ-* 'INTR; MID', *tɣ-* 'NVOL' and *pɣ-* 'CAUS; FCT' only emerge before vowel-initial bases. They are reduced to *b-, t-, p-* respectively before bases with a single initial consonant, with which they combine to form a consonant cluster. Bases with initial complex clusters, on the other hand, generally do not undergo prefixation. Since the clusters resulting from prefixation need to be phonologically well-formed, the reduced single-segment prefix mainly appears before bases with an equally or more sonorous initial consonant. In other cases, the prefix assimilates to the base-initial consonant, creating morphologically complex word-initial

geminates. I have demonstrated that while these complex initial geminates may arise from the assimilation of an underlying prefix to a base-initial consonant, this unified analysis based on allomorphic alternations cannot comprehensively account for all occurrences of complex geminates. Therefore, a morphophonological process of initial gemination must be acknowledged. Initial gemination can also operate as a locative marker, which is best treated as a special type of preposition cliticisation. The nasal prefixes NN_1- 'IPFV' and NN_2- 'NMLS' are underlyingly geminated, and they have a restricted distribution. They only attach to bases with an initial vowel or *ɣ*-, appearing as *ŋŋ*-, or to bases with an initial voiceless obstruent, wherein the underspecified nasal element takes the place of articulation of the base-initial segment and substitutes it.

In additional to prefixation and initial gemination, compounding and reduplication are other word-formation processes in NEPMs. The majority of compounds are attributive compounds, constituted by a noun + a noun or a noun + a stative verb. Reduplication is a productive process that can be applied to bases from various word classes, performing a variety of grammatical functions. Formally, reduplication may be considered a special type of compounding, as it is restricted to full reduplication and a few instances of echo reduplication, which produce complex words with two roots.

As far as word classes are concerned, NEPMs have two open word classes, namely nouns and verbs, which are primarily differentiated based on their syntactic properties rather than morphological features. NEPMs lack a distinct category of adjectives. Semantic adjectives are subsumed as a subclass of verbs, referred to as stative verbs, on account of their shared grammatical properties with dynamic verbs. Stative verbs may also function as manner adverbs without any formal changes; thus, adverbs are best viewed as a closed word class with a limited number of members. Other word categories that may occur in an NP include quantifiers, numerals and classifiers, which as a rule precede the head noun. On the other hand, possessive pronouns, relative clauses and demonstratives follow the head noun within an NP.

Clauses in NEPMs can be classified into verbal and non-verbal ones. Verbal clauses can be further categorised as dynamic and stative, or as intransitive, transitive and ditransitive clauses based on the verb's valency. Notably, NEPM verbs do not have active/passive voice-marking morphology. In transitive verbal clauses, the differentiation between active and

passive constructions is primarily indicated by word order. An active transitive clause has the basic word order of AVP, whereas the a passive construction is marked by the order of PV(A), often combined with an adversative auxiliary. The agent is introduced by a preposition as an adjunct, and when the adversative auxiliary is absent, the expression of an agent is compulsory.

Overall, NEPMs exhibit more similarities in their morphosyntactic structures, but there are also some distinct features that set ITM apart from KM and CTM. For example, while ITM distinguishes exclusivity in the first-person plural pronouns as many other Malayic varieties do, such as distinction is neutralised in KM and CTM. ITM also stands out in having a reduced set of demonstratives and the prevalent usage of an anaphoric pronoun *ŋə*. Non-verbal clauses with prepositional predicates are typically attested in ITM, but not in KM and CTM.

The synchronic descriptions form the foundation for investigating the historical development of NEPMs. By comparing the phonological and morphological systems of NEPMs with existing PM reconstructions, the diachronic changes that have taken place in the history of NEPMs are established. The evolution from PM to NEPMs can primarily be characterised as showing a tendency towards reduction, both in phonology and morphology. Word-final consonants underwent various mergers and losses: final stops merged to /ʔ/, final nasals merged to /ŋ/, *-s underwent lenition and merged with *-h, and all final approximants were eliminated. Additionally, word-medial consonant sequences consisting of a nasal + a homorganic voiceless obstruent were reduced to the obstruent component.

The most prominent change in NEPMs is the process of syllable reduction, which affected both disyllables and trisyllables. As a consequence of losing initial *(h)ə, some PM disyllables have become monosyllabic. More significantly, trisyllables, encompassing morphologically simple words and prefixed derivatives, have been reduced to disyllables through vowel contraction or antepenultimate vowel syncope. These shortened disyllables often have initial clusters, some of which have undergone cluster assimilation, leading to the formation of geminate clusters.

These changes in the consonant system are shared by all three NEPM varieties (with some minor differences in the processes involving syllable reduction), which gives the impression that NEPMs resemble each other to a great extent. However, the development of the vowel systems reveals signi-

ficant divergence, which again distinguishes ITM from KM and CTM. In certain aspects, there is a trend of change towards complexification rather than reduction, as NEPMs have acquired more vowel phonemes, including phonemic nasal vowels. KM and CTM underwent the lowering of high vowels in both penultimate and ultimate syllables, whereas ITM has retained high vowels in the penultimate syllables and diphthongised high vowels in ultimate syllables. Some subvarieties of ITM have undergone a two-stage diphthongisation process with particularly complex conditions. While there are some common vowel changes across NEPM varieties, such as the raising and rounding of ultimate *a, the environments in which such changes occurred are not identical.

The establishment of a relative chronology of sound changes demonstrates that the seemingly shared development in the consonant systems must have been preceded by distinct changes in the vowel system of the individual languages. That is to say, the common changes in the consonant systems cannot be attributed to shared innovations in a single ancestral language, but more likely have emerged as the result of diffusion.

The morphological history of NEPMs reveals an even clearer pattern of reduction. Out of more than sixteen affixes that could be reconstructed to PM, only five are preserved in KM and CTM, and merely four in ITM. PM prefixes are typically retained, whereas all suffixes and circumfixes have either been lost or have become fossilised in a few instances. Since original prefixes mostly fell on the antepenultimate syllables which were the target of syllable reduction, the retained prefixes are phonologically reduced. Some prefixes have transformed into geminated segments, and such contrasts between a base form with an initial singleton and a corresponding complex form with an initial geminate have served as the basis for generalising initial gemination as a morphophonological operation, presumably through analogy. Other prefixes have only been retained in a restricted set of phonological environments.

Furthermore, I have showed that the fossilisation of certain suffixes was not random, and the loss of other suffixes and circumfixes was not a wholesale process. Both processes can be understood as being influenced by the structural pressure of maintaining a canonical disyllabic structure. If a disyllabic target had been achieved through other means of reduction, or if the original suffixed forms were semantically disassociated from their bases, suffixes tended to become fossilised. The observation that the reduction of

morphology followed a specific pattern suggests that internal phonological changes were the impetus behind the morphological reduction in NEPMs.

Lastly, I discussed the potential role of language contact and substratal influences in morphological reduction, and examined probable indications of external interference in various aspects of NEPM grammars. Clear traces of contact-induced change are not found in the lexicon, phonology or morphology of NEPMs. Therefore, there is little evidence supporting a contact scenario, and it is not immediately obvious that external influences played a significant role in motivating the morphological reduction.

To sum up, the perception of NEPMs as some of the most aberrant Malayic varieties, whether expressed by laypeople or earlier general studies on Malayic varieties, is not unfounded. This study has demonstrated that NEPMs exhibit many distinct features, both from Malayic and cross-linguistic perspectives. While many general characteristics are common to all three NEPM varieties, KM and CTM appear to be more closely related to each other, while ITM stands out as more distinctive.

9.3 Implications for the Malayic migration history

It is generally presumed that the dispersal of Malayic languages followed a route via Sumatra from the homeland in West Borneo before reaching the Malay Peninsula (see a summary in §1.2). As the ancestors of Malayic-speaking populations on the peninsula arrived from overseas, they likely settled initially along the coasts and river mouths before spreading inland along the riverine systems (§1.4.1.1). Importantly, previous studies also suggest that the Malayic varieties spoken on the east coast of the peninsula form a distinct subgroup within Malayic (§1.4.2). Assuming that the pattern of language dispersal reflects the prehistoric movements of the ancestors of those who speak them (cf. linguistic migration theory, Sapir 1968 [1916]; Dyen 1956), it can be inferred from the existing literature that the ancestors of NEPM speakers migrated into the region in a single group, presumably from Sumatra.

However, in light of the findings from the current study, a revision and adjustment of the aforementioned scenario is required. It appears improbable that the ancestors of NEPM speakers reached the Malay Peninsula as a singular group. Instead, I propose a two-wave migration pattern based on

the present-day distribution and differentiation of NEPMs.

Several important points should be highlighted here. First, despite the superficial similarities across NEPMs, there is a fair amount of variation among the three varieties. In §9.2, I drew attention to the similarities between KM and CTM on the one hand, and the distinctiveness of ITM on the other hand. The closer relationship between KM and CTM is evident in various aspects of the grammar, including phonology, morphology and certain syntactic structures. ITM stands out with both conservative and innovative features that are not attested in KM or CTM (e.g., the retention of penultimate high vowels and the diphthongisation of ultimate high vowels). Second, ITM displays an exceptionally high degree of internal variation (see Collins 1983a), which suggests more significant differentiation *in situ* for a longer period. Third, it is also noteworthy that ITM shares some of its peculiarities with non-Peninsular Malayic varieties such as Kerinci, the most prominent one being the diphthongisation of earlier ultimate high vowels (see early remarks in van Reijn 1974). All these observations imply that ITM is the older variety among NEPMs, and it may have closer relationships with other non-Peninsular Malayic varieties. Moreover, as I argued in §7.6, NEPMs do not form a discrete subgroup within Malayic, and they do not have an immediate common ancestor. This indicates that the ancestors of NEPM-speakers did not migrate to the region as a unified group.

Summing up the linguistic evidence, I suggest that ITM represents an older variety, presumably an earlier offshoot of PM, whose speakers settled inland in Terengganu at an early stage. KM and CTM, on the other hand, reflect a migration of more recent origin, and their speakers spread primarily along the coastline before advancing inland into Kelantan.

Similar remarks were previously made by Mohd Tarmizi (2018a), who contends that the inland Malayic varieties of the east coast of the peninsula (Kelantan, Terengganu and Pahang) have a longer history compared to the coastal varieties. He further suggests that the east coast varieties originated in the hinterland and then dispersed downstream towards the coast. However, this downstream migration pattern is unconvincing for two reasons. First, the direction of migrating from the inland to the coasts contradicts the commonly accepted pattern of Malayic settlements, which typically began at river mouths. Second, in the case of Terengganu, there is no evidence indicating that CTM represents an offshoot of ITM as a result of the

speakers' migration. The two varieties spoken in Terengganu do not form a dialect continuum, but are separated by a number of clear isoglosses (see Collins 1983a). The geographic and political proximity between inland Terengganu and coastal Terengganu does not suggest a closer linguistic relationship between CTM and ITM (against KM).

The migration scenario proposed here is also supported by non-linguistic data, as historical records demonstrate early settlements in the inland areas of the east coast. Chinese records documented the existence of several inland kingdoms, such as Chi'tu around the sixth century, which was likely situated in inland Kelantan, and Fo-lo-an around the twelfth century, possibly located in inland Terengganu (§1.4.1.2). According to *Sui-shu*, it took a month's journey to reach the capital of Chi'tu from the coast (Wheatley 1973: 36), suggesting that the kingdom was situated in the hinterland. Moreover, the establishment of human settlements in the inland must have begun thousands of years ago before the arrival of the Malays (and possibly pre-Malayic Austronesian speaking people), likely driven by the abundant mineral resources such as gold and tin on the peninsula's interior (Benjamin 1987; Bellwood 1993). The discovery of the Terengganu inscription stone further underscores the historical significance of the Hulu Terengganu region. Sheppard (1949: 3–4) postulated that a (probably Buddhist) Malay kingdom might have already existed in inland Terengganu in the eighth century, predating the foundation of Malacca by several centuries. The founding legend of the Patani Kingdom, on the other hand, suggests that its ancestors arrived from overseas, probably in the early fifteen century (§1.4.1.2).

This timeline is consistent with the two-wave migration pattern. It is plausible that ITM descends from a group of Malayic-speaking people who had already settled in the inland area during the first millennium, though the exact date cannot be determined. Subsequently, the ruler of the inland kingdom converted to Islam and erected the Terengganu inscription stone in the early fourteenth century. The ancestors of KM and CTM speakers (as well as Patani Malay speakers) arrived in the region during a later period, either from other parts of the Malay Peninsula or possibly from Sumatra. These people primarily inhabited the coasts and established kingdoms in the nearby regions before eventually moving inland.

Based on the historical scenario proposed above, it remains unclear how and why the languages spoken by these groups of people underwent sub-

stantial phonological changes, which further drove morphological reduction. It is plausible that the ancestors of NEPM speakers came into contact with pre-Malayic speaking people who were already present on the Malay Peninsula before the arrival of the Malays. However, as discussed in §8.4.3, there are no clear indications in present-day NEPMs that speak in favour of such a contact scenario. Further comprehensive research is required to thoroughly investigate this aspect.

9.4 Limitations and directions for future research

Given the limited time and data available, compounded by the negative impact of the Covid-19 pandemic on field trips, it is important to acknowledge that the present study has several limitations. There are also a number of unresolved issues and questions that require further research in the future.

First, more data, especially additional naturalistic conversational data, would be beneficial to supplement the corpora and synchronic description. The data for the present study were collected during two field trips. However, due to the requirement of visiting at least three field sites for collecting data from three varieties, the limited duration of stay at each site has resulted in a reduced amount of data collected for each variety. A fair amount of data used in this study was obtained through elicitation, which is arguably less representative of the natural language use. However, efforts were made to minimise the influence of the intermediate language (SM) during data collection by employing non-linguistic stimuli and prompting informants to generate sentences rather than translate them from SM. Additionally, a more optimal approach would also involve recording data from a larger number of speakers, so that inter-speaker variations can be taken into account.

Second, a few issues in the synchronic description have been glossed over or left untreated for future studies. The stress patterns of NEPMs, for instance, have only been described in a mostly impressionistic manner, and the acoustic correlates of stress have not been examined in detail. This aspect of research requires future phonetic study. Another noteworthy phenomenon that has not been explored in the present study is the tendency to shorten many disyllabic words to monosyllables in connected speech. This type of syllable reduction generally deletes the first syllable or only the

first consonantal segment, which is quite different from the historical processes of reducing disyllables to monosyllables discussed previously. This phenomenon appears to be particularly prevalent in CTM. Unfortunately, there is a limited amount of naturalistic data available for this variety. In addition to common reductions such as *buwɔh* → *wɔh* 'fruit', *buwi* → *wi* 'to give' and *buwaʔ* → *waʔ* 'to do; CAUS' (see Table 7.17 in Chapter 7), other examples are provided below in (1). Both full forms and shortened forms are attested, but typically only the full forms were given by the speakers in word list elicitation, which may be considered citation forms. It is important to note that syllable reduction as such is not limited to function words like prepositions and conjunctions, but also applies to content words, as seen in *pitih* → *itih* 'money', *ɣumɔh* → *umɔh~mɔh* 'house' and *buboh* → *uboh~boh* 'to put'. Further investigation is needed to uncover the patterns of this process and the potential factors that motivate it.

(1) CTM

buleh	→	*uleh~leh*	'can; be able to'
dəŋaŋ	→	*ŋaŋ*	'with; and'
ləpah	→	*pah*	'after; then'
tapi	→	*pi*	'but'
səbaʔ	→	*baʔ*	'because'
taʔdɔʔ	→	*aʔdɔʔ*	'NEG.EXIST'
taʔdi	→	*aʔdi*	'just now'
saŋaʔ	→	*ŋaʔ*	'very'
dulu	→	*lu*	'first'
lagi	→	*agi*	'again; more; still; yet'
pitih	→	*itih*	'money'
buboh	→	*uboh~boh*	'to put'
makaŋ	→	*akaŋ*	'to eat'
katə	→	*atə*	'to say'
ɣumɔh	→	*umɔh~mɔh*	'house'

Third, in examining of the history of NEPMs, this dissertation has focused on establishing the changes that have taken place from PM to NEPMs and comparing the three varieties against each other. In order to gain a fuller understanding of the dispersal of NEPMs and the migration history of their speakers, it is necessary to examine the relationship between NEPMs and other

Peninsular Malayic varieties, as well as non-Peninsular Malayic varieties. Before achieving this goal, however, more descriptive data must be collected. This dissertation has taken one step forward in advancing Malayic descriptive and historical linguistics, but further research is required to achieve a detailed internal subgrouping of the Malayic languages.

Lastly, in order to comprehensively study the prehistory of the Malay Peninsula, it is essential to consider all language groups present in the region and their relationships. This includes investigating language contact between NEPMs and neighbouring Aslian languages in both present and past contexts, as well as potential contact between Malayic and Mon-Khmer languages. Towards this end, further research with a more specific focus on uncovering traces of language contact is warranted. Moreover, it is crucial to combine linguistic data with insights from history, archaeology, ethnography and genetic studies. An interdisciplinary approach will provide a more comprehensive understanding of the languages in the region and the migration history of their speakers.

APPENDIX A

Basic word lists

No.	Meaning	KM	CTM	Dusun	SM
1.	hand	*taŋɛ*	*taŋaŋ*	*taŋaŋ*	*taŋan*
2.	leg/foot	*kaki*	*kaki*	*kakɛi*	*kaki*
3.	head	*ppalɔ*	*ppalə*	*ppalɛ*	*kəlapa*
4.	eye	*matɔ*	*matə*	*matɛ*	*mata*
5.	ear	*tliŋɔ~lliŋɔ*	*lliŋə*	*tliŋɛ*	*təliŋa*
6.	nose	*idoŋ*	*idoŋ*	*idəʊŋ*	*hiduŋ*
7.	mouth	*muloʔ*	*muloʔ*	*muluʔ*	*mulut*
8.	tooth	*gigi*	*gigi*	*gigɛi*	*gigi*
9.	tongue	*lidɔh*	*lidɔh*	*lidɔh*	*lidah*
10.	hair	*ɣamboʔ*	*ɣambuʔ*	*ɣambuʔ*	*rambut*
11.	neck	*təkɔʔ*	*təkɔʔ*	*lahi~lihi*	*leher*
12.	shoulder	*bau*	*bau*	*bau*	*bahu*
13.	chest	*dadɔ*	*dadə*	*dadɛ*	*dada*
14.	breast	*kɔpɛʔ*	*kɔpɛʔ*	*pɛiʔ*	*buah dada*
15.	belly	*pəɣoʔ*	*pəɣoʔ*	*pəɣuʔ*	*pərut*
16.	back	*blakɛ*	*blakaŋ*	*blakɔŋ*	*bəlakaŋ*

No.	Meaning	KM	CTM	Dusun	SM
17.	buttocks	*puŋgoŋ*	*puŋgoŋ*	*puŋgəʊŋ, jubu taiʔ*	*puŋguŋ*
18.	knee	*lutuʔ*	*lutuʔ*	*lutuʔ*	*lutut*
19.	skin	*kuleʔ*	*kuleʔ*	*kuliʔ*	*kulit*
20.	guts	*usuh, isi pəɣoʔ*	*usuh, tali pəɣoʔ*	*pəɣuʔ liliʔ*	*usus, isi pərut*
21.	liver	*ati*	*ati*	*atɛi*	*hati*
22.	heart	*jatoŋ*	*jatoŋ*	*jatəʊŋ*	*jantuŋ*
23.	bone	*tulɛ*	*tulaŋ*	*tulɔŋ*	*tulaŋ*
24.	blood	*daɣɔh*	*daɣɔh*	*daɣɔh*	*darah*
25.	meat, flesh	*dagiŋ*	*dagiŋ*	*dagɛiŋ*	*dagiŋ*
26.	to breath	*ɲɲawɔ, nnapah*	*ɲɲawə, nnapah*	*ɲɲawɛ*	*bər-nafas*
27.	to vomit	*mutɔh*	*mutɔh*	*mutɔh*	*muntah*
28.	to spit	*ludɔh*	*ludɔh*	*luwɔh*	*ludah*
29.	to bite	*kəkɔh*	*xxəkɔh*	*kəkɔh*	*gigit*
30.	to yawn	*ŋŋuwaʔ*	*ŋŋuwaʔ*	*ŋŋuwaʔ*	*məŋ-<k>uap*
31.	to sleep	*tido*	*tido*	*tidu*	*tidur*
32.	to dream	*mmipi*	*mmipi*	*mmipɛi*	*bər-mimpi*
33.	to kill	*bunoh*	*bunoh*	*bunuh*	*bunuh*
34.	to die	*mati*	*mati*	*matɛi*	*mati*
35.	to live	*idoʔ*	*idoʔ*	*iduʔ*	*hidup*
36.	to scratch	*gaɣu*	*gaɣu*	*gaɣəʊ*	*garu*
37.	to swell	*bəkɔʔ*	*bəkɔʔ*	*bəkɔʔ*	*bəŋkak*
38.	ill, sick	*sakeʔ*	*sakeʔ*	*sakiʔ*	*sakit*
39.	feather	*bulu*	*bulu*	*buləʊ*	*bulu*
40.	wing	*sayaʔ*	*sayaʔ*	*sayaʔ*	*sayap*
41.	horn	*tandoʔ*	*tandoʔ*	*tandəʊʔ*	*tanduk*
42.	tail	*ɛkɔ*	*ɛkɔ*	*iku*	*ekor*
43.	sky	*laŋiʔ*	*laŋiʔ*	*laŋiʔ*	*laŋit*
44.	moon	*bulɛ*	*bulaŋ*	*bulaŋ*	*bulan*
45.	star	*bitɛ*	*bitaŋ*	*bitɔŋ*	*bintaŋ*
46.	cloud	*awɛ*	*awaŋ*	*awaŋ*	*awan*
47.	fog	*kabuh*	*kabuh*	*kabuh*	*kabus*
48.	rain	*ujɛ*	*ujaŋ*	*ujaŋ*	*hujan*

No.	Meaning	KM	CTM	Dusun	SM
49.	thunder	*plaka*	*guɣoh*	*litɔ*	*guruh*
50.	lightning	*kilaʔ*	*kilaʔ*	*kilaʔ*	*kilat*
51.	wind	*aŋiŋ*	*aŋiŋ*	*aŋiŋ*	*aŋin*
52.	earth, soil	*tanɔh*	*tanɔh*	*tanɔh*	*tanah*
53.	stone	*batu*	*batu*	*batəʊ*	*batu*
54.	sand	*pase*	*pase*	*pasi*	*pasir*
55.	water	*ae*	*ae*	*ai*	*air*
56.	river	*suŋa*	*suŋa*	*ai, suŋa*	*suŋai*
57.	sea, ocean	*lauʔ*	*lauʔ*	*lauʔ*	*laut*
58.	lake	*taseʔ*	*tasiʔ*	*tasɛiʔ*	*tasik*
59.	wood	*kayu*	*kayu*	*kayəʊ*	*kayu*
60.	forest	*utɛ*	*utaŋ*	*utaŋ*	*hutan*
61.	fire	*api*	*api*	*apɛi*	*api*
62.	dust	*dəbu*	*dəbu*	*dəbəʊ*	*dəbu*
63.	smoke	*asaʔ*	*asaʔ*	*asaʔ*	*asap*
64.	(fireplace) ash	*abu*	*abu*	*abəʊ*	*abu*
65.	island	*pula*	*pula*	*pulɔ*	*pulaw*
66.	mountain	*gunoŋ*	*gunoŋ*	*gunuŋ*	*gunuŋ*
67.	to burn	*baka*	*bakɔ*	*bakɔ*	*bakar*
68.	night	*malɛ*	*malaŋ*	*malaŋ*	*malam*
69.	day	*aɣi*	*aɣi*	*aɣɛi*	*hari*
70.	year	*taoŋ*	*taoŋ*	*tauŋ*	*tahun*
71.	tomorrow	*ɛsɔʔ*	*ɛsɔʔ*	*isəʊʔ*	(*b*)*esok*
72.	old	*tuwɔ*	*tuwə*	*tuwɛ*	*tua*
73.	new	*baɣu*	*baɣu*	*baɣəʊ*	*baru*
74.	young	*mudɔ*	*mudə*	*mudɛ*	*muda*
75.	animal	*nnatɛ*	*nnataŋ*	*nnatɔŋ*	*binataŋ*
76.	dog	*aɲiŋ*	*aɲjiŋ*	*aɲjɛiŋ*	*aɲjiŋ*
77.	cat	*kuciŋ*	*kuciŋ*	*kucɛiŋ*	*kuciŋ*
78.	bird	*buɣoŋ*	*buɣoŋ*	*buɣəʊŋ*	*buruŋ*
79.	rat	*tikuh*	*tikuh*	*tikuh*	*tikus*
80.	snake	*ula, cɛwɛ*	*ulɔ*	*ulɔ*	*ular*
81.	worm	*caciŋ*	*caciŋ*	*cacɛiŋ*	*caciŋ*
82.	louse	*gutu*	*kutu*	*gutəʊ*	*kutu*

No.	Meaning	KM	CTM	Dusun	SM
83.	mosquito	*ɲamoʔ*	*ɲamoʔ*	*ɲamuʔ*	*ɲamuk*
84.	spider	*llabɔ*	*llabə*	*glabɛ*	*laba-laba*
85.	fish	*ikɛ*	*ikaŋ*	*ikaŋ*	*ikan*
86.	shrimp	*udɛ*	*udaŋ*	*udɔŋ*	*udaŋ*
87.	buffalo	*kuba*	*kuba*	*kubɔ*	*kərbaw*
88.	to plant	*tanɛ*	*tanaŋ*	*tanaŋ*	*tanam*
89.	to dig	*gali*	*gali*	*galɛi*	*gali*
90.	tree	*pɔhoŋ, pɔkɔʔ*	*pɔhoŋ, pɔkɔʔ*	*pahuŋ, pukəʊʔ*	*pohon, pokok*
91.	branch	*dɛhɛ*	*dahaŋ*	*dahaŋ*	*dahan*
92.	bark (of tree)	*kuleʔ pɔhoŋ, kuleʔ pɔkɔʔ*	*kuleʔ pɔkɔʔ*	*kuliʔ kayəʊ*	*kulit pohon*
93.	leaf	*daoŋ*	*daoŋ*	*dauŋ*	*daun*
94.	root	*aka*	*akɔ*	*akɔ*	*akar*
95.	flower	*buŋɔ*	*buŋə*	*buŋɛ*	*buŋa*
96.	grass	*ɣupuʔ*	*ɣupuʔ*	*upuʔ*	*rumput*
97.	seed	*bəneh, bute*	*bəneh, bute*	*bənih, buti*	*bənih, biji*
98.	to eat	*makɛ*	*makaŋ*	*makaŋ*	*makan*
99.	to drink	*minoŋ*	*minoŋ*	*minuŋ*	*minum*
100.	to cook	*mmasɔʔ*	*mmasɔʔ*	*mmasɔʔ*	*mə-masak*
101.	to chew	*mamɔh*	*kuɲɔh*	*mamɔh*	*kuɲah*
102.	to suck	*isaʔ*	*isaʔ*	*isaʔ*	*hisap*
103.	egg	*təlɔ*	*təlɔ*	*təlu*	*təlur*
104.	rotten	*busũʔ*	*busũʔ*	*busũʔ*	*busuk*
105.	fruit	*buwɔh~wɔh*	*buwɔh~wɔh*	*buwɔh~wɔh*	*buah*
106.	ripe	*masɔʔ, ləbu*	*masɔʔ*	*masɔʔ, ilaŋ*	*masak, mataŋ*
107.	cooked rice	*nasiʔ*	*nasiʔ*	*nasiʔ*	*nasi*
108.	salt	*gaɣɛ*	*gaɣaŋ*	*gaɣaŋ*	*garam*
109.	oil	*miɲɔʔ*	*miɲɔʔ*	*miɲɔʔ*	*miɲak*
110.	fat, grease	*ləmɔʔ*	*ləmɔʔ*	*ləmɔʔ*	*ləmak*
111.	house	*ɣumɔh*	*ɣumɔh*	*umɔh*	*rumah*
112.	roof	*ataʔ, ɣapoŋ*	*ataʔ*	*ataʔ*	*atap*
113.	village	*kapoŋ*	*kapoŋ*	*kapəʊŋ*	*kampuŋ*
114.	one	*sɔ*	*sə*	*sɛ*	*satu*

No.	Meaning	KM	CTM	Dusun	SM
115.	two	*duɯwɔ*	*duɯwə*	*duɯwɛ*	*dua*
116.	three	*tigɔ*	*tigə*	*tigɛ*	*tiga*
117.	four	*ppaʔ*	*ppaʔ*	*ppaʔ*	*əmpat*
118.	five	*limɔ*	*limə*	*limɛ*	*lima*
119.	six	*nnɛ*	*nnaŋ*	*nnaŋ*	*ənam*
120.	seven	*tujoh*	*tujoh*	*tujəʊh*	*tujuh*
121.	eight	*lapɛ*	*lapaŋ*	*lapaŋ*	*lapan*
122.	nine	*smilɛ*	*smilaŋ ~mmilaŋ*	*smilaŋ*	*səmbilan*
123.	ten	*s=puloh*	*s=puloh*	*s=puləʊh*	*sə=puluh*
124.	eleven	*s=bəlah*	*s=bəlah*	*s=bəlah*	*sə=bəlas*
125.	to count	*kiɣɔ*	*kiɣə, bilaŋ*	*kiɣɛ, itəʊŋ*	*hituŋ, kira*
126.	many, much	*baɲɔʔ, ɣama*	*baɲɔʔ, ɣama*	*baɲɔʔ, ɣama*	*baɲak, ramai*
127.	how many, how much	*bɣapɔ*	*wwapə*	*bɣapɛ*	*bərapa*
128.	few	*sikiʔ*	*sikiʔ*	*sikĩʔ, ikĩʔ*	*sədikit*
129.	other	*laiŋ*	*laiŋ*	*laiŋ*	*lain*
130.	all	*ssəmɔ*	*ssəmə*	*smuwɛ*	*səmua*
131.	full	*pənoh*	*pənoh*	*pənuh*	*pənuh*
132.	rope	*tali*	*tali*	*talɛi*	*tali*
133.	knife	*pisa*	*pisa*	*pisɔ*	*pisaw*
134.	stick	*tukaʔ*	*tɔkaʔ*	*tukaʔ*	*toŋkat*
135.	to do	*buɯaʔ*	*buɯaʔ*	*buɯaʔ*	*buat*
136.	to work	*xxijɔ~kkijɔ*	*xxijə~kkijə*	*xxəjɛ~xxəjɛ*	*bə-kərja*
137.	to take	*ambiʔ*	*ambeʔ*	*ambɛiʔ*	*ambil*
138.	to pull	*taɣeʔ*	*taɣeʔ*	*taɣɛiʔ*	*tarik*
139.	to tie up	*ikaʔ*	*ikaʔ*	*ikaʔ*	*ikat*
140.	to stab	*tikɛ*	*tikaŋ*	*tikaŋ*	*tikam*
141.	to cut	*kəɣaʔ*	*kəɣaʔ, pɔtoŋ*	*kəɣaʔ, putəʊŋ*	*potoŋ*
142.	to hit	*ttɛ, katɔʔ*	*bahaŋ, katɔʔ*	*bahaŋ, katəʊʔ*	*pukul*
143.	to split	*bəlɔh*	*bəlɔh*	*bəlɔh*	*bəlah*
144.	to squeeze	*pəɣɔh*	*pəɣɔh*	*pəɣɔh, ramah*	*pəras, pərah*
145.	to pound	*tumboʔ*	*tumboʔ*	*tumbəʊʔ*	*tumbuk*

No.	Meaning	KM	CTM	Dusun	SM
146.	to rub	*gɔsɔʔ, gɔcɔh*	*gɔsɔh, gɔɲɔh*	*gusəʊʔ, kəsaʔ*	*gosok*
147.	to wipe	*laʔ*	*laʔ*	*laʔ*	*lap*
148.	needle	*jaɣoŋ*	*jaɣoŋ*	*jaɣuŋ*	*jarum*
149.	to sew	*jaiʔ*	*jaiʔ, jaheʔ*	*jaiʔ*	*jahit*
150.	to weave	*tənoŋ*	*tənoŋ*	*tənuŋ*	*tənun*
151.	road	*jalɛ*	*jalaŋ*	*jalaŋ*	*jalan*
152.	to walk	*jjalɛ*	*jjalaŋ*	*jjalaŋ*	*bər-jalan*
153.	to go	*gi*	*gi*	*gɛi*	*pərgi*
154.	to come	*maɣi*	*maɣi*	*maɣɛi*	*dataŋ*
155.	to turn	*pusiŋ*	*pusiŋ, bɛlɔʔ*	*pusɛiŋ*	*pusiŋ*
156.	to climb	*naiʔ*	*naiʔ*	*naiʔ*	*naik*
157.	to swim	*bɣənɛ*	*bɣənaŋ*	*bunɔŋ*	*bə-rənaŋ*
158.	to throw	*lɛpa*	*lɛpɔ, tɔhɔʔ*	*lipɔ, capɔ, kaləʊŋ*	*lempar*
159.	to fall	*jatoh*	*jatoh*	*jatəʊh*	*jatuh*
160.	to fly	*tɣəbɛ*	*tɣəbaŋ*	*tubɔŋ*	*tərbaŋ*
161.	to flow	*ŋŋale*	*ŋŋale*	*ŋŋali*	*alir*
162.	to blow	*tiyuʔ, mbuh*	*tiyuʔ, mbuh*	*tiyuʔ, mbuh*	*tiup*
163.	to push	*tɔlɔʔ*	*tɔlɔʔ*	*tulɔʔ*	*tolak*
164.	canoe	*pɣau*	*pɣau*	*pɣau*	*pərahu*
165.	to float	*nnapoŋ*	*timbo*	*timbu*	*tər-apuŋ*
166.	sail	*laya*	*layɔ*	*layɔ*	*layar*
167.	to hunt	*bbuɣu*	*bbuɣu*	*bbuɣəʊ*	*bər-buru*
168.	to shoot	*tɛmbɔʔ, mmanɔh*	*tɛmbɔʔ*	*timbɔʔ*	*tembak*
169.	fishing hook	*matɔ kaẽ*	*matə kaẽ*	*matɛ kaĩ*	*mata kail*
170.	to steal	*cuɣi*	*cuɣi*	*cuɣɛi*	*curi*
171.	to give	*buwi~wi*	*buwi~wi*	*buwɛi~wɛi*	*bəri*
172.	to hold	*pəgɛ*	*pəgaŋ*	*pəgɔŋ*	*pəgaŋ*
173.	to buy	*bəli*	*bəli*	*bəlɛi*	*bəli*
174.	to know	*tau*	*tau*	*tau*	*tahu*
175.	to think	*mmike*	*mmike*	*mmiki*	*pikir, fikir*
176.	what	*ggapɔ*	*məndə ~pəndə*	*məndɛ*	*apa*
177.	who	*sapɔ, pdiyɔ*	*sapə, piyə*	*pdiyɛ~piyɛ*	*siapa*

No.	Meaning	KM	CTM	Dusun	SM
178.	when	*bilɔ*	*bilə*	*bilɛ*	*bila*
179.	where	*mmanɔ*	*mmanə*	*dwanɛ*	*di mana*
180.	how	*gganɔ, gwanɔ*	*gganə*	*gwanɛ*	*bagay-mana*
181.	and; with	*dəŋɛ~dəŋa ~ŋɛ~ŋa*	*dəŋaŋ ~ŋaŋ~ŋə*	*dəŋaŋ~ŋaŋ ~ŋə*	*dan, dəŋaŋ*
182.	if	*kalu*	*kalu*	*kaləʊ*	*kalaw*
183.	no, not	*dɔʔ, tɔʔ*	*dɔʔ*	*dɔʔ*	*tidak*
184.	no	*bukɛ*	*bukaŋ*	*bukaŋ*	*bukan*
185.	to hear	*dəŋa*	*dəŋɔ*	*dəŋɔ*	*dəŋar*
186.	to see	*napɔʔ*	*napɔʔ*	*napɔʔ, liyaʔ*	*nampak, lihat*
187.	to sniff	*siyoŋ*	*ciyoŋ*	*ciyuŋ*	*cium*
188.	sharp	*tajɛ*	*tajaŋ*	*tajaŋ*	*tajam*
189.	dull	*tupo*	*tupo*	*tupu*	*tumpul*
190.	smooth	*liciŋ*	*liciŋ*	*liciŋ*	*licin*
191.	bitter	*pahiʔ*	*paiʔ, paheʔ*	*paiʔ*	*pahit*
192.	warm	*panah*	*panah*	*panah*	*panas*
193.	cold	*səjoʔ*	*səjoʔ*	*səjəʊʔ*	*səjuk*
194.	dry	*kəɣiŋ*	*kəɣiŋ*	*kəɣɛiŋ*	*kəriŋ*
195.	wet	*basɔh*	*basɔh*	*basɔh*	*basah*
196.	heavy	*bəɣaʔ*	*bəɣaʔ*	*bəɣaʔ*	*bərat*
197.	dirty	*cəma*	*kɔtɔ*	*kutu*	*kotor*
198.	black	*itɛ*	*itaŋ*	*itaŋ*	*hitam*
199.	white	*puteh*	*puteh*	*putɛih*	*putih*
200.	red	*mɛɣɔh*	*mɛɣɔh*	*miɣɔh*	*merah*
201.	yellow	*kuniŋ*	*kuniŋ*	*kuniŋ*	*kuniŋ*
202.	green	*ija*	*ija*	*ijɔ*	*hijaw*
203.	blue	*biɣu*	*biɣu*	*biɣəʊ*	*biru*
204.	good	*baiʔ, juɣuh, mɔlɛʔ*	*baiʔ, juɣuh, mɔlɛʔ*	*baiʔ, mulɛiʔ*	*baik*
205.	bad	*jahaʔ*	*jahaʔ*	*jahaʔ*	*jahat*
206.	true	*bəto*	*bəto*	*bənɔ, bətu*	*bənar*
207.	to laugh	*sukɔ, ŋŋila*	*ssukə*	*sukɛ*	*kətawa*
208.	to cry	*tiyɔʔ, jjəɣiʔ*	*nnaŋih*	*nnaŋih*	*mən-<t>aŋis*
209.	to play	*mmaiŋ*	*mmaiŋ*	*mmaiŋ*	*bər-main*

No.	Meaning	KM	CTM	Dusun	SM
210.	to choose	*pileh*	*pileh*	*pilɛih*	*pilih*
211.	ashamed	*malu*	*malu*	*maləʊ*	*malu*
212.	be afraid	*takoʔ*	*takoʔ*	*takuʔ*	*takut*
213.	left side	*kiɣi*	*kiɣi*	*kiɣɛi*	*kiri*
214.	right side	*kanɛ*	*kanaŋ*	*kanaŋ*	*kanan*
215.	small	*kəcɛ̃ʔ*	*kəciʔ*	*kəcĩʔ*	*kəcil*
216.	big	*bəsa*	*bəsɔ*	*bəsɔ*	*bəsar*
217.	short	*pandɔʔ, pɛndɛʔ*	*pɛndɛʔ*	*pandɔʔ, pindɛiʔ*	*pendek*
218.	long	*paɲjɛ*	*paɲjaŋ*	*paɲjɔŋ*	*paɲjaŋ*
219.	thin	*nipih*	*nipih*	*nipih*	*nipis, tipis*
220.	thick	*təba*	*təba*	*təba*	*təbal*
221.	narrow	*səpeʔ*	*səpeʔ*	*səpiʔ*	*səmpit*
222.	wide	*lɛba*	*lɛbɔ, luwah*	*libɔ*	*lebar*
223.	near	*dəkaʔ*	*dəkaʔ*	*dəkaʔ*	*dəkat*
224.	far	*jaoh*	*jaoh*	*jauh*	*jauh*
225.	round	*bulaʔ*	*bulaʔ*	*bulaʔ*	*bulat*
226.	straight	*bəto*	*stəɣeʔ*	*luɣuh*	*lurus*
227.	to sit	*dudoʔ*	*dudoʔ*	*dudəʊʔ*	*duduk*
228.	to stand	*ddiɣi*	*ddiɣi*	*ddiɣɛi*	*bər-diri*
229.	to lie down	*guliŋ, nnətɛ*	*bbaɣiŋ, guliŋ, gɔlɛʔ*	*bbaɣɛiŋ, tlətɔŋ*	*bər-bariŋ*
230.	to grow	*tumboh*	*tumboh*	*tumbəʊh*	*tumbuh*
231.	to hide	*susuʔ ~nnusuʔ*	*ɲnusuʔ ~nnusuʔ*	*ɲnusuʔ, nnapəʊʔ*	*səmbuɲi*
232.	to open	*bukɔ*	*bukə*	*bukɛ*	*buka*
233.	at	*di*	*di*	*də*	*di*
234.	inside	*dalɛ*	*dalaŋ*	*dalaŋ*	*dalam*
235.	top	*atah*	*atah*	*atah*	*atas*
236.	bottom	*bɔwɔh*	*bɔwɔh*	*bawɔh*	*bawah*
237.	language	*basɔ*	*bahasə*	*bahasɛ*	*bahasa*
238.	name	*namɔ*	*namə*	*namɛ*	*nama*
239.	say	*kɛcɛʔ*	*katə, kabɔ*	*cakaʔ*	*bər-kata, cakap*
240.	to sing	*ɲɲaɲi*	*ɲɲaɲi*	*ɲɲaɲiŋ*	*mə-ɲaɲi*

No.	Meaning	KM	CTM	Dusun	SM
241.	I	*aku, kawɛ, ambɔ, kitɔ*	*aku, ambə, sayə*	*akəʊ*	*aku, saya*
242.	you	*moŋ, dɛmɔ*	*moŋ*	*məʊŋ*	*kamu*
243.	(s)he	*diyɔ, i̇yɔ*	*yə, diyə*	*diyɛ*	*dia*
244.	we (INCL)	*kitɔ*	*kitə*	*kitɛ*	*kita*
245.	we (EXCL)	*kitɔ*	*kitə*	*kamiŋ*	*kami*
246.	you (PL)	*dɛmɔ (ɣama), moŋ ssəmɔ*	*moŋ (ssəmə)*	*mikɛ*	*kamu səkalian*
247.	they	*dɛmɔ, diyɔ, i̇yɔ*	*(sɛʔ-sɛʔ) yə*	*dimɛ, mikɛ*	*məreka*
248.	person; human being	*ɔɣɛ*	*ɔɣaŋ*	*uɣɔŋ*	*oraŋ*
249.	male	*jatɛ*	*llaki*	*llakɛi*	*ləlaki*
250.	female	*ttinɔ*	*ppuwaŋ*	*ppuwaŋ*	*pərəmpuan*
251.	mother	*mɔʔ, maʔ, mɛʔ, ibu, umi*	*mɔʔ*	*mɔʔ*	*əmak*
252.	father	*ayɔh, pɔʔ, abah*	*ayɔh, baʔ*	*ayɔh*	*ayah*
253.	child	*anɔʔ*	*anɔʔ*	*anɔʔ*	*anak*
254.	husband	*laki*	*laki*	*lakɛi*	*suami*
255.	wife	*bini~biniŋ*	*bini~biniŋ*	*biniŋ*	*isteri, bini*
256.	sister	*kakɔʔ, kaʔ*	*kakɔʔ, kaʔ*	*kakɔʔ, abɔŋ*	*kakak*
257.	this	*ni~niŋ*	*ni~niŋ*	*nɛiŋ*	*ini*
258.	that	*tu*	*tu*	*təʊ*	*itu*
259.	here	*ssini~ssiniŋ*	*ssini~ssiniŋ*	*sinɛiŋ*	*sini*
260.	there	*ssitu, dənoŋ*	*ssitu, dənoŋ*	*sinəʊŋ, sitəʊ, dinəʊŋ, ditəʊ*	*sana, situ*

APPENDIX B

Sample texts

B.1 Sample text of KM

Conversation: Making Sambal fish

The following text is an excerpt of a daily conversation in KM, which took place between two main speakers (labelled as A and B) in Kampung Kusial Bharu, Tanah Merah, Kelantan (citation code: KM_180820_cv03). A third speaker (labelled as C) also took part in the conversation.

(1) A. *umi*[1] *bɛsɔ* *makɛ ikɛ* ... *ikɛ* *diyɔ waʔ* *gganɔ, paŋge ikɛ*
mother usually eat fish ... fish 3 make how call fish
samba.
sambal
'Umi do you eat the fish ... how do they make it ... they call it Sambal fish.'

(2) A. *tapi diyɔ gunɔ ladɔ-ija* *sajɔ, ladɔ-sɔlɔʔ.*
but 3 use chilli-green only chilli-solok
'But they use green chilli only, solok chilli.'

[1] Used as a form of address.

(3) A. *diyɔ blɛndə ladɔ-sɔlɔʔ tu dəŋa asɛ-lima.*
3 blend(ENG) chilli-solok DEM.DIST with sour-citrus
'They blend the solok chilli with lime.'

(4) A. *hɔʔ tu jah, sədaʔ, umi tra.*
REL DEM.DIST only delicious mother try(ENG)
'Only that, but it's tasty, umi you need to try.'

(5) A. *ikɛ-kɔŋɛ hɔʔ ikɛ suŋa, ikɛ-kɔŋɛ.*
torpedo.scad REL fish river torpedo.scad
'Torpedo scad, the river fish, torpedo scad.'

(6) A. *kaʔti iŋaʔ tɔʔ, kaʔti hɔʔ bəsa-bəsa jjabaʔ kawɛ?*
? remember NEG ? REL RDP-big office 1SG
'Kakti, do you remember the big kakti in my office?'

(7) B. *deh, b-bəlɔh ikɛ=lah supɔ ggitu?*
SFP CAUS-split fish=FOC like like.that
'So you half the fish like that?'

(8) A. *ikɛ tu hɔʔ diyɔ gɔɣeŋ tawa=lah, diyɔ gɔɣeŋ tawa dulu*
fish DEM.DIST REL 3 fry plain=SFP 3 fry plain first
ataupoŋ boh gaɣɛ sikiʔ dulu lagi masiŋ-masiŋ.
or put salt little first more RDP-salty
'The fish is fried plain, it's fried without any spices first, or (you can) add some salt to make it a bit salty.'

(9) A. *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.'

(10) B. *hɔ̃ jadi ae asɛ-lima jah.*
AFF so water sour-citrus only
'Ah so it's only lime juice.'

(11) A. *ae asɛ-lima jah.*
water sour-citrus only
'Only lime juice.'

(12) A. *pah tu gulɔ gaɣɛ.*
after DEM.DIST sugar salt
'Then sugar and salt.'

(13) B. *iyɔ tɔʔ buboh bawɛ s=ulah ggapɔʔ*
3 NEG put onion one=CLF what
'They don't put a clove of garlic or things like that?'

(14) A. *dɔʔ, tɔʔ buboh bawɛ-puteh tɔʔ buboh bawɛ-bəsa tɔʔ buboh.*
NEG NEG put onion-white NEG put onion-big NEG put
'No, they don't put garlic or onion.'

(15) A. *diyɔ tumih hɔʔ tu ... cumɔ gaʔ sɛ waʔ*
3 sauté REL DEM.DIST ... only EMPH someone(?) make
ni hɔ̃=lah kitɔ ... mugɔ kitɔ nɔʔ cpəɣɔʔ=jə,
DEM.PROX AFF=FOC 1PL ... thing 1PL want juicy(?)=just
sɛ masoʔ ae sikiʔ.
someone(?) add water little
'The sauté they make ... some people just make it like that. But in case we want it to be juicier, we add a bit of water.'

(16) A. *tapi bəna tɔʔ masoʔ ae poŋ.*
but real NEG add water also
'But actually it's also fine not to add water.'

(17) A. *diyɔ ssəɣɔ masɛ pədah manih, ɔlɔh səda?=lah.*
3 feel sour spicy sweet INTERJ delicious=SFP
'It tastes sour, spicy and sweet, ah delicious.'

(18) B. *pəɣɔh asɛ-lima=kɛ, diyɔ baɲɔʔ.*
squeeze sour-citrus=TAG, 3 much
'When you squeeze the lime, you know, there's a lot (of juice).'

(19) B. *nɔʔ tra buwaʔ eh.*
want try(ENG) make INTERJ
'I want to try to make it.'

(20) A. *mɛmɛ sətɔra waʔ d=umɔh, pakaʔ katɔ sədaʔ*
indeed as.soon.as(SM) make LOC=house agree say delicious
blakɔ, walaupoŋ pədah tapi tageh nɔʔ makɛ.
all although spicy but addicted want eat
'When we make it at home, everyone agrees that it's tasty. Even though it's spicy, people can't stop eating it.'

(21) B. *nɔʔ tra waʔ eh, ikɛ-kɔŋɛ=lah?*
want try(ENG) make INTERJ torpedo.scad=SFP
'I want to try to make it, torpedo scad, right?'

(22) A. *ikɛ-kɔŋɛ.*
torpedo.scad
'Torpedo scad.'

(23) C. *ikɛ manih-manih ggitu=lah isi samba.*
fish RDP-sweet like.that=FOC content sambal
'The fish that's kind of sweet, filled with sambal.'

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

(25) B. *hɔ̃=lah kitɔ tumih jadi kɔko tu=lah,*
AFF=FOC 1PL sauté become brown DEM.DIST=SFP
kitɔ m-masoʔ gulɔ gaɣɛ.
1PL CAUS-add sugar salt
'So we sauté it until it becomes brown, then we add sugar and salt.'

(26) A. *bau diyɔ b-bakeʔ, ladɔ-sɔlɔʔ=kɛʔ*
smell 3 MID-rise chilli-solok=TAG
'The smell rises abruptly, (because of) the solok chilli, right?'

(27) A. *mɛmɛ masɔ kitɔ bləndə tu kuwaʔ bau,*
indeed when 1PL blend(ENG) DEM.DIST strong smell
kitɔ nn-<t>umih poŋ bakeʔ tapi mɛmɛ sədaʔ.
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.'

(28) B. *byəsiŋ=lah, hɔ̃?*
sneeze=SFP AFF
'Gonna sneeze, huh?'

(29) A. *kalu tu gaʔ, buboh blacɛ siki ʔ,*
if DEM.DIST EMPH put shrimp.paste little
kalu tɔʔ boh blacɛ poŋ taʔpɔ dɔh.
if NEG put shrimp.paste also no.problem already
'In that case, you can also add a bit of shrimp paste, but if not it's also fine.'

(30) B. *paka sumbaʔ=jə=lah?*
use stuff=just=SFP
'So you just stuff it?'

(31) A. *diyɔ macɛ bilɔ kitɔ bəlɔh=kɛ, kitɔ bəlɔh.*
3 like when 1PL split=TAG 1PL split
'It's like, we half it, right, we half it.'

(32) A. *hɔ̃ kitɔ tumih pah gaʔ kitɔ m-masoʔ=lah siki ʔ dalɛ ...*
AFF 1PL sauté then EMPH 1PL CAUS-enter=FOC little inside ...
pah kitɔ gɔlɛʔ-gɔlɛʔ ggitu ah.
then 1PL RDP-flip like.that INTERJ
'The seasoning we make, we stuff it a bit inside (the fish), then we keep flipping (the fish) like that.'

(33) A. *maʔsoʔ ɲɔ diyɔ bukɛ kənɔ k-kuwɔh,*
meaning ANAPH 3 NEG must INTR-gravy
diyɔ maʔsoʔ ɲɔ diyɔ ikɛ kəɣiŋ=lah.
3 meaning ANAPH 3 fish dry=SFP
'It means we don't want to make gravy; it's like dry fish.'

(34) B. *diyɔ kənɔ pəkaʔ bəto.*
3 must stick right
'So (the seasoning) should stick properly.'

(35) A. *hɔ̃ tu poŋ maʔsoʔ ɲɔ bilɔ kitɔ makɛ*
AFF DEM.DIST also meaning ANAPH when 1PL eat
hɔʔ ləkaʔ-ləkaʔ tu kə badɛ dɔh.
REL RDP-stick DEM.DIST to body already
'Yeah in that way, it means when we eat, (the seasoning) sticks to the fish.'

(36) A. *hɔ̃ diyɔ ɣasɔ tɔʔ bɣ-ae dɔh.*
AFF 3 feel NEG INTR-water already
'So it does not taste watery.'

(37) A. *ɔlɔh mɛmɛ puwah-ati=lah katɔ.*
INTERJ indeed satisfied-liver=SFP say
'Ah that's really satisfying.'

(38) B. *ah tra waʔ s=kali ikɛ-kɔŋɛ.*
INTERJ try(ENG) make one=time torpedo.scad
'I want to try to make this fish once.'

B.2 Sample text of CTM

Narrative: Pear story

The following text is a narrative of the pear story in CTM (with visual stimuli), recorded in Kampung Gong Sentul, Kuala Nerus, Terengganu (citation code: CTM_181025_n02).

(1) *dalaŋ s=buɯɔh kapoŋ, adə tanamaŋ wɔh pɛ.*
inside one=CLF village EXIST plantation(SM) fruit pear

'In a village, there was a pear plantation.'

(2) *adə s=ɔɣaŋ ləlaki tu, yə xxijə kuteʔ wɔh pɛ.*
EXIST one=CLF male(SM) DEM.DIST 3 work pick fruit pear

'There was a guy, whose job was to pick pears.'

(3) *kuteʔ niŋ, wɔh pɛ niŋ, yə ambeʔ upɔh ah,*
pick DEM.PROX fruit pear DEM.DIST 3 take wage INTERJ
mukiŋ yə ambeʔ upɔh ah kuteʔ wɔh pɛ niŋ.
maybe 3 take wage INTERJ pick fruit pear DEM.DIST

'The task of picking pears ... maybe he does that for a living.'

(4) *yə gunə taŋgə.*
3 use ladder

'He used a ladder.'

(5) *yə gunə kɛndoŋ, ɔɣaŋ paŋge, kɛndoŋ, kaiŋ,*
3 use handkerchief person call handkerchief cloth
kɛndoŋ boh dəpaŋ.
handkerchief put front

'He used a handkerchief, we call it "kendong", which he wore in the front.'

(6) *yə ambeʔ dɛɣɔh s=puloh biji ... s=puloh ute,*
3 take approximately(?) one=ten CLF ... one=ten CLF
limə-bəlah ute, yə waʔ tuɣoŋ b=bɔwɔh, buboh dalaŋ bako.
five-teens CLF 3 CAUS go.down LOC=below put inside basket

'He picked about ten or fifteen pears, then he descended and poured them in a basket.'

(7) *yə təŋɔh-təŋɔh uboh niŋ, təŋɔh uboh, yə kuteʔ-kuteʔ uboh,*
3 RDP-PROG put DEM.PROX PROG put 3 RDP-pick put
kuteʔ-kuteʔ uboh, yə naiʔ.
RDP-pick put 3 go.up

'As he poured the pears ... he picked the placed them into a basket, then he went up again.'

(8) *pah tu, yə paka niŋ, kaiŋ niŋ,*
after DEM.DIST 3 wear DEM.PROX cloth DEM.PROX
tudoŋ muloʔ yə niŋ, kuteʔ wɔh pɛ, tudoŋ hɔ̃.
cover mouth 3 DEM.PROX pick fruit pear cover AFF

'Then, he was also wearing this ... the cloth ... the mouth cover, as he picked pears.'

(9) *hɔ̃ niŋ, hɔ̃ yə buwaŋ kaiŋ panah, yə laʔ, yə laʔ wɔh pɛ.*
AFF DEM.PROX AFF 3 throw cloth hot 3 wipe 3 wipe fruit pear

'He took off the cloth (because it was) hot, and he wiped the pear.'

(10) *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.'

(11) *kaiŋ, hɔʔ kitə kɔʔ kaiŋ-twala=lah, yə tutoʔ siniŋ,*
cloth REL 1PL call(?ENG) cloth-towel=SFP 3
tutoʔ mukə muloʔ yə.
close here close face mouth 3

'The cloth is like what we call a towel, and he fastened it here, covering his face and mouth.'

(12) *utoʔ ni ah, yə naiʔ pulɔʔ.*
close DEM.PROX INTERJ 3 go.up again

'He fastened the cloth, then he climbed up again.'

(13) *pah tu, yə təŋɔh-təŋɔh naiʔ tu, adə ɔɣaŋ*
after DEM.PROX 3 RDP-PROG go.up DEM.DIST EXIST person
gmalə kambiŋ bɔwɔʔ lalu, kambiŋ lalu yə niŋ ah.
shepherd goat take pass.by goat pass.by 3 DEM.DIST INTERJ

'As he was climbing (the tree), a shepherd with a goat passed by.'

(14) *yə təŋɔh nai̍ʔ wɔh pɛ niŋ, tapi kambiŋ ni dɔʔ ...*
3 PROG go.up fruit pear DEM.PROX but goat DEM.PROX NEG ...
dɔʔ makaŋ buwɔh pɛ.
NEG eat fruit pear

'He was climbing up the pear (tree), but the goat didn't eat pears.'

(15) *lalu sajə=jə, ikoʔ pɔkɔʔ yə.*
pass.by only=just follow tree 3

'It just passed by, alongside the tree.'

(16) *yə ambeʔ agi hɔ̃, s=ɛkɔ kambiŋ lalu, mməɣaŋ.*
3 take again AFF one=CLF goat pass.by bleat

'The guy went on picking pears, and a goat passed by, bleating.'

(17) *yə dɔʔ kaca wɔh pɛ, nai̍ʔ bukeʔ, kambiŋ.*
3 NEG disturb fruit pear go.up hill goat

'The goat didn't disturb the pears and headed up to a hill.'

(18) *hɔʔ ɔɣaŋ niŋ ambeʔ agi wɔh pɛ.*
REL person DEM.PROX take again fruit pear

'And the man carried on picking pears.'

(19) *baɲɔʔ yə ... yə pəteʔ wɔh pɛ, yə kuteʔ ah,*
much 3 ... 3 pick fruit pear, 3 pick INTERJ
ambeʔ upɔh kuteʔ wɔh pɛ, tuɣoŋ lagi.
take wage pick fruit pear go.down again

'He picked many pears. He was picking pears for a living. He went down again.'

(20) *lalu=lah s=ɔɣaŋ budɔʔ llaki ŋə basika.*
pass.by=FOC one=CLF kid male with bike

'A boy with a bike passed by.'

(21) *pɔʔciʔ niŋ, ə̃, ə̃, təŋɔh kuteʔ wɔh pɛ atah pɔhoŋ.*
uncle DEM.PROX HES HES PROG pick fruit pear top tree

'The guy was still picking pears up in the tree.'

(22) *yə lalu ŋə basika.*
3 pass.by with bike
'The boy passed by with a bike.'

(23) *pɔʔciʔ dɔʔ sədɔ adə budɔʔ llaki lalu.*
uncle NEG realise EXIST kid male pass.by
'The guy was unaware that a boy passed by.'

(24) *yə lalu, yə ambeʔ s=bako wɔh pɛ niŋ,*
3 pass.by 3 take one=basket fruit pear DEM.PROX
yə waʔ naiʔ basika diyə.
3 CAUS go.up bike 3
'The boy passed by; he took a basket of pears and put it on his bike.'

(25) *yə cuɣi ah, s=bako yə ambeʔ, yə ita, yə tɛŋɔʔ hɔʔ tukaŋ*
3 steal INTERJ one=basket 3 take 3 peep 3 look REL craftsman
kuteʔ wɔh pɛ tu dɔʔ napɔʔ, yə poŋ ambeʔ.
pick fruit pear DEM.PROX NEG see 3 also take
'Ah he was stealing; he took a basket (of pears). He glanced around and saw that the pear picker didn't seem to notice; he took a basket.'

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

(27) *hɔ̃, yə nɔʔ ambeʔ hɔ̃, yə ambeʔ s=bako.*
AFF 3 want take AFF 3 take one=basket
'He wanted to take (the pears), so he grabbed a basket.'

(28) *səbaʔ hɔʔ ni dɔʔ sədɔ, yə ambeʔ, yə tɛŋɔʔ dɔh,*
because REL DEM.PROX NEG realise 3 take 3 look already
hɔʔ pəteʔ wɔh pɛ dɔʔ sədɔ, yə poŋ ambeʔ.
REL pick fruit pear NEG realise 3 also take
'Since the pear picker didn't realise – the boy had noticed this – he went ahead and grabbed the pears.'

(29) *hɔ̃ yə ambeʔ=lah bako, uboh atah basika.*
AFF 3 take=FOC basket put top bike
'He took the basket, and put it on his bike.'

(30) *dəɣah-dəɣah, yə nɔʔ, nɔʔ, nɔʔ waʔ laɣi ah.*
RDP-fast 3 want want want CAUS run INTERJ

'He wanted to run away in a hurry.'

(31) *yə uboh atah basika yə, budɔʔ llaki ni hɔ̃.*
3 put top bike 3 kid male DEM.PROX AFF

'The boy put the basket on his bike.'

(32) *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.'

(33) *paɣə blakaŋ yə dɔʔ uboh, yə uboh atah paɣə dəpaŋ,*
rack back 3 NEG put 3 put top rack front
ɔɣaŋ paŋge paɣə ah basika niŋ, uboh atah paɣə.
person call rack INTERJ bike DEM.PROX put top rack

'He didn't put the pears on the back carrier, he put them on the front carrier. We call it "pare". He put it on the rack.'

(34) *yə poŋ gi j-jalaŋ.*
3 also go INTR-road

'He went on his way.'

(35) *waʔ laɣi dɔh ni, pɔʔciʔ ni dɔʔ sədɔ yaŋ ...*
CAUS run already DEM.PROX uncle DEM.PROX NEG realise REL(SM) ...
buwɔh pɛ yə kənə cuɣi ɛ̃h, s=bako ɛ̃h, dɔʔ sədɔ.
fruit pear 3 ADVS steal INTERJ one=basket INTERJ NEG realise

'The boy ran away (with the pears), and the man didn't realise that his pears had been stolen.'

(36) *budɔʔ ni poŋ waʔ laɣi, waʔ laɣi hɔ̃, waʔ laɣi, waʔ laɣi.*
kid DEM.PROX also CAUS run CAUS run AFF CAUS run CAUS run

'The boy ran away, he ran, ran, and ran (away with the pears).'

(37) *təŋɔh-təŋɔh jalaŋ adə batu adə pəndə.*
RDP-middle road EXIST stone EXIST what

'In the middle of the road there was a stone or something.'

(38) *yə təŋɔh waʔ layi ni, adə=lah bəlɔh noŋ, yə ssimboŋ,*
3 PROG CAUS run DEM.PROX EXIST=FOC side there 3 encounter(?)
jupə ŋə hɔʔ ppuwaŋ naiʔ basika ɣɔʔ, dəkaʔ-dəkaʔ tu
meet with REL female ride bike rustle(?) RDP-near DEM.DIST
hɔ̃.
AFF

'While he was running away, on the opposite side (of the street) there was ... he encountered ... he crossed paths with a girl riding a squeaky bike, they were quite near to each other.'

(39) *yə poŋ b-laŋgɔ ɣəbɔh, yə b-laŋgɔ ɣəbɔh hɔ̃,*
3 also MID-crash fall 3 MID-crash fall AFF
c-cabuʔ tɔpi yə.
NVOL-pull.out hat 3

'He crashed (with the stone) and fell, and his hat was blown away.'

(40) *likuʔ abih=lah jatoh wɔh pɛ atah tanɔh.*
overturn finished=FOC fall fruit pear top ground

'The pears cascaded onto the ground.'

(41) *budɔʔ ppuwaŋ aʔdi layi.*
kid female just.now run

'The girl ran away.'

(42) *yə poŋ kuteʔ=lah buwɔh pɛ niŋ, dɔh budɔʔ ppuwaŋ*
3 also pick=FOC fruit pear DEM.PROX already kid female
tu layi, sapu-sapu sakeʔ lutuʔ, sakeʔ məndə hɔ̃, uwaŋ-uwaŋ
DEM.PROX run RDP-wipe hurt knee hurt what AFF RDP-throw
stɔkiŋ, sakeʔ.
sock hurt

'He collected the pears after the girl went away, dusting himself off. His knee hurt, something hurt ... then he was dusting his socks.'

(43) *yə poŋ napɔʔ adə tigə ɔɣaŋ llaki doʔ situ,*
3 also see EXIST three person male LOC there
tuloŋ ah kuteʔ wɔh pɛ yə ni.
help INTERJ pick fruit pear 3 DEM.PROX

'Then he saw three boys helping him pick pears.'

(44) *tuloŋ kuteʔ wɔh pɛ, tuloŋ ... tuloŋ sapu-sapu ... baju yə niŋ,*
help pick fruit pear help ... help RDP-wipe ... shirt 3 DEM.DIST
yə jatoh hɛ̃.
3 fall INTERJ

'They helped him pick pears, and helped him dust himself off ... the boy fell down.'

(45) *tuloŋ kuteʔ, tuloŋ bɛʔki basika ha ... tuloŋ kuteʔ blakə,*
help pick help repair bike INTERJ ... help pick all
tuloŋ kuteʔ wɔh pɛ.
help pick fruit pear

'They helped him pick pears and fix the bike ... they helped to pick up the pears and everything.'

(46) *yə dɔʔ wi, yə dɔʔ wi pəndə, ɔɣaŋ tuloŋ yə niŋ,*
3 NEG give 3 NEG give what person help 3 DEM.PROX
dɔʔ wi pəndə.
NEG give what

'He didn't offer anything to the guys who helped him.'

(47) *tuloŋ sajə=jə=lah, dɔʔ upɔh apə-apə, tuloŋ sajə=jə.*
help only=just=SFP NEG employ RDP-what(SM) help only=just

'They just helped him and didn't expect anything in return; they just helped him.'

(48) *sɛpaʔ batu, yə poŋ j-jalaŋ, ləpah ɔɣaŋ j-jalaŋ.*
kick stone 3 also INTR-road then person INTR-road

'One guy kicked the stone, and went on his way. The guys went on their way.'

(49) *ləpah j-jalaŋ hɔʔ tigə ɔɣaŋ budɔʔ niŋ,*
then INTR-road REL three CLF kid DEM.PROX
napɔʔ=lah adə tɔpi yə jatoh.
see=FOC EXIST hat 3 fall
'While the guys were walking, they saw the hat that had fallen.'

(50) *hɔʔ ... hɔʔ cuɣi aʔdi.*
REL ... REL steal just.now
'It belonged to the boy who stole (the pears).'

(51) *tɔpi yə jatoh, pah tu budɔʔ llaki ni poŋ waʔ*
hat 3 fall after DEM.DIST kid male DEM.PROX also make
siyu, waʔ buŋni swaɣə macaŋ gadə ni bəpusiŋ.
whistle make sound voice like club DEM.PROX MID-turn(SM?)
'His hat fell ... then the boy whistled, making a sound like a swinging club.'

(52) *yə kabɔ=lah tɔpi yə t-tiŋga.*
3 tell=FOC hat 3 NVOL-leave
'He told the boy that his hat was left behind.'

(53) *yə poŋ wi tɔpi smula kə budɔʔ hɔʔ cuɣi ah.*
3 also give hat back to kid REL steal INTERJ
'Then he gave the hat back to the boy who stole (the pears).'

(54) *baga balasaŋ yə poŋ wi tigə bute wɔh pɛ,*
as(SM) reply(SM) 3 also give three CLF fruit pear
s=ute s=ɔɣaŋ.
one=CLF one=CLF
'In return, the boy handed them three pears, one pear for each person.'

(55) *yə wi ah tigə buwɔh pɛ hah, səbaʔ hɔʔ tu tuloŋ*
3 give INTERJ three fruit pear INTERJ because REL DEM.DIST help
wi baleʔ tɔpi diyə, yə wi tigə bute wɔh pɛ.
give return hat 3 3 give three CLF fruit pear
'He gave them three pears, because they helped to return his hat; he gave them three pears.'

(56) *yə poŋ təɣuh agi ah pəjalanaŋ waʔ agi mmanə*
3 also continue again INTERJ journey(SM) make again where
ɛ̃h wɔh pɛ ni ɛ̃h.
INTERJ fruit pear DEM.PROX INTERJ

'He continued on his journey, not sure where he was going (with) the pears.'

(57) *hɔ̃ təɣuh jalaŋ, hɔʔ ni tigə ɔɣaŋ poŋ gi ...*
AFF continue road REL DEM.PROX three person also go ...
taŋ jalaŋ=lah kapoŋ ni ah.
? road=FOC village DEM.PROX INTERJ

'He went on his way ... the three boys also went on their way to the village.'

(58) *hɔ̃, wi b-bagi s=ute s=ɔɣaŋ wɔh pɛ hɔʔ kawaŋ yə wi*
AFF give CAUS-share one=CLF one=CLF fruit pear REL friend 3 give
aʔdi.
just.now

'One of them shared a pear given by the boy with each of his friends.'

(59) *hɔ̃, yə poŋ j-jalaŋ, j-jalaŋ tigə ɔɣaŋ.*
AFF 3 also INTR-road INTR-road three person

'The three boys continued walking.'

(60) *j-jalaŋ, j-jalaŋ, j-jalaŋ naiʔ bukeʔ, j-jalaŋ, j-jalaŋ,*
INTR-road INTR-road INTR-road go.up hill INTR-road INTR-road
j-jalaŋ.
INTR-road

'Walk, walk, walk, up to a hill, walk, walk, walk.'

(61) *lalu=lah, jupə=lah pɔʔciʔ hɔʔ tuɣoŋ,*
pass.by=FOC meet=FOC uncle REL go.down
hɔʔ kuteʔ buwɔh pɛ ni baɣu tuɣoŋ taŋgə.
REL pick fruit pear DEM.PROX have.just go.down ladder

'As they were passing by, they came across the man who was picking pears and just climbed down the ladder.'

(62) *yə tɛŋɔʔ-tɛŋɔʔ, wɔh pɛ aʔdɔʔ s=bako,*
3 RDP-look fruit pear NEG.EXIST one=basket
tigə bako tiŋga sə, duwə, hɔʔ tu kɔsoŋ.
three basket leave one two REL DEM.PROX empty

'He glanced over and noticed that one basket of pears went missing. Only two of three baskets remained, and the other one was empty.'

(63) *pah mmanə yə katə wɔh pɛ yə, sapə cuɣi yə?*
then where 3 say fruit pear 3 who steal 3

'Where did the pears go? Who stole them?'

(64) *mmike yə, lalu ah tigə ɔɣaŋ budɔʔ ɛ̃h.*
think 3 pass.by INTERJ three CLF kid INTERJ

'He was pondering, as the three boys walked by.'

(65) *tigə ɔɣaŋ budɔʔ ni dɔʔ ccaɣə baʔ yə dɔʔ cuɣi,*
3 CLF kid DEM.PROX NEG speak because 3 NEG steal
yə lalu sajə=jə hɔ̃.
3 pass.by only=just AFF

'The three boys didn't say anything, because they didn't steal the pears. They just passed by.'

(66) *lalu sajə, yə dɔʔ cuɣi, hɔʔ saʔdi cuɣi.*
pass.by only 3 NEG steal REL just.now steal

'They just passed by. They didn't steal, it was the other boy who stole (the pears).'

(67) *lalu, j-jalaŋ tigə ɔɣaŋ.*
pass.by INTR-road three person

'The three guys passed by and kept walking.'

(68) *j-jalaŋ tigə ɔɣaŋ, yə dɔʔ hɛɣaŋ pɔkɔʔ manə ɛh.*
INTR-road three person 3 NEG wonder tree which INTERJ

'The three guys went on their way, and they didn't give any thought to the trees.'

B.3 Sample text of ITM

Conversation/Monologue: Old days

The following text is an excerpt of a daily conversation between the speaker and me, which turned out to be almost a monologue of her telling stories. The conversation took place in Kampung Dusun, Ulu Terengganu, Terengganu (citation code: ITM_220920_cv01). A few sentences of my attempted speech in ITM have been removed from this sample text.

(1) *kaləʊ ubɛi təʊ mulɛ, jəɣəʊŋ təʊ ubɛi, jəɣəʊŋ*
TOP tuber DEM.DIST beginning corner DEM.DIST tuber corner
nɛiŋ jagəʊŋ dalaŋ umɔh uɣɔŋ təʊ, baɲɔʔ makanaŋ.
DEM.PROX corn inside house person DEM.DIST much food(SM)

'As for tubers (i.e., yam and sweet potato) during those days ... (there were) tubers in one corner and corn in the other corner of people's houses – lots of food.'

(2) *lɛnɛiŋ adɛ bəɣah smatɛ-matɛ bəɣah dalaŋ umɔh, taʔdɔʔ,*
now EXIST rice only rice inside house NEG.EXIST
taʔdɔʔ makanaŋ.
NEG.EXIST food(SM)

'Now there is only rice in the house, and (there is) no food.'

(3) *kacɔŋ-tanɔh təʊ, musɛiŋ-musɛiŋ ujaŋ gaʔ,*
bean-ground DEM.DIST RDP-reason rain EMPH
guɣɛiŋ ŋə kuliʔ təʊ sədaʔ.
fry with skin DEM.DIST delicious

'The peanuts during the rainy season, when fried with their pod, were delicious.'

(4) *mɔʔ mɔʔciʔ*[2] *nɛiŋ, diɣɛ sukɛ buwaʔ makaŋ.*
mother auntie DEM.PROX 3SG like make food

'My mother liked to cook.'

[2] Used as a form of address to refer to the speaker herself.

(5) *s=kalɛi diyɛ guɣɛiŋ ikaŋ, sə ... s=kutɔʔ bəsɔ nɛiŋ.*
one=time 3SG fry fish one(SM) ... one=box big DEM.PROX
'When she fried fish, she would make a big box like this.'

(6) *diyɛ wɛi makaŋ.*
3SG give food
'Then she shared the food (with others).'

(7) *diyɛ guɣɛiŋ cəluʔ təpəʊŋ.*
3SG fry dip flour
'She would coat the fish with batter before frying.'

(8) *kaləʊ jagəʊŋ təʊ, s=puɣəʊʔ bəsɔ-bəsɔ nɛiŋ.*
TOP corn DEM.DIST one=pot RDP-big DEM.PROX
'When it comes to corn, she would make a large pot like this.'

(9) *gɛi skulɔh puŋ dɔʔ bwɔʔ, dɔʔ bwɔʔ pitih mulɛ.*
go school also NEG bring NEG bring money beginning
'When we went to school, we didn't bring any money.'

(10) *taʔdɔʔ kɛtiŋ, skulɔh kapəʊŋ nɛh.*
NEG.EXIST canteen school village DEM.PROX
'There was no canteen at the village school.'

(11) *diyɛ haʔ puku s=pulɔh təʊ, balɛiʔ k=umɔh.*
3SG REL o'clock one=ten DEM.DIST return to=house
'It was like we needed to go back home at 10 o'clock.'

(12) *balɛiʔ umɔh, makaŋ nasiʔ=kə, makaŋ=lah.*
return house eat rice=or eat=SFP
'We went back home, ate rice or ... just ate.'

(13) *mɔʔ puŋ taʔdɔʔ d=umɔh, gɛi ... k=umɛ=lah byasɛ diyɛ,*
mother also NEG.EXIST LOC=house go ... to=paddy=FOC usually 3SG
gɛi kəbuŋ-kəbuŋ nuŋ.
go RDP-farm there
'My mother was not at home either; she usually went to the paddy field or the farms there.'

(14) *kaləʊ ... kaləʊ adɛ musɛiŋ-musɛiŋ buwɔh təʊ,*
TOP ... TOP EXIST RDP-season fruit DEM.DIST
puwah caɣɛi buwɔh duləʊ, baɣəʊ gɛi skulɔh nuŋ.
satisfied search fruit first only.then go school there

'During the fruit season, we would search for fruits first before heading back to the school.'

(15) *nɛiŋ kuciʔ ... kuciʔ bajəʊ nɛh,*
DEM.PROX pocket ... pocket shirt DEM.PROX
kaŋ=paka skəʔ mulɛ=kaŋ, gɛi skulɔh paka skəʔ.
TAG=wear 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.'

(16) *apa kəɣɔŋ, kəɣah kuciʔ bajəʊ nɛh.*
until hard hard pocket shirt DEM.PROX

'(We'd stuff the fruits in the pocket) till the pocket became hard.'

(17) *dɔh, dəŋaŋ jagəʊŋ guɣɛiŋ ŋə, ubɛi ɣəbuh ŋə sumbaʔ dalaŋ*
well with corn fry ANAPH tuber boil ANAPH stuff inside
kuciʔ nɛh ... kuciʔ bajəʊ skulɔh tah.
pocket DEM.PROX ... pocket shirt school DEM.DIST

'Well, (we would also) stuff fried corn and boiled sweet potatoes in the pocket ... in the school uniform pocket.'

(18) *dɔʔ napɔʔ balah.*
NEG see reply

'We could not look back.'

(19) *dɔh kaləʊ panah-panah ginɛiŋ dɔh,*
well TOP RDP-hot like.this already
skulɔh nuŋ mulɛ manɛ adɛ kipah-kipah.
school there beginning where EXIST RDP-fan

'Well during hot days like this, the school didn't have fans back then.'

(20) *waya?=lah ci?gəʊ, nɔ? gɛi biɣɔ? k=ai nuŋ, padaha gɛi mandɛi.*
tell=FOC teacher want go poo to=water there in.fact go shower
'We would tell the teacher that we wanted to poo in the river, but in fact we went there for a shower.'

(21) *panah-panah.*
RDP-hot
'(It was) very hot.'

(22) *mulɛ manɛ adɛ tandah skulɔh, gɛi suŋa ah,*
beginning where EXIST toilet school go river INTERJ
gɛi k=ai nuŋ.
go to=water there
'There was no toilet at school back then, so we just went to the river.'

(23) *lɛnɛiŋ panah, panah lagɛi.*
now hot hot more
'Nowadays it's hot, much hotter.'

(24) *mulɛ, ta?dɔ? kipah bə-s=buti puŋ umɔh uɣɔŋ.*
beginning NEG.EXIST fan INTR-one=CLF also house person
'Back then, there was no fan in any house.'

(25) *hɔ̃, sədaʔ yə dudəʊʔ a, lɛnɛiŋ,*
AFF nice 3SG sit INTERJ now
dudəʊʔ dalaŋ ɛkɔŋ puŋ dɔʔ jadɛi.
sit inside aircon(ENG) also NEG work.out
'Yeah while it was nice to sit there back then, nowadays, even sitting in an air-conditioned room won't do.'

(26) *hɔ̃, panah lɛnɛiŋ. mulɛ taʔdɔʔ, kipah taʔdɔʔ.*
AFF hot now beginning NEG.EXIST fan NEG.EXIST
'Yeah it's hot now. There was no fan back then.'

(27) *nɔʔ kipah gwanɛ, ɛlɛtriʔ puŋ taʔdɔʔ ... taʔdɔʔ.*
want fan how electricity(ENG) also NEG.EXIST ... NEG.EXIST
'How could we have fans, there was not even any electricity.'

(28) *kaləʊ bulɛih dəŋɔ ɣɛdiyu haʔ paka bɛtəɣi tah,*
if can listen radio(ENG) REL use battery(ENG) DEM.PROX
s=kəɣaʔ malaŋ dəŋɔ uɣɔŋ məɲaɲi dalaŋ ɣɛdiyu
one=chunk night listen person sing(SM?) inside radio(ENG)
tah, ɔlɔh sədaʔ saŋaʔ dɔh ... hɔ̃.
DEM.DIST INTERJ nice much already ... AFF

'If we could listen to a battery-powered radio, (we would) listen to people singing on the radio all night long ... ah that was so enjoyable.'

(29) *dɣama-dɣama dalaŋ ɣɛdiyu tah, sədaʔ.*
RDP-drama(SM) inside radio DEM.DIST nice

'The dramas on the radio, so nice.'

(30) *tiŋuʔ tibi tah, adɛ tigɛ buwɔh tibi, tigɛ buwɔh dalaŋ*
watch TV DEM.DIST EXIST three CLF TV three CLF inside
kapəʊŋ nɛh masɛ təʊ tibi ... umɔh-umɔh ujəʊŋ nuŋ.
village DEM.PROX time DEM.DIST TV ... RDP-house end there

'When it comes to watching TV, there were only three TVs in the village back then, at the houses on the far end of the village.'

(31) *kaʔ slikɔh təʊ s=buwɔh, tapi bukaŋ umɔh diyɛ.*
near corner DEM.DIST one=CLF but NEG house 3SG

'There was one at the corner, but it was not in their house.'

(32) *nuŋ ... atah nuŋ s=buwɔh, ujəʊŋ kapəʊŋ nuŋ s=buwɔh.*
there ... top there one=CLF end village there one=CLF

'Another quite far down, and another on the far end of the village.'

(33) *haʔ ɣajiŋ gɛi umɔh nuŋ ŋaŋ umɔh nuŋ.*
REL often go house there and house there

'The ones we often went to were that house and that house.'

(34) *gɛi umɔh nuŋ, kənɛ dudəʊʔ bɔwɔh umɔh,*
go house there must sit below house
mulɛ umɔh uɣɔŋ paka tiyɔŋ=kaŋ, tiŋgɛi.
beginning house person use pole=TAG tall

'When we went to that house, we had to sit beneath the house; houses in those days were raised on tall poles, you know?'

(35) *dudəʊʔ=lah bɔwɔʔ xxətah-xxətah, dudəʊʔ bɔwɔh umɔh.*
sit=FOC bring RDP-paper sit below house
'We'd just sit down, bringing papers and sitting beneath the house.'

(36) *diyɛ b-kəda, jadɛi diyɛ təbəʊʔ=lah lubɔŋ ... lubɔŋ tikaʔ,*
3SG INTR-shop so 3SG poke=FOC hole ... hole window
pitəʊ-malɛiŋ diyɛ təʊ lətɔʔ tibi.
door-thief 3SG DEM.DIST put TV
'The owner opened a shop there, so he made a hole, a hole in the window to put the TV.'

(37) *hɔ̃, tiŋuʔ ditəʊ=lah, s=kapəʊŋ tiŋuʔ ŋə.*
AFF watch there=SFP one=village watch ANAPH
'Yeah we would watch TV there with the whole village.'

(38) *nɔʔ-nɔʔ kaləʊ ... uɣɔŋ maiŋ tiŋju ... tiŋju ... bɔʔsiŋ.*
especially if ... person play boxing ... boxing ... boxing(ENG)
'Especially if people played "tinju"... boxing.'

(39) *pakaʔ gɛi, lagɛi ... paka nɛh, paka jənərɛtə ... hɔ̃*
agree go again ... use DEM.PROX use generator(ENG) ... AFF
baɣəʊ ... diyɛ paka jənərɛtə.
only.then ... 3SG use generator(ENG)
'We would go there together. He used a generator ... only then ... he used a generator.'

(40) *tapi diyɛ ... diyɛ dɔʔ mitɔʔ duwiʔ,*
but 3SG ... 3SG NEG request money
adɛ kapəʊŋ-kapəʊŋ diyɛ mitɔʔ duwiʔ.
EXIST RDP-village 3SG request money
'But he didn't ask for money; some villages would charge.'

(41) *pɣənɔh mɔʔciʔ tidu umɔh tuwaŋ tibi təh sapa kə pagɛi.*
ever auntie sleep house owner TV DEM.DIST until to morning
'I once slept overnight at the house with the TV until morning. '

(42) *hɔ̃, tiŋuʔ tibi dəŋaŋ kawaŋ, duwɛ uɣɔŋ.*
AFF watch TV with friend two CLF
'Yeah I was watching TV with a friend, the two of us.'

(43) *umɔh pɔʔ ... pɔʔ sdaɣɛ diyɛ=lah, tibi... haʔ adɛ tibi təʊ.*
house uncle ... uncle sibling 3=FOC TV REL EXIST TV DEM.DIST

'That was her uncle's house, the one with the TV.'

(44) *lagɛi dɔʔ, diyɛ dɔʔ ... tɯwaŋ ɣumɔh təʊ dɔʔ staʔ,*
again NEG 3SG NEG ... owner house DEM.DIST NEG start(ENG)
əh jənərɛtə təʊ ... ibəʊ tibi uɣɔŋ paŋgi ...
HES generator(ENG) DEM.DIST ... mother TV person call ...
tuŋgəʊ=lah ditəʊ.
wait=FOC there

'One more thing, when the TV owner didn't start the generator, which we called the "TV head", everyone would just wait.'

(45) *bukaŋ s=uɣɔŋ dɯwɛ, bə-pulɔh-pulɔh uɣɔŋ.*
NEG one=person two INTR-RDP-ten(SM?) person

'It was not only one or two people, but a whole bunch of people.'

(46) *tuŋgəʊ ŋə, nɔʔ suɣɔh tɯwaŋ umɔh təʊ bukɛ tibi,*
wait ANAPH want request owner house DEM.DIST open TV
nɔʔ tiŋuʔ tibi.
want watch TV

'We'd wait and ask the TV owner to turn on the TV, (because) we wanted to watch TV.'

(47) *tauŋ ... tujəʊh puluhaŋ ... tauŋ lapaŋ puluhaŋ təʊ,*
year ... seven ten's ... year eight ten's DEM.DIST
adɛ tibi waɣənɛ-waɣənɛi ... hɔ̃, tibi kalə dɔh.
EXIST TV RDP-colour(SM) ... AFF TV colour(ENG) already

'In the 70s ... in the 80s, there were colour TVs already ... yes, there were TVs with colours already.'

(48) *asa tubiʔ gambɔ=yə, tiŋuʔ ah.*
as.soon.as come.out picture=just watch INTERJ

'As soon as the image came out, we would just watch it.'

(49) *tapi mɔʔciʔ dəɣah bulɛih mutusika.*
but auntie fast can motorcycle(ENG)

'But I learnt to ride a motorcycle pretty early.'

(50) *uɣɔŋ ppuwaŋ haʔ ptamɛ skalɛi=lah bulɛih naiʔ mutusika*
person female REL first most=FOC can ride motorcycle(ENG)
mɔʔciʔ.
auntie

'The first woman who could ride a motorcycle was me.'

(51) *mulɛ-mulɛ ... baʔ ayɔh adɛ mutu,*
RDP-beginning ... because father EXIST motorcycle(ENG)
mɔʔciʔ bulɛih naiʔ mutu.
auntie can ride motorcycle(ENG)

'At first ... because my father had a motorcycle, I got the chance to ride it.'

(52) *maiŋ puŋ cuɣɛi dəŋaŋ ayɔh təʊ, takuʔ ...*
play also steal with father DEM.DIST afraid ...
takuʔ diyɛ maɣɔh nuŋ.
afraid 3SG angry there

'Whenever I did ride it, I did it secretly, fearing that he would be angry with me.'

(53) *gɛi galɔh, naiʔ pɣau ... bidu, kapəʊŋ sbəɣɔŋ nuŋ,*
go pole ride boat ... Bidur village across there
diyɛ tiŋga mutu diyɛ dinuŋ.
3SG leave motorcycle 3SG there

'I used to row a boat ... (to) Bidur, the village across the river; my dad would leave his motorcycle there.'

(54) *diyɛ balɛiʔ xxəjɛ pətaŋ-pətaŋ,*
3SG return work RDP-afternoon
mɔʔciʔ gɛi maiŋ ah bəɣɔŋ nuŋ, dəŋaŋ kawaŋ.
auntie go play INTERJ side there with friend

'(After) he came back from work in the afternoon, I would go there to ride the motorcycle, with a friend.'

(55) *masɛ təʊ taʔdɔʔ kənaiʔkaŋ ah jalaŋ nɛiŋ,*
time DEM.DIST EXIST vehicle(SM) INTERJ road DEM.PROX
adɛ bɣapɛ uɣɔŋ=yə mutu kapəʊŋ nɛiŋ,
NEG.EXIST few person=just motorcycle(ENG) village DEM.DIST
adɛ duwɛ tigɛ uɣɔŋ=yə.
EXIST two three person=just

'At that time there weren't any vehicles or roads like this. There were only a few people with a motorcycle in this village, two or three at most.'

(56) *tapi mɔʔciʔ=lah, duləʊ bulɛih naiʔ mutu uɣɔŋ ppuwaŋ,*
but auntie=FOC first can ride motorcycle(ENG) person female
kapəʊŋ nɛiŋ.
village DEM.PROX

'But it was me, who was the first girl that could ride a motorcycle in this village.'

(57) *jənih nɔʔ jugɛ, nɔʔ jugɛ ... nɔʔ jugɛ! slagɛi dɔʔ bulɛih.*
kind want also want also ... want also as.long.as NEG can

'(I was) the kind of person who would insist on doing things, as long as I didn't have the chance.'

(58) *tapi alhamdulilah, dɔʔ pɣənɔh jatəʊh dalaŋ mutu.*
but Alhamdulillah NEG ever fall inside motorcycle(ENG)

'But thank goodness, I never fell off a motorcycle.'

Bibliography

Abdul Hamid Mahmood. 1971. *Fonem dialek Kelantan*. Kuala Lumpur: Universiti Malaya. (BA thesis.)

Abdul Hamid Mahmood. 1990. Dialek Terengganu: Satu tinjauan ringkas. *Dewan Bahasa* 34(3). 217–225.

Abdul Hamid Mahmood. 1994. *Sintaksis dialek Kelantan*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Abdullah bin Abdul Kadir. 1949. *The voyage of Abdullah* (*A translation from the Malay*), *with notes and appendices*. Trans. by A. E. Coope. Singapore: The Malaya Publishing House.

Abdullah bin Mohamed (Nakula). 1981. *Keturunan raja-raja Kelantan dan peristiwa-peristiwa bersejarah*. Kota Bharu: Perbadanan Muzium Negeri Kelantan.

Abdullah Junus. 1977. *Dialek Ulu Telemong: Satu huraian fonologi dan leksikal*. Kuala Lumpur: Universiti Malaya. (BA thesis.)

Abramson, Arthur S. 1986. The perception of word-initial consonant length: Pattani Malay. *Journal of the International Phonetic Association* 16. 8–16. https://doi.org/10.1017/S0025100300003054.

Abramson, Arthur S. 1987. Word-initial consonant length in Pattani Malay. In Thomas V. Gamkrelidze (ed.), *Proceedings of the 11th International Congress of Phonetic Sciences, Vol. 6*, 68–70. Tallinn: Academy of Sciences of the Estonian USSR.

Abramson, Arthur S. 2003. Acoustic cues to word-initial stop length in Pattani Malay. In Maria-Josep Solé, Daniel Recasens & Joaquin Romero

(eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (*ICPhS XV*), 387–390. Barcelona: Universitat Autònoma de Barcelona.

Adelaar, Alexander. 1983. Malay consonant-harmony: An internal reconstruction. In James T. Collins (ed.), *Studies in Malay Dialects. Part 1* (NUSA 16), 57–67.

Adelaar, Alexander. 1984. Some Proto-Malayic affixes. *Bijdragen tot de Taal-, Land- en Volkenkunde* 140(4). 402–421.

Adelaar, Alexander. 1988. More on Proto-Malayic. In Mohd. Thani Ahmad & Zaini Mohamed Zain (eds.), *Rekonstruksi dan cabang-cabang Bahasa Melayu Induk*, 59–77. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Adelaar, Alexander. 1992. *Proto Malayic: The reconstruction of its phonology and parts of its lexicon and morphology*. Canberra: Department of Linguistics, Research School of Pacific Studies, The Australian National University.

Adelaar, Alexander. 1993. The internal classification of the Malayic subgroup. *Bulletin of the School of Oriental and African Studies* 56(3). 566–581. https://doi.org/10.1017/S0041977X00007710.

Adelaar, Alexander. 1994. The history of thing, animal, person and related concepts in Malay. In A. K. Pawley & M. D. Ross (eds.), *Austronesian terminologies: Continuity and change* (Pacific Linguistics C-127), 1–20. Canberra: Department of Linguistics, Research School of Pacific Studies, The Australian National University.

Adelaar, Alexander. 1995. Borneo as a cross-roads for comparative Austronesian linguistics. In Peter Bellwood, James J. Fox & Darrell Tryon (eds.), *The Austronesians: Historical and comparative perspectives*, 81–102. Canberra: Department of Anthropology, Research School of Pacific and Asian Studies, The Australian National University.

Adelaar, Alexander. 1997. An exploration of directional systems in West Indonesia and Madagascar. In Gunter Senft (ed.), *Referring to space: Studies in Austronesian and Papuan languages* (Oxford Studies in Anthropological Linguistics). Oxford: Clarendon Press.

Adelaar, Alexander. 2004a. À la recherche d'affixes perdus en malais. In Elizabeth Zeitoun (ed.), *Fait de langues: Revue de linguistique n° 23-24: Les langues austronésiennes*, 165–176. Paris: OPHRYS.

Adelaar, Alexander. 2004b. Where does Malay come from? Twenty years of discussions about homeland, migrations and classifications. *Bijdragen tot de Taal-, Land- en Volkenkunde* 160(1). 1–30.

Adelaar, Alexander. 2005a. Much ado about *di-*. *Bijdragen tot de Taal-, Land- en Volkenkunde* 161(1). 127–142.

Adelaar, Alexander. 2005b. *Salako or Badameá: Sketch grammar, text and lexicon of a Kanayatn dialect in West Borneo*. Wiesbaden: Otto Harrassowitz KG.

Adelaar, Alexander. 2005c. Structural diversity in the Malayic group. In Alexander Adelaar & Nikolaus P. Himmelmann (eds.), *The Austronesian languages of Asia and Madagascar*, 202–226. London: Routledge.

Adelaar, Alexander. 2005d. The Austronesian languages of Asia and Madagascar: A historical perspective. In Alexander Adelaar & Nikolaus P. Himmelmann (eds.), *The Austronesian languages of Asia and Madagascar*, 1–42. London: Routledge.

Adelaar, Alexander. 2008. On the classifiability of Malayic. In Yuri A. Lander & Alexander K. Ogloblin (eds.), *Language and text in the Austronesian world. Studies in Honour of Ülo Sirk*, 1–22. Munich: Lincom Europa.

Adelaar, Alexander. 2018. Dialects of Malay/Indonesian. In Charles Boberg, John Nerbonne & Dominic Watt (eds.), *The handbook of dialectology*, 571–581. John Wiley & Sons, Inc. https://doi.org/10.1002/9781118827628.ch36.

Adelaar, Alexander & D. J. Prentice. 1996. Malay: Its history, role and spread. In Stephen A. Wurm, Peter Mühlhäusler & Darrell T. Tryon (eds.), *Atlas of languages of intercultural communication in the Pacific, Asia and the America*, vol. II.1 (Trends in Linguistics. Documentation 13), 673–693. Berlin: De Gruyter Mouton.

Adi Haji Taha. 1985. The re-excavation of the rockshelter of Gua Cha, Ulu Kelantan, West Malaysia. *Federation Museums Journal* 30.

Adi Yasran Abdul Aziz. 2005. Inventori konsonan dialek Melayu Kelantan: Satu penilaian semula. *Jurnal Pengajian Melayu* 16. 200–217.

Adi Yasran Abdul Aziz. 2010. Inventori vokal dialek Melayu Kelantan: Satu penilaian semula. *Jurnal Linguistik* 11. 1–19.

Adi Yasran Abdul Aziz. 2011. Suku kata dasar dialek Kelantan berdasarkan Teori Optimaliti. *GEMA Online® Journal of Language Studies* 11(2). 121–136.

Adi Yasran Abdul Aziz. 2012. Analisis koda berdasarkan kekangan dalam dialek Kelantan. *GEMA Online® Journal of Language Studies* 12(4). 1127–1145.

Aikhenvald, Alexandra Y. 2006. Serial verb constructions in typological perspective. In Alexandra Y. Aikhenvald & R. M. W. Dixon (eds.), *Serial verb constructions: A cross-linguistic typology* (Explorations in Linguistic Typology), 1–68. Oxford: Oxford University Press.

Aikhenvald, Alexandra Y. 2007. Typological distinctions in word-formation. In Timothy Shopen (ed.), *Language typology and syntactic description*, Second edition, vol. III: Grammatical categories and the lexicon, 1–65. Cambridge: Cambridge University Press.

Aikhenvald, Alexandra Y., R. M. W. Dixon & Nathan M. White. 2020. The essence of 'word'. In Alexandra Y. Aikhenvald, R. M. W. Dixon & Nathan M. White (eds.), *Phonological word and grammatical word: A cross-linguistic typology* (Explorations in Linguistic Typology), 1–24. Oxford: Oxford University Press.

Ajid Che Kob. 1985. *Dialek geografi Pasir Mas* (Monograf 3, Institut Bahasa Kesusasteraan dan Kebudayaan Melayu). Bangi: Penerbit Universiti Kebangsaan Malaysia.

Ajid Che Kob. 1997. Word final nasal in Malay dialects. In Cecilia Odé & W. A. L. Stokhof (eds.), *Proceedings of the Seventh International Conference on Austronesian Linguistics, Leiden, 22-27 August 1994*, 35–43. Amsterdam: Rodopi.

Ajid Che Kob. 2008. Subklasifikasi dialek Melayu Patani-Kelantan-Terengganu: Satu analisis kualitatif. *Jurnal Pengajian Melayu* 3. 1–10.

Andaya, Barbara Watson. 1986. The making of a tributary state: Siam and Terengganu in the eighteenth century. In Ronald D. Renard (ed.), *Anuson Walter Vella*, 156–192. Honolulu: Southeast Asia Papers, Center for Asian and Pacific Studies, University of Hawaii at Manoa.

Andaya, Barbara Watson & Leonard Y. Andaya. 2017. *A history of Malaysia*. Third edition. London: Palgrave.

Andaya, Barbara Watson & Yoneo Ishii. 1992. Religious developments in Southeast Asia, c. 1500–1800. In Nicholas Tarling (ed.), *The Cambridge history of Southeast Asia*, vol. 1: From early times to c. 1800, 508–571. Cambridge: Cambridge University Press.

Andaya, Leonard Y. 2001. The search for the 'origins' of Melayu. *Journal of Southeast Asian Studies* 32(3). 315–330.

Anderbeck, Karl. in print. Historical linguistics of the Malayic subgroup. In Antoinette Schapper & Alexander Adelaar (eds.), *Oxford companion to*

Malayo-Polynesian languages in Southeast Asia. Oxford: Oxford University Press.

Anderbeck, Karl. 2012. The Malayic-speaking Orang Laut. *Wacana* 14(2). 265–312.

Anderbeck, Karl. 2019. Rethinking Proto-Malayic ultimate *ə: Geographical evidence. Paper presented at the Twenty-Ninth Meeting of the Southeast Asian Linguistics Society (SEALS 29), Tokyo, Japan, 25-27 May 2019.

Anderbeck, Karl & Doug Cooper. 2017. Malayic compiled wordlist database. Computer database. Kuala Lumpur.

Anderson, Stephen R. 2005. *Aspects of the theory of clitics* (Oxford Studies in Theoretical Linguistics). Oxford: Oxford University Press.

Arai, Takayuki. 2004. Formant shift in nasalization of vowels. *The Journal of the Acoustical Society of America* 115(5). 2541–2541. https://doi.org/10.1121/1.4783589.

Asmah Haji Omar. 1977. *Kepelbagaian fonologi dialek-dialek Melayu*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Asmah Haji Omar. 1985. *Susur galur bahasa Melayu*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Asmah Haji Omar. 1991. *Kepelbagaian fonologi dialek-dialek Melayu*. Edisi kedua. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Asmah Haji Omar. 1995. *Rekonstruksi fonologi bahasa Melayu induk* (Siri Monograf Sejarah Bahasa Melayu). Kuala Lumpur: Dewan Bahasa dan Pustaka.

Asmah Haji Omar. 2008. *Susur galur bahasa Melayu*. Edisi kedua. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Bauer, Christian. 1992. Mon-Aslian contacts. *Bulletin of the School of Oriental and African Studies* 55. 532–537.

Bauer, Laurie. 2003. *Introducing linguistic morphology*. Second edition. Edinburgh: Edinburgh University Press.

Beddor, Patrice Speeter. 1983. *Phonological and phonetic effects of nasalization of vowel height*. Minneapolis: University of Minnesota. (Doctoral dissertation.)

Beddor, Patrice Speeter, Rena Arens Krakow & Louis M. Goldstein. 1986. Perceptual constraints and phonological change: A study of nasal vowel height. *Phonology Yearbook* 3. 197–217.

Bellwood, Peter. 1993. Cultural and biological differentiation in Peninsular Malaysia: The last 10,000 years. *Asian Perspectives* 32(1). 37–60.

Benjamin, Geoffrey. 1987. Ethnohistorical perspectives on Kelantan's prehistory. In Nik Hassan Shuhaimi bin Nik Abd Rahman (ed.), *Kelantan zaman awal: Kajian arkeologi dan sejarah di Malaysia*, 108–153. Kota Bharu: Perbadanan Muzium Negeri Kelantan.

Benjamin, Geoffrey. 1997. Issues in the ethnohistory of Pahang. In Nik Hassan Shuhaimi bin Nik Abd. Rahman, Mohamed Mokhtar Abu Bakar, Ahmad Hakimii Khairuddin & Jazamuddin Baharuddin (eds.), *Pembangunan arkeologi pelancongan negeri Pahang*, 82–121. Pekan: Lembaga Muzium Negeri Pahang.

Benjamin, Geoffrey. 2012. The Aslian languages of Malaysia and Thailand: An assessment. In Stuart McGill & Peter K. Austin (eds.), *Language Documentation and Description*, vol. 11, 136–230. School of Oriental and African Studies, University of London.

Blench, Roger. 2006. Why are Aslian-speakers Austronesian in culture? Paper presented at the Preparatory meeting for the Third International Conference on Austronesian Linguistics (ICAL-3), Siem Reap, Cambodia, 28-29 June 2006.

Blevins, Juliette. 1995. Syllable in phonological theory. In John A. Goldsmith (ed.), *The handbook of phonological theory*, 206–244. Cambridge, MA: Blackwell.

Blevins, Juliette. 2004. *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Blust, Robert. 1979. Proto-Western Malayo-Polynesian vocatives. *Bijdragen tot de Taal-, Land- en Volkenkunde* 135(2). 205–251. https://doi.org/10.1163/22134379-90002556.

Blust, Robert. 1985. The Austronesian homeland: A linguistic perspective. *Asian Perspectives* 26(1). 45–67.

Blust, Robert. 1988. Malay historical linguistics: A progress report. In Mohd. Thani Ahmad & Zaini Mohamed Zain (eds.), *Rekonstruksi dan cabang-cabang Bahasa Melayu Induk*, 1–33. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Blust, Robert. 1990. Patterns of sound change in the Austronesian languages. In Philip Baldi (ed.), *Linguistic change and reconstruction methodology* (Trends in Linguistics Studies and Monographs Studies and Monographs 45), 231–267. Berlin: Mouton de Gruyter.

Blust, Robert. 1994. The Austronesian settlement of Mainland Southeast Asia. In Karen L. Adams & Thomas John Hudak (eds.), *Papers from the*

Second Annual Meeting of the Southeast Asian Linguistics Society 1992, 25–83. Tempe: Arizona State University.

Blust, Robert. 1995. Notes on Berawan consonant gemination. *Oceanic Linguistics* 31(1). 123–138.

Blust, Robert. 1996. The Neogrammarian hypothesis and pandemic irregularity. In Mark Durie & Malcolm Ross (eds.), *The comparative method reviewed: Regularity and irregularity in language change*, 135–156. New York: Oxford University Press.

Blust, Robert. 1997. Nasals and nasalization in Borneo. *Oceanic Linguistics* 36(1). 149. https://doi.org/10.2307/3623074.

Blust, Robert. 1998. In defense of Dempwolff: Austronesian diphthongs once again. *Oceanic Linguistics* 37(2). 354. https://doi.org/10.2307/3623415.

Blust, Robert. 2004. Austronesian nasal substitution: A survey. *Oceanic Linguistics* 43(1). 73–148. https://doi.org/10.1353/ol.2004.0004.

Blust, Robert. 2007. Disyllabic attractors and anti-antigemination in Austronesian sound change. *Phonology* 24(1). 1–36. https://doi.org/10.1017/S0952675707001108.

Blust, Robert. 2009. *The Austronesian languages* (Pacific Linguistics 602). Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, Australian National University.

Blust, Robert. 2013. *The Austronesian languages*. Revised edition. Canberra: Asia-Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Blust, Robert. 2017. Odd conditions: Context-sensitive sound change in unexpected contexts. *Journal of Historical Linguistics* 7(3). 322–371. https://doi.org/10.1075/jhl.15013.blu.

Blust, Robert & Stephen Trussel. 2023. *The Austronesian Comparative Dictionary*. Web edition. (https://www.trussel2.com/ACD/).

Booij, Geert. 2007. *The grammar of words: An introduction to linguistic morphology*. Second edition. Oxford: Oxford University Press.

Bougas, Wayne. 1990. Patani in the beginning of the XVII century. *Archipel* 39. 113–138. https://doi.org/10.3406/arch.1990.2624.

Braddell, Roland. 1936. An introduction to the study of ancient times in the Malay Peninsula and the Straits of Malacca. *Journal of the Malayan Branch of the Royal Asiatic Society* 14(3). 10–71.

Brown, C. C. 1921. *Perak Malay* (Papers on Malay Subjects Second series). Calcutta: The Committee for Malay Studies, Federated Malay States.

Brown, C. C. 1927. *Kelantan Malay* (Papers on Malay Subjects Second series). Singapore: Government Printing Office.

Brown, C. C. 1935. Trengganu Malay. *Journal of the Malayan Branch of the Royal Asiatic Society* 13(3). 1–111.

Brown, C. C. 1956. *Studies in Country Malay*. London: Luzac & Company.

Burenhult, Niclas. 2005. *A grammar of Jahai*. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Bybee, Joan L., William Pagliuca & Revere D. Perkins. 1990. On the asymmetries in the affixation of grammatical material. In William A. Croft, Suzanne Kemmer & Keith Denning (eds.), *Studies in typology and diachrony: Papers presented to Joseph H. Greenberg on his 75th birthday* (Typological Studies in Language 20), 1–42. Amsterdam: John Benjamins Publishing Company. https://doi.org/10.1075/tsl.20.04byb.

Chambers, J. K. & Peter Trudgill. 1998. *Dialectology*. Second edition. Cambridge: Cambridge University Press.

Cheng, Siok Hwa. 1969. The rice industry of Malaya: A historical survey. *Journal of the Malaysian Branch of the Royal Asiatic Society* 42(2). 130–144.

Chrétien, C. Douglas. 1965. The statistical structure of the Proto-Austronesian morph. *Lingua* 14. 243–270. https://doi.org/10.1016/0024-3841(65)90044-6.

Chung, Siaw-Fong. 2005. *Kena* as a third type of Malay passive. *Oceanic Linguistics* 44(1). 194–214. https://doi.org/10.1353/ol.2005.0017.

Clifford, Hugh. 1897. A journey through the Malay states of Trengganu and Kelantan. *The Geographical Journal* 9(1). 1–37.

Clifford, Hugh & Frank Athelstane Swettenham. 1894. *A dictionary of the Malay language: Malay-English*. Taiping, Perak: The Goverment Printing Office.

Clynes, Adrian. 1997. On the Proto-Austronesian "Diphthongs". *Oceanic Linguistics* 36(2). 347–361. https://doi.org/10.2307/3622989.

Clynes, Adrian. 1999. Rejoinder: Occam and the Proto-Austronesian "Diphthongs". *Oceanic Linguistics* 38(2). 404–408. https://doi.org/10.2307/3623300.

Cœdès, George. 1968. *The Indianized states of Southeast Asia*. Walter F. Vella (ed.). Trans. by Susan Brown Cowing. Honolulu: East-west Center Press.

Cole, Peter, Gabriella Hermon & Yanti. 2008. Voice in Malay/Indonesian. *Lingua* 118(10). 1500–1553. https://doi.org/10.1016/j.lingua.2007.08.008.

Collins, James T. 1980. *Ambonese Malay and creolization theory*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Collins, James T. 1983a. *Dialek Ulu Terengganu*. Kuala Lumpur: Penerbit Universiti Kebangsaan Malaysia.

Collins, James T. 1983b. Diftong dan diftongisasi dialek Ulu Terengganu. *Dewan Bahasa* 27. 562–581.

Collins, James T. 1985. The phonology of Tioman Malay and the reconstruction of Proto-Malay. In Suriya Ratanakul, David Thomas & Suwilai Premsrirat (eds.), *Southeast Asian linguistic studies presented to André-G. Haudricourt*, 541–566. Bangkok: Mahidol University.

Collins, James T. 1987. *Dialek Melayu Sarawak*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Collins, James T. 1989. Malay dialect research in Malaysia: The issue of perspective. *Bijdragen tot de Taal-, Land- en Volkenkunde* 145(2/3). 235–264.

Collins, James T. 1994. Sumbangan dialek Brunei dalam pengkajian sejarah Bahasa Melayu. In *Tinggal landas ke abad 21*, 62–75. Bandar Seri Begawan: Dewan Bahasa dan Pustaka.

Collins, James T. 2001. Contesting straits-Malayness: The fact of Borneo. *Journal of Southeast Asian Studies* 32(03). https://doi.org/10.1017/S0022463401000212.

Collins, James T. 2006. Homelands and the homeland of Malay. In James T. Collins & Awang Sariyan (eds.), *Borneo and the homeland of the Malays*, 1–44. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Collins, James T. & Naseh Hassan. 1981. Kajian dialek Ulu Terengganu: Pemetaan dan kesimpulan awal. *Akademika* 18. 1–35.

Comrie, Bernard. 1976. *Aspect: An introduction to the study of verbal aspect and related problems*. Cambridge: Cambridge University Press.

Crawfurd, John. 1852. *A grammar and dictionary of the Malay language, with a preliminary dissertation*. Vol. 1, dissertation and grammar. London: Smith, Elder and Co.

Davis, Stuart. 1994. Geminate consonants in moraic phonology. *Proccedings of the West Coast Conference on Formal Linguistics* 13. 32–45.

Davis, Stuart & Natsuko Tsujimura. 2014. Non-concatenative derivation: Other processes. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford handbook of derivational morphology*, 190–218. Oxford: Oxford University Press.

de Swart, Henriëtte. 2012. Verbal aspect. In Robert I. Binnick (ed.), *The Oxford handbook of tense and aspect*, 752–780. Oxford: Oxford University Press.

Diffloth, Gérard. 2005. The contribution of linguistic palaeontology to the homeland of Austro-Asiatic. In Laurent Sagart, Roger Blench & Alicia Sanchez-Mazas (eds.), *The peopling of East Asia: Putting together archaeology, linguistics and genetics*, 77–80. London: RoutledgeCurzon.

Dixon, R. M. W. 1988. *A grammar of Boumaa Fijian*. Chicago: The University of Chicago Press.

Dixon, R. M. W. & Alexandra Y. Aikhenvald. 2002. Word: A typological framework. In R. M. W. Dixon & Alexandra Y. Aikhenvald (eds.), *Word: A cross-linguistic typology*, 1–41. Cambridge: Cambridge University Press.

Dobby, E. H. G. 1942. Settlement patterns in Malaya. *Geographical Review* 32(2). 211–232. https://doi.org/10.2307/210271.

Dobby, E. H. G. 1951. The Kelantan Delta. *Geographical Review* 41(2). 226–255. https://doi.org/10.2307/211021.

Donohue, Mark. 1999. *A grammar of Tukang Besi* (Mouton Grammar Library 20). Berlin: Mouton de Gruyter.

Donohue, Mark. 2009. Dental discrepancies and the sound of Proto Austronesian. In Bethwyn Evans (ed.), *Discovering history through language: papers in honour of Malcolm Ross*, 271–287. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Downs, R. E. 1960. A rural community in Kelantan, Malaya. *Studies on Asia* 1(1). 51–62.

Dryer, Matthew S. 2007. Noun phrase structure. In Timothy Shopen (ed.), *Language typology and syntactic description*, Second edition, vol. II: Complex constructions, 151–205. Cambridge: Cambridge University Press.

Dunn, Michael, Nicole Kruspe & Niclas Burenhult. 2013. Time and place in the prehistory of the Aslian languages. *Human Biology* 85(1-3). 383–400. https://doi.org/10.3378/027.085.0318.

Durie, Mark. 1985. *A grammar of Acehnese on the basis of a dialect of north Aceh* (Verhandelingen van het Koninklijk Instituut voor Taal-, Land- en Volkenkunde 112). Dordrecht: Foris Publications.

Dyen, Isidore. 1956. Language distribution and migration theory. *Language* 32(4). 611. https://doi.org/10.2307/411084.

Eberhard, David M., Gary F. Simons & Charles D. Fennig (eds.). 2023. *Ethnologue: Languages of the World*. Twenty-fifth edition. Dallas, Texas: SIL International. (http://www.ethnologue.com).

El Zarka, Dina. 2005. On the borderline of reduplication: Gemination and other consonant doubling in Arabic morphology. In Bernhard Hurch (ed.), *Studies on reduplication* (Empirical Approaches to Language Typology 28), 369–394. Berlin: Mouton de Gruyter.

Erbaugh, Mary S. 2001. *The Chinese Pear Stories - Narratives across Seven Chinese Dialects*. http://pearstories.org/.

Fabb, Nigel. 2001. Compounding. In Andrew Spencer & Arnold M. Zwicky (eds.), *The handbook of morphology*, 66–83. Oxford: Blackwell.

Farid Mohamed Onn. 1976. *Aspects of Malay phonology and morphology*. University of Illinois at Urbana-Champaign. (Doctoral dissertation.)

Fedden, Sebastian, Dunstan Brown & Greville G. Corbett. 2010. *Conditions on pronominal marking: A set of 42 video stimuli for field elicitation*. University of Surrey.

Firth, Raymond. 1943. The coastal people of Kelantan and Trengganu, Malaya. *The Geographical Journal* 101(5/6). 193–205. https://doi.org/10.2307/1789626.

Firth, Raymond. 1966. *Malay fishermen: Their peasant economy*. Second edition. Routledge & Kegan Paul Ltd.

Gerini, Colonel G. E. 1909. *Researches on Ptolemy's geography of Eastern Asia* (*further India and Indo-Malay Archipelago*). London: Royal Asiatic Society and Royal Geographical Society.

Gil, David. 2002. The prefixes *di-* and *N-* in Malay/Indonesian dialects. In Fay Wouk & Malcolm Ross (eds.), *The history and typology of western Austronesian voice systems*, 241–283. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Gil, David. 2020. What does it mean to be an isolating language? The case of Riau Indonesian. In David Gil & Antoinette Schapper (eds.), *Austronesian undressed: How and why languages become isolating* (Typological Studies in Language 129), 9–95. Amsterdam: John Benjamins Publishing Company.

Gosling, L. A. Peter. 1978. Contemporary Malay traders in the Gulf of Thailand. In Karl L. Hutterer (ed.), *Economic exchange and social interaction in Southeast Asia: Perspectives from prehistory, history, and ethnography*,

73–95. Ann Arbor: The University of Michigan Center for South and Southeast Asian Studies.

Graham, W. A. 1908. *Kelantan, a state of the Malay Peninsula: A handbook of information*. Glasgow: James Maclehose and sons.

Greenberg, Joseph H. 1957. Order of affixing: A study in general linguistics. In *Eassys in linguistics*, 86–94. New York: Wenner-Gren Foundation for Anthropological Research.

Hajek, John & Rob Goedemans. 2003. Word-initial geminates and stress in Pattani Malay. *The Linguistic Review* 20(1). 79–94. https://doi.org/10.1515/tlir.2003.003.

Haji Abdul Jalil bin Haji Anuar, Abd. Hamid bin Mahmood, Tengku Fauziah binti Ku Mat, Haji Abdullah bin Haji Salleh & Ismail bin Salleh (eds.). 2016. *Glosari dialek Kelantan*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Hammarström, Harald, Robert Forkel, Martin Haspelmath & Sebastian Bank. 2023. *Glottolog 4.7*. Jena: Max Planck Institute for the Science of Human History. (http://glottolog.org).

Hardy, Heather K. & Timothy R. Montler. 1988. Imperfective gemination in Alabama. *International Journal of American Linguistics* 54(4). 399–415.

Hashim bin Musa. 1974. *Morfemik dialek Melayu Kelantan*. Kuala Lumpur: Universiti Malaya. (MA thesis.)

Haspelmath, Martin. 2001. Word classes and parts of speech. In Neil J. Smelser & Paul B. Baltes (eds.), *International encyclopedia of the social and behavioral sciences*, 16538–16545. Amsterdam: Pergamon.

Haspelmath, Martin & Andrea D. Sims. 2010. *Understanding morphology*. Second edition (Understanding Language Series). London: Hodder Education.

Hayes, Bruce. 1989. Compensatory lengthening in moraic phonology. *Linguistic Inquiry* 20. 253–306.

Heine, Bernd & Tania Kuteva. 2004. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.

Henderson, Eugénie J. A. 1965. The topography of certain phonetic and morphological characteristics of South East Asian languages. *Lingua* 15. 400–434.

Hill, A. H. 1951. Kelantan padi planting. *Journal of the Malayan Branch of the Royal Asiatic Society* 24(1). 56–76.

Himmelmann, Nikolaus P. 2005. The Austronesian languages of Asia and Madagascar: Typological characteristics. In Alexander Adelaar & Nikolaus P. Himmelmann (eds.), *The Austronesian languages of Asia and Madagascar*, 110–181. London: Routledge.

Hogan, David W. 1988. *Urak Lawoi': Basic structures and a dictionary* (Pacific Linguistics C-109). Canberra: Department of Linguistics, Research School of Pacific Studies, The Australian National University.

Höhle. 1988. Vorwort und Nachwort zu Verumfokus. *Sprache und Pragmatik* 5(1). 1–7.

Höhle, Tilman N. 1992. Über Verum-Fokus im Deutschen. In Joachim Jacobs (ed.), *Informationsstruktur und Grammatik*, 112–141. Opladen: Westdeutscher Verlag.

Hoogervorst, Tom. 2015. Detecting pre-modern lexical influence from South India in Maritime Southeast Asia. *Archipel* 89. 63–93.

Hsu, Yun-Ts'iao. 1947. Notes on Tan-Tan. *Journal of the Malayan Branch of the Royal Asiatic Society* 20(1). 47–63.

Hsu, Yun-Ts'iao. 1961. *Nanyang shi* [*The history of South Seas*]. Vol. 1. Singapore: Shijie Shuju.

Hudson, Alfred B. 1970. A note on Selako: Malayic Dayak and Land Dayak languages in western Borneo. *Sarawak Museum Journal* 18(36-37). 301–318.

Ibrahim Syukri. 1985. *History of the Malay Kingdom of Patani* (*Sejarah Kerajaan Melayu Patani*). Trans. by Conner Bailey & John N. Miksic (Monographs in International Studies, Southeast Asia Series 68). Athens, Ohio: Ohio University Center for International Studies.

Inglese, Guglielmo. 2022. Towards a typology of middle voice systems. *Linguistic Typology* 26(3). 489–531. https://doi.org/10.1515/lingty-2020-0131.

Inkelas, Sharon. 2005. Morphological Doubling Theory: Evidence for morphological doubling in reduplication. In Bernhard Hurch (ed.), *Studies on reduplication* (Empirical Approaches to Language Typology 28), 65–88. Berlin: Mouton de Gruyter.

Inkelas, Sharon. 2014. Non-concatenative derivation: Reduplication. In Rochelle Lieber & Pavol Štekauer (eds.), *The Oxford handbook of derivational morphology*, 169–189. Oxford: Oxford University Press.

Inkelas, Sharon & Cheryl Zoll. 2005. *Reduplication: Doubling in morphology*. Cambridge: Cambridge University Press.

Ismail bin Dahaman, Abdul Jalil bin Haji Anuar, Nor Hazizan bin Talib, Nawi bin Ismail, Hashim bin Abdul Rahman, Haji Alias bin Haji Omar, Abdullah bin Muhammad & Mohd Ali bin Hitam (eds.). 1997. *Glosari dialek Terengganu*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Ismail Hussein. 1973. Malay dialects in the Malay Peninsula. *Nusantara* 3. 69–79.

Jacq-Hergoualc'h, Michel. 2002. *The Malay Peninsula: Crossroads of the maritime silk road* (*100 CD - 1300 AD*). Trans. by Victoria Hobson. Leiden: Brill.

Jones, Russel (ed.). 2007. *Loan-words in Indonesian and Malay*. Leiden: KITLV Press.

Joseph, Brian D. 2001. Diachronic morphology. In Andrew Spencer & Arnold M. Zwicky (eds.), *The handbook of morphology*, 351–373. Oxford: Blackwell.

Kemmer, Suzanne. 1993. *The middle voice* (Typological Studies in Language 23). Amsterdam: John Benjamins Publishing Company.

Kenstowicz, Michael J. 1994. *Phonology in generative grammar*. Cambridge: Basil Blackwell Publishers.

Klamer, Marian. 1998. *A grammar of Kambera* (Mouton Grammar Library 18). Berlin: Mouton de Gruyter.

Klamer, Marian, Owen Edwards, Hanna Fricke, Zoi Gialitaki, Francesca Moro, Axel Palmér, George Saad, Yunus Sulistyono, Eline Visser & Jiang Wu. 2021. Practicalities of language data collection and management in and around Indonesia. *Wacana* 22(2). 467–521.

Klamer, Marian & Francesca R. Moro. 2020. What is "natural" speech? Comparing free narratives and Frog stories in Indonesia. *Language Documentation & Conservation* 14. 238–313.

Koh, Ann Sweesun. 1990. *Topics in colloquial Malay*. Melbourne: University of Melbourne. (Doctoral dissertation.)

Kraehenmann, Astrid. 2011. Initial geminates. In Marc van Oostendorp, Colin J. Ewen, Elizabeth Hume & Keren Rice (eds.), *The Blackwell companion to phonology*, vol. II, Suprasegmental and prosodic phonology, 1124–1146. Malden, MA: Wiley-Blackwell.

Kroeger, Paul R. 2005. *Analyzing grammar: An introduction*. Cambridge: Cambridge University Press.

Kruspe, Nicole, Niclas Burenhult & Ewelina Wnuk. 2015. Northern Aslian. In Mathias Jenny & Paul Sidwell (eds.), *The handbook of Austroasiatic languages*, vol. 1, 419–474. Leiden: Brill.

Lieber, Rochelle. 2010. *Introducing morphology* (Cambridge Introductions to Language and Linguistics). Cambridge: Cambridge University Press.

Linehan, W. 1936. A history of Pahang. *Journal of the Malayan Branch of the Royal Asiatic Society* 14(2). 1–257.

Linehan, W. 1951. The identifications of some of Ptolemy's place-names in the Golden Khersonese. *Journal of the Malayan Branch of the Royal Asiatic Society* 24(3). 86–98.

Lloyd, R. J. 1921. On the phonology of the Malay and Negrito dialects spoken in the Malay states of Lower Siam. *Bulletin of the School of Oriental Studies, London Institution* 2(1). 27–38.

Lohnstein, Horst. 2016. Verum focus. In Caroline Féry & Shinichiro Ishihara (eds.), *The Oxford handbook of information structure*, 290–313. Oxford: Oxford University Press.

Lyons, John. 1977. *Semantics*. Vol. 2. Cambridge: Cambridge University Press.

Mackey, William F. 1988. Geolinguistics: Its scope and principles. In Colin H. Williams (ed.), *Language in geographic context*, 20–46. Clevedon: Multilingual Matters.

Mahdi, Waruno. 2005. Old Malay. In Alexander Adelaar & Nikolaus P. Himmelmann (eds.), *The Austronesian languages of Asia and Madagascar*, Second edition, 182–201. London: Routledge.

Marriott, H. 1916. A fragment of the history of Trengganu and Kelantan. *Journal of the Straits Branch of the Royal Asiatic Society* 72. 3–23.

Matisoff, James A. 1975. Rhinoglottophilia: The mysterious connection between nasality and glottality. In Charles A. Ferguson, Larry M. Hyman & John J. Ohala (eds.), *Nasálfest: Papers from a symposium on nasals and nasalization*, 265–287. Stanford: Department of Linguistics, Stanford University.

Matisoff, James A. 2003. Aslian: Mon-Khmer of the Malay Peninsula. *Mon-Khmer Studies* 33. 1–58.

Mayer, Mercer. 1969. *Frog, where are you?* New York: Dial Books for Young Readers.

McCarthy, John J. & Alan S. Prince. 1986. *Prosodic morphology*. University of Massachusetts and Rutgers University. (Manuscript.)

McCarthy, John J. & Alan S. Prince. 1995. Prosodic morphology. In John A. Goldsmith (ed.), *The handbook of phonological theory* (Blackwell Handbooks in Linguistics), 318–366. Cambridge, MA: Basil Blackwell.

McDonnell, Bradley. 2016. *Symmetrical voice constructions in Besemah: A usage-based approach*. Santa Barbara: University of California. (Doctoral dissertation.)

McDonnell, Bradley, Jiang Wu, Timothy Mckinnon & Alexander Adelaar. in print. Malayic languages. In Antoinette Schapper & Alexander Adelaar (eds.), *Oxford companion to Malayo-Polynesian languages in Southeast Asia*. Oxford: Oxford University Press.

McKerron, P. A. B. 1931. A Terengganu vocabulary. *Journal of the Malayan Branch of the Royal Asiatic Society* 9(1). 123–128.

Means, Nathalie. 1998. *Temiar-English, English-Temiar dictionary*. Gordon P. Means (ed.). Saint Paul, Minnesota: Hamline University Press.

Miksic, John N. 1978. Archaeology and palaeogeography in the Straits of Malacca. In Karl L. Hutterer (ed.), *Economic exchange and social interaction in Southeast Asia: Perspectives from prehistory, history, and ethnography*, 155–175. Ann Arbor: The University of Michigan Center for South and Southeast Asian Studies.

Mills, J. V. 1930. Eredia's description of Malaca, Meridional India, and Cathay (Translated from the Portuguese with notes). *Journal of the Malayan Branch of the Royal Asiatic Society* 8(1). 1–288.

Mohd Hilmi Hamzah, Janet Fletcher & John Hajek. 2016. Closure duration as an acoustic correlate of the word-initial singleton/geminate consonant contrast in Kelantan Malay. *Journal of Phonetics* 58. 135–151. https://doi.org/10.1016/j.wocn.2016.08.002.

Mohd Hilmi Hamzah, John Hajek & Janet Fletcher. 2018. Non-durational acoustic correlates of word-initial consonant gemination in Kelantan Malay: The potential roles of amplitude and f0. *Journal of the International Phonetic Association*. 1–38. https://doi.org/10.1017/S0025100318000142.

Mohd Tarmizi Hasrah. 2018a. *Analisis sejarah fonologi dan pengelompokan dialek Melayu hulu di bahagian timur Semenanjung Malaysia*. Penang: Universiti Sains Malaysia. (Doctoral dissertation.)

Mohd Tarmizi Hasrah. 2018b. Dialek Melayu di Hulu Kelantan: Laporan awal. *Jurnal Bahasa* 18(2). 219–252.

Mohd Tarmizi Hasrah. 2018c. Dialek Melayu hulu dan hilir di timur Semenanjung Malaysia. *Jurnal Bahasa* 18(1). 65–102.

Mohd Tarmizi Hasrah. 2020. Dialek Pasir Raja: Ciri fonologi dan pengelompokan. *Jurnal Bahasa* 20(2). 173–202.

Moro, Francesca R. & Hanna Fricke. 2020. Contact-induced change in Alorese *give*-constructions. *Oceanic Linguistics* 59(1-2). 116–147. https://doi.org/10.1353/ol.2020.0008.

Muller, Jennifer S. 2001. *The phonology and phonetics of word-initial geminates*. Columbus: The Ohio State University. (Doctoral dissertation.)

Naseh Hasan. 1981. *Fonologi subdialek Terengganu: Kajian kes di Mukim Jenagor, Ulu Terengganu.* Bangi: Universiti Kebangsaan Malaysia. (BA thesis.)

Nash, Manning. 1974. *Peasant citizens: Politics, religion, and modernization in Kelantan, Malaysia* (Papers in International Studies Southeast Asia Series 31). Athens, Ohio: Ohio University Center for International Studies.

Newbold, T. J. 1839. *Political and statistical account of the British settlements in the Straits of Malacca, viz. Pinang, Malacca, and Singapore; with a history of the Malayan states on the Peninsula of Malacca*. Vol. 2. London: John Murray.

Nik Safiah binti Haji Abdul Karim. 1965. *Loghat Melayu Kelantan: Suatu cherakinan kajibunyi bahasa*. Kuala Lumpur: Universiti Malaya. (MA thesis.)

Nik Safiah binti Haji Abdul Karim. 1966. Loghat Kelantan: Huraian fonoloji dan chetetan pendek mengenai sifat-sifat umumnya. *Dewan Bahasa* 10(6). 258–264.

Nik Safiah binti Haji Abdul Karim. 1967. Meninjau beberapa aspek phonologi loghat Kelantan. *Dewan Bahasa* 11(8). 357–362.

Nik Safiah Karim, Farid M. Onn, Hashim Hj. Musa & Abdul Hamid Mahmood. 1986. *Tatabahasa dewan*. Vol. 1: Ayat. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Nomoto, Hiroki. 2013. On the optionality of grammatical markers: A case study of voice marking in Malay/Indonesian. In Alexander Adelaar (ed.), *Voice variation in Austronesian languages of Indonesia* (NUSA 54), 121–143.

Nomoto, Hiroki & Abd. Wahab Kartini. 2012. *Kena* adversative passives in Malay, funny control, and covert voice alternation. *Oceanic Linguistics* 51(2). 360–386. https://doi.org/10.1353/ol.2012.0017.

Nor Hashimah Jalaluddin, Wan Athirah Adilah Wan Abdul Halim & Harishon Radzi. 2021a. Sorotan dialek di Terengganu selepas tiga dekad: Per-

spektif geolinguistik. *GEMA Online® Journal of Language Studies* 21(3). 77–102.

Nor Hashimah Jalaluddin, Wan Athirah Adilah Wan Abdul Halim & Khairul Ashraaf Saari. 2021b. The continuum of Terengganu dialect along the East Coast Malaysia: A geolinguistic study. *Journal of Nusantara Studies* 6(1). 176–198.

Noraien Mansor, Noor Rohana Mansor & Normaliza Abd Rahim. 2013. *Dialek Melayu Terengganu Pesisir Pantai*. Kuala Terengganu: Penerbit Universiti Malaysia Terengganu.

Norfatiha Othman & Nor Hayati Sa'at. 2022. Nelayan laut dalam di sektor perikanan Malaysia. In Khoirul Rosyadi & Nor Hayati Sa'at (eds.), *Masyarakat nelayan Malaysia dan Indonesia*, 147–168. Yayasan Pustaka Obor Indonesia.

Nothofer, Bernd. 1996. Migrasi orang Melayu Purba: Kajian awal. *Sari* 14. 33–52.

Nothofer, Bernd. 1997. *Dialek Melayu Bangka*. Bangi: Penerbit Universiti Kebangsaan Malaysia.

Nur Adibah Hasan & Sharifah Raihan Syed Jaafar. 2017. Struktur suku kata terbitan dialek Terengganu. *Akademika* 87(1). 165–175.

Ogloblin, Alexander. 2018. Notes on structural distinctions in Malay dialects. *Wacana* 19(2). 327. https://doi.org/10.17510/wacana.v19i2.706.

Ohala, John J. 1975. Phonetic explanations for nasal sound patterns. In Charles A. Ferguson, Larry M. Hyman & John J. Ohala (eds.), *Nasálfest: Papers from a symposium on nasals and nasalization*, 289–316. Stanford: Department of Linguistics, Stanford University.

Paauw, Scott H. 2008. *The Malay contact varieties of Eastern Indonesia: A typological comparison*. The State University of New York at Buffalo. (Doctoral dissertation.)

Paterson, H. S. 1924. An early Malay inscription from Trengganu. *Journal of the Malayan Branch of the Royal Asiatic Society* 2(3). 252–258.

Payne, Thomas Edward. 1997. *Describing morphosyntax: A guide for field linguists*. Cambridge: Cambridge University Press.

Pepys, W. E. 1916. A Kelantan glossary. *Journal of the Straits Branch of the Royal Asiatic Society* 74. 303–321.

Pigeaud, Theodore G. Th. 1960. *Java in the 14th century, a study in cultural history: The Nāgara-Kĕrtāgama by Rakawi Prapañca of Majapahit, 1365 A.D.* Third edition. Vol. III: Translations (Koninklijk Instituut voor Taal-,

Land- en Volkenkunde Translation Series 4). The Hague: Martinus Nijhoff.

Ras, J. J. 1970. Lange consonanten in enige Indonesische talen II. *Bijdragen tot de Taal-, Land- en Volkenkunde* 126(4). 429–447.

van Reijn, E. O. 1974. Some remarks on the dialects of North Kerintji: A link with the Mon-Khmer languages. *Journal of the Malaysian Branch of the Royal Asiatic Society* 47. 130–138.

Rentse, Anker. 1934. History of Kelantan. I. *Journal of the Malayan Branch of the Royal Asiatic Society* 12. 44–62.

Richard, McGinn. 1982. *Outline of Rejang syntax* (NUSA 14). Jakarta: Universitas Atma Jaya.

Richards, Anthony. 1981. *An Iban-English dictionary*. Oxford: Clarendon Press.

Rockhill, W. W. 1915. Notes on the relations and trade of China with the Eastern Archipelago and the coast of the Indian Ocean during the fourteenth Century. Part II. *T'oung Pao* 16(1). 61–159.

Rodriguez, Béatrice. 2009. *The chicken thief*. Wellington: Gecko Press.

Ross, Malcolm. 2004. Notes on the prehistory and internal subgrouping of Malayic. In John Bowden & Nikolaus Himmelmann (eds.), *Papers in Austronesian subgrouping and dialectology*, 97–110. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Ross, Malcolm. 2013. Diagnosing contact processes from their outcomes: The importance of life stages. *Journal of Language Contact* 6(1). 5–47. https://doi.org/10.1163/19552629-006001002.

Rubino, Carl. 2005. Reduplicatino: Form, function and distribution. In Bernhard Hurch (ed.), *Studies on reduplication* (Empirical Approaches to Language Typology 28), 11–29. Berlin: Mouton de Gruyter.

Rubino, Carl. 2013. Reduplication. In Matthew S. Dryer & Martin Haspelmath (eds.), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (http://wals.info/chapter/27).

Ruslan Uthai. 2011. *Keistimewaan dialek Melayu Patani*. Bangi: Penerbit Universiti Kebangsaan Malaysia.

Sapir, Edward. 1968. Time perspective in aboriginal American culture: A study in method. In David G. Mandelbaum (ed.), *Selected writings of Edward*

Sapir in language, culture and personality, 398–462. Berkeley: University of California Press.

Schachter, Paul & Timothy Shopen. 2007. Parts-of-speech systems. In Timothy Shopen (ed.), *Language typology and syntactic description*, Second edition, vol. I: Clause structure, 1–60. Cambridge: Cambridge University Press.

Scott, N. C. 1956. *A dictionary of Sea Dayak*. School of Oriental and African Studies, University of London.

Seidlitz, Eric. 2005. *A study of Jakun Malay: Some aspects of its phonology*. Bangi: Universiti Kebangsaan Malaysia. (MA thesis.)

Selkirk, Elisabeth O. 1995. The prosodic structure of function words. In Jill N. Beckman, Laura W. Dickey & Suzanne Urbanczyk (eds.), *Papers in Optimality Theory*, vol. 18 (University of Massachusetts Occasional Papers in Linguistics), 439–470. Amherst, MA: Graduate Linguistics Student Association (GLSA), University of Massachusetts.

Sharifah Raihan Syed Jaafar. 2018. Nasal substitution and the limited role of *NÇ̥ in Malay dialects. *Journal of the Southeast Asian Linguistics Society* 11(1). 35–46.

Sheikh Othman bin Sheikh Salim (ed.). 2007. *Kamus Dewan*. Edisi keempat. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Shellabear, W. G. 1916. *An English-Malay dictionary*. Singapore: The Methodist Publishing House.

Sheppard, M. C. 1949. A short history of Trengganu. *Journal of the Malayan Branch of the Royal Asiatic Society* 22(3). 1–74.

Shorto, Harry. 2006. *A Mon-Khmer comparative dictionary*. Paul Sidwell (ed.) (Pacific Linguistics 579). Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Sieveking, G. de G. 1954. Excavations at Gua Cha, Kelantan, 1954. Part I. *Federation Museums Journal* 1/2. 75–138.

Siti Nadiah binti Nuwawi. 2020. *An acoustic study of the monophthongs of Ulu Terengganu Malay dialect*. University of Malaya. (MA thesis.)

Skeat, Walter William & Charles Otto Blagden. 1906. *Pagan races of the Malay peninsula*. Vol. 2. London: Macmillan and Co.

Smith, Alexander D. 2017. *The languages of Borneo: A comprehensive classification*. Honolulu: University of Hawai'i at Mānoa. (Doctoral dissertation.)

Smith, Carlota S. 1997. *The parameter of aspect*. Second edition. Dordrecht: Springer Science+Business Media.

Sneddon, James Neil. 2010. *Indonesian reference grammar*. Alexander Adelaar, Dwi Noverini Djenar & Michael C. Ewing (eds.). Second edition. Sydney: Allen & Unwin.

Soh, Hooi Ling. 2013. Voice and aspect: Some notes from Malaysian Malay. In Alexander Adelaar (ed.), *Voice variation in Austronesian languages of Indonesia* (NUSA 54), 159–173.

Soh, Hooi Ling & Hiroki Nomoto. 2009. Progressive aspect, the verbal prefix *meN-* and stative sentences in Malay. *Oceanic Linguistics* 48(1). 148–171. https://doi.org/10.1353/ol.0.0026.

Soh, Hooi Ling & Hiroki Nomoto. 2015. Degree achievements, telicity and the verbal prefix *meN-* in Malay. *Journal of Linguistics* 51(1). 147–183. https://doi.org/10.1017/S002222671400019X.

Spencer, Andrew. 2001. Morphophonological operations. In Andrew Spencer & Arnold M. Zwicky (eds.), *The handbook of morphology*, 123–143. Oxford: Blackwell.

Stargardt, Janice. 1983. *Satingpra I: The environmental and economic archaeology of south Thailand*. Oxford: British Archaeological Reports.

Steinhauer, Hein. 2002. More (on) Kerinci sound changes. In Alexander Adelaar & Robert Blust (eds.), *Between worlds: Linguistic papers in memory of David John Prentice*, 149–176. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Steinhauer, Hein & A. Hakim Usman. 1978. Notes on the morphemics of Kerinci (Sumatra). In S. A. Wurm & Lois Carrington (eds.), *Second International Conference on Austronesian Linguistics: Proceedings. Fascicle I. Western Austronesian* (Pacific Linguistics C-61), 483–502. Canberra: Department of Linguistics, Research School of Pacific Studies, The Australian National University.

Štekauer, Pavol, Salvador Valera & Lívia Körtvélyessy. 2012. *Word-formation in the world's languages: A typological survey*. Cambridge: Cambridge University Press.

Sturrock, A. J. 1912. Some notes on the Kelantan dialect, and some comparisons with the dialects of Perak and Central Pahang. *Journal of the Straits Branch of the Royal Asiatic Society* 62. 1–7.

Sutlive, Vinson H & Joanne Sutlive (eds.). 1994. *A handy reference dictionary of Iban and English*. Kuching: Tun Jugah Foundation.

Swettenham, Frank Athelstane. 1885. Journal kept during a journey across the Malay Peninsula. *Journal of the Straits Branch of the Royal Asiatic Society* 15. 1–37.

Tadmor, Uri. 1995. *Language contact and systemic restructuring: The Malay dialect of Nonthaburi, Central Thailand*. Honolulu: University of Hawai'i at Mānoa. (Doctoral dissertation.)

Tadmor, Uri. 2003. Final /a/ mutation: A borrowed areal feature in western Austronesian languages. In John Lynch (ed.), *Issues in Austronesian historical phonology*, 15–35. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Teeuw, A. 1957. Review of *Studies in Country Malay*, by C. C. Brown. *Bijdragen tot de Taal-, Land- en Volkenkunde* 113. 293–297.

Teeuw, A. 1961. *A critical survey of studies on Malay and Bahasa Indonesia* (*with the assistance of H. W. Emanuels*) (Koninklijk Instituut voor Taal-, Land- en Volkenkunde, Bibliographical Series 5). 's-Gravenhage: Martinus Nijhoff.

Teeuw, A. & D. K. Wyatt. 1970. *Hikayat Patani* (*The story of Patani*). Vol. 1. The Hague: Martinus Nijhoff.

Teoh, Boon Seong. 1994. *The sound system of Malay revisited*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Thomason, Sarah Grey. 2001a. Contact-induced typological change. In Martin Haspelmath, Ekkerhard König, Wulf Oesterreicher & Wolfgang Raible (eds.), *Language typology and language universals/Sprachtypologie und sprachliche Universalien: An international handbook*, 1640–1648. Berlin: Walter de Gruyter.

Thomason, Sarah Grey. 2001b. *Language contact: An introduction*. Edinburgh: Edinburgh University Press.

Thomason, Sarah Grey. 2010. Contact explanations in linguistics. In Raymond Hickey (ed.), *The handbook of language contact*, 31–47. Malden, MA: Blackwell.

Thomason, Sarah Grey & Terrence Kaufman. 1988. *Language contact, crealization, and genetic linguistics*. Berkeley: University of California Press.

Thurgood, Elzbieta. 1998. *A description of nineteenth century Baba Malay: A Malay variety influenced by language shift*. Honolulu: University of Hawai'i at Mānoa. (Doctoral dissertation.)

Thurgood, Graham. 1993. Geminates: A cross-linguistic examination. In Joel A. Nevis, Gerald McMenamin & Graham Thurgood (eds.), *Papers in honor of Frederick H. Brengelman on the occasion of the 25th anniversary of the Department of Linguistics*, 129–139. Fresno: Department of Linguistics, California State University, Fresno.

Thurgood, Graham. 1999. *From ancient Cham to modern dialects: Two thousand years of language contact and change* (Oceanic Linguistics Special Publication No. 28). Honolulu: University of Hawai'i Press.

Tjia, Johnny. 2007. *A grammar of Mualang: An Ibanic language of Western Kalimantan, Indonesia*. Utrecht: LOT.

Topintzi, Nina. 2008. On the existence of moraic onset geminates. *Natural Language & Linguistic Theory* 26(1). 147–184. https://doi.org/10.1007/s11049-008-9034-4.

Topintzi, Nina & Stuart Davis. 2017. On the weight of edge geminates. In Haruo Kubozono (ed.), *The phonetics and phonology of geminate consonants*, 260–282. Oxford: Oxford University Press.

Trudgill, Peter. 2010. Contact and sociolinguistic typology. In Raymond Hickey (ed.), *The handbook of language contact*, 299–319. Malden, MA: Blackwell.

Wake, Christopher H. 1964. Malacca's early kings and the reception of Islam. *Journal of Southeast Asian History* 5(2). 104–128. https://doi.org/10.1017/S0217781100000958.

Wales, H. G. Quaritch. 1974. Langkasuka and Tambralinga: Some archaeological notes. *Journal of the Malaysian Branch of the Royal Asiatic Society* 47(1). 15–40.

Wallace, Stephen. 1976. *Linguistic and social dimensions of phonological variations in Jakarta Malay*. Ithaca, NY: Cornell University. (Doctoral dissertation.)

Welch, David J. & Judith R. McNeill. 1989. Archaeological investigations of Pattani history. *Journal of Southeast Asian Studies* 20(1). 27–41. https://doi.org/10.1017/S0022463400019822.

Wheatley, Paul. 1955. The Malay Peninsula as known to the Chinese of the third century A.D. *Journal of the Malayan Branch of the Royal Asiatic Society* 28(1). 1–23.

Wheatley, Paul. 1956. Langkasuka. *T'oung Pao* 44(4/5). 387–412.

Wheatley, Paul. 1973. *The golden Khersonese: Studies in the historical geography of the Malay Peninsula before AD 1500*. Westport, Connecticut: Greenwood Press.

Wilkinson, R. J. 1959. *A Malay-English dictionary*. London: Macmillan.

Winstedt, Richard O. 1923. *Malaya: The Straits Settlements and the Federated and Unfederated Malay States*. London: Constable & Co. Ltd.

Winstedt, Richard O. 1935. A history of Malaya. *Journal of the Malayan Branch of the Royal Asiatic Society* 13(1). iii–270.

Winstedt, Richard O. (ed.). 1938. The Malay Annals or Sejarah Melayu. *Journal of the Malayan Branch of the Royal Asiatic Society* 16(3). 1–226.

Wolff, John U. 2009. The Proto Austronesian laryngeal. In Alexander Adelaar & Andrew Pawley (eds.), *Austronesian historical linguistics and culture history: A festschrift for Robert Blust*, 115–128. Canberra: Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University.

Wolff, John U. 2010. *Proto-Austronesian phonology with glossary*. Vol. I. Ithaca, NY: Cornell University Press.

Wolters, O. W. 1958. Tambralinga. *Bulletin of the School of Oriental Studies, University of London* 21. 587–607.

Wu, Jiang. 2023. *Corpora of Kelantan Malay, Coastal Terengganu Malay and Inland Terengganu Malay*. DataverseNL, V1. https://doi.org/10.34894/HWUVLM.

Wyatt, David K. 1974. Nineteenth century Kelantan: A Thai view. In William R. Roff (ed.), *Kelantan: Religion, society and politics in a Malay state*, 1–21. Kuala Lumpur: Oxford University Press.

Yanti. 2010. *A reference grammar of Jambi Malay*. Newark: University of Delaware. (Doctoral dissertation.)

Yunus Maris, M. 1980. *The Malay sound system*. Kuala Lumpur: Penerbit Fajar Bakti Sdn. Bhd.

Yupho, Nawanit. 1989. Consonant clusters and stress rules in Pattani Malay. *Mon-Khmer Studies* 15. 125–137.

Zaharani Ahmad. 1993. *Fonologi generatif: Teori dan penerapan*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Zaharani Ahmad. 2006. Kepelbagaian dialek dalam bahasa Melayu: Analisis tatatingkat kekangan. *e-Bangi: Jurnal Sains Sosial dan Kemanusiaan* 1(1). 1–26.

Zaharani Ahmad, Nor Hashimah Jalaluddin & Shariffulizan Malek. 2011. Fonologi rangkap vokal dan kepelbagaian dialek Melayu: Analisis Teori Optimaliti. *GEMA Online® Journal of Language Studies* 11(3). 5–30.

Zorc, R. David. 1982. Where, O where, have the laryngeals gone? Austronesian laryngeals re-examined. In Amran Halim, Lois Carrington & S. A. Wurm (eds.), *Papers rom the Third International Conference on Austronesian Linguistics*, vol. 2 (Pacific Linguistics C-75), 111–144. Canberra: Department of Linguistics, Research School of Pacific Studies, The Australian National University.

Zorc, R. David. 1996. The reconstruction and status of Austronesian glottal stop - chimera or chameleon. In Bernd Nothofer (ed.), *Reconstruction, classification, description : Festschrift in honor of Isidore Dyen*, 41–72. Hamburg: Abera-Verl.

Zwicky, Arnold M. 1977. *On clitics*. Bloomington, IN: Indiana University Linguistics Club.

Zwicky, Arnold M. 1985. Clitics and particles. *Language* 61(2). 283–305.

Zwicky, Arnold M. & Geoffrey K. Pullum. 1983. Cliticization vs. inflection: English *n't*. *Language* 59(3). 502–513.

English summary

This dissertation presents a synchronic description and an analysis of the historical developments of the Malayic varieties spoken in Kelantan and Terengganu, two Malaysian states situated on the northeast coast of the Malay Peninsula. Specifically, it examines three distinct varieties, namely Kelantan Malay (KM), Coastal Terengganu Malay (CTM) and Inland Terengganu Malay (ITM). These belong to the Malayic subgroup within the Austronesian language family and are collectively referred to as Northeastern Peninsular Malayic varieties (NEPMs).

Chapter 1 offers an overview of Malayic languages with more detailed background information on NEPMs, as well as an overview of the geo-historical settings of the area where these languages are spoken. It also explains the methodology and data collection of the present study. This dissertation is furthermore divided into two parts. Part one provides descriptive grammar sketches of NEPMs, comprising five chapters. Chapters 2 to 4 focus on the phonology of KM, CTM and ITM respectively. Chapters 5 and 6 describe the basic morphosyntax of NEPMs. Building on the description in Part one, Part two delves into the historical development of NEPMs. Chapters 7 and 8 draw comparisons between NEPMs and the pre-existing reconstructions in their common ancestral language, namely Proto Malayic, investigating the historical phonology and morphology respectively. Sound changes and their chronological order are established, and the potential factors contributing to the morphological reduction observed in these languages are discussed. Chapter 9 is a concluding chapter which draws upon the linguistic data and historical records to infer the migration patterns of

NEPM speakers.

While NEPMs share many common sound patterns, particularly in terms of syllable and word structure, as well as the distribution of consonants, each NEPM variety has its own distinct phonological system. Notably, the presence of initial geminate consonants, both in morphologically simple and complex forms, represents a remarkable and typologically unusual feature within NEPM phonology. The morphosyntactic aspects of NEPMs also display significant similarities, characterised by a largely isolating-analytic profile. Diachronically, the evolution of NEPMs is marked by substantial phonological and morphological reduction. Word-final consonants have often undergone mergers and losses, and only a limited number of prefixes are retained from Proto Malayic. Furthermore, original prefixation has given its way to initial gemination, which is a polyfunctional morphophonological operation. While some previous studies suggest that the morphological reduction may be attributed to language contact, this dissertation argues that the primary driving force behind the morphological reduction is internal phonological changes, and no clear traces of substrate influences have been found.

Among the three NEPM varieties, this study highlights the similarities between KM and CTM on the one hand, and the distinctiveness of ITM on the other hand, which is manifested in various aspects of its grammar including phonology, morphology and specific syntactic structures. The shared changes observed in the consonant systems of NEPMs must also have followed distinct developments in the vowel systems. It is suggested that ITM represents a separate variety, distinct from KM and CTM, and that NEPMs do not form a discrete subgroup within the Malayic languages despite their commonalities. On the basis of the geographical distribution and linguistic differentiation of NEPMs, a two-wave migration pattern is hypothesised. It is posited that the ancestors of ITM speakers arrived earlier and settled in the inland region of Terengganu, while those of KM and CTM speakers arrived at a later stage, primarily dispersing along the coastal areas.

Samenvatting in het Nederlands

Dit proefschrift bevat een synchrone beschrijving en een analyse van de historische ontwikkelingen van de Maleise variëteiten die gesproken worden in Kelantan en Terengganu, twee Maleisische staten aan de noordoostkust van het Maleisisch schiereiland. Drie verschillende variëteiten in het bijzonder worden onderzocht, namelijk Kelantan Maleis (KM), Terengganu Kust Maleis (CTM) en Terengganu Land Maleis (ITM). Ze behoren alle drie tot de Malayische subgroep van de Austronesische taalfamilie en worden ook wel de Noordoostelijk Schiereiland Malayische variëteiten (NEPMs) genoemd.

Hoofdstuk 1 biedt een overzicht van Maleise talen met meer gedetailleerde achtergrondinformatie over NEPMs, en van de geo-historische setting van het gebied waar deze talen gesproken worden. Het legt verder ook de methodologie en gegevensverzameling van deze studie uit. Het overige gedeelte van dit proefschrift bestaat uit twee delen. Deel één bestaat uit vijf hoofdstukken en geeft beschrijvende grammaticale schetsen van de NEPMs. Hoofdstukken 2 tot en met 4 richten zich op de fonologie van respectievelijk KM, CTM en ITM. Hoofdstukken 5 en 6 beschrijven de basale morfosyntaxis van NEPMs. Voortbouwend op de beschrijving in deel één, gaat deel twee in op de historische ontwikkelingen van NEPMs. Voortbouwend op de beschrijving in deel één, gaat deel twee in op de historische ontwikkeling van NEPMs. Hoofdstukken 7 en 8 trekken vergelijkingen tussen de NEPMs en de reeds bestaande reconstructies in hun gemeenschappelijke vooroudertaal, namelijk het Proto Malayisch, en ze onderzoeken de fonologie en morfologie vanuit historisch perspectief. Klankveranderingen en hun chronologische volgorde worden vastgesteld, en een bespreking volgt van de mo-

gelijke factoren die bijgedragen hebben tot de morfologische reductie die in deze talen waargenomen wordt. Hoofdstuk 9 dient als afsluitend hoofdstuk, waarin op basis van de linguïstische gegevens en historische documenten de migratiepatronen van de NEPM-sprekers worden afgeleid.

Elke NEPM variëteit heeft haar eigen fonologische systeem, maar onderling hebben ze veel gemeenschappelijke klankpatronen gemeen, met name in termen van lettergreep- en woordstructuur en in de verdeling van medeklinkers. Vooral de aanwezigheid van gegemineerde beginmedeklinkers is een opmerkelijke en typologisch ongebruikelijke eigenschap binnen de fonologie van de NEPMs, zowel in morfologisch eenvoudige als complexe vormen. De morfosyntactische aspecten van NEPMs vertonen ook significante overeenkomsten, die gekenmerkt worden door een grotendeels isolerend-analytisch profiel. Diachroon wordt de evolutie van NEPMs gekenmerkt door substantiële fonologische en morfologische reductie. Medeklinkers aan het einde van woorden zijn vaak samengevoegd en verloren gegaan en slechts een beperkt aantal voorvoegsels uit het Proto Malayisch is bewaard gebleven. Bovendien heeft oorspronkelijke prefigering plaatsgemaakt voor initiële geminatie, hetgeen een polyfunctioneel-morfofonologische operatie is. Hoewel sommige eerdere studies suggereren dat de morfologische reductie kan worden toegeschreven aan taalcontact, dit proefschrift betoogt dat interne fonologische veranderingen de primaire drijfveer zijn van de morfologische reductie. Bovendien vertonen NEPMs geen duidelijke sporen van substraatinvloed.

Van de drie NEPM-variëteiten belicht deze studie enerzijds de overeenkomsten tussen KM en CTM en anderzijds de distincte eigenschappen van ITM, die tot uiting komen in fonologie, morfologie en specifieke syntactische structuren. De gemeenschappelijke veranderingen die in de medeklinkersystemen van NEPMs worden waargenomen moeten ook verschillende ontwikkelingen in de klinkersystemen hebben gevolgd. Er wordt gesuggereerd dat ITM een aparte variëteit vertegenwoordigt, die verschilt van KM en CTM, en dat, ondanks hun onderlinge overeenkomsten, NEPMs geen aparte subgroep vormen binnen de Malayische talen. De huidige verspreiding en differentiatie van de NEPMs suggereren een migratiepatroon in twee golven. Er wordt verondersteld dat de voorouders van ITM-sprekers eerder aankwamen en zich vestigden in het binnenland van Terengganu, terwijl de voorouders van KM- en CTM-sprekers later arriveerden en zich voornamelijk langs de kustgebieden hebben verspreid.

Ringkasan dalam Bahasa Melayu

Disertasi ini membentangkan deskripsi sinkronik dan analisis perkembangan sejarah dialek-dialek Melayu yang dituturkan di Kelantan dan Terengganu, dua negeri yang terletak di pantai timur laut Semenanjung Tanah Melayu. Secara khususnya, disertasi ini meneliti tiga dialek, iaitu Dialek Kelantan (KM), Dialek Terengganu Pesisir (CTM) dan Dialek Ulu Terengganu (ITM). Ketiga-tiga dialek ini tergolong dalam rumpun Malayik dalam keluarga bahasa Austronesia. Ketiga-tiga dialek tersebut juga dirujuk sebagai Dialek-dialek Malayik Semenanjung Timur Laut (NEPMs) secara kolektifnya dalam disertasi ini.

Bab 1 memberikan gambaran keseluruhan tentang bahasa-bahasa Malayik, serta maklumat latar belakang yang lebih terperinci tentang NEPMs dan maklumat geografi dan sejarah kawasan di mana dialek-dialek tersebut dituturkan. Bab 1 juga menerangkan metodologi dan kaedah pengumpulan data dalam kajian ini. Isi utama disertasi ini terbahagi kepada dua bahagian. Bahagian Pertama terdiri daripada lima bab yang melakarkan tatabahasa deskriptif NEPM. Antaranya, Bab 2 hingga 4 masing-masing memfokuskan fonologi KM, CTM dan ITM. Bab 5 dan 6 menerangkan aspek-aspek morfosintaksis asas NEPM. Berdasarkan deskripsi dalam Bahagian Pertama, Bahagian Kedua menyelidiki perkembangan sejarah NEPM. Bab 7 dan 8 membandingkan NEPMs dengan bahasa purba yang sama, iaitu Proto Malayik, di samping masing-masing menghuraikan sejarah fonologi dan morfologi. Perubahan bunyi dan susunan kronologinya diwujudkan, dan faktor-faktor yang berpotensi menyumbang kepada pengurangan morfologi yang boleh diperhatikan dalam dialek-dialek tersebut juga dibincangkan. Sebagai bab

penutup, Bab 9 menggunakan data linguistik dan rekod sejarah untuk membuat penggulungan corak migrasi penutur NEPM.

Setiap dialek NEPM mempunyai sistem fonologi tersendiri sambil berkongsi banyak pola bunyi yang serupa, terutamanya dalam hal suku kata dan struktur perkataan, serta taburan konsonan. Keistimewaan utama terletak pada kehadiran konsonan panjang di awal perkataan, sama ada kata tunggal mahupun kata terbitan. Ciri ini adalah luar biasa dari segi tipologi fonologi. Aspek morfosintaksis NEPM juga memaparkan persamaan yang ketara. Pada keseluruhannya, ketiga-tiga dialek tersebut mempunyai profil morfologi yang isolatif dan analitik. Evolusi diakronik NEPM ditandai dengan pengurangan fonologi dan morfologi yang ketara. Konsonan akhir kata sering mengalami penggabungan dan kehilangan. Di samping itu, hanya sejumlah kecil imbuhan awalan dikekalkan daripada Proto Malayik. Tambahan pula, awalan asal telah memberikan laluan kepada pemanjangan konsonan awal, satu operasi morfofonologi yang berbagai fungsi. Walaupun beberapa kajian terdahulu telah mencadangkan bahawa pengurangan morfologi mungkin dipengaruhi oleh pertembungan dengan bahasa-bahasa lain, disertasi ini berpendapat bahawa faktor utama yang menyebabkan pengurangan morfologi ialah perubahan fonologi dalaman, dan tidak ada kesan jelas pengaruh substrat telah ditemui.

Kajian ini juga mengetengahkan persamaan antara KM dan CTM, dan keistimewaan ITM di antara tiga dialek NEPMs. Hal ini dimanifestasikan dalam pelbagai aspek tatabahasa, termasuk fonologi, morfologi dan struktur sintaksis tertentu. Perubahan bersama yang diperhatikan dalam sistem konsonan NEPM mestilah juga mengikuti perkembangan yang berbeza dalam sistem vokal. Disertani ini mencadangkan bahawa ITM mewakili satu dialek yang berasingan daripada KM dan CTM. NEPMs tidak membentuk subkumpulan diskret dalam rumpun bahasa Malayik, sesungguhnya terdapat persamaan. Berdasarkan taburan geografi dan pembezaan linguistik NEPM pada masa kini, disertasi ini menghipotesiskan corak migrasi dua gelombang: nenek moyang penutur ITM tiba lebih awal dan menetap di kawasan pedalaman di Terengganu, manakala penutur KM dan CTM tiba kemudian dan kebanyakannya tersebar di sepanjang kawasan pesisir pantai.

Curriculum vitae

Jiang Wu (吴疆) was born on 20th June, 1992, in the city of Rui'an in Zhejiang province in China. He grew up in Hangzhou and completed his primary and secondary education there. He attended Beijing Foreign Studies University from 2010 to 2014, where he obtained a bachelor degree in Malay language and literature. During his undergraduate studies, he spent one exchange semester at the Academy of Malay Studies at the University of Malaya. In February 2015, he came to the Netherlands to pursue a research master degree in linguistics at Leiden University, for which he received a Leiden University Excellence Scholarship. He graduated *cum laude* in January 2017. He was also a research assistant in Prof. Marian Klamer's VICI project *Reconstructing the past through languages of the present: the Lesser Sunda Islands* from 2016 to 2017. In 2017, he was awarded a NWO grant under the scheme "PhDs in the Humanities" for his project titled *Tracking Malay migration history through the genealogy of vernacular Malayic varieties*. This dissertation is the outcome of that project. At the time of his defence, he has taken up a postdoctoral researcher position at Leiden University, in the NWO-funded project *On the representation of quantity: how our brains shape language* led by Prof. Jenny Doetjes. He also works as a freelance editor for the Linguistic Bibliography at Brill.