

# The genesis of Sri Lanka Malay

ed. Sebastian Nordhoff

July 7, 2012



## Table of contents

Part I Overview

Chapter 1 Introduction  
Sebastian Nordhoff

Chapter 2 Synchronic grammar of Sri Lanka Malay  
Sebastian Nordhoff

Part II Sociology, history, and demography

Chapter 3 Sri Lanka Malay: new findings on contacts  
Peter Bakker

Chapter 4 Known, inferable, and discoverable in Sri Lankan Malay research  
Peter Slomanson

Chapter 5 Issues of power and privilege in the maintenance of Sri Lanka Malay: A  
sociolinguistic analysis  
Romola Rassool

Part III Language contact

Chapter 6 The Lexical Sources of Sri Lanka Malay Revisited  
Scott Paauw

Chapter 7 Sri Lankan Languages in the South-South Asia Linguistic Area: Sinhala  
and Sri Lanka Malay  
James W. Gair

Chapter 8 Hijacked constructions in Second Language Acquisition: Implications for  
Sri Lanka Malay  
Ian Smith

Chapter 9 The Semantics of Serial Verb Constructions in Sri Lanka Malay  
Mohamed Jaffar

Chapter 10 The genesis of Sri Lanka Malay as a multi-layered process  
Sebastian Nordhoff



## List of tables

1.1 Comparison of a prototypical Creole setting with the Sri Lankan setting

- 2.1 SLM consonant phonemes
- 2.2 Frequencies of dental and retroflex stops
- 2.3 SLM vowel phonemes
- 2.4 Pronouns
- 2.5 SLM interrogative pronouns
- 2.6 Deictics
- 2.7 SLM numerals
- 2.8 Case marking postpositions
- 2.9 Verbal affixes in SLM
- 2.10 SLM vector verbs and their values
- 2.11 Coordinating clitics
- 2.12 Overview of case frames for predicates of different arity
- 2.13 Negation patterns for various predicate types and tenses

3.1 Abbreviations of creole names in Figure 3.1

3.2 Abbreviations and affiliations of non-creoles

5.1 SLM speech communities

- 6.1 Malay(ic) Varieties
- 6.2 Features collected by Adelaar 1991
- 6.3 Gil's feature sets for strong correlations for SLM with Maluku
- 6.4 Gil's feature sets for strong correlations for SLM with Java
- 6.5 Origin of SLM lexemes in two studies
- 6.6 Importance of the three possible regions of origin for the lexicon
- 6.7 Comparison of the vocabulary of the different varieties
- 6.8 1st and 2nd person pronouns in a number of varieties
- 6.9 Comparison of pronouns between SLM, Java, and Maluku
- 6.10 Loan words from Sri Lankan languages in SLM
- 6.11 Loanwords from other languages in SLM

7.1 The nine features typical of the SSA

- 8.1 The sources of Sourashtra tense-mood-aspect markers
- 8.2 The sources of Sri Lanka Portuguese tense-mood-aspect markers
- 8.3 Summary of verbal noun functions in the five languages

10.1 The different stages of language contact in Sri Lanka Malay

10.2 'Sri Lankan' features in a variety of linguistic domains and the most likely point

- of their emergence
- 10.3 ‘Sri Lankan’ features in phonology
- 10.4 ‘Sri Lankan’ features in morphology
- 10.5 ‘Sri Lankan’ features in syntax
- 10.6 ‘Sri Lankan’ features in semantics
- 10.7 ‘Sri Lankan’ feature in discourse
- 10.8 ‘Sri Lankan’ lexical features
- 10.9 Stage 0 features
- 10.10 Stage 1 features
- 10.11 Stage 2 features
- 10.12 Stage 3 features
- 10.13 Stage 4 features
- 10.14 Stage 5 features
- 10.15 The three phases of dialect levelling according to Trudgill 1986

## List of figures

- 1.1 Malay mosques during the colonial period
- 1.2 Migration during the colonial era
- 1.3 Timeline of Sri Lanka Malay history
  
- 2.1 Coding of semantic roles in SLM
  
- 3.1 Genetic distances between Sri Lankan populations
- 3.2 A tree of Sri Lanka populations with Europeans and Malays
- 3.3 18 creoles and Sri Lanka Malay
- 3.4 Sri Lanka Malay among creoles and noncreoles
  
- 4.1 A Malay *kadutham* from 1919
- 4.2 A page from a Malay marriage register
- 4.3 A page from a Moorish marriage register in a Malay mosque, in which almost half of the marriages are Moor-Malay marriages
- 4.4 Glossary page from a nineteenth century text used to teach Arabic in Qur'anic schools attended by Malays
  
- 5.1 Path from perceived language loss to identity re-alignment
  
- 8.1 Sourashtra cities and towns

## List of contributors

**Ian Smith** holds a doctorate in Linguistics from Cornell University. He has taught at Monash University, the National University of Singapore, the University of Sydney, and at York University, where he is Associate Professor of Linguistics. His primary research field is language contact. The bulk of his work focuses on South Asian contact languages, particularly Sri Lanka Portuguese, Sri Lanka Malay, and Sourashtra. He has also published on English and Kugu Nganhcara.

**Scott Paauw** is an Assistant Professor of Linguistics at the University of Rochester, where he has taught since 2005. He has a PhD in Linguistics from the University at Buffalo. He has worked on Sri Lanka Malay since 2003. His primary areas of interest are language contact, language description and documentation, pidgins and creoles, bilingual issues, historical linguistics, typology, applied linguistics and sociolinguistics. He is particularly interested in these topics as they relate to Malay/Indonesian, Austronesian languages, the languages of Papua, and the languages of South Asia.

**Romola Rassool** is presently working towards the completion of her doctoral thesis at the Asia Institute, University of Melbourne, Australia. Her thesis, which is on sociolinguistic aspects of the Sri Lanka Malay language, is based on extensive fieldwork she undertook in the Western, Eastern, Central and Southern Provinces of Sri Lanka. She is presently on study leave from the University of Kelaniya, Sri Lanka, where she holds the position of senior lecturer in English as a Second Language at the English Language Teaching Unit.

**Sebastian Nordhoff** acquired his PhD from the University of Amsterdam with *A Grammar of Upcountry Sri Lanka Malay*. His main interest are language description and documentation, language contact, and language change. He is also the author of the grammar authoring platform galoes.org as well as of glottolog.org, a knowledgebase of references to linguistic literature and genealogical affiliations, which was constructed at his current workplace, the Max Planck Institute for Evolutionary Anthropology.

**Mohamed Jaffar**, a native Sri Lankan Malay, has steeped himself in the culture, lore and language of his community since his early years. He counts himself among a fortunate few, among Sri Lankan Malays, that can still decipher and read with a great deal of fluency old extant texts in Ceylon Malay ‘gundul’, the modified Arabic script also known as ‘huruf Jawi’. Jaffar is a founder member of the old *Alliance Française de Ceylan* and is a self-confessed Francophile. While in Sri Lanka he worked for *La Compagnie des Messageries Maritimes*, and later on, up to his emigration to Canada in 1974, for *L’Union des Transports Aériens*. He became active in the *Sri Lanka Malay Association of Toronto* and served as its president. In June 2012 he graduated from Toronto’s York University with a specialist B.A. majoring in Linguistics and French. His particular interests comprise reviving lost lexicon, devising a system of Romanized spelling that will prove user-friendly, and, eventually, a primer in the English medium for autodidacts.

**Peter Slomanson** received his Ph.D. in 2005 from the City University of New York. He conducted post-doctoral fieldwork in Sri Lanka, funded by the American Institute for Sri Lankan Studies, and subsequently taught at the College of Staten Island. He has lectured since 2009 at the University of Aarhus, in Denmark. Dr. Slomanson is also affiliated with the Information Structuring and Typology research group of the French *Centre National de Recherche Scientifique*. His research involves modeling non-canonical contact language development, and the roles of discourse-pragmatic accommodation and reanalysis in the development of contact language and second language grammars.

## Acknowledgments

This book is a result of the workshop ‘Past, Present, and Future of Sri Lanka Malay’ held at the Max-Planck Institute for Evolutionary Anthropology in Leipzig in November 2010. I would like thank the director of the linguistics department, Bernard Comrie, for financial support and Claudia Schmidt for taking care of the logistics. I would furthermore like to thank Hugo Cardoso, Susanne Michaelis, John Peterson, David Gil, Umberto Ansaldi, and Lisa Lim for agreeing to review papers of this volume. Finally, I would like to thank the series editors John Peterson and Anju Saxena as well as the staff at Brill, Liesbeth Kanis, Sasha Goldstein, Wendy Shamier, and Jasmin Lange.

## Abbreviations

ABL	ablative	INDEF	indefinite
ACC	accusative	INF	infinitive
ADDIT	additive	INSTR	instrumental
ADHORT	adhortative	INTERR	interrogative
ADJ	adjective	INVOL	involitive
ADJZ	adjectivizer	IPFV	imperfective
ADVZ	adverbializer	IRR	irrealis
ALTER-BEN	alterbenefactive	LOC	locative
ANIM	animate	MASC	masculine
APPL	applicative	NEG	negative
AUG	augment	NONV	nonverbal
AUX	auxiliary	NHON	nonhonorific
BEN	benefactive	NHUM	nonhuman
CAUS	causative	NMLZ	nominalizer
CNJ	conjunction	ORD	ordinal
COMIT	comitative	P	patient
COMPL	completive	PER	permissive
COND	conditional	PFV	perfective
CP	conjunctive participle	PHAT	phatic
DAT	dative	PL	plural
DEB	debitive	POSS	possessive
DEF	definite	PRED	predicate
DES	desiderative	PRON	pronominal
DET	determiner	PROX	proximal
DIST	distal	PRS	present
DUR	durative	PTCP	participle
EMPH	emphatic	QUOT	quotative
EVID	evidential	RED	reduplication
EXPER	experiential	REFL	reflexive
EXPL	expletive	REL	relative
FEM	feminine	REP	reportative
FIN	finite	SELF-BEN	self-benefactive
FOC	focus	SG	singular
FUT	future	SIMIL	similitive
GEN	genitive	SIMULT	simultaneous
HABIT	habitual	SPEC	specific
HBL	habilitive	TNS	tense
HON	honorific	TR	transitive
IMP	imperative	V	verb
INANIM	inanimate		

# **Part I**

# **Overview**



# Chapter 1

## Introduction

Sebastian Nordhoff

*Max Planck Institute for Evolutionary Anthropology*

This book details the genesis of Sri Lanka Malay from a variety of angles. Before delving into contributions from typology, creole studies, Indian studies, Malay studies, historical demography and sociolinguistics, this introduction will give an overview of the context of this fascinating language.

The Malays in Sri Lanka number about 50,000 people. Not all ethnic Malays are speakers of the language, and at present there is a noticeable shift from SLM towards English and Sinhala. The majority of the Malays live in the Colombo metropolitan area and in the smaller towns in the central Upcountry. Another important area is the southeastern coastal area, where we find the highest percentages of Malays on the island in the town of Hambantota (30%) and the village of Kirinda (close to 100%). The Malays – with the exception of Kirinda – are clearly an urban population, which is related to their close ties to the colonial administration.

The Malays were mainly brought as soldiers during the Dutch (1656-1796) and British (1796-1948) rule of Ceylon, as the island was known then. The center of recruitment for the Dutch was Batavia on Java (today Jakarta). The recruits consisted of Javanese, Sundanese, Moluccans, Balinese, Buginese, and a variety of other peoples from the Indonesian archipelago. This group was complemented by royal exiles, convicts, and slaves. Looking at the ethnic makeup of the immigrant population, the term ‘Malay’ is actually a misnomer. The Dutch referred to them as ‘Javanen’ or ‘Oosterlingen’ (Easterlings). The connection with Java is also found in the Sinhala (*jaa minissu*) and Tamil (*cāvakar, jāvā manusar*) ethnonyms. The immigrants came from a variety of ethnic backgrounds and spoke many different languages, but used Trade Malay as a lingua franca for interethnic communication. When the British took over from the Dutch in 1796, they noticed that this group spoke a variety of Malay, so they chose to call them ‘Malays’, although almost none of them were ethnic Malays. The British continued recruiting in Jakarta until Indonesia was given back to the Dutch; the recruitment center shifted to the Straits Settlement then. It was only at this point in time that the first ethnic Malay immigrants came to Sri Lanka, but their linguistic varieties failed to have a significant impact (see [Paauw this volume](#)).

Sri Lanka Malay culture is based on the military. The vast majority of immigrants were mercenaries, who fought in the jungle for the colonial powers, whose own sol-

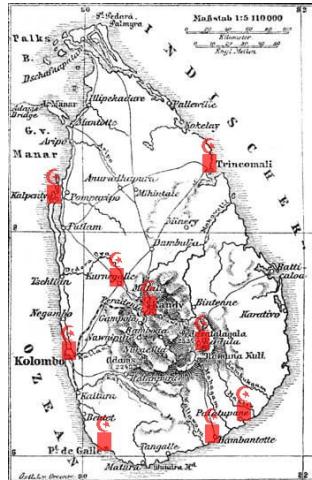


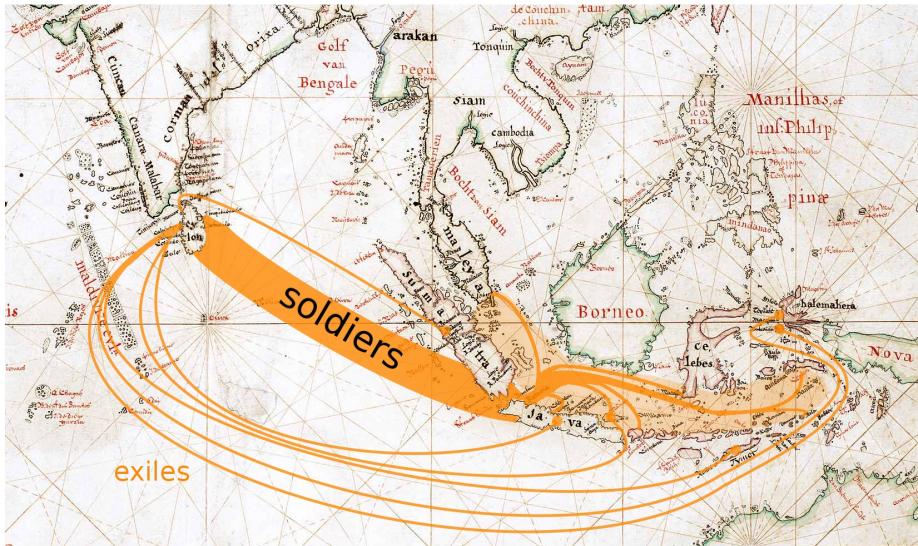
Figure 1.1: Malay mosques during the colonial period. These are still indicative of the distribution of the Malay population today.

diers were not fit for this climate and terrain. The bravery of the Malay soldiers is mentioned frequently, as is their addiction to gambling.

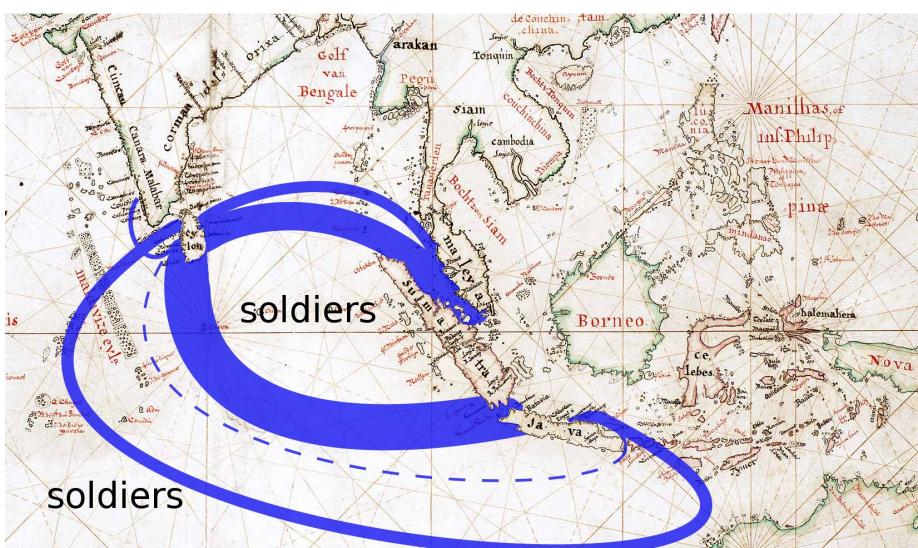
Due to the large number of Malay soldiers, the British chose to set up a special Malay regiment, which turned out to be crucial for the development of Malay culture on the island (Hussainmiya 1987, 1990) as it provided the small group with the proximity and the close-knit network needed to maintain a vibrant diaspora community. It is important to note that the 19th century Malays became the most educated of the native groups, second only to the Burghers (of European ancestry). After the disbandment of the regiment in 1873, the Malays kept their inclination towards the security business and became policemen, firemen, overseers, or prison guards, but were dispersed all over the country. Their command of English remained much higher than that of the other groups due to their close ties with the colonial administration.

The turn of the century saw the rise of communalist tendencies among the Sinhalese, Tamil, and Moor populations, which led to the formation of a number of political organizations based on ethnicity. The Malays also set up their own organization. It was at that time that the Malay hero, TB Jayah, joined the government and played a substantial role during the struggle for independence. Ceylon became independent from the UK in 1948. In 1956, the majoritarian Sinhalese population voted the Sinhala-Only law, which made schooling in Sinhala mandatory. The Malays had up to then used Sri Lanka Malay as the home language and English as the medium of schooling. Given that Sinhala replaced English in school, and that the Malays wanted to keep their comparatively better command of English, they opted for English as the home language. This meant that Malay was pushed out of the home domain, which is where the attrition of the language began.

Ethnic upheaval between Tamils and Sinhalese followed the Sinhala-Only law and dominated the decades to come, eventually leading to the Sri Lankan Civil War (1983-2009). In this war, the Malays were loyal to the powers that be, i.e. the Sinhalese, and joining the army is still a common career choice for Sri Lanka Malays. Their command of Tamil was an asset in this domain, as they could be used behind enemy lines.



(a) Dutch rule



(b) British rule

Figure 1.2: Migration during the colonial era.

1503-1656	Portuguese rule
1656	Start of Dutch rule
-1700	exiles from the Moluccas and the Lesser Sunda Islands; Java; and Baca-n, Tidore, Timor, Madura, and Sumatra.
17th and 18th c.	convicts, slaves
-1796	soldiers (Ambonese, Bandanese, Balinese, Buginese, Javanese, Madurese, Malays)
1783	existence of a Malay mosque in Sri Lanka
1796-1798	British take over from the Dutch
-1816	English recruit soldiers from Jakarta
1819-	English start recruiting on the Malaysian peninsula
1827	foundation of the Ceylon Rifle Regiment
1873	disbandment of the Regiment
1922	foundation of the <i>All Ceylon Malay Association</i>
1924	Malays receive an appointed member of parliament
1948	Independence of Ceylon
1956	Sinhala Only Law
1965	Malays lose their appointed member of parliament
1972	Ceylon changes its name to Sri Lanka
1983	start of the Sri Lankan Civil War
1985	foundation of SLAMAC (Sri Lankan Malay Confederation)
1989	MH Amith is elected MoP for the Sri Lanka Muslim Congress
26 Dec 2004	Tsunami
2009	End of the Sri Lankan Civil War, rebels defeated

Figure 1.3: Timeline of Sri Lanka Malay history.

The 2004 Tsunami struck the community hard. Especially the coastal settlements of Kirinda, Hambantota and parts of Colombo were affected, with Hambantota paying the highest toll: about 1/3 of the Malay population there died on December 26, 2004.

The Upcountry was naturally not directly affected by this disaster, but was still shaken by the consequences of many friends and relatives' hardship.

Detailed sociological information about the Malays is lacking. They normally work in law enforcement and related areas or in white collar jobs. While [Bichsel-Stettler \(1989\)](#) refers to discrimination on the job market, unemployment does not seem to be a greater concern for the Malays today than for all Sri Lankans in general. Malays work as supreme court judges, national sportsmen, film stars, singers, media presenters, important journalists, or high representatives of Sri Lankan Airlines, so that we are not dealing with a systematic discrimination here. It is common to work for several years in the Gulf countries in order to earn enough money to buy a house. Sri Lankans generally live on their own property, or with their (extended) family.

The Malays are Sunni Muslims and often have a very liberal interpretation of their faith. Marriage partners are either other Malays or members of the 'Moor' group, who share their Islamic faith. Marriages to Sinhalese or Burghers do occur, but in these cases, the marriage partner is required to convert to Islam. Based on [Nordhoff \(2009\)](#), we can estimate that about 20% of marriages are with Moors, but more thorough demographic research is required.

The relationship to the Moors is a difficult one. On the one hand, both groups share the Islamic faith, on the other hand, the Malays tend to be much more liberal and inclined towards education than the Moors, who have more traditional views and are more interested in commerce. The Malays resent the Moors' claim to speak for all Muslims —The term 'Muslim' is occasionally taken to mean 'Moor'— and find it important to mark the differences. As a result, the differences between Malays and Moors are occasionally highlighted and the commonalities tend to be played down.

The existence of Sri Lanka Malay was essentially ignored until the end of the 1980s, when [Hussainmiya \(1987, 1990\)](#) and [Bichsel-Stettler \(1989\)](#) started publishing their work on the language. The 1990s saw some interest from senior European scholars ([Adelaar 1991](#), [Bakker 1995](#), [Adelaar & Prentice 1996](#)), and since the early years of this century, the study of Sri Lanka Malay can be regarded as established, both in the fields of Malay studies as well as in the field of contact linguistics.

While in the 20<sup>th</sup> century, basic problems of description and social history were addressed, linguists working on the language in the 21<sup>st</sup> century have focused on problems of genesis, where the relative importance of Tamil on the one hand and Sinhala on the other was the major point of contention. Ramifications of this debate include the existence and extent of intermarriage between Malays and Moors, the shape of the language in the early 19th century and the discussion which features could actually be used to determine the differential input from Sinhala and Tamil.

Sri Lanka Malay is very relevant for the field of contact linguistics because of a certain number of rare characteristics:

- There is no Indo-European or West African input. This distinguishes it from most Creole languages.
- Sri Lanka Malay has mostly retained its lexicon but modified its grammar. This is the opposite of the process of relexification often assumed for Creole genesis.
- Sri Lanka Malay became more complex during its development. It developed bound morphology such as non-finite forms and case markers absent in other

	Creole	SLM
A language spoken	+	+
by displaced people	+	+
from Africa	(+)	-
on an island	(+)	+
in the tropics	+	+
who work for the colonizers	+	+
, who had forced their migration,	+	(+)
on plantations	+	-
under appalling conditions	+	-
converting to Christianity	+	-
changing their original language	+	+
towards the colonizers' language	+	-
retaining the grammar	+	-
and adopting the lexicon,	+	-
leading to a 'simpler' language	+	-
within three to four centuries.	+	+

Table 1.1: Comparison of a prototypical Creole setting with the Sri Lankan setting.

varieties of Malay. Most theories predict that language contact should lead to simplification.

- Sri Lanka Malay changed its grammar almost completely within only 300 years. This speed of development is extraordinary.

A schematic comparison between Sri Lanka Malay and the languages normally studied in creole studies is given in Table 1.1.

In (1), we see examples of the changes undergone by the language, comparing a simple sentence in Sri Lanka Malay with its Colloquial Indonesian translation.

- (1) a. *Saya tinggal di Colombo* COLLOQUIAL INDONESIAN  
     1SG live LOC Colombo  
     'I live in Colombo.' (p.c. David Gil)

b. *Se Kluuṁbu=ka arā-duuduk* SRI LANKA MALAY  
     1SG Colombo=LOC NONPAST-EXIST.ANIM  
     'I live in Colombo.'

We see that the order of constituents in the sentence is verb-medial in Indonesian but verb-final in Sri Lanka Malay. Similarly, Indonesian has a preposition *di*, where Sri Lanka Malay has a postposition *=ka*. This postposition is furthermore encliticized, showcasing the development of bound morphology in Sri Lanka Malay. Bound morphology can also be found in the verbal domain, where Sri Lanka Malay uses the obligatory prefix *arà-*, whereas tense marking in Indonesian is optional and often left omitted.

In the domain of phonology, we can see the development of quantity distinction in the two vowels of *duuduk* and the phonemicization of prenasalised stops in *Kluuňbu*. As for semantics, we can note the development of an animacy distinction in the existential, where animate *duuduk* contrasts with *aada*, which is underspecified for animacy. Both differ from *tinggal* as used in the Indonesian example.

# Bibliography

- Adelaar, K. A., 1991. “Some Notes on the Origin of Sri Lanka Malay.” In *Papers in Austronesian Linguistics*, edited by H. Steinhauer. Canberra: Australian National University, 23–37.
- Adelaar, K. A. & Prentice, D. J., 1996. “Malay: its History, Role and Spread.” In *Atlas of Languages of Intercultural Communication in the Pacific, Asia and the Americas*, edited by Stephen A. Wurm, Mühlhäusler Peter, & Darrell T. Tryon. Berlin: Mouton de Gruyter, 673–693.
- Bakker, Peter, 1995. “Over convergentie in een sprachbund.” In *Artikelen van de Tweede Sociolinguïstische conferentie*, edited by Erica Huls & Jetske Klatter-Folmer. Delft: Eburon, 37–51.
- Bichsel-Stettler, Anne, 1989. *Aspects of the Sri Lanka Malay Community and their Language*. Master’s thesis, University of Bern.
- Hussainmiya, B. A., 1987. *Lost cousins: The Malays of Sri Lanka*. Bangi: Institut Bahasa, Kesusaasteraan dan Kebudayaan Melayu.
- Hussainmiya, B. A., 1990. *Orang Rejimen: The Malays of Sri Lanka*. Bangi: Universiti Kebangsaan Malaysia, 1st edition.
- Nordhoff, Sebastian, 2009. *A Grammar of Upcountry Sri Lanka Malay*. Ph.D. thesis, University of Amsterdam, Utrecht.  
URL <http://www.lotpublications.nl/publish/articles/003744/bookpart.pdf>
- Paauw, Scott, this volume. “The Lexical Sources of Sri Lanka Malay Revisited.”



## Chapter 2

# Synchronic grammar of Sri Lanka Malay

Sebastian Nordhoff

*Max Planck Institute for Evolutionary Anthropology*

### 2.1 Introduction

This paper presents a synchronic sketch of current Sri Lanka Malay as spoken in Kandy and surroundings. It serves as a basic introduction to the language, and as a backdrop for the diachronic chapters, which discuss the development of this language in more detail. In order not to preempt the historical papers, diachronic observations are limited to a minimum in this introduction. The breadth and depth of this sketch are limited by the available space. For more elaboration of the grammatical domains mentioned, the reader is referred to [Nordhoff \(2009\)](#).

### 2.2 Phonology

#### 2.2.1 Consonants

Sri Lanka Malay has 24 native consonants and 3 consonants only used in loanwords (see Table 2.1). The most notable sounds are the sounds traditionally referred to as ‘retroflexes’. These are in opposition to the dentals, which is unique in the Malay world. The term ‘retroflex’ is chosen here in analogy to a similar phonological distinction found in Tamil and Sinhala, even if the SLM articulation is more to the front than what we find in the other languages. Example (1) gives (near-)minimal pairs.

- (1) *baatok baatok*      *daatay daapur*  
cough coconut.shell come oven

The voiceless dental variant is far more common than the voiced dental variant, while for the retroflex consonants this is reversed, with the voiced stop much more common than the voiceless stop. Table 2.2 gives a breakdown of the occurrences of these phonemes in a subset of my corpus. This skewed distribution is due to the asymmetric realization of /t/ and /d/ in the Malay varieties which gave rise to SLM: /t/ is

	labial	dental	retroflex	palatal	velar	glottal
stops						
voiceless	p	t̪	t̪	c	k	
voiced	b	d̪	d̪	ɟ	g	
prenasalized	m̪b		n̪d̪	n̪ɟ	n̪g	
nasals	m		n	n̪	n̪	
fricatives	(f)		s (z) (ʃ)			h
approximants	v			j		
liquids			r l			

Table 2.1: SLM consonant phonemes, including approximants. Palatal stops are phonetically affricated. Parentheses indicate phonemes only found in loanwords.

	dental		apical			
voiceless	186	(64%)	18	(6%)	182	(70%)
voiced	11	(4%)	77	(26%)	77	(30%)
		197 (68%)	95 (32%)			292 (100%)

Table 2.2: Frequencies of dental and retroflex stops.

normally realized as dental, while /d/ is realized as alveolar. This was served as a default articulation in SLM.

The second remarkable feature is the presence of a phonemic series of prenasalized stops. These count as one segment in phonotactics. Near minimal pairs, with indication of syllabification, are given below.

- (2) *baa.pa baa.bet̪<sub>1</sub> braa.ṁbat̪<sub>2</sub>*  
father triples spread

Note that the long vowel in /braaṁbat̪/ is synchronically not an instance of phonetic compensatory lengthening (Wetzels & Wezer 1986), since there are words like *ambel* ‘take’, which do not show a long vowel in front of the sequence NC. There is independent evidence for vowel lengthening in open penultimate syllables, so that an account based on different syllabifications of [am.bel] and [bra:.ṁbat̪] can provide a satisfactory explanation.

### 2.2.2 Vowels

SLM has a six-vowel system (Table 2.3). The five full vowels /a,e,i,o,u/ are easily established and present no problem for the analysis. It can be noted, however, that /e/ and /o/ are remarkably less frequent than /i/ and /u/ which are in turn less frequent than /a/, in a ratio of about 1:1:3:3:9.

i		u	
e	ə	o	
a			

Table 2.3: SLM vowel phonemes.

The status of schwa presents more problems for the analysis. The first descriptions (Hussainmiya 1987, Bichsel-Stettler 1989) note that historical schwa is realized as [i]

or [u] and question its phonemic status. There is also a good deal of variation with regard to the realization as [i] or [u]. The analysis of Tapovanaye (1986, 1995), elaborated by Nordhoff (2009), is able to shed light on this: there is a phonemic schwa, which is realized as [ə] in the final syllable as well as in the antepenultimate. In the penultimate syllable, schwa is raised to [ɪ] or [ʊ] to make up for a lack in moraic weight (see below). The raised allophone of schwa depends on a variety of circumstances, such as phonological environment, but also idiolect. Presence of labials or [u] favours [ʊ], presence of [i] or palatals/velars favours [ɪ]. Cases in point are [d̪uppen] ‘before’ and [d̪ɪkkat̪] ‘vicinity’. Certain speakers raise /ə/ even further and arrive at realization like [i,u], where the sound is indistinguishable from realizations of /i,u/, explaining the initial confusion as to the status of /ə/. Independent of phonetic realization, the moraic structure of the word nearly always allows us to ascertain whether we are dealing with an underlying schwa (which triggers gemination of the following consonant in most contexts) or with an underlying /i/ or /u/.

### 2.2.3 Syllable structure

The SLM syllable structure is

$$(3) \quad (s)(C)(L)V(X)$$

where *s* is extrasyllabic; *C* is any consonant, except for *ŋ* or *t̪* in the onset; *L* is a liquid or a glide; and *X* is either the second part of a long vowel or a coda consonant. The coda consonant on the penultimate is typically either the first part of a geminate or a nasal homorganic to a following stop. Exceptions to this rule are *k* in [seksa] ‘problem’ and *r* in [t̪irban] ‘fly’.

### 2.2.4 Structure of the phonological word

The phonological word typically consists of two syllables. There are only a handful of monosyllabic content words like *tee* ‘tea’, *pon* ‘bride’, *pii* ‘go’. Trisyllabic words are very often derived from a disyllabic stem, most commonly be the nominalizer *-an*. A case in point is *makanan* ‘food’ from *maakang* ‘eat’. Tetrasyllabic words are normally compounds of two disyllabic stems, such as *anak klaaki* ‘child’+‘male’=‘boy’.

The penultimate of disyllables is lengthened when open. This can be explained by extrametricality of the final syllable and a bimoraic foot requirement, as shown in (4), which presents the prototypical structure of the SLM word.

$$(4) \quad (\mu\mu) < \sigma >$$

I will now discuss how the schema in (4) applies to SLM lexemes and how it explains certain phonological phenomena, most notably vowel lengthening, vowel raising, and gemination.

The simplest case is a word which already has sufficient material to build a bimoraic foot without the final syllable in the underlying form. A case in point would be /bantu/ ‘help’, which can be parsed without problems, as shown in (5).

$$(5) \quad (ba_\mu n_\mu) < \mathring{tu} >$$

/ə/ is not able to project a mora in SLM, and would lead to a defective foot. In such cases, underlying /ə/ is realized as [i] or [ʊ] when a mora is required. This is shown for the lexeme /mənt̪a/ ‘raw’.

(6)  $*mən_μ < \underline{t}a >$

(7)  $(m_μ n_μ) < \underline{t}a >$

If there is no coda in the lexeme, the vowel is lengthened if it is /a/, /e/, /i/, /o/, /u/. The extra length of the vowel provides the mora required for the bimoraic foot. This is exemplified with the word /makanŋ/ ‘eat’

- (8) a.  $*ma_μ < kaŋ >$   
b.  $(ma_μ : _μ) < kaŋ >$

In case the underlying vowel of the penult is /ə/, two things happen: first, the vowel is raised as in (7). Second, the onset of the final syllable is geminated, thereby yielding a coda for the penultimate. This is shown for the lexeme /pərang/ ‘war’. Note that b. and c. contain operations which add one mora, but only when they are combined as in d. is there sufficient material to build a foot.

- (9) a.  $*pə < rəŋ >$   
b.  $*p_μ < rəŋ >$   
c.  $*pər_μ < rəŋ >$   
d.  $(p_μ r_μ) < rəŋ >$

As for what concerns trisyllables, this typically contain more segmental material to begin with, so that mora-generating processes are less often required. A typical example is /makanan/ ‘food’, which parses as follows:

(10)  $(ma_μ ka_μ) < naŋ >$

As compared to *maakang* ‘eat’ discussed above, there is no need for vowel lengthening since there are already two moras available. Trisyllables only undergo vowel-lengthening if the first syllable contains a schwa.<sup>1</sup> A case in point is /kəpala/ ‘head’. In order to create an additional mora for the creation of the bimoraic foot, the vowel-lengthening strategy is chosen.

- (11) a.  $*kəpa_μ < la >$   
b.  $kə(p_μ a_μ) < la >$

Raising /ə/ to [i] or [ʊ] would be another theoretical possibility, but this is not employed in the antepenultimate syllable. A theoretical explanation for this is found in [Apoussidou & Nordhoff \(2009\)](#).

Phonological parsing becomes more unstable when dealing with compound words, which are the only ones to have four or more syllables. In phonologically integrated compounds, only one long vowel can be found. This is illustrated for /kaca maŋa/ ‘spectacles’

(12)  $(ka_μ ca_μ)(ma_μ a_μ) < \underline{t}a >$

Not all compounds are phonologically integrated however, and parsings of the type  $(μμ) < σ > (μμ) < σ >$  can also be found.

---

<sup>1</sup>Schwa is not found in the second syllables of trisyllables.

### 2.2.5 Orthography

The orthography of SLM used in this article is based on the Indonesian and Malaysian standard with the following additions and modifications:

- ‘Retroflex’ stops are represented by plain consonants ⟨d,t⟩.
- Dental stops are represented by adding an ⟨h⟩, yielding ⟨dh,th⟩.
- Unraised schwa is written ⟨à⟩; raised schwa is written ⟨ì⟩ or ⟨ù⟩, depending on the feature [back].
- Vowel length is represented by doubling the vowels.
- Consonant length is represented by doubling the consonants.
- For the digraphs ⟨dh, th, ny, ng⟩ only the first part is doubled when geminated, yielding ⟨ddh,<sup>2</sup> tth, nny, nng⟩.
- Prenasalized consonants may optionally be indicated with a breve over the nasal part if this is needed for the topic under discussion ⟨ṁb, ŋd, ŋj, ŋg⟩. In most cases, this is not done.

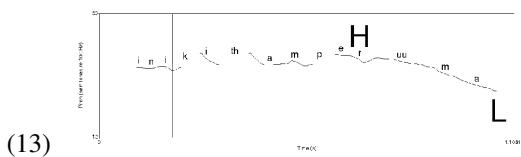
The differences between RUMI and SLM conventions can be exemplified with the following cognates:

1. Indonesian *tembak* ‘shoot’ is rendered in SLM as ⟨theembak⟩, with an ⟨h⟩ for the dental stop, a long vowel, and a prenasalized stop.
2. Indonesian *banyak* ‘many’ is rendered in SLM as ⟨bannyak⟩ with two ⟨n⟩ indicating the geminate status of /ɲ/. The sequence ⟨nny⟩ is dispreferred as it could lead to the pronunciation as [nini] by speakers familiar with English orthography.

### 2.2.6 Intonation

SLM has three main intonation contours: declarative (13), progradient (14), and interrogative (15). SLM declarative sentences show a fall in intonation towards the end, starting at the beginning of the last lexical word. Interrogatives show a rise in pitch towards the end. The most salient feature of SLM intonation is the steep LH% pattern at the end of subordinate clauses (progradient). The following three examples illustrate these patterns.

The assertive contour is characterized by a fall from H to a low tone target. The exact position of H needs more research, but appears to be towards the beginning of the last lexical word of the assertion.

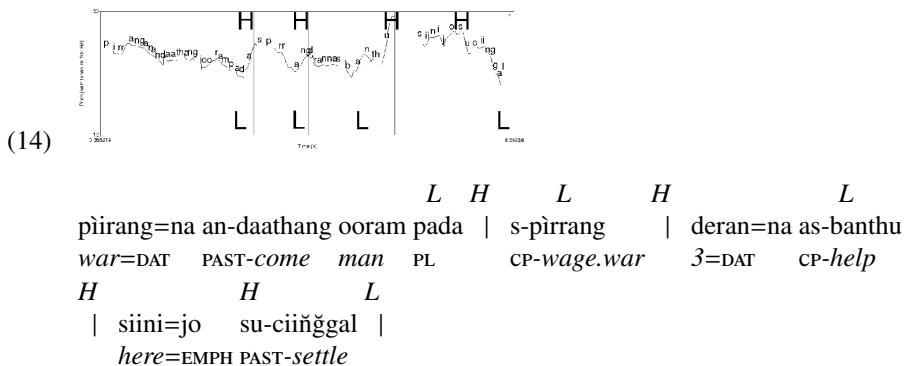


<sup>2</sup>⟨ddh⟩ is a theoretical possibility, although no word with this sequence has been found as of yet.

*H L*  
ini kitham-pe ruuma  
PROX 1PL=POSS *house*

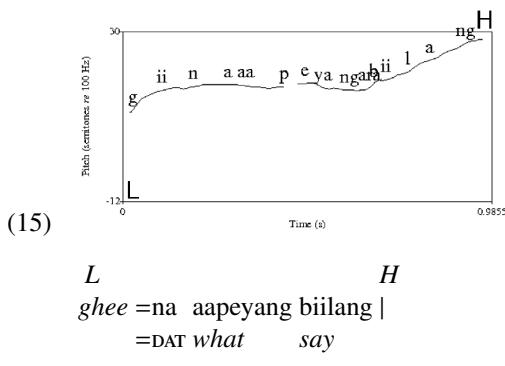
‘This is our house.’

The progredient contour has a low tone on the beginning of the last syllable and a high boundary tone, leading to a steep rise.



‘The men who had come, after having waged war and helped him, settled down.’

The interrogative contour resembles the progredient contour in having a steep rise, but the preceding dip is either less pronounced or non-existent.



‘How do you say for “Ghee”?’

## 2.3 Word classes

### 2.3.1 Main word classes

In contradistinction to most other varieties of Malay, SLM has solid word class distinctions, most prominently between nouns, verbs, and adjectives (Nordhoff [forthcominga](#)). A complication in this is that adjectives can convert to nouns or verbs and thus conceal their true nature to the superficial observer. *Bissar* ‘big’ can be used as a noun to mean ‘the big one’ or as a verb ‘to become big’. It is still possible to tell the adjectival nature of *bissar*, though: neither nouns nor verbs can be used in all the three

meanings given above. It is thus the flexibility of the adjectives which singles them out. See Nordhoff (*forthcominga*) for a fuller discussion.

Nouns and verbs can be identified by their (im)possibility to combine with verbal morphology as listed below in Section 2.4.2.

Two subclasses are of special interest: Within the verbs, we can single out two existentials: *aada*, which can combine with any referent, and *duuduk*, which can only be combined with animate referents.

- (16) a. *ruuma aada/\*arà-duuduk*  
house exist/NONPAST-exist.ANIM  
'There are houses.'
- b. *moonyeth aada/arà-duuduk*  
monkey exist/NONPAST-exist.ANIM  
'There are monkeys.'

Among the nouns, a special subclass is the so-called ‘relator nouns’ which indicate a relation between two referents, most often spatial, although other relations are also found. An example is given in (17).

- (17) *Ruuma=pe duppang pohong aada*  
house=Poss front tree exist  
'There are trees in front of the house.' (K081103eli02)

In the example above, the relator noun is linked to the ground by the possessive enclitic *=pe*. This is the most common case, but absence of any linker, as well as the use of the dative marker *=nang* instead of *=pe* are also found. The dative marker is especially used for temporal relations. For a relator noun like *blaakang* ‘behind/after’, the spatial meaning emerges when combined with *=pe* whereas the temporal meaning is obtained by using *=nang*.

- (18) *Kandi=ka Malay mosque<sub>space</sub>=pe blaakang=ka incayang=pe zihaarath aada.*  
Kandi=Loc Malay mosque=POSS back=LOC 3SG.POLITE=POSS shrine exist.  
'In Kandy behind the Malay Mosque, there is his shrine.'
- (19) *Dr Draaman duuva thaau<sub>time</sub>=nang blaakang incayang su-mniin̄ggal.*  
Dr Draaman two year=DAT after 3SG.POLITE PAST-die  
'After two years, Dr Draman died.' (K051213nar08)

### 2.3.2 Pronouns

#### 2.3.3 Personal pronouns

SLM distinguishes two numbers, three persons, and two levels of politeness in pronouns (Table 2.4). First and second person singular impolite pronouns are originally Chinese loans and have pejorative connotations in other Malay varieties as well. The plural pronouns all share the formative *-ang*, which grammaticalized from *oorang* ‘man, people’. The plural pronouns can additionally be marked by the plural word *pada* but this is optional.

SLM pronouns can combine with enclitic case markers in a quite regular way. There are two slight twists: the monosyllabic pronouns *see*, *goo*, *luu*, and *dee* shorten

the vowel when combining with a case marker. For the same pronouns, the dative allomorph is *=dang* instead of *=nang*, and the possessive allomorph is *=ppe* instead of *=pe*. An overview of SLM pronouns are given in Table 2.4.

### Interrogative pronouns

The basic SLM pronouns are given in Table 2.5. They can be combined with postpositions to yield a wider array of meanings, e.g. *saapa+pe* ‘who+POSS’=‘whose’.

### 2.3.4 Indefinite pronouns

Indefinite pronouns are formed by combining a suitable interrogative pronoun with one of the following enclitics

- additive *=le*: every-
- similitive *=ke*: any-
- indeterminate *=so*: some-
- negative *=pon*: negation

Examples include *saapale* ‘everyone’ *manake* ‘anywhere’ *dhraapso* ‘however many’ *kaapangpon* ‘never’.

### 2.3.5 Deictics

SLM distinguishes two distances in its deictic system, proximal and distal. The base forms *in(n)i* and *it(t)hu* are transparently related to a number of adverbs of location, source, manner and amount, as shown in Table 2.6.

### 2.3.6 Numerals

The numerals from one to twenty are given in Table 2.7. Productive numeral affixes are *blas* ‘-teen’, *-pulu* ‘-ty’ and *kà-* ‘ORD’.

### 2.3.7 Modals

A very small but important word class are the modals. There are five of them:

1. *boole* ‘can’
2. *thàrboole* ‘cannot’
3. *(kà)maau(van)* ‘want’
4. *thàr(kà)mauvan* ‘NEG.want’
5. *thussa* ‘NEG.want’

The modals can be used on their own to form a full utterance (e.g. *Boole*. ‘It is possible.’). The complement of the modal takes the infinitive (e.g. *Ini mà-miinung boole*. ‘It is possible to drink this.’). If the participant to whom the modal applies shall be indicated, it takes the dative marker *=nang* (e.g. *Lorang=nang ini mà-miinung boole* ‘You can drink this’).

---

<sup>3</sup> Adelaar (1991:32) has *lurang*.

pronoun	=yang [ACC]	=nang [DAT]	=pe [POSS]	meaning	origin
goo	goyang	godang	goppe	1S.FAMILIAR (Southern dialect)	Hokkien ( <a href="#">Adelaar 1991, Ansaldi 2005</a> )
see	seeyang	sedang	seppe	1S.POLITE	
luu	lu(u)yang	hudang	luppe	2S.FAMILIAR	Hokkien ( <a href="#">Adelaar 1991, Ansaldi 2005</a> )
dee	deeyang	dedang	deppe	3S.INANIMATE/ 3S.FAMILIAR	
diya	diyayang	diyanang	diyape	3S.	
siaanu	sianuyang	siaanunang	siaanupe	3S.PROXIMAL	
incayang	incayangyang	incayangnang	incayangpe	3SG.POLITE	<i>encik ia</i> + addition of velar nasal ( <a href="#">Sloman- son 2008</a> )
spaaman	spaamanyang	spaamannang	spaamanpe	3SG.POLITE	uncle?/ (Std, Malay, Javanese <i>paman</i> ( <a href="#">Adelaar 1985:141</a> ))
kithang	kithangyang	kithannang	kithampe	1PL	<*kita orang ('1PL man')
lorang <sup>3</sup>	lorangyang	lorangnang	lorampe	2PL/2S.POLITE	<*lu orang
derang	derangyang	derangnang	derampe	3PL	<*de orang, neutral incayang + <i>pada</i> 'pl.'
incayang pada	incayang padayang	incayang padanang	incayang padape	3PL.POLITE	

Table 2.4: Pronouns. The etymology of the plural pronouns includes a contraction of *orang* ‘man’ to *rang* ([Adelaar & Prentice 1996, Adelaar 2005, 1991](#)).

## 2.4 Morphology

### 2.4.1 Nominal morphology

The SLM noun takes very few affixes. The causativizer *-king* can marginally be combined with it, e.g. *kafan* ‘shroud’ → *kafanking* ‘enshroud’. Furthermore, the suffix *-an*, which is normally a nominalizer, can also be combined with nouns in certain cases, which are lexicalized. Examples are *rajahan* ‘government’ from *raaja* ‘king’ or *thumpathan* ‘seat in parliament’ from *thumpath* ‘place’. Last but not least, kin terms can take the suffix *-yang*, which means that the kin relation does not involve either speaker or addressee. An example would be *bapayang* ‘X’s father’.

Most nominal categories are expressed by enclitics rather than suffixes. The clitic status is clear from the fact that material can intervene between a noun and these markers, most typically an adjective. Concepts covered by enclitics are for instance the plural (*pada*), the indefiniteness marker *hatthu*, which can attach on either side, and the enclitic postpositions which indicate semantic roles and can be seen as case markers. Table 2.8 gives an overview of those.

#### The indefiniteness marker *hatthu*

The indefiniteness marker *hatthu* can be realized as *hat(t)hu* or *at(t)hu*, and can be either proclitic or enclitic. Occasionally, speakers use both the pro- and the enclitic, which yields double morphosyntactic encoding of semantic content. Care must be taken to distinguish the free numeral *sat(t)hu* ‘one’ from the clitic indefinite marker, especially because the numeral has two allomorphs *hatthu* and *atthu*, which are homophonous to the two forms of the clitic. However, the form with initial *s-* is not possible for the clitic, so that numerals can often be identified in that manner.

Example (20) shows the different realizations of the clitic. In this section, the cliticized nature of *hatthu* will be indicated by an equal sign (=). This is not done in other sections for reasons of legibility.

- (20) *Se atthu=aade, se hatthu=aade.*  
 1SG INDEF=younger.sibling 1SG INDEF=younger.sibling  
 ‘I am a younger sibling, I am a younger sibling.’ (K061120nar01)

The following example shows the use of the numeral *satthu*, identifiable by the *s-*. This *s-* could not occur in the example above, but the two forms used in (20) could also occur in (21).

- (21) *Mlaayu=dring satthu oorang=jo se.*  
 Malay=ABL one man=EMPH 1SG  
 ‘I am one of those Malays.’ (K060108nar02)

The following examples show the proclitic use (22), the enclitic use (23), and the doubled use (24) when referring to the concept of story.

- (22) *Hathu=oorang=pe muuluth=dering hatthu=criitha kal-dhaathang.*  
 INDEF=man=POSS mouth=ABL INDEF=story when-come  
 ‘When a story comes out of a man’s mouth.’ (B060115prs15)
- (23) *Giini criitha=hatthu=le aada.*  
 like.this story=INDEF=ADDIT exist  
 ‘There is also a story like that.’ (K051206nar07)

aapa	'what'	mana	'where', 'which'	craapa	'how'
saapa	'who'	kaapang	'when'	dhraapa	'how many'
				kànaapa	'why'

Table 2.5: SLM interrogative pronouns.

base form	place	source	manner	amount	gloss
ini	siini	sindari	giini	sgiini	proximal
ithu	siithu	sithari	gitthu	sgiithu	distal

Table 2.6: Deictics.

satthu,hatthu	1	subblas	11	duvapulsatthu	21
duuva	2	dooblas	12	duvalpulduuva	22
thiiga	3	thigablas	13	thigapulu	30
(ù)mpath	4	ùmpathblas	14	sraathus	100
liima	5	limablas	15	duuva raathus	200
(i)nnam	6	ìnnamblas	16	thiiga raathus	300
thuuju	7	thujublas	17	sriibu	1,000
dhlaapan	8	dhlapamblas	18	duuva riibu	2,000
s(m)biilan	9	s(m)bilamblas	19	thiiga riibu	3,000
spuulu	10	duvapulu	20	lakh [læk]	100,000

Table 2.7: SLM numerals.

form	meaning	form	meaning
Ø	nominative	=kapang	'when'
=yang	accusative	=sangke	'until'
=nang	dative	=thingka	'during, while'
=ka	locative	=apa	'after'
=dering	ablative	=sikin	'because (of)'
=pe	possessive	=lanthran	'because (of)'
=(sà)saama	comitative	=subbath	'because (of)'

Table 2.8: Case marking postpositions.

- (24) *Itthu attu=story=attu.*  
 DIST INDEF=story=INDEF  
 ‘This is a story.’ (B060115nar05)

Next to its function of marking indefiniteness, *hatthu* is also used to integrate English loanwords, a calque of the Sinhala form *eka* ‘one’ (Nitz & Nordhoff 2010). The following example shows a loanword integrated by *hatthu*. Crucially, we are dealing with a *definite* context here, showing that *hatthu* is not used in its primary function here.

- (25) *Inni* mock wedding=*hatthu mas-gijja.*  
 PROX mock wedding=INDEF must-make  
 ‘I have to do this mock wedding.’ (K060116nar10)

### The plural marker *pada*

*Pada* is a cliticized plural word (Dryer 1989). Ansaldi (2005:14) gives *pada* as a suffix, but the fact that it can combine with hosts from a variety of word classes, like nouns, adjectives, quantifiers, numerals and pronouns and even headless relative clauses suggests that it is a clitic. The clitic nature of *pada* is furthermore underscored by (26), where we find metalinguistic material intervening between *pada* and its host.

- (26) *Derang samma jaau uuistik* – village area – ***pada=nang su-pii.***  
 3PL all far village PL=DAT PAST-go  
 ‘They all went to remote villages.’ (K051222nar04)

The most common use of *pada* is on nouns, as in (27)

- (27) *Suda skaarang kitham=pe aanak pada laayeng pukurjan pada*  
 thus now 1PL=POSS child PL other work PL  
*arà-girja.*  
 NON.PAST-make.  
 ‘So now our children work in other professions.’ (K051222nar05)

Nominalized clauses can also host *pada*. Note that there is no overt signalling of the derivation.

- (28) *[[[Neene pada] anà-biilang] Ø pada]=jo itthu.*  
 grandmother PL PAST-say PL=EMPH DIST  
 ‘What the grandmothers said was this.’ (K051206nar03)

*Pada* optionally combines with plural pronouns (29) and quantifiers (30). Since these are all inherently already plural, *pada* only serves to emphasize that fact.

- (29) *Itthu=nam blaakang=jo, kitham pada anà-bissar.*  
 DIST after=EMPH 1PL PL PAST-big  
 ‘After that, we grew up.’ (K060108nar02)
- (30) *Spaaru pada bannya baae=nang anthi-duuduk.*  
 some PL very good=DAT IRR-stay  
 ‘Some are well off.’ (K061122nar01)

This emphasizing function can also be used in enumerations, as in (31).

- (31) *Mr Dole=pe Mr Samath=pe Mr Yusu pada=le* interview=nya thraa.  
 Mr Dole=POSS Mr Samath=POSS Mr Yusu PL=ADDIT interview=DAT NEG  
 ‘Mr Dole, Mr Samath and Mr Yusu were not selected for the interview.’ (K060116nar05)

### The semantics of *pada* and *hatthu*

*Pada* might actually be better analyzed as expressing *collective nominal aspect*, rather than nominal number. Collective aspects signals ‘that the set consists of multiple individual entities which together form a collective’ (Rijkhoff 2002:102). This collective interpretation of a set is supported by (31), where a group of three gentlemen should give interviews. The three gentlemen are named and coordinated, but then the plural marker is added. It is clear that there is only one specimen of each of the named persons, and the group only exists once, so *pada* cannot indicate cardinality greater than one in the strict sense. Rather, it emphasizes that we are dealing with a collectivity, the group is not seen as monolithic, but as composed of several members ('multiple individual entities'), and the cardinality of the members is greater than one. This interpretation as collective actually fits well with the optionality of *pada* according to Rijkhoff’s presentation of *set nouns*.

A logical extension of the analysis of *pada* as a ‘collective aspect marker’ would be to analyze *hatthu* (called ‘indefiniteness marker’ here) as a ‘singulative aspect marker’, indicating that the set is conceptualized as a whole, and not as ‘consisting of multiple individual entities’. SLM *ruuma* ‘house’ would then be transnumeral, *ruuma pada* ‘house COLLECTIVE’ would mean ‘the concept “house” interpreted as consisting of multiple entities’, and *hatthu ruuma* ‘SINGULATIVE house’ would mean ‘the concept “house” interpreted as consisting of a singleton entity’. This ‘singularizing’ function of *hatthu* finds support in example (32), where the concept of ‘parents’, which is ontologically necessarily of cardinality greater than one, is modified with *hatthu*, indicating that it should be conceptualized holistically, and that the internal constituency of the concept does not matter.

- (32) *Kithang samma hatthu umma+baapa=pe aanak pada, kithang samma*  
 1PL all INDEF mother+father=poss child PL 1PL all  
*sudaara pada.*  
 brother PL

‘We are all the same parents’ children, we are all brothers’ (B060115cvs01)

A reanalysis of the indefinite article as singulative has been proposed for Turkish by Schroeder (1999), and can be used to explain the uncommon order of the ‘indefinite article’ *bir* in this language (Rijkhoff 2002:319). In Turkish, *bir* can intervene between an adjective and a noun as in *meshur bir şair* ‘famous a poet’. If *bir* is an indefinite article, this order would violate a universal that the order ADJ INDEF N does not exist (Greenberg 1963, Hawkins 1994). If it is analyzed as a nominal aspect marker, on the other hand, the universal would not apply.

### 2.4.2 Verbal morphology

SLM has limited derivational morphology for verbs, but a comparatively large array of inflectional affixes as far as Malay varieties go.

#### Derivational morphology

Verbs can take the nominalizer *-an* and the causativizer *-king*.

- (33) *Se=ppe laayeng omong-an samma see anà-biilang.*  
 1SG=POSS other speak-NMLZ all 1SG PAST-say  
 ‘I had said everything in my other speech.’ (K061127nar03)

arà-	nonpast	thàrà-	past negative
anà-	past	thama-	nonpast negative
su-	past	thus-	want.not
mas-	debitive	jamà-	subordinate negative
anthi-	irrealis	kal-	conditional
mà-	infintive	kapang-	when
asà-	conjunctive participle	marà-	adhortative
bole=	can		
-la	imperative	-de	impolite imperative

Table 2.9: Verbal affixes in SLM.

- (34) *Baaye meera caaya kapang-jaadi, thurung-king.*  
 good red colour when-become, descend-caus  
 ‘When [the food] has turned to a nice rose colour, remove (it) [from the fire].’  
 (K060103rec02)

Both can attach to any verb regardless of transitivity. Furthermore, an ‘involitive verb’ can be derived with the prefix *kànà-*. This implies lack of control on the part of the agent as shown in (35) and (36).

- (35) *Se naasi arà-maakang.*  
 1SG rice NONPAST-eat  
 ‘I eat rice.’ (K081104eli03)
- (36) *Se=dang naasi arà-kànà-maakang.*  
 1S=DAT rice NONPAST-INVOL-eat  
 ‘I eat rice without wanting it/I compulsively eat rice/I was forced to eat rice by something beyond my control.’ (K081104eli03)

### Inflectional morphology

Inflection is preverbal in SLM (with the exception of imperatives) (Smith et al. 2004, Slomanson 2006). Table 2.9 lists the forms. Verbs have only one preverbal slot for inflection. This means that some meanings cannot be combined in morphology. This is most notably the case for the debititive *mas-* and the habilitive *bole-*, which block the expression of tense on the same verb. The same is true of the prefix *kapang-* ‘when’, which means any of *when X verbs*, *when X verbed*, *when X will verb*.

For the glosses of these prefixes, some explanation is in order: *arà-* is often referred to as present, but it can also refer to the future as in (37), so that ‘nonpast’ is the more precise gloss.

- (37) *Laskalli arà-maakang.*  
 other.time NONPAST-eat  
 ‘Next time you will eat.’ (B060115cvs16)

This contrasts with the past morphemes *su-* and *anà-*. These have the same temporal value, but differ in information structure. As a broad generalization, *su-* can only be used in pragmatically neutral contexts (predicate focus or sentence focus), whereas *anà-* is required with argument focus, although it can be used in neutral contexts as

well. Example (38) illustrates *anà-* in an argument focus context, a constituent question in this case. Example (39) shows that in a normal declarative sentence, both prefixes can be used

- (38) *Mana binaathang lorang=yang anà/\*su-giigith.*  
 which animal 2PL=ACC PAST-bite  
 ‘Which animal bit you?’ (K081105eli02)

- (39) *Itthu binaathan lorang=yang anà-/su-giigith.*  
 DIST animal 2PL=ACC PAST-/PAST-bite  
 ‘That animal bit you.’ (K081105eli02)

The morpheme *anthi-* is often glossed as ‘future’, but can also be used for the apodosis of conditionals. This suggests that it rather has a modal value of ‘irrealis’, encompassing situations which have not happened yet, regardless of whether one thinks that they are likely to happen (future) or not (irrealis).

- (40) *Cingga blaajar katha biilang thingka, ithu=kapang=jo gaaji*  
 Sinhala learn QUOT say when DIST=when=EMPH salary  
*athi-livath-king.*  
 IRR-much-CAUS  
 ‘When I say learning Sinhala, that means that your salary increases then.’  
 (K051222nar06)

### Periphrases

SLM does not make use of nominal reduplication, however, verbal reduplication is used to inflect verbs for present participle status, a calque of a Sinhala structure.

- (41) *Kancil lompath lompath arà-laari.*  
 rabbit jump~RED NONPAST-run  
 ‘The rabbit runs away jumping.’ (K081104eli06)

Besides the morphologically marked tenses, SLM has a periphrastic tense, the perfect. This is formed by the so-called conjunctive participle of a verb and the existential *aada* in the affirmative. In the negative, *thraa* replaces *aada*.

- (42) *Itthu=le kitham=pe mlaayu pada=jo itthu thumpath samma asà-kaasi aada.*  
 but 1PL=POSS Malay PL=EMPH DIST place all CP-give exist  
 ‘Therefore it was our Malays who have all given away those lands.’ (B060115cvs04)
- (43) *Hatthu dhaatha asà-kaaving thraa.*  
 one elder.sister CP-marry NEG  
 ‘One elder sister has not married.’ (K061019prs01)

SLM has two types of verb serialization, which are very similar to each other: Vector Verb Serialization and Motion Verb Serialization (Nordhoff forthcomingb, Jaffar this volume). Vector Verb Serialization follows a general South Asian model of marking aspectual and attitudinal values by a second verb following the primary verb. In this case, this second verb loses its literal meaning and expresses aspectual, adverbial or attitudinal shades of meaning.

verb	literalmeaning	aspectualmeaning	verb	literalmeaning	aspectualmeaning
ambel	take	inchoative	duuduk	sit	progressive
thaaro	put	hostilitive	abbis	finish	completive
simpang	keep	prospective			

Table 2.10: SLM vector verbs and their values.

- (44) *Kanabisan=ka=jo duva oorang=le anà-thaau ambel* [Andare duva last=LOC=EMPH two man=ADDIT PAST-know take Andare two oorang=yang=le asà-enco-kang aada] katha.  
 man=ACC=ADDIT CP-fool-CAUS exist QUOT  
 ‘At the very end, both women understood that Andare had fooled both of them.’  
 (K070000wrt05)

A list of these vector verbs and their aspectual values is given in Table 2.10. In contradistinction to the comparable structures in most South Asian languages, none of the verbs is morphologically marked as being subordinate to the other; in most Indian languages which have a similar construction, the first verb is either a participle or an infinitive.

The second serial verb construction involves two motion verbs, broadly construed. The first denotes the manner of motion and the second, the path. The path is typically expressed by *pii* ‘go’ or *dhaathang* ‘come’, but more complex paths like *kluuling* ‘roam’ are also possible.

- (45) *Hathu haari, hathu oorang thoppi mà-juval=nang kampong=dering*  
 INDEF day INDEF man hat INF-sell=DAT village=ABL  
*kampong=nang su-jaalang pii.*  
 village=DAT PAST-walk go  
 ‘One day, a man went and walked from village to village to sell hats.’ (K070000wrt01)
- (46) *Aanak su-laari kluuling.*  
 child PAST-run roam  
 ‘The child went astray.’ (K061019sng01)

### 2.4.3 Adjectival morphology

Adjectives can take the causativizer *-king*, the nominalizer *-an*, and the superlative marker *anà-* (homophonous with the past tense marker). Adjectives can furthermore convert to nouns or verbs and then have the entire inflectional potential of those classes.

### 2.4.4 Coordinating clitics

Coordinating clitics play an important role in SLM morphosyntax. They can attach to NPs, Verbs, or clauses and are used to express a number of coordinations, quantifications and referential statuses. Their meaning varies depending on whether they are used on one host or on multiple hosts. A special meaning is obtained when combining them with interrogative pronouns. See Table 2.11.

	simple use	multiple use	indefinite pronoun	gloss
=le	X also	X and Y	everyone	additive
=si	X ?	X or Y ?	n/a	interrogative
=so	n/a	X or Y	someone	indeterminate
=ke	like X	X or Y	anyone	similitive
=pon	not any X	neither X nor Y	not anyone	any

Table 2.11: Coordinating clitics.

### 2.4.5 Discourse clitics

SLM has an ubiquitous ‘emphatic’ clitic *=jo*, which is used for various senses of highlighting, e.g. argument focus (47), contrastive topic (48), and specificational/identificational predication (49), among others.

- (47) *Guunung=ka=jo kithang arà-duuduk; guunung=nang=jo kithang arà-pii.*  
mountain=LOC=EMPH 1PL      NONPAST-stay mountain=DAT=EMPH 1PL  
NONPAST-go  
‘It is in the hills that we live, it is to the hills that we go.’ (B060115prs01)
- (48) *[Pon=pe ruuma=ka=jo thaama duuva thiiga aari athi-duuduk.*  
bride=POSS house=LOC=EMPH earlier two three day IRR-stay  
‘As for at the bride’s house, then, they would stay two, three days back then.’  
(K061122nar01)
- (49) *Itthu oorang pada=jo kithang.*  
DIST man PL=EMPH 1PL  
‘Those people are us’ (K051222nar03)

A less common morpheme is *=jona*, which is glossed as ‘phatic’ here. It has about the same domain of usage as *y’know* in English, i.e. it is used to express that the speaker expects the hearer to be already aware of the propositional content conveyed.

- (50) *Punnu mlaayu pada kàthaama English=jona anthi-oomong.*  
many Malay PL earlier English=PHAT IRR=speak  
‘Many Malays would speak English, y’know, in former times.’ (K051222nar06)

## 2.5 Syntax

As a general rule, SLM is a right-headed language. The verb is normally the last element of the sentence, adpositions follow nouns, and nominal heads follow their modifiers. There are exceptions to all these rules but the overall tendency is very clear in all domains.

### 2.5.1 Syntax of the NP

The SLM NP is generally right-headed, although exceptions occur. Some nominal modifiers, like deictics, possessors, quantifiers, and relative clauses always precede the NP. Examples for this are given in (51) to (54).

- (51) *[Se=ppe naama] Mohomed Imran Salim.*  
 1SG=POSS name Mohomed Imran Salim  
 ‘My name is Mohomed Imran Salim.’ (K060108nar01)
- (52) *[Ithu jaalang]=ka mà-pii thàràboole.*  
 DIST road=LOC INF-go cannot  
 ‘You could not take that road.’ (B060115nar05)
- (53) *[[Seelon=nang dhaathang aada 0]RELC mlaayu] oorang ikkang.*  
 Ceylon-DAT come exist Malay man fish  
 ‘The Malays who came to Sri Lanka were fishermen.’ (K060108nar02)
- (54) *Inni sudaari=pe femili=ka [bannyak oorang] tsunami=dang s-puukul su-pii.*  
 PROX sister=POSS family=LOC many people tsunami=DAT CP-hit PAST-go  
 ‘In this sister’s family, many people were swept away by the tsunami.’ (B060115nar02)

It is possible for quantifiers to float outside of their NP. This is shown in the following example for *samma* ‘all’.

- (55) *[Kafan kaayeng pada] samma asà-ambel.*  
 shroud cloth PL all CP-take  
 ‘Having taken all the tissue for the shroud, ...’ (B060115nar05)

For some other modifiers, prenominal position is still the norm, but postnominal position is also allowed. This is true for numerals (56), the indefiniteness marker *hatthu* (57), and adjectives (58).

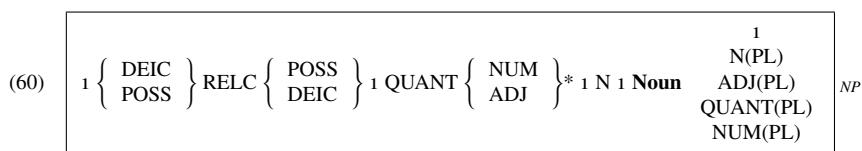
- (56) a. *Duva-pulu innam riibu ùmpath raathus lima-pulu duuva* votes  
 two-ty six thousand four hundred five-ty two votes  
*incayang=nang anà-daapath.*  
 3SG.POLITE=DAT PAST-get  
 ‘He got 26,452 votes.’ (N061031nar01)
- b. *[Panthes rooja kumbang pohong komplok] duuva asà-jaadi su-aada.*  
 beautiful rose flower tree bush two CP-grow PAST-exist  
 ‘Two beautiful rose bushes had grown.’ (K070000wrt04)
- (57) a. *Hatthu avuliya aada kitham=pe ruuma dikkath.*  
 INDEF saint exist 1PL=POSS house vicinity  
 ‘There is a saint close to our house.’ (K060108nar02)
- b. *See avuliya atthu su-jaadi.*  
 1SG saint INDEF PAST-become  
 ‘I have become a saint.’ (K051220nar01)
- (58) a. *Mintha<sub>ADJ</sub> daaging<sub>N</sub>=yang cuuci.*  
 raw beef=ACC wash  
 ‘Wash the raw beef.’  
 ...
- b. *asà-giiling daaging<sub>N</sub> mùntha<sub>ADJ</sub>=yang baathu=ka asà-thaaro, giccak.*  
 CP-grind beef raw=ACC stone=LOC CP-put smash  
 ‘Having ground and having put the raw beef on a stone, smash it.’ (K060103rec02)

There is one nominal modifier which can only occur at the very right edge. This is the plural marker *pada*.

- (59) [Spaaru oorang **pada**] su-pii.  
          some     man     PL     PAST-go  
          ‘Some men left.’ (B060115nar01)

Furthermore, NPs can contain two or more nouns, which can be analysed as compounds or adnominal modification. These sequences are either left-headed like *oorang ikkang* ‘man’ + ‘fish’ = ‘fisherman’ or right-headed like *paadi+pohong* ‘unharvested rice’ + ‘tree’ = ‘rice plant’.

The order of elements in the NP is schematized as below. All elements are optional. This also applies to the head noun.



This reads as follows: elements to the left precede elements to the right. Elements in vertical alignment offer a choice of any one of the vertically aligned elements, but not more than one. The asterisk indicates that there may be several instances of this modifier. 1 represents the position of the indefiniteness marker, which is particularly interesting and is discussed in more detail in the following section.

### 2.5.2 The position of the indefiniteness marker

The indefiniteness marker *hatthu* has some very peculiar morphosyntactic properties. First, it can occur before (61), after (62), or on both sides of the noun (63). Second, its position is normally between prenominal modifiers and the head noun, violating the principle of scope. This is found e.g. for the sequence of adjective and noun in (61), which is interrupted by *hatthu*.

- (61) *Itthu* [bannyak [laama]<sub>ADJ</sub> [hathu]<sub>INDEF</sub> [ruuma]<sub>N</sub>].  
 DIST very old INDEF house  
 ‘That one was a very old house.’ (K070000wrt04)

(62) *Hathu muusing=ka* ... [hathu]<sub>INDEF</sub> [kiccil]<sub>ADJ</sub> [ruuma]<sub>N</sub> su-aada  
 INDEF time=LOC ... INDEF small house PAST-exist  
 ‘Once upon a time, there was a small house.’ (K07000wrt04)

(63) *Sithu=ka* hathu maccan hathu duuduk aada.  
 there=LOC INDEF tiger INDEF stay exist  
 ‘A tiger stayed there.’ (B060115nar05)

The ‘interrupting’ position of *hatthu* is also attested with N+N sequences.

- (64) *Ini poongan atthas=ka moonyeth hathu=kavanana su-aada.*  
 PROX tree top=LOC monkey INDEF=group PAST-exist  
 ‘On top of this tree was a group of monkeys.’ (K070000wrt01)

### 2.5.3 Simple declarative clause

The simple declarative clause is verb final and has SOV word order.<sup>4</sup> The order in the preverbal field is free, but due to frequent dropping of elements, the question of the relative ordering of elements is normally moot since only one NP is realized.

An example of a simple declarative clause is given below.

- (65) *Se=ppe baapa incayang=nang ummas=Ø su-kaasi.*  
 1s=POSS father 3SG.POLITE=DAT gold PAST-give  
 ‘My father gave him gold.’ (K070000wrt04)

It is possible to find NPs after the verb. The postverbal position can be occupied by heavy elements (66), spatial arguments of location or goal (67), and elements in focus (68).

- (66) *Kitha=nang<sub>NP</sub> maaupred [kitham=pe mlaayu lorang blaajar lorang=pe mlaayu kitham blaajar]<sub>NP</sub>.*  
 1PL=DAT want 1PL=POSS Malay 2PL learn 2PL=POSS Malay 1PL learn  
 ‘We want that you learn our Malay and that we learn your Malay.’ (K060116nar02)
- (67) *see anà-laaher Navalapitiya=ka<sub>loc</sub>*  
 1s PAST-be.born Nawalapitiya=LOC  
 ‘I was born in Nawalapitiya.’ (K051201nar01)
- (68) *Hindu arà-maakang kambing.*  
 Hindu NONPAST-eat goat  
 ‘Hindus eat GOAT.’ (K060112nar01)

### 2.5.4 Copular clauses

One noteworthy morphosyntactic difference between SLM and other varieties of Malay is the presence of a copula, (*asà*)*dhaathang(apa)*. This copula derives from the conjunctive participle of the verb *dhaathang* ‘come’, and probably went from a resultative meaning similar to English *a dream come true* to a stative meaning mainly used for predicates of group membership. A fuller account can be found in Nordhoff (2011). The three forms given below are in free variation.

- (69) *Se=ppe naama asàdhaathang Cintha Sinthani.*  
 1SG=POSS name COPULA Chinthia Sinthani.  
 ‘My name is Chinthia Sinthani.’ (B060115prs04)
- (70) *Se=ppe baapa dhaathangapa Jinaan Samath.*  
 1SG=POSS father COPULA Jinaan Samath.  
 ‘My father was Jinaan Samath.’ (N060113nar03)
- (71) *Estate=pe field officer asàdhaathangapa kithang=pe kaake*  
 estate=POSS field officer COPULA 1PL=POSS grandfather.  
 ‘The estate field officer was our grandfather.’ (N060113nar03)

---

<sup>4</sup>As usual in word order typology, S and O have semantic definitions. The label ‘SOV’ is used for convenience here. Neither subject nor object have straightforward definitions in SLM. The descriptor ‘NP NP V’ would be more appropriate than ‘SOV’, but preference is given here to the more familiar term ‘SOV’.

### 2.5.5 Relative clause

Relative clauses are preposed and not signalled by morphology. There is no verbal marker of relative status nor is there a relative pronoun. Relative clauses are not restricted in any way as compared to main clauses. Any morphology, any tense, and any predication type may be used in relative clauses.

The following examples show relative clauses where the relativized element is the agent (72), a place (73), a point in time (74), and a possessor (75).

- (72) *thoppi arà-daagang oorang*  
hat NONPAST-trade man  
'The hat seller' (K070000wrt01)
- (73) *Siihu [nya-duuduk thumpath]=ka baapa su-jaatho.*  
there PAST-stay place=LOC father PAST-fall  
'There, at the place he was staying, my father fell.' (K051205nar05)
- (74) *Suda [puthri=le biini=le arà-caañda haari]=le su-dhaathang.*  
so queen=ADDIT wife=ADDIT SIMULT-meet day=ADDIT PAST-come  
'So then the day came when the wife and the queen were to meet.' (K070000wrt05)
- (75) *[Laaki anà-mniiñggal hathu pompong].*  
husband PAST-die INDEF woman  
'A woman whose husband had died.' (K070000wrt04)

SLM also has the possibility to form so-called 'headless relative clauses'. The relative clause in (76) and the headless relative clause in (77) share the same structure, but the head is missing in (77).

- (76) *[Lorang anà-maasak ikkang] eenak.*  
2PL PAST-cook fish tasty  
'The fish you cooked is tasty.' (K081105eli02)
- (77) *[Lorang anà-maasak Ø] eenak.*  
2PL PAST-cook tasty  
'What you cooked is tasty.' (K081105eli02)

### 2.5.6 Argument clause

Clauses used as arguments of e.g. perception verbs do not take any special morphological marking. There is no complementizer, no special verb form, etc. The subordinate clause is simply put in the place that another NP would have in the predication. Argument clauses can take case enclitics like any other NP. This is shown below for the dative.

- (78) a. *lorang suurath=yang mlaayu=dering anà-thuulis*  
2PL letter=ACC Malay=ABL past=write  
'You wrote the letter in Malay.'
- b. *Kitham=pe baapa su-biilang [[lorang suurath=yang mlaayu=dering*  
1PL=POSS father PAST-say 2PL letter=ACC Malay=ABL  
*anà-thuulis]CLAUSE=nang bannyak arà-suuka].*  
past=write=DAT much SIMULT-like  
'Daddy said that he liked very much that you wrote the letter in Malay.'  
(Letter 26.06.2007)

Utterance verbs like *biilang* ‘say’ or *butharak* ‘shout’ use the quotative *katha* to mark their arguments. These arguments are not clauses, but full utterances. This can be seen from the fact that non-clausal utterances like *iiya* ‘yes’ or onomatopoeic expressions like *dam dam dam* can also be used with *katha*.

- (79) *Skarang biini arà-iingath [puthri thuuli katha]; Puthri arà-iingath*  
now wife NONPAST-think queen deaf QUOT queen NONPAST-think  
*[biini thuuli katha].*  
wife deaf QUOT  
‘Now the wife thought the queen was deaf, and the queen thought the wife was deaf.’ (K070000wrt05)
- (80) **Yes *katha*** *m-biilang.*  
yes QUOT PAST-say  
‘He said “yes”.’ (K060116nar11)
- (81) *Aanak pom pang duuva=nang [slaamath] katha su-biilang.*  
child girl two=DAT goodbye QUOT PAST-say  
‘He said “Goodbye” to the two girls.’ (K070000wrt04)
- (82) **Dam dam dam *katha*** *su-aada.*  
dam dam dam QUOT PAST-exist  
‘(The rain) went like “dam dam dam”.’ (overheard)

Purposive clauses are formed with the infinitive *mà-* and the dative marker *=nang*. They can precede the matrix verb (83) but are often dislocated because of their heaviness (84).

- (83) *Derang [dikkath=ka aada laapang]=nang [mà-maayeng]purp=nang su-pii.*  
3PL vicinity=LOC exist ground=DAT INF-play=DAT PAST-go  
‘They went to play on the ground nearby.’ (K070000wrt04)
- (84) *Itthu cave=nang kithang=le pii aada [mà-liyath]purp=nang.*  
DIST cave=DAT 1PL=ADDIT go exist INF-look=DAT  
‘We have also gone to that cave to have a look.’ (K051206nar02)

### 2.5.7 Interrogative clause

#### Polar questions

Polar questions are normally marked by the interrogative clitic *=si* and a particular intonation contour. Normally, both of these signs are present, but one only can also be sufficient. The interrogative clitic attaches to the verb when the whole sentence is questioned.

- (85) *Se=pe uumur masà-biilan=si?*  
1=POSS age must-tell=INTERR  
‘(Do I) have to tell my age?’ (B060115prs01)

The clitic attaches to an NP or other element if only this portion of the sentence is questioned. In this case, the choice of past tense morphemes is restricted to *anà-*; *su-* is not possible.

- (86) *Daging baabi=si \*su-/anà- billi???*  
 pork=INTERR su-anà- buy  
 ‘Did you buy PORK???’ (K081105eli02)

It is possible to attach =*si* to more than one element, resulting in an alternative question.

- (87) *Piisang=si maañgga=si maa?*  
 plantain=INTERR mango=INTERR want  
 ‘Is it plantain or mango that you want?’ (K081105eli02)

### Content Questions

Content questions are formed by putting a suitable interrogative pronoun (see Section 2.3.3) at the place of the questioned element. This is thus *in situ*, although interrogative pronouns also frequently occur in initial position. Since this position could also be occupied by normal NPs, it can still be analyzed as *in situ*.

- (88) *Mana nigiri=ka arà-duuduk?*  
 which country=LOC NONPAST-stay  
 ‘In which country do you live?’ (B060115cvs16)

### 2.5.8 Imperative clause

Imperative clauses are formed by either attaching one of the suffixes *-la* ‘IMP.POLITE’ and *-de* ‘IMP.IMPOLITE’, by using the preverbal particle *mari*, or both.

- (89) a. *Saayang se=ppe thuan mari laari-la.*  
 love 1SG=POSS sir come.IMP run-IMP  
 ‘Come my beloved gentleman, come running here.’  
 b. *See=samma kumpul mari thaanđak-la.*  
 1SG=COMIT gather come.IMP dance-IMP  
 ‘Come and dance with me.’ (N061124sng01)

Additionally, *mari(-la)* used without a verb means ‘Come!’. The use of a bare verb is also possible (90).

- (90) *Aajuth thaakuth=ka su-naangis, “See=yang Ø-luppas-Ø”.*  
 dwarf fear=LOC PAST-cry 1SG=ACC leave  
 ‘The dwarf screamed in fear: “Leave me!” ’ (K070000wrt04)

Adhortatives are formed with the prefix *marà*, etymologically related to *mari*.

- (91) *Kitham marà-maayeng.*  
 1PL ADHORT-play  
 ‘Let’s play.’ (K081104eli06)

Negative imperatives are formed with the verbal prefix *jamà-*.

- (92) *Hatthu=le jamà-gijja, baapa, ruuma=ka duuduk!*  
 INDEF=ADDIT NEG.IMP-do father house=LOC stay  
 ‘Don’t do anything, daddy, stay at home!’ (B060115nar04)

one-place	two-place			three-place
NOM V	NOM NOM	DAT NOM	INSTR NOM	NOM NOM DAT
DAT V	NOM ACC	DAT ACC	INSTR DAT	NOM ACC DAT
ACC V	NOM DAT	DAT DAT	INSTR ACC	NOM NOM ABL
INSTR V	NOM INSTR	DAT INSTR	INSTR INSTR	NOM ACC ABL

Table 2.12: Overview of case frames for predicates of different arity.

## 2.6 Cases

Case relations are realized as enclitic postpositions in SLM. They can attach to a noun, a pronoun, a quantifier, a deictic, a numeral, or an argument clause. There is a many-to-many relationship between case postpositions and semantic roles. The most frequent morphemes are *=nang* ‘DAT’, *=yang* ‘ACC’, *=ka* ‘LOC’, *=dering* ‘ABL’, and *=pe* ‘POSS’. In addition, there are a number of temporal and causal postpositions, which are also enclitics, but are much less frequent than the morphemes just mentioned.

## 2.7 Valency

SLM does not have very solid valency distinctions. Addition and deletion of arguments are constrained by semantic plausibility, but only marginally by the lexical class of a lexeme. Typical case frames are given in Table 2.12.

### 2.7.1 One-place predicates

SLM is a role-dominated language. The notions of ‘subject’ and ‘object’ are not needed to describe its syntax. The semantic role of a referent is directly reflected in morphosyntax through the case marker selected (See Section 2.4.1). One-place-predicates can have their arguments marked with either nominative, accusative, dative, or instrumental. The discussion of case frames will therefore be restricted to these four cases here, setting aside spatial and temporal cases, among others. As for one-place predicates, nominative is the most common case (93). Dative is assigned for experiencers (94), or if a modal is present in the clause (95). Accusative-marking is very restricted in one-place predicates and has only been found on one verb *thinggalam* ‘sink’ (96). Finally, the instrumental is not conditioned by semantic role, but by the nature of the referent: if it is an institution, such as a government or a committee, use the instrumental (97), otherwise, the postposition is assigned according to semantic role.

- (93) *Itthukapang Tony Hassan=Ø<sup>5</sup> su-pii.*  
 then           Tony Hassan       PAST-go  
 ‘Then Tony Hassan left.’ (K060116nar09)

- (94) *Go=dang karang bannyak thàràsìggar.*  
 1S.FAMILIAR=DAT now very sick  
 ‘I am now very sick.’ (B060115nar04)

<sup>5</sup>I use Ø to explicitly indicate the absence of a postposition.

- (95) *Kithang=nang* [...] two o'clock=*ke=sangke bole=duuduk*.  
 1PL=DAT two o'clock=SIMIL=until can-stay  
 'We can stay up until two o' clock.' (K061026rcp04)
- (96) Titanic *kappal=yang su-thìnggalam*.  
 Titanic ship=ACC PAST-sink  
 'The ship "Titanic" sank.' (K081104eli05)
- (97) Police=*dering su-dhaathang*.  
 police=ABL PAST-come  
 'The police came.' (K081105eli02)

## 2.7.2 Two-place predicates

With two place predicates, we have to distinguish the actor argument and the undergoer argument. The actor will normally be either in the nominative (most verbs) or in the dative (experiencer verbs and modals). It will be in the instrumental under the same conditions that hold for one-place predicates.

The non-actor argument will normally be in the accusative, although the accusative marker =*yang* is often dropped (98). Recipients and beneficiaries (99), as well as some rare patients (of the verbs *puukul* 'hit', *thiikam* 'stab' *theembak* 'shoot') will also be in the dative (100). The person being addressed in questions appear in the instrumental.

In case of experiencer verbs, the stimulus will normally be in the nominative. In some cases, most notably *thaau* 'know', the stimulus/theme can also be in the accusative.

If a modal is present in a clause with a verb which already assigns dative to the non-actor, both actor and non-actor will be marked with the dative, giving rise to ambiguity.

- (98) *Ithukapang lorang=pe leher=(yang) kithang=Ø athi-poothong*.  
 then 2PL=POSS neck=ACC 1PL IRR-cut  
 'Then we will cut your neck.' (K051213nar06)
- (99) *Derang pada=Ø arà-banthu cinggala raaja=nang*.  
 3PL PL NONPAST-help Sinhala king=DAT  
 'They help the Sinhalese king.' (K051206nar03)
- (100) *Rose-red=Ø buurung=nang su-puukul*.  
 Rose-red bird=DAT PAST-hit  
 'Rose-red hit the bird.' (K070000wrt04)
- (100) shows one possibility of assigning zero and dative to two arguments. The actor is Ø-marked while the undergoer receives dative marking. Another possibility, which exists for experiencer verbs, is to mark the experiencer with the dative =*nang* and the stimulus with zero. This is shown in (101) (and is a common South Asian construction (Masica 1976:159ff)).
- (101) *[svaara hatthu]=Ø derang=nang su-dìnngar*.  
 noise INDEF 3PL=DAT PAST-hear  
 'They heard a noise.' (K070000wrt04)

Institutional actors take the instrumental as discussed above. The other argument can be marked with =*yang* (102), or bear no marking (103).

- (102) *See=yang police=dering nya-preksa.*  
 1SG=ACC police=ABL PAST-enquire  
 ‘I was questioned by the police.’ (K051213nar01)
- (103) British Government=**dering** Malaysia Indonesia, [inni nigiri pada]= $\emptyset$   
 British Government=ABL Malaysia Indonesia PROX country PL  
*samma peegang.*  
 all catch  
 ‘The British Government captured Malaysia and Indonesia, those countries.’  
 (K051213nar06)

With modal proclitics, the actor can receive dative marking. In (104), *kithang* ‘we’ receives dative marking and the theme of reading, *mulbar* ‘Tamil’, is zero-marked.

- (104) *Kithang=nang baaye=nang mulbar=* $\emptyset$  *bole=baaca.*  
 1PL=DAT good=DAT Tamil can=read  
 ‘We can read Tamil well.’ (K051222nar06)

The presence of a modal proclitic does not preclude accusative marking on the undergoer. This is found for instance in (105).

- (105) *aathi=yang sajja hatthu oorang=nang bole=ambel.*  
 liver=ACC only one man=DAT can-take  
 ‘Only one person can take the liver.’

When using a verb which normally assigns the dative case to the undergoer a modal proclitic results in both arguments being marked with the dative.

- (106) *Se=dang Farook=nang bole=puukul.*  
 1S=DAT Farook=DAT can=hit  
 ‘I can hit Farook.’ (K081104eli05)

In these cases, the actor is normally associated with the leftmost argument, while the undergoer is the other argument.

To sum up, two-place predicates normally have zero-marked actors, and undergoers are either marked for accusative or dative. In special cases, actors can be marked for instrumental or dative.

### 2.7.3 Three-place predicates

Three-place predicates are typically predicates of transfer, i.e. giving and taking. As such, they include an agent, a theme, and a goal (or source). The agent is typically in the nominative (107)-(108), although modals in the sentence can change this to dative (109), and institutional agents trigger instrumental marking(110). The theme is either unmarked (=nominative) (107)-(109) or in the accusative (110). Recipients, beneficiaries and goals are marked by the dative, and sources by the ablative. The following examples illustrate this

- (107) *Se=ppe baapa=* $\emptyset$  *incayang=nang ummas=* $\emptyset$  *su-kaasi.*  
 1SG=POSS father 3SG.POLITE=DAT gold PAST-give  
 ‘My father gave him gold.’ (K070000wrt04)

- (108) *Derang=pe umma=Ø derang=nang [jaithan=Ø=le, jaarong pukurjan=Ø=le]*  
 3PL=POSS mother 3PL=DAT sewing=ADDIT needle work=ADDIT  
*su-aajar.*  
 PAST-teach  
 ‘Their mother taught them sewing and needle work.’ (K070000wrt04)
- (109) *Kithang=nang miskin pada=nang duvith=Ø bole=kaasi.*  
 1PL=DAT poor PL=DAT money can=give  
 ‘We can give money to the poor.’ (K081104eli05)
- (110) *Police=dering see=yang remand=nang su-kiiring.*  
 police=ABL 1SG=ACC remand=DAT PAST-send  
 ‘The police sent me into custody.’ (K081105eli02)

#### 2.7.4 Summary of valency structure

To sum up the distribution of zero, accusative, dative and instrumental, on the roles of S, A, P and R (=Recipient), the following can be said.<sup>6</sup>

- The dative marker can be found on R and P. Additionally, it can be found on S and A if they are experiencers. Furthermore, modal proclitics can assign the dative to S or A
- The accusative marker can be found on P and in rare instances on S
- The instrumental marker can be found on S and A when they are institutional
- Zero can be found on S, A and P. Zero is never found on R.

This distribution can be illustrated as in Figure 2.1.

## 2.8 Negation

SLM has a highly developed negation system, which distinguishes predicate type, tense, and syntactic status, as well as a number of lexicalized forms. The first important distinction is between verbal negators, nonverbal negators, and existential negators. *Thraa* is used to negate existence and *bukang* is used to negate any other nonverbal predicates. Within the verbal predicates, there are a number of further choices. If the predicate is found in a subordinate clause which would take the infinitive or the conjunctive participle in the affirmative, *jamà-* is used to negate this predicate. If the predicate is in another type of clause, the choice of negator depends on the time reference. For past reference, *thàra-* is used, for nonpast reference, *thama-*. A special case is the negation of the perfect tense, which is *V thraa*, instead of the affirmative *V aada*.

Adjectives fall into two lexical classes. The larger one is negated by postverbal *thraa*. The smaller one is negated by the prefix *thàrà-*. This prefix is used for all tenses. This contrasts with the use of the same prefix with verbs, where it encodes past reference.

The lexicalized negations include *thàrboole* ‘cannot’ for *boole* ‘can’ and *thussa* ‘want not’ for *maau* ‘want’. The other meaning of *maau*, ‘need’, is negated by *thàrka-mauvan*.

---

<sup>6</sup>The Sinhala facts are very similar to this. See Gair (1976, 1991), Henadeerage (2002).

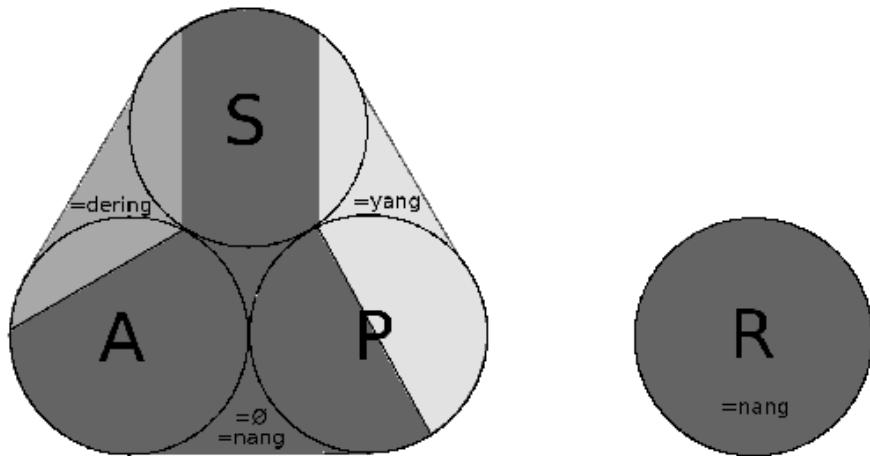


Figure 2.1: The coding of semantic roles in SLM. The instrumental *=dering* can code S and A, but is marginal, as seen by the small portion of the circles it covers. The accusative *=yang* can be used for P, where it is common, and for S, where it is marginal. This difference in frequency is represented by the different sizes of the covered surface. The dative marker *=nang* can be used for S, A and P, as can zero. For expository reasons, dative and zero are not distinguished in the illustration. Zero is much more frequent than the dative in all three roles. R, finally, can only be marked by the dative.

When answering questions in the negative, *thraa* is used as well. These facts are summarized in Table 2.13.

## 2.9 Discourse

A very salient feature of SLM discourse is the frequent dropping of participants. It is possible to drop any argument of a clause. So, *Sukaasi*. ‘Gave.’ is a full sentence with the meaning ‘X gave Y to Z’, where X, Y, and Z are instantiated with referents from the common ground between speaker and hearer.

Another salient feature is the frequent use of conjunctive participle clauses to indicate sequences of events. In a chain of events, all but the last are marked by the conjunctive participle *asà-*. The last event then receives the normal tense marking, thereby establishing reference time.

- (111) *Oorang pada asà-pìrrang, derang=nang asà-banthu, siini=jo su-ciin̊ggal.*  
 man PL CP-wage.war 3PL=DAT cp-help here=EMPH PAST-settle

‘The men, having waged war, having helped them, settled down right here.’  
 (K051222nar03)

Finally, the distal deictic *itthu* can combine with a number of clitics. These combinations are then used as a variety of connectors, e.g.

- *itthu* ‘DIST’ + *nang* ‘DAT’ = *ithunam* ‘then’

	past	perfect	present	future
fin. verbal clause	thàrà-V	V thraa		thama-V
inf. verbal clause			jamà	
existential			thraa	
nominal		bukang		thama-jaadi/bukang
adjectival1		ADJ thraa		thama-ADJ
adjectival2		thàrà-ADJ		thama-ADJ
circumstantial		bukang		
animate locational	thàràduuduk	duuduk thraa		thama-duuduk
(ka)maau(van)			thàrkamauvan/thussa	
boole			thàrboole	
constituent neg.			bukang	

Table 2.13: Negation patterns for various predicate types and tenses.

- *itthu* ‘DIST’ + *kapang* ‘when’ = *i(tthu)ka(pa)ng* ‘then’ (112)
- *itthu* ‘DIST’ + *le* ‘ADDIT’ = *itthule* ‘but’ (113)

- (112) a. *Kandi=ka Malay mosque=pe blaakang=ka incayang=pe zihaarath aada.*  
           Kandi=LOC Malay mosque=POSS back=LOC 3SG.POLITE=POSS shrine  
           exist  
           ‘In Kandy behind the Malay Mosque, there is his shrine.’
- b. *I(tthu)ka(pa)ng derang=pe sudara pompang=jo aada Hanthane=ka.*  
       then           3SG=POSS sibling sister=EMPH exist Hantane=LOC  
       ‘Then, his sister is in Hantane.’ (K060108nar02)
- (113) *Spaaru mlaayu pada arà-oomong itthule mulbar arà-oomong.*  
       some Malay PL NONPAST-speak but Tamil NONPAST-speak  
       ‘(Only) few Malays speak (Malay), but they speak Tamil.’ (G051222nar04)



# Bibliography

- Adelaar, K. A., 1991. "Some Notes on the Origin of Sri Lanka Malay." In *Papers in Austronesian Linguistics*, edited by H. Steinhauer. Canberra: Australian National University, 23–37.
- Adelaar, K. A., 2005. "Structural diversity in the Malayic subgroup." In *The Austronesian Languages of Asia and Madagascar*, edited by K. Alexander Adelaar & Nikolaus P. Himmelmann. London, New York: Routledge, 202–226.
- Adelaar, K. A. & Prentice, D. J., 1996. "Malay: its History, Role and Spread." In *Atlas of Languages of Intercultural Communication in the Pacific, Asia and the Americas*, edited by Stephen A. Wurm, Mühlhäusler Peter, & Darrell T. Tryon. Berlin: Mouton de Gruyter, 673–693.
- Adelaar, K. Alexander, 1985. *Proto-Malayic; The reconstruction of its phonology and parts of its morphology and lexicon*. Ph.D. thesis, Leiden University.
- Ansaldo, Umberto, 2005. "Typological admixture in Sri Lanka Malay. The case of Kirinda Java." Unpublished manuscript.
- Apossidou, Diana & Nordhoff, Sebastian, 2009. "Feet in Sri Lanka Malay – no stress please." Manuscript University of Amsterdam.
- Bichsel-Stettler, Anne, 1989. *Aspects of the Sri Lanka Malay Community and their Language*. Master's thesis, University of Bern.
- Dryer, Matthew, 1989. "Plural Words." *Linguistics* 27:865–895.
- Gair, James W., 1976. "Is Sinhala a Subject Language?" In *The Subject in South Asian Languages*, edited by Manindra Verma. Madison: University of Wisconsin South Asia Program, 39–63.
- Gair, James W., 1991. "Subjects, Case and INFL in Sinhala." In *Experiencer Subjects in South Asian Languages*, edited by M. Verma & K. P. Mohanan. Palo Alto: Center for the Study of Language and Information, 13–41. Reprinted in Gair (1998), pp. 65–86.
- Gair, James W., 1998. *Studies in South Asian linguistics; Sinhala and other South Asian Languages*. New York; Oxford: Oxford University Press.
- Greenberg, Joseph Harold, 1963. "Some universals of grammar with particular reference to the order of meaningful elements." In *Universals of language : report of a conference held at Dobbs Ferry, New York, April 13-15, 1961*, edited by Joseph H. Greenberg. Cambridge, MA: The M.I.T. Press, 73–113.
- Hawkins, John A., 1994. *A performance theory of order and constituency*. Cambridge: CUP.
- Henadeerage, Kumara, 2002. *Topics in Sinhala Syntax*. Ph.D. thesis, Australian National University.
- Hussainmiya, B. A., 1987. *Lost cousins: The Malays of Sri Lanka*. Bangi: Institut

- Bahasa, Kesusasteraan dan Kebudayaan Melayu.
- Jaffar, Mohamed, this volume. “The Semantics of Serial Verb Constructions in Sri Lanka Malay.”
- Masica, Colin P., 1976. *Defining a Linguistic Area: South Asia*. Chicago: University of Chicago Press.
- Nitz, Eike & Nordhoff, Sebastian, 2010. “Subtractive Plural Morphology in Sinhala.” In *Rara & Rarissima. Collecting and interpreting unusual characteristics of human languages*, edited by Jan Wohlgemuth & Michael Cysouw. Berlin: Mouton de Gruyter, 247–266.
- Nordhoff, Sebastian, 2009. *A Grammar of Upcountry Sri Lanka Malay*. Ph.D. thesis, University of Amsterdam, Utrecht.
- URL <http://www.lotpublications.nl/publish/articles/003744/bookpart.pdf>
- Nordhoff, Sebastian, 2011. “Having come to be a copula in Sri Lanka Malay – an unusual grammaticalization path.” *Folia Linguistica* 45(1):103–126.
- Nordhoff, Sebastian, forthcominga. “Jack-of-all-trades: The Sri Lankan Malay property word.” In *Flexible word classes: a typological study of underspecified parts-of-speech*, edited by Eva van Lier & Jan Rijkhoff. Oxford: OUP.
- Nordhoff, Sebastian, forthcomingb. “Multi-verb constructions in Sri Lanka Malay.” *Journal of Pidgin and Creole Languages*.
- Rijkhoff, Johannes Nicolaas Maria, 2002. *The Noun Phrase*. Oxford: OUP.
- Schroeder, Christoph, 1999. *The Turkish Noun Phrase in Spoken Discourse*. Wiesbaden: Harrassowitz.
- Slomanson, Peter, 2006. “Sri Lankan Malay morphosyntax: Lankan or Malay?” In *Structure and Variation in Language Contact*, edited by Ana Deumert & Stephanie Durreiman. Amsterdam: John Benjamins, 135–158.
- Slomanson, Peter, 2008. “Progressive Aspect in Sri Lankan Malay and the Syntactic Status of *ambe(l)*.” Paper presented at ISMIL 12, 26-27 June 2008 University of Leiden.
- Smith, Ian R., Paauw, S., & Hussainmiya, B. A., 2004. “Sri Lanka Malay: the state of the art.” *Yearbook of South Asian Languages* 2004:197–215.
- Tapovanaye, Suthadara, 1986. *Analysis of the segmental phonemes of the Sri Lanka Malay language*. Master’s thesis, Univeristy of Kelaniya. Written in Sinhala.
- Tapovanaye, Suthadara, 1995. *Vowel Length in Sri Lankan Malay*. Master’s thesis, University of Hawaii.
- Wetzels, Leo & Wezer, E. (eds.), 1986. *Studies in Compensatory Lengthening*. Dordrecht: Foris.

## **Part II**

# **Sociology, history, and demography**



# Chapter 3

## Sri Lanka Malay: new findings on contacts

Peter Bakker

*Aarhus University*

### 3.1 Introduction

The purpose of this paper is threefold. One is historical. It will shed a light on the vexing question of contacts and **intermarriages** between the Sri Lanka Malays and the Moors that have divided the students of Sri Lanka Malay (SLM). Evidence will be discussed on the basis of biological literature that sheds light on these questions, namely a previously uncited study on the molecular genetics of Sri Lankan populations (Section 3.2).

The second vexing question is just as controversial. It has become known as the so-called Tamil bias. Some people, including myself, have suggested Tamil as a source for the typological conversion of some form of colloquial, more or less isolational Malay towards the rather agglutinative and verb-final form of Sri Lanka Malay (SLM) that we find today. Others, notably Ansaldo and Nordhoff, claim that Sinhala was an important source, if not necessarily the main source, of the changes in SLM. I will provide background material that will shed light on the question, both from historical, and linguistic vantage points. On the basis of historical sources, most of them not previously quoted in the discussion of SLM among linguists, I will show that the connections between the Malays and the Sinhalese were rather tense in most of the historical period (Section 3.3). I will also show that, from a linguistic perspective, the influences from Tamil and Sinhala are compatible with early contact with Tamil and later contact with Sinhala (Section 3.4). Sociolinguistically also, the evidence is overwhelming for contacts with Tamil rather than with Sinhala. All of these more or less historical questions are important for a third question, and that is the genesis of the SLM language, to be discussed in Section 3.6. Before that, in Section 3.5, I will discuss whether we can call SLM a creole language – based on a historical and typological arguments. In order to be able to say anything about the genesis of the language, any answer should be informed by the historical data, even if they are limited. This scenario is not new, but it is based

on the appropriate contemporary sources. The final section (3.6) will place SLM in a wider perspective.

### 3.2 Malay intermarriage: genetic data on Sri Lanka Malays

One point of controversy is whether or not there was no, some or considerable intermarriage between Tamil speakers (mostly Moors, of partial Arabic descent) and Sri Lanka Malays. See [Slomanson \(this volume\)](#). In the travel literature I have not encountered a great deal of discussion on that theme (see Section 3.3). In this section I will summarize an article (perhaps the only one on the subject that includes data on Sri Lanka Malays) dealing with intermarriage between the major population groups of Sri Lanka, based on human genetics data.

[Papiha et al. \(1996\)](#) gathered genetic data from 508 individuals representing the Sinhalese, Tamils, Moors, Burghers and Malays of Sri Lanka, and from each of these five groups at least 100 persons were investigated. For this project, the authors studied a large number of loci of genetic variation. They compared the groups with each other and with other groups such as the Portuguese, Dutch, Arabs, and Malays of Malaysia and Singapore, in order to “clarify the range of genetic variation among the populations of Sri Lanka” ([Papiha et al. 1996:707](#)).

Before going into their data, it may be useful to briefly introduce the groups studied (see also [Coperahewa \(2009\)](#) for sociolinguistic data on Sri Lanka languages). The Sinhalese are the largest community of Sri Lanka, with ca. 74% of the population in 1981 (apparently the most recent census data available before the sampling of subjects at that time); they are mostly Buddhists and speak an Indo-Aryan language. The Tamils consist of two groups, the Sri Lanka Tamils (12.8%) and Indian Tamils (5.4%). The first group arrived in waves at an unspecified time, up to millennia ago, and the Indian Tamils arrived in recent centuries. The Tamils are mostly Hindus. Due to frequent intermarriage, the Papiha study groups two Tamil populations together. Tamil is a Dravidian language. The Moors comprise 7.1% of the Sri Lankan population. They descend from Arab traders who came to Sri Lanka in the 9<sup>th</sup> and 10<sup>th</sup> centuries, and intermarried with local people. The native language of the Moors is a non-standard variety of Tamil. The Moors are Muslims. The Burghers are descendants of Dutch and Portuguese settlers who intermarried with Sinhalese (Tamils are not mentioned, but as Nordhoff, p.c., 2011, points out, Tamils as well), or locals who identified with them. They are Christians. The Malays came first from Java and other parts of what is now Indonesia, and subsequently from Malaysia. They are Muslims. The Burghers and Malays comprise less than one percent of the population. The Veddas, the aboriginal inhabitants of Sri Lanka, and other minor groups, were not studied (an earlier study had confirmed the genetic distinctiveness of the Veddas, [Papiha et al. \(1996:708\)](#); see also [Stoudt \(1961\)](#)).

Papiha et al. conclude, confirming earlier studies, that the Tamils and Sinhalese are quite similar to each other genetically, and that the Moors, the Burghers and the Malays are dissimilar from the two major groups. They state that “[o]verall, genetic heterogeneity among the five populations was statistically significant”, and also that there is “significant variation among subpopulations” (p. 726). The distinctiveness of the Malays with respect to the other groups suggests that there was not so much intermarriage of the Malays with other groups, even though Papiha et al. remark that

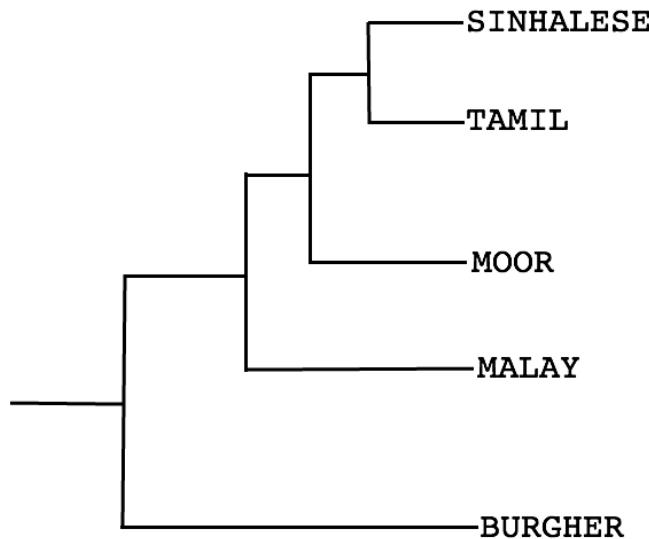


Figure 3.1: A tree of genetic distances between Sri Lankan populations (from Papiha et al. 1996:732).

“The Malays have married Sinhalese Tamil and Muslim women” (p. 711), but this statement does not seem to be based on the genetic research they report on. See Figure 3.1, which is Figure 3 from Papiha’s article (1996:732).

Papiha et al. have studied several frequencies of genetic signals that can be found in blood samples. Looking at their findings about the Malays in more detail, they found that the Malays completely lacked one type of haplotype (a set of distinctive genetic markers in DNA sequences), whose lack is “a typical characteristic of most Asian populations of Southeast Asia” (Papiha et al. 1996:719). This suggests that the Sri Lanka Malays have remained relatively close to their pre-Sri Lanka Malay genetic profile. In fact, data show that “the genetic composition of the Sri Lankan Malays has been preserved and still resembles the original populations.” (p. 733-734). Another marker also “indicates their affinity with Asian populations” (p. 723) and the mean heterozygosity was lowest among the Malays, which suggests “some degree of isolation” and preferred marriage partners among their own group (p. 726). Among the Malays, “the effect of admixture is not as pronounced” (p. 733). For another marker, “the Malays differed significantly from the Sinhalese” (p. 720), and for yet another one the “differences between the Malays and Tamils were also statistically significant” (p. 724). For another marker, the Malays differed significantly from the Tamils and Burghers (p. 721). Elsewhere, the Malays differ significantly from all the other groups (p. 723-724), or from the Sinhalese and the Tamils (p. 725), or from the Moors (p. 725).

On the basis of these data it is possible to estimate how endogamous the different Sri Lankan groups are. The Malays appeared to be the most endogamous, “possibly resulting from social, professional and religious isolation” (p. 728) of all the Sri Lanka

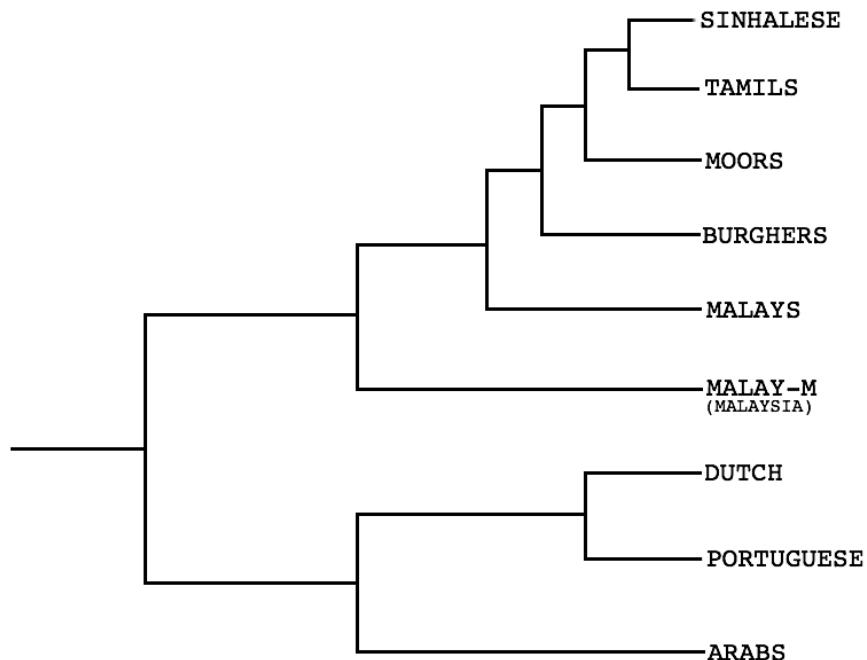


Figure 3.2: A tree of Sri Lanka populations with Europeans and Malays (from [Papiha et al. 1996:734](#)).

subpopulations. An earlier study ([Stoudt 1961:43](#)), on the basis of measurement of physiological characteristics, had concluded the same: “The Moors and Malays, as Muslims, are both endogamous groups and have virtually no marital and few extra-marital sexual relations with the other races of the island”.

Figure 3.1 showed the genetic differentiation between the five Sri Lankan groups. [Papiha et al. \(1996\)](#) also compared them with groups from outside the island. When Arabs, Dutch, and Portuguese genetic data are included, the result is a tree as in Figure 3.2 (taken from [Papiha et al. \(1996:734, Fig. 4\)](#)). Of the five groups, the Malays are the most genetically distant, then the Burghers (no doubt due to their European-Sri Lankan admixture), then the Moors and then the Tamils and Sinhalese. The presence of these other groups pulls the Malays and Burghers away from the other Sri Lankan populations, and closer to the Malaysian and European populations respectively, as one might expect.

When we try to interpret these data, we find that the Malays are the group with the least admixture with the other Sri Lankan groups. They are the most deviant of the Sri Lankan populations, and they show the closest genetic affinity with the ancestral

population of (Malaysian) Malays. On the other hand, there is a clear chronological pattern in the degree of admixture: the two oldest populations, the Tamils and the Sinhalese, both present on the island since millennia, display the highest degree of admixture. The Moors (whose presence on the island began a thousand years ago) are the next, then the Burghers (whose ethnogenesis can be estimated to date from the early 1600s or earlier) and the Malays (whose collective presence on the island dates primarily from the late 1600s). The fact that the Malays show the lowest degree of admixture may be due to the fact that they were the most recent population to arrive at the island.

In a sense, these data are inconclusive. Proponents of the claim that Malays and Moors intermarried may point out that the closest affinity of the Sri Lanka Malays among the island's population is with the Moors. Opponents may point out that the Malays show the least admixture of all the island populations, and maintain a strong East Asian genetic profile (i.e. Malaysian; no Javanese or Indonesian data were taken into consideration).

Even though intermarriage may have been relatively rare, this does not mean that the social, economic and religious contacts of the Sri Lanka Malays were exclusively in-group, exclusively with other Malays. In the next section we provide historical data about contacts with other groups.

### 3.3 Malays, Tamils and Sinhalese: a historical survey of relations

The genetic data discussed above suggest limited gene flow between Malays and other Sri Lankan groups. Who were the main contacts of the Sri Lankan Malays outside their own community? Some have claimed that these were Tamil, others that these were Sinhalese. Here I will provide data on contacts between the Sri Lanka Malays and the other Sri Lankan groups from contemporary sources.

I did an informal survey of several dozen early books on Sri Lanka, in order to find out what was said about the languages of the Sri Lanka Malays. The Malays have always been a small minority on the island, and the chances that travelers or scientists actually encountered and identified them are small, and even less frequent are descriptions of the group, or remarks on their language, especially on the basis of informed, direct observation. Here I list all those that I have encountered, without being selective in a certain direction, distinguishing between the pre-British period, and the colonial period of the island, keeping the independent indigenous Kandyan Kingdom distinct.

#### 3.3.1 The Portuguese and Dutch period

The oldest source on Malays and their languages in Ceylon/Sri Lanka mentioned in the linguistic literature is the Schweitzer quotation dating from 1670s, which mentions Ambonese, Sinhalese and Tamil marriage partners (e.g. [Nordhoff \(2009:14\)](#), [Smith & Paauw \(2006\)](#)). This is the only source known to me in which Malay-Sinhalese marriages are mentioned. Schweitzer, a German in the service of the in Dutch, met a party of "Ambonese" in Sitiwaka (Dutch territory, at that time), whose wives are described as follows: "The wives, who in part are Ambonese, in part Sinhalese and Malabar-

ian may not say anything [against the stripping of their ornaments].”<sup>1</sup> (Schweitzer 1931[1680]:106), [Nordhoff's translation from German (2009:14)] These Ambonese were polyglots: “Beside their own Language, they generally speak Maleysh, Cingulaish, Portuguese and Dutch”. The Sinhalese were said to be “mighty afraid” of the Ambonese, which suggests a social distance between the groups.

Schweitzer met Tamils “under the influence of Hollanders” (Schweitzer 1931[1680]), and his party also used Tamil guides. He also mentions “Cingulaish soldiers or inhabitants”, including those fighting with the Dutch against the Kandyans. Thus, Sinhalese soldiers fought both on the side of the Dutch and for the Kandyan King.

Powell (1973:26) mentions Portuguese sources claiming that already in 1587 the King of Kandy and Sitawaka, used soldiers who were not Sinhalese but “Javas, Caffirs and of other nationalities”. I have not been able to locate a contemporary source for this claim. In fact, the Sinhalese are frequently mentioned as not using arms, and as not engaging in trade. For their armies, they used non-Sinhalese forces. This habit may go back several centuries, as Marco Polo had observed the same in the 13<sup>th</sup> century: “The inhabitants of Ceylon are not fighting men but paltry and mean-spirited creatures. If they have the need of soldiers, they hire them abroad, especially Saracens.”<sup>2</sup> (Book 3, Chapter 14). These Saracens are likely to be Arabs, and probably ancestors of the Sri Lanka Moors. Emerson Tennent (1859) quotes 16<sup>th</sup> century Portuguese sources suggesting that the Moors combined Arabic and Tamil in their speech. With reference to the Sri Lanka Moors, Odoardo Barbosa “describes their language as a mixture of Arabic and Malabar, and states that numbers of their co-religionists from the Indian coast resorted constantly to Ceylon, and established themselves there as traders.” (p. 617). These Muslim immigrants from India may be the incentive for the Moors to shift to Tamil.

Robert Knox spent 20 years in the Kandyan kingdom from 1660 to 1679, after the crew of the ship, on which he was a sailor, was captured. He left a detailed description of the island and its inhabitants, printed in London in 1681, and an account of his escape (Winterbottom 2009). Knox does not mention the Malays in his writings, but as he appears familiar both with events around the Kandy court and every-day life in the countryside, it is likely that Malays were not a particularly prominent group, if not completely absent, in Kandy at the time. Knox spoke Portuguese (p. 171) and Sinhala, besides English, and he spoke Portuguese with the Tamils near Mannar, because he could not understand Tamil (p. 166-167). Knox met Tamils in the Kandyan kingdom who did not speak Sinhala (p. 159). This quotation may indicate a lack of bilingualism: “The Language they [Tamils, P.B.] speak is peculiar to themselves, so that a *Chingulays* cannot understand them, nor they a *Chingulays*. “ (p. 175.). The Kandyan King at the time knew Sinhala and Portuguese (p. 136, 175), but not Tamil.

Knox gives first hand descriptions not only of the Sinhalese, Moors and Tamils (p. 61), but also the Veddas and even the tiny Gypsy-like group of the Rodi or Rodiyas (“Roudeahs”, p. 71), and Muslim Moorish beggars who have their own “temple” in Kandy (p. 187). All this suggests a keen interest in ethnic differences, and it makes the absence of references to Malays even more remarkable.

The King’s armed forces were of diverse origin. The King’s body guards include Africans, called Kaffirs: “Next his own Person Negro’s watch. He hath also a Guard of

<sup>1</sup> ‘Malabarian’ means Tamil in this case.

<sup>2</sup> Two alternative translations: “the people are averse to a military life, abject and timid, and when they have occasion to employ soldiers, they procure them from other countries in the vicinity of the Mahometans.” (translation Emerson Tennent); “The people of Seilan are no soldiers, but poor cowardly creatures. And when they have need of soldiers they get Saracen troops from foreign parts. (translation Henry Yule)

*Cofferries or Negro's*, in whom he imposeth more confidence, then in his own People.” (p. 35). He also had “white soldiers”, commanded by Dutch and Portuguese people (p. 187), “*Europæans*; making them his great Officers, accounting them more faithful and trusty than his own People” (p. 187). It is not clear what the ratio of White, Black and indigenous soldiers was.

[Wickremesekera \(2004\)](#) studied the military operations of the Kandyan kingdom. She showed that the kings used local conscripts (presumably mostly Sinhala-speaking, with some Tamil, if they reflect the general population of Kandy) and foreign mercenaries. In and before the Portuguese period the latter were mostly from India, speakers of both Dravidian and Indic languages. Later also Europeans and Africans were pressed into service after being taken prisoner (p. 139-140). Malay troops in particular “played a prominent role in the Kandyan army” (p. 140). In the early 1800s an estimated 400-500 Malays were observed with Kandyan troops by British spies, but foreigners made up only a fraction of the soldiers in battle (p. 140). Modest numbers of Tamils functioned as personal guards to the Kandyan kings between 1740 and 1815 (p.139). The Kandyan Malay soldiers could have met speakers of both Tamil and Sinhala.

### 3.3.2 Malays in the coastal areas

Until 1815, Europeans only exercised power over the coastal areas, while having diplomatic or military contacts with the kingdoms of the interiors. Here I offer initial remarks about language use and intermarriage from the colonial area, thereafter from the Kandyan Kingdom.

The oldest source I found is from a Dutch account of 1672, accessed through an English translation of the original Dutch ([Baldaeus 1958-1959](#)). The author spent ten years on the island, between 1656 and 1665. The book contains detailed accounts of the military conflicts between the Dutch and Portuguese. Javanese are mentioned several times. One assault described for 1656 mentions the employment of Javanese, “Bandanezen”, “Lascarijnen” (local soldiers, probably Sinhalese) and “Toepassen” (free indigenous Christians, or descendants of Portuguese fathers and indigenous mothers; [Veth 1889](#)) and Europeans (p. 225/116).<sup>3</sup> Also earlier, during the Dutch siege on Colombo, mixed groups of soldiers had been used, consisting of people from Banda (Dutch possessions in Indonesia), Javanese and Mardijkers<sup>4</sup> (free indigenous people, probably from Dutch possessions; [Veth 1889:110ff](#)) for transport of equipment, and “Ceylonsche” for bearing away the wounded (p. 148/71). Similar mixed groups are mentioned elsewhere (p. 147/71: Javanese, Bandanese, Mardijkers, Dutch; p. 162: Bandanese and Mardijkers); Javanese (p. 226/116); Lascarijns from Matara, probably Sinhala speakers (p. 258/133); Dutch, Javanese and Lascarijns in 1653 (p. 275/144); Javanese and Lascarijns in 1650 (p. 279-80/145); Portuguese, Toepassen and Sinhalese (p. 280/146). The account also makes clear that desertion was common and that many soldiers shifted their loyalties. This source suggests cooperation and contact between the Malays and Sinhalese under Dutch rule in the 1650s and 1660s.

The next source is from 1797-1799. A Swiss regiment called *Régiment De Meuron* consisted of mercenaries. At the time of their arrival in Sri Lanka they worked for the Dutch, but they soon shifted their allegiance to the British in 1795. In a book about the Regiment, referring to the late 1700s, it was claimed that the Malays and the

<sup>3</sup>The page numbers refer to the 1958-9 translation of Baldaeus, followed by the original page in the Dutch 1672 edition.

<sup>4</sup>The ethnonym *Mardijkers* would be derived from the Malay word *merdeka* ‘free’. See e.g. [Veth \(1889\)](#).

Muslims (Moors) did not speak Sinhala, and also that the Malays treated the Sinhalese badly. This is a second hand source, and a source for this claim is not given ([Meuron 1982:155](#)).

Robert Percival, a captain in the British army, wrote a description of Sri Lanka, published in 1803, and a second edition appeared in 1805. He had arrived on the island in 1796, and spent several years on the coast and in the interior. Percival mentions intermarriage of the Malays with especially the Moors:

The Malays of the various islands and settlements also differ among themselves, according to the habits and appearance of the nations among whom they are dispersed. Yet still they are all easily distinguished to be of the Malay race. For, although they intermarry with the Moors and other casts, particularly in Ceylon, and by this means acquire a much darker colour than is natural to a Malay." ([Percival 1805:168](#))

The clothing of the Sri Lanka Malays was quite distinctive, but they had also adopted some habits from the Moors:

"The Malays of a higher rank wear a wide Moorish coat or gown, which they call *badjour*; not unlike our dressing-gowns. (...) The slipper or sandal in use among them, is the same with that worn by the Moors." ([Percival 1805:169](#))

According to Percival, "Low Portuguese" (presumably a pidginized or creolized variety) was the language spoken in communication between Europeans and Sinhalese, including in the army at the time, and Sinhala ("Ceylonese") was not used, but knowledge of a little Tamil ("Malabar") was necessary in the armed forces and in the domestic sphere.

"The Ceylonese language is so harsh and disagreeable to an European that few or none ever attempt to speak it; nor indeed is it at all necessary. The officers of the regiments stationed here have little opportunity and little occasion to learn it, few or none of the natives of the island but who have some other dialect being in our domestic service. Some little smattering of the Moor and Malabar language is necessary to be able to speak to the black servants of that description. **The low Portuguese is the universal language spoken amongst the Cinglēse in our settlements**, and indeed amongst all the natives who have any intercourse or connexion with Europeans; and it is also spoken by the Moor and Malabar servants." ([Percival \(1805:203\)](#), my emphasis).

Vicomte George Valentia, who traveled through the Orient in the early 1800s, visited Sri Lanka, and described *inter alia* the Muslim population. The Muslim Malays "can be contrasted with the natural inhabitants of the country, with whom they have no connection" ([Valentia 1813](#)).<sup>5</sup>

In 1845, the Malays were associated with the Tamils and the Moors by the Danish-French geographer Malte-Brun:

---

<sup>5</sup>All translations are mine. Unfortunately I had no access to the original English text of Valentia. The complete text in French is given here. "Il y a aussi deux castes de Musulmans, qui sont très nombreuses. La première est celle des *lebbies* ou des marchands africains que les Hollandais considéraient comme étrangers, et qui étaient taxés chacun à vingt-quatre schellings par an, taxe que le gouvernement anglais a supprimée. Les lebbies sont actifs et industriels. M. North leur avait donné un mufti, pour juger leurs procès; mais, comme ses confrères, il se laissait corrompre par des présens. En conséquence il a été destitué et ses fonctions

“These places are home to a population of 8,000-10,000 individuals, whose physical characteristics and religion are similar to those of the Arabs, whereas their language is derived from Malay. They are called Moplaïs by the Tamils, and are subjected to a chief who considers himself a vasal of the British”<sup>6</sup> (1845-1847: 316)

In the mid 1800s, the Sri Lanka Malays served in a military expedition against the Sinhalese of the interior, under the command of the British. They had killed many Sinhalese, and naturally this had done considerable damage to the rapport between the two groups, which lasted for some time. Pearson mentions 2,000 Sinhalese refugees:

Neither the Dutch nor the Portuguese had ever conquered the whole of the island, which was accomplished by the British in 1815. Since then there have been a few rebellions, which, however, were easily suppressed. During the last one, in 1848, some 2,000 up-country Sinhalese were put to flight by thirty Malays who wore the British uniforms, a proof that the ancient warlike spirit of the Kandyans is practically extinct (Pearson 1904; second hand source)

A contemporary Russian observer describes the atrocities of the Malays against the Sinhalese in some detail:

It is not too long ago that the British took possession of the interior of the island of Ceylon. They had a lot of trouble to succeed, because of the extreme thickness of the forests and the innumerable spider webs that fill these woodland mazes, where the British only advanced with a lot of difficulty, and where the Sinhalese kept hiding to attack them suddenly and destroy them. Therefore they employed the Malays, whom they brought from Java and who, like ferocious animals, have penetrated into those depths, tearing up the Sinhalese with their poisoned krises for the British.<sup>7</sup> (Soltykoff 1853:27-28)

Not surprisingly, it was also reported that the Sinhalese feared the Malays:

---

de juge sont exercées par le gouverneur lui-même. La seconde classe de Musulmans est celle des Malais, qu'on peut diviser en princes, en militaires et en brigands, quoique cette dernière dénomination puisse sans injustice leur être appliquée à tous. Les princes sont des souverains déposés, soit de Java, soit des îles adjacentes, soit de la presqu'île de Malacca, que les Hollandais ont bannis à Ceylan. Les militaires sont à la solde des Anglais et font de très-bons soldats. Quoiqu'ils soient si sensibles à l'honneur, que leur courroux est fatal lorsqu'ils se croient outragés, ils se soumettent sans murmure aux punitions infligées en vertu des règlements. On peut les opposer aux naturels du pays, avec lesquels ils ne contractent aucune liaison. Cependant, leur nombre n'est pas assez grand pour qu'ils soient dangereux. La conduite qu'ils ont tenue durant la guerre de Candy leur a fait infiniment d'honneur. Ils n'ont abandonné leurs postes qu'après que les officiers anglais leur en ont eu donné l'exemple; et même leurs chefs ont préféré être morts à l'ignominie.”

<sup>6</sup>“Ces lieux nourrissent une population de 8 à 10.000 individus, que leur caractère physique et leur religion rapprochent des Arabes, tandis que leur langue dérive du malais. Ils sont appelés Moplaïs par les Malabares, et soumis à un chef qui se reconnaissait vassal des Anglais.”

<sup>7</sup>Il n'y a pas longtemps que les Anglais ont pris possession de l'intérieur de l'île de Ceylan, qui était inconnu, et ils ont eu beaucoup de peine à y réussir, à cause de l'extrême épaisseur des forêts et des inénormables toiles d'araignée qui remplissent ces dédales sylvestres, où les Anglais n'avancent qu'avec beaucoup de difficultés, et où se tenaient cachés les Cingalis pour tomber sur eux à l'improviste et les détruire. Alors on a employé les Malais, qu'on a fait venir de Java, et qui, comme des bêtes féroces, ont pénétré dans ces profondeurs, et, déchirant de leurs *criss* empoisonnés les Cingalis, ont conquis ainsi l'intérieur de Ceylan pour les Anglais.

Since that time Malay regiments were created that are here. All these informations I have from the Malay who works for me. But, what is positive is that the Sinhalese, the indigenous inhabitants of Ceylon, dread the Malays extremely.”<sup>8</sup> ([Soltykoff 1853:27-28](#))

Even in the late 1800s, the Malay soldiers working for the British would not encounter any Sinhalese while on duty, only Europeans and Africans:

The troops stationed there by the British consist for two thirds of European soldiers. The remainder are Kafirs [Africans brought to Sri Lanka by the Portuguese, P.B.] and Malays. No Sinhalese was part of the army.<sup>9</sup><sup>10</sup> ([Cotteau 1889:412](#))

In the late 1800s, Tamils and Malays were reported to be employed in transportation in the same location. [Bruyas \(1898:65\)](#) described “voyageurs”, people who transport passengers between Trincomalee and Batticaloa, as belonging to two ethnic groups, Tamils and Malays, who are distinguishable by their clothing.

Many observers, by the way, describe the different way of dressing of the Moors, Malays, Tamils, Portuguese, Sinhalese (e.g. [Saltikov \(Soltykoff\)](#), [Bruyas \(1898:65\)](#), [Madrolle \(1926:89\)](#), writing about 1902). This shows that the ethnic groups valued their distinctness.

This selection of quotations, covering almost 250 years, point to two directions: In the Dutch period, at least in the period 1650-1667, Javanese, Bandanese, and other Malay speaking soldiers in the service of the Dutch were often associated with local soldiers, often Sinhala speakers. A century later, the Malays seem to have had almost exclusively contacts with Tamil speakers. The Malays associate with Moors and speakers of Tamil, sometimes even marrying them, while no connections with the Sinhalese are mentioned, except in the Dutch period, which may have extra weight as it took place in the formative period. On the contrary, the Malays in the service of the British massacred Sinhalese, who after this hated the Malays. This suggests potential influence from both Sinhala and Tamil on SLM.

### 3.3.3 The Malays in Kandy, 1800-1815

. All the sources in Section 3.3.2 relate to the coastal areas conquered by the Europeans – first the Portuguese, then the Dutch and finally the British. The interior of the island was an independent kingdom, virtually impenetrable by the Portuguese and Dutch armies, but finally conquered by the British in through several “wars” (1803-1805, 1815, 1817-1818). But even before the British period, there were numerous contacts – both diplomatic and military confrontations - between the Europeans and the Kandyans. Some were mentioned in .

SLM is also spoken in what used to be an independent kingdom, as is evident from [Nordhoff \(2009\)](#), a description of the Kandy variety for instance. What was the

---

<sup>8</sup>Depuis ce temps on a formé des régiments de Malais qui sont ici. Tous ces renseignements, c'est de mon Malais que je les tiens; mais, ce qu'il y a de positif, c'est que les Cingalis, habitants aborigènes de Ceylan, redoutent extrêmement les Malais.

<sup>9</sup>Les troupes que l'Angleterre y entretient sont pour les deux tiers composées de soldats européens; le reste est cafre ou malais; aucun Cingalais ne fait partie de l'armée.

<sup>10</sup>Sebastian Nordhoff (p.c.) points out that other indigenous groups such as the Tamils and the Veddas were likewise excluded from the British army. Indeed the Brits typically used soldiers from other parts of the empire, in this case mostly India and Malacca, not local ones.

language situation there before the conquest, and what in particular do we know about the Malays in the area before approximately 1850?

Sri Wickrama, the last King of Kandy, used Tamil soldiers in the early 1800s (Powell 1973:88), by 1800 also Malay mercenaries (*id.*), and by 1803 “Kaffirs” (p. 87, 89, 123, 147). The Kandyan chief communicated in Tamil with the British (Powell 1973:125). Also in the second Kandy war, an invasion on behalf of the Kandyan king was conducted by “1,000 men, many of whom were Tamils or Malays” (Powell 1973:204), and at a later stage “only 200 Malabars, the same number of Kandyman militia and a few Moormen, Kaffirs and Malays” (p. 215, 218), and another group of 50 Malabars (p. 219). Kandyan representatives wrote their names “in a mixture of Sinhalese, Nagari [Indian] and Tamil scripts” (Powell 1973:231). It seems that Tamils dominated the Kandyan army: in the 1815 convention “all male people of the Malabar caste” (Powell 1973:284) were expelled from the Kandyan provinces, presumably to prevent a military action against the newly established British regime.

All this seems to point to the predominance of Tamil speakers even in the army of the Kandyan king, and thus a more likely source language for the Malay soldiers fighting for the Kandyan king.

Why the Sinhalese were “as little interested in trade as they were in manual labour” (Powell 1973:56,225) may have to do with the presence of three other populations that performed both functions: Tamil-speaking Hindus from Sri Lanka and South India, and Muslims of Malay and partial Arabic descent (Tamil-speaking Moors) (*id.*). The Moors are also regularly mentioned as important in trade.

### 3.3.4 The Malays and the Tamil and Sinhala languages

These quotations (which have not been selected in order to promote a particular point of view in the discussion of the “Tamil bias”) attest to the historical importance of contacts with Tamil speakers in the later period and with Sinhala speakers in the Dutch army. If there has been further influence from the Sinhala language on Sri Lanka Malay, this must date from after 1900, perhaps mostly from after 1950. However, also in the 1600s Malays may have been in direct contact with Sinhalese women (one quotation) and before 1796 they were certainly enlisted in armies in which Sinhalese were also present. There is only one primary source on language contact, and that involves Sinhala. In trade, Tamil may have been more important, and certainly in religion, Sinhala played no role for the Malays. We can conclude that the Malays preserved a clear separate identity, also in dress, from all other groups. This does not imply that the Sri Lanka Malays were a socially isolated group. Malays had historically more documented contact with Tamils and Moors than with Sinhala-speakers, with the possible exception of the Dutch period in the army. Religion, location, and employment patterns may all have played a role here. Tamil and Portuguese, not Sinhala, were the main contact languages of the Sri Lanka Malays until at least the early 20th century. In the army, Portuguese was probably a lingua franca, alongside the main language, English.

Religious allegiances must have been strong, especially when religions enforce conversion, or when members of a religious community expect all marriage partners to have the same religion. Both of these considerations apply in the case of Muslims, and as the Tamil Moors and Sri Lanka Malays share a religion, in fact one that requires marriages with Muslims only, it is to be expected that this form of contact existed. Ethnic bonds were also strong for the different groups.

We could say that the future is here, and these results point to a Moorish Tamil direction as far as the historical facts are concerned with a considerable quantity of quotations on Malay-Moor contact, and some source mentioning Sinhalese – though there is also some indirect evidence of contacts with the Sinhalese in the pre-British period. In the next section we will look at the linguistic and sociolinguistic data, and see that those also point to a historical Tamil influence, where Sinhala influence became important only recently.

### 3.4 Linguistics: the Tamil bias – or the Sinhala bias?

No researcher on SLM would contend that the radical changes in the structure of SLM from SVO to SOV, from prepositional to postpositional, the development of case-marking, and so on, have developed spontaneously, were already in place in Indonesia, or were the result of a process of drift. No doubt these processes of change were contact-induced. The same process took place in Sri Lanka Portuguese, with an almost identical outcome. The largely parallel changes in the two languages (cf. [Bakker 2006](#)) must have taken place under the influence of the major Sri Lankan languages, and that is probably also uncontroversial.

There are two matters that have been a subject of debate with regard to contact. The first matter is, **when** did the radical changes take place; Was it in the 20<sup>th</sup> century, or in the 18<sup>th</sup> century? Alternatively, could the changes have take place gradually? The second matter is the question as to **which Sri Lankan language** triggered the changes that led to SLM differing so radically from all other varieties of Malay? Was it the minority language, Tamil, or was it the majority language Sinhala? I will only discuss the second question here.

Many authors had routinely assumed that the major or only source of influence was Tamil, until this question was taken up by Ansaldi (2008, and earlier), who claimed that it was actually Sinhala, or at least more Sinhala than Tamil. [Ansaldi \(2005, 2008\)](#) referred to an ostensible “Tamil bias” in the study of Sri Lanka Malay. [Nordhoff \(2009\)](#) devotes section 2.7 to the plausibility of contact with the Tamil-speaking Muslim Moors and the Sinhala speaking Buddhists, providing a balanced overview of the different positions.

Let us go review what Nordhoff, in his generally excellent book on Kandy SLM, has to say with respect to the respective potential source languages (also see [Nordhoff \(2008\)](#) for an updated survey).

#### 3.4.1 Structural influences from Tamil or Sinhala

##### Loanwords

From which language are there more lexical borrowings into SLM? Nordhoff writes (2009:257):

There are many Tamil loanwords for basic vocabulary terms, like *kattil* ‘bed’, as well as many animal names (*vanaati* ‘butterfly’, *vavval* ‘bat’). The Tamil loans are phonologically integrated in the SLM system, and also used by speakers who do not know Tamil.

The fact that quite intimate and basic words are borrowed, plus words for local fauna, suggests that Tamil lexical influence is rather deep and old. The loans are also

phonologically integrated, which is usually indicative of greater age. Plus, people who do not speak Tamil also use them.

There are also loans from Sinhala. These are treated in a very different manner by SLM speakers:

The Sinhalese borrowings seem to be nonce borrowings, and the speakers are aware of the code-switch. (p. 163)

Any Sinhala word can be code-switched, possibly not, or not always, phonologically integrated. Still, this is quite marginal, as Sinhala is stigmatized (S. Nordhoff, p.c.). Both the occurrence and its stigmatization are typical of code-switching contexts. However, no conclusion can be drawn with certainty about the age of this pattern. It is theoretically possible that speakers have been mixing Malay and Sinhala in this way for centuries, keeping Sinhala words phonologically distinct. In general, however, the lack of permanent loans and the apparent lack of phonological integration, suggest recent influence in a context involving expanded knowledge of Sinhala as a second language. [Paauw \(this volume\)](#) identified 71 words of Tamil origin and 19 of Sinhala origin. This also indicates a stronger lexical influence from Tamil than from Sinhala. The lexical data suggest early influence from Tamil, and late influence from Sinhala.

### **Phonology**

With regard to phonology, Nordhoff discusses two aspects in relation to the source of contact. A number of specific points where Tamil influence had been suggested are discussed, and Nordhoff concludes that these do not hold:

In phonology, the phoneme inventories do not suggest Tamil influence, and the purported constraints on the occurrence of initial retroflexes do not hold either. (p. 60)

One of the properties of phonology is stress, and according to [Nordhoff \(2009:131\)](#), SLM stress is hard to be determined, if it exists at all. This elusive nature of stress is shared with Tamil, but not with Sinhala, where stress is assigned based on weight and position of the syllables.

On the whole, however, there are influences from both languages: “Phonology is a mix of Sinhala and Malay features” (p. 60). Phonology thus appears inconclusive, or at least, there is not one clear source for the phonological properties of SLM.

### **Semantics**

With regard to semantics, Nordhoff points to influence from both languages: “*Anthi* is very common. The conflation of future and habitual is of Tamil influence.” (2009:294) and “Tamil has a similar use of the numeral *one* for indicating vagueness ([Schiffman 1999:135](#), cited on p. 322).” Sinhala is mentioned as an adstrate influence: both SLM and Sinhala have two existential forms, one for animate and one for inanimate subjects (the former being *duuduk* in SLM). Nordhoff writes “that the grammaticalization path of *duuduk* can be explained by this [Sinhala] adstrate influence” (2009:169).

Within the noun phrase, a number of controversies have surfaced. The interplay of marking accusative and animacy has been attributed to Tamil ([Smith \(2003\)](#); see [Nordhoff \(2009:59\)](#)). Nordhoff claims that SLM falls between Sinhala and Tamil (p. 60). The identity of the ablative and instrumental cases reported for SLM is attributed to Sinhala

by Ansaldo in an unpublished paper: “Ansaldo (2005) notes that this morpheme shows an ablative/instrumental syncretism, as also found in Sinhala.” (Nordhoff 2009:345). Further the marking of (in)definiteness is obligatory in Sinhala, not in SLM, and is hence closer to Tamil according to Smith. Nordhoff contradicts this and claims that indefiniteness marking is obligatory, and hence closer to Sinhala.

In the verb, the “infinitive combined with the interrogative clitic =*si* is used to request permission” in SLM and Sinhala, not in Tamil (Nordhoff 2009:280). This, however, appears to be present in Tamil as well (S. Nordhoff, p.c.). Hence, semantic data point in both directions, and are inconclusive.

### Morphology: plural marking

According to Smith, number marking is optional in Tamil, obligatory in Sinhala, and SLM is like Tamil. Nordhoff claims that number marking is also optional in SLM, inherited from trade Malay. On the other hand, the “reinforcing use of *pada* parallels the use of -*gal* in Tamil” (Mr. X, Mr. Y, Mr. Z *pada*, Nordhoff 2009:324).

### In short

For phonology, there is no evidence pointing to a historical shift that must have been from Tamil (or Sinhala, for that matter) to Malay (but some features such as retroflex consonants and distinctive vowel length are definitely Lankan). The lexicon shows an early dominance of Tamil, and (late) twentieth century dominance of Sinhalese. As for semantics, there is evidence of more recent Sinhala influence in the case syncretism of SLM (this is a pragmatically salient structure), and the existence of divergent grammaticality judgments on the irrealis (p. 294).

For semantics, there is evidence of more recent Sinhala influence (case syncretism), e.g. pragmatically salient structures, and existence of divergent judgments of grammaticality (p. 294, 1st p. irr.). There is also Tamil influence, for instance in the merger of future and habitual.

If we only had the linguistic data at our disposal, we would have to conclude that there is early influence from Tamil. Tamil was the first, or at least the most important, second language learned in (most of?) the SLM communities. This is also found in historical sources. Only later, perhaps as late as the 20<sup>th</sup> century, was Sinhala added to the repertoire for the *general* community. Sinhala influence is mostly semantic (convergence), and probably recent. There is also textual evidence that support such a claim, as in example (1) (Nordhoff 2009:257):

- (1) *Itthu muusing=ka cinggala thraa*  
 DIST time=LOC Sinhalese NEG  
 ‘At that time there was no Sinhala.’

This is corroborated by the history of education among the Sri Lanka Malays (cf. Bichsel-Stettler 1989). Schools for the Malays were established from 1812, where the curriculum included Malay and Tamil until the last quarter of the 19<sup>th</sup> century (Hussainmiya 1990:96-99), and Sinhala became a language of instruction for them only in the 20<sup>th</sup> century.

### 3.4.2 Educational and religious history of the Sri Lanka Malays

The Sri Lanka Malays had a strong connection with the armed forces throughout their history in the Dutch and British periods. There were military schools for Malay boys, from 1815-1860. [Hussainmiya \(1990:96\)](#) writes the following about the languages of instruction at these schools: “The education provided was excellent and included Malay language, Tamil and English, but no Christian religious education”.

According to Bichsel-Stettler, by the end of the 19<sup>th</sup> century: “Generally, Malays either opted for Tamil-medium or, most often, for English-medium schools ([Bichsel-Stettler \(1989:24\)](#), also cited by [Nordhoff \(2009:26\)](#)). Sinhala-medium education was not chosen until the middle of the [20<sup>th</sup>] century ([Bichsel-Stettler 1989:27f](#))”.

The Sri Lanka Malays were Muslims when they arrived in Sri Lanka. Malays had their own mosques from 1783 ([Nordhoff 2009:47](#)). Mosque services took place in Malay and Tamil ([Hussainmiya 1990:124](#)), according to the above-mentioned sources (which are not always specific as to places and periods), but Sinhala is never mentioned.

### 3.4.3 Recent developments: From quadrilingualism to bilingualism

A recurrent theme in [Nordhoff \(2009\)](#) and in the texts, is the pride in the multilingualism of the older generation, and regret with respect to language loss among the younger generations. The older people are generally quadrilingual (SLM, Tamil, Sinhala, English), whereas many young people are losing the connection with their Malay heritage and only speak Sinhala and English.

Nordhoff describes this as follows:

While the older generations have very good command of grammar, style and register in all four languages, this is not necessarily the case with the younger generation, who often only have full command of Sinhala and a local variety of English. Knowledge of Tamil is often absent, and full command of Sri Lanka Malay (which would permit to convey any message) is as well. ([Nordhoff 2009:32](#))

For a sociolinguist, this would point to a situation of shift, away from Tamil and SLM, and towards English and Sinhala.

### 3.4.4 Tamil or Sinhala?

If we now compare the linguistic data presented in this and the preceding section, then everything (historical, educational, religious, structural, sociolinguistic) appears compatible with a scenario of early influence from Tamil, and later influence (from the mid-1900s?) from Sinhala. If Tamil were the most important L2 for the Malays, then the founder principle should predict, from the historical data, a much more significant impact from Tamil than from Sinhala, and that is indeed what we find.

However, there is growing evidence for Sinhala knowledge by Malays in the Dutch period. There is the Schweitzer observation, and additional indirect evidence of early Sinhala competence in that Malays often belonged to the same army units as the Sinhalese in the Dutch period. However, Portuguese may have been their lingua franca ([Baldaeus 1958-1959:346 and passim](#)). Most of the Sinhala influence is found in areas typically found in current language contact (cf. [Matras 2008](#)). There may indeed be a “Tamil bias” in SLM studies, but rightfully so: it is the only language (aside from

Malay itself) that is mentioned for education, outgroup contacts and religion from the late 1700s to the mid 1900s.

### 3.5 Is SLM a creole?

SLM and Sri Lanka Portuguese (SLP) are treated as falling under the rubric of creole studies ([Nordhoff this volume](#)). But are these languages creoles? Both languages underwent structural influence, but hardly any lexical influence from other languages, to the extent that they started to resemble one another structurally; both languages converged to a great extent with the local languages (cf. [Bakker 2006](#)). SLP was traditionally called a creole, SLM was not (at least not until recently, and in fact only by some of the linguists involved). Perhaps there is a Eurocentric bias here, in that linguists have been more aware of creolized Portuguese in South Asia than of Malay.

If one takes a certain structural prototype when defining what a creole is, then SLM is very far from a prototypical creole. Derek Bickerton, for instance, who worked on structural properties of creoles, would not have considered SLM a creole, and probably no one else would, on the basis of its structural features. If creoles have SVO word order, no inflection and preverbal TMA ([Seuren 1998:292-293](#)), then SLM could not be classified as a creole, as only TMA is creole-like, remotely resembling prototypical creole TMA systems. SLM is, however, an SOV language with inflectional morphology, and therefore ought not to be classified as a creole. This is a structural-typological definition.

One can also take another point of departure. If one says that a creole language is a language that inherited, for historical reasons such as slavery or migration, only part of the lexicon and grammar of a language, which also underwent influence from other languages, then SLM would fall under that definition, and hence be classified as a creole. Thus, if one defines a creole as a language whose speakers developed a partly new grammar on top of the limited set of features inherited from the combination of lexifier and substrates, then SLM would qualify as a creole. In contrast to other creoles, however, the structural innovations in SLM are not mainly from unknown or unidentified sources, but partial copies of existing structures. The innovations in SLM are not creative solutions triggered by the communicative needs in situations of interethnic communication, but changes triggered by the frequent use of several languages, in a situation where grammatical norms are loosened (cf. [Matras 2008](#)).

[Nordhoff \(2009\)](#) avoids the term “creole” for SLM, whereas [De Silva Jayasuriya \(2002\)](#) and [Smith et al. \(2004\)](#) call the language a “creole”. [Robuchon \(2003\)](#) discusses creolization processes, but he does not consider the language to be a creole. ??[Slomansson \(2011\)](#) does not consider the language to be a creole either. Also [Ansaldi \(2008\)](#) argues against the use of the label “creole” for SLM.

In my view SLM is indeed a restructured language, but not a creole from a structural or historical point of view. At no point in its history was Malay in Sri Lanka reduced lexically and grammatically to a medium of interethnic communication that one could call a pidgin or a basic variety (cf. [Becker & Veenstra \(2003\)](#) for discussion of the latter in the context of creole studies) – which would be a necessary part of my definition of a creole language.

In recent work, we have shown that creoles form a distinct type of language, with a number of properties that in combination set creoles apart from non-creoles ([Bakker et al. 2011](#)). If we take existing sets of features as formulated as typical for creoles as a point of departure, and we sample a set of non-creoles and compare them, then

creoles are clearly different from the non-creoles – despite the fact that no creole has all the relevant features. Similarly, if we take existing sets of features selected by typologists for the languages of the world, and add creoles to the set of languages, we get the same results: creoles stand out as a distinct group among the languages of the world (see Bakker et al. 2011 for details). The findings by [Szmrecsanyi & Kortmann \(2009\)](#), based only on English varieties, also show that creoles are structurally distinct varieties, different from both first language varieties and second language varieties of English. Unpublished preliminary findings by Michael [Cysouw \(2009\)](#), based on a comparison of 46 features common between an early version of the *Atlas of Pidgin and Creole Structures* APICS (which includes SLM) and the typological *World Atlas of Language Structures* (WALS; [Haspelmath et al. 2005](#)), made him conclude that: “APICS languages are clearly different from WALS languages”. In Cysouw’s findings, SLM does not cluster with the creoles, but with the non-creole WALS languages.

We obtained similar results for SLM compared to a sample of creole and non-creole languages. For our study, we chose *inter alia* the dataset gathered for the *Comparative Creole Syntax* project, published in [Holm & Patrick \(2007\)](#). Those authors tried to cover a broad range of categories ( $n = 97$ ) that were reputed to be characteristic of creoles. To give an impression of the range of features covered, here are the category headings, where different types were distinguished under each heading:

- the verb and stative/dynamic distinctions,
- tense-mood-aspect,
- complementizers,
- dependent clauses,
- negation,
- passive,
- adjectives as verbs,
- copula,
- noun phrase structure,
- the expression of possession,
- pronouns and case,
- coordinating conjunctions,
- prepositions,
- miscellaneous.

In brief, they cover a wide range of morphological and syntactic features that a language may or may not have. All features have been scored as either present or absent – and not all of the features appear to be recurrent in all creoles. In fact, none of the 97 selected features are shared by all the sampled creoles, and a fair number of the features were only found in a minority of the sampled creoles. This shows the relative variety of the creoles (they are by no means a homogenous set). Apparently many

abbreviation	name	lexifier	location
AN	Angolar	Portuguese	Gulf Guinea
BD	Berbice Dutch	Dutch	Caribbean
CV	Cape Verdean	Portuguese	Gulf of Guinea
DM	Dominican	French	Caribbean
HA	Haitian	French	Caribbean
JA	Jamaican	English	Caribbean
KO	Korlai	Portuguese	South Asia
KR	Krio	English	West Africa
KR	Kriol	English	Guinea coast
NB	Nubi, Kinubi	Arabic	Interior Africa
ND	Ndyuka	English	Caribbean
NG	Nagamese	Indic	South Asia
NH	Negerhollands	Dutch	Caribbean
PL	Palenquero	Spanish	South America
PP	Papiamentu	Span/Port	Caribbean
SC	Seselwa (Seychelles)	French	Indian Ocean
SL	Sri Lanka Malay	Malay	South Asia
TP	Tok Pisin	English	Pacific
ZM	Zamboangueno	Spanish	Pacific

Table 3.1: Abbreviations of creole names in Figure 3.1.

creoles do not conform to the prototype that the (Atlantic-oriented) editors presumed to be typical for creoles.

Holm and Patrick asked the contributors to their book to collect the same information on 18 different creoles, thus providing directly comparable data on 18 creoles. The syntactic categories studied by the contributors to Holm & Patrick are clearly biased in the direction of a creole profile. Table 3.1 lists the contact languages (18 creoles, plus SLM), used in the comparison, and the abbreviations used, the respective lexifiers and the approximate location of the creoles. It is immediately clear that there is a large range of lexifiers (Arabic, Dutch, Indic, English, French, Portuguese, Spanish) and they span the globe. It is a balanced sample, with geographical, lexical and social differences, be it somewhat biased towards Caribbean creoles.

Figure 3 shows a phylogenetic network comprising the 18 creoles as discussed in Holm & Patrick, to which Sri Lanka Malay was added. The SLM scores were added by me, on the basis of Nordhoff’s (2009) description. The longer the lines are that radiate from the center, the more deviant a language is. Transverse lines point to conflicting signals. The star-like Figure 1 suggests that, at least on the basis of the 97 morphosyntactic features selected by Holm & Patrick, these creoles have just as much in common as they differ from one another, because of the roughly equal length of the spokes – with the notable exception of SLM.

Note that the creole languages cluster only loosely around lexifiers in that the Spanish and Portuguese creoles cluster, but otherwise there is no significant clustering of regions, age or type of creole (see Bakker et al. 2011 for details, where statistical tests confirm the findings). The odd one out is clearly SLM, with its much longer spoke than what we find for the other languages, all of them creoles, suggesting it does not fit in with the other languages. On the other hand, SLM does cluster with the other two South Asian languages, Korlai Creole Portuguese and Nagamese (also called “Pidgin

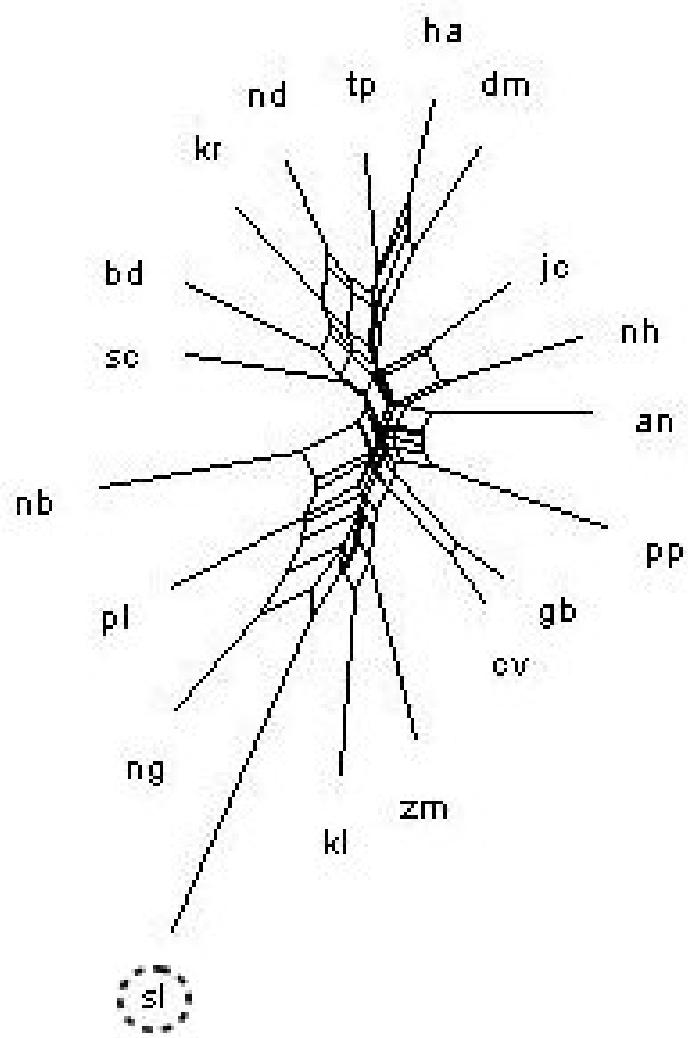


Figure 3.3: 18 creoles and Sri Lanka Malay (SL). SLM is found at the bottom, along with the non-creoles.

abbreviation	name	affiliation	location
AIN	Ainu	Isolate	East Asia/Japan
BRA	Brahui	Dravidian	South Asia/N. India
IND	Indonesian	Austronesian	Pacific, Indonesia
KOL	Yukhagir	Isolate	Asia, Siberia
KOY	Koyra Chiini	Nilosaharan, Songhay	Africa, Mali
MAN	Mandarin	Sino-Tibetan	Asia, China
MIN	Mina	Afro-Asiatic, Chadic	Africa, Cameroon
PIR	Pirahã	Mura, Amerind	S. America, Amazon
SLM	Sri Lanka Malay	mixed?	South Asia

Table 3.2: Abbreviations and affiliations of non-creoles.

Assamese”, actually a creole). These three languages apparently share a number of properties, most likely being those that developed under the influence of Dravidian and Indo-Aryan languages.

What happens if we add more non-creoles to the creoles? We added eight non-creoles from eight phyla, after having extracted features from published grammatical descriptions. These are listed in Table 3.2. All of these languages are spoken in regions far removed from one another, excluding direct contact or areal influence. From each family, we have selected a member that is most known for its more analytic profile (often regarded as a typical profile for creoles). This was done in order to skew the data towards creole-like structures. For instance, we chose from the Afro-Asiatic family the isolating Mina language rather than morphologically quite fusional Bedouin Arabic. Other non-creoles have been selected for their simplicity score in Parkvall (2008)’s matrix, because of the supposed lesser complexity of creoles, thus further biasing the sample towards a creole profile. As a 9<sup>th</sup> language we added SLM. The results can be seen in Figure 3.4.

Figure 3.4 shows two quite clear things. Creole languages, all starting with C- followed by the two letter code used in Table 3.1, all cluster together, and the non-creoles, all marked with three capitals (see Table 3.2 for explanation), also cluster together. This means that creoles, at least on the basis of these 97 features, are very different from this selection of the languages of the world, intentionally skewed in the direction of some sort of creole profile. SLM appears among the non-creoles and relatively far from its lexifier Indonesian (IND). The language is most closely connected to two other continental Asian languages, Brahui (BRA, Dravidian, Pakistan) and Kolyma Yukaghir (KOL, isolate, Siberia).

This corroborates one thing that has been suggested before by several researchers, namely that SLM is typologically far removed from creoles, and much closer to Dravidian languages. If we can take the 97 creole features as a litmus test, then SLM does not qualify as a creole. From this structural-typological perspective, SLM is not a creole. Note that this is based on a predefined set of creoles and their typological profile.

### 3.6 How did SLM come into being?

SLM is interesting because it is such an extreme outcome of language contact – but it is far from unique in the world. The combination of its Malay lexicon, with its Tamil-Sinhala structure, without the direct borrowing of morphemes, has inspired all those

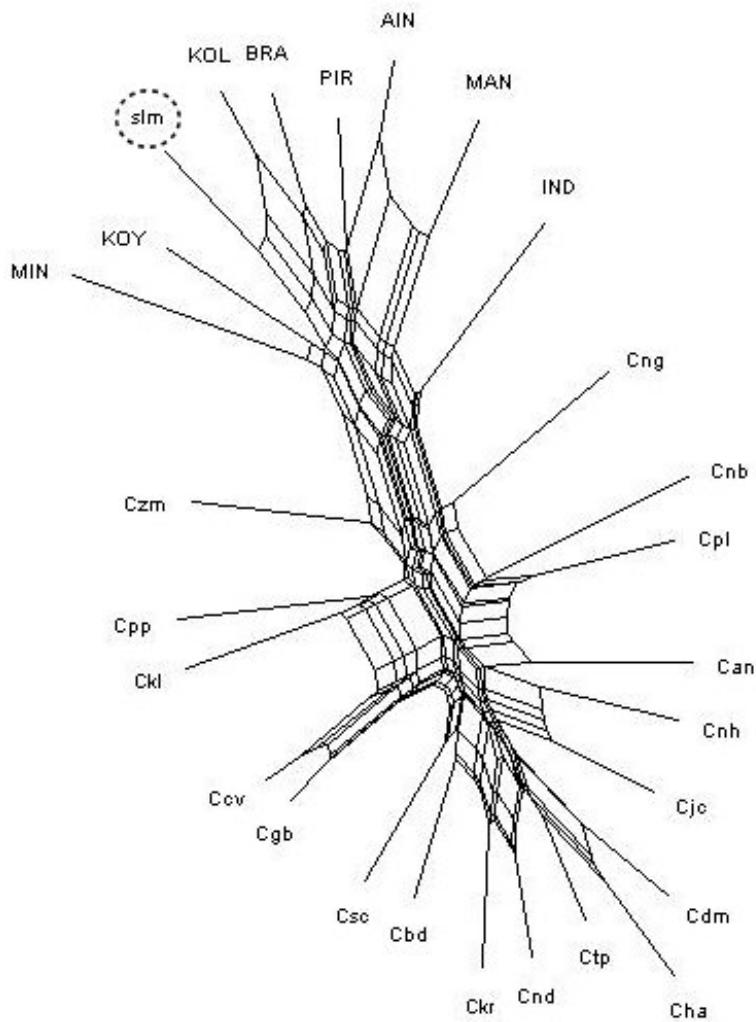


Figure 3.4: Sri Lanka Malay among creoles and noncreoles (creoles marked with initial C, and noncreoles in CAPS; slm = Sri Lanka Malay) Sri Lanka Malay is found at the top left, among the non-creoles and not close to IND (=Indonesian). Out of the nine non-creoles, Sri Lanka Malay is typologically closest to Kolyma Yukhagir and Brahui.

linguists and others who have investigated the language to speculate on its genesis. I will outline my view of that matter in this section.

Some simplification processes of Austronesian took place in the genesis of Proto-Malay from its Austronesian predecessor. When Malay developed into a language of interethnic communication in Southeast Asia, more simplification took place, leading to vernacular forms of Malay and pidginized forms such as the set of contact vernaculars called Bazaar Malay. Vernacular forms of Malay, including possibly pidginized varieties, were subsequently transplanted from insular South East Asia and Malaysia to Sri Lanka.

Once in Sri Lanka, the Malays also learned one or more of the local languages. Most or all of them also knew other languages, such as some form of Portuguese, Sinhala and Tamil, or one additional language, Tamil. Tamil was a more likely choice because of religious affiliations as well as the political, educational and military situation. This led to a situation in which both Malay and Tamil were used (in some cases perhaps also Sinhala, but historical evidence is scanty thus far, and linguistic evidence is thin). Most likely there was even a double-nested diglossia, where both languages also had their L and H varieties (colloquial Sinhala, formal Sinhala; vernacular Malay and Indonesian/Malaysian Malay). At some point in time, most likely rapidly, this multilingual situation led to a fairly radical convergence toward “Lankan”, in which SLM developed into the language it is now.

The resulting language SLM shows, summarized in a simplified way, the following properties. At the level of the sentence, SLM has developed Lankan syntax: the language is verb-final. In the verb phrase we find preservation of the position of creole/vernacular Malay preverbal marking, not marking with stem modification (Ablaut, consonant changes) and verbal suffixes as in Tamil and Sinhala. It has been claimed that TMA marking in creoles shows remarkable similarities in the semantic and syntactic properties (some of which are also found in vernacular forms of Malay), but SLM deviates significantly from those semantic categories. The meanings of the vernacular forms of the TMA show a semantic merger with the Lankan categories. Under influence of the Lankan languages, the number of preverbal markers was expanded, to fit the same categories as found in Tamil.

In the noun phrase, the convergence towards Lankan languages was more radical. The noun phrases became noun-initial, i.e. instead of prepositions, the language developed postpositions, postclitics and case markers. The semantic categories found in the NP show much overlap with Lankan languages. The result is a language that is semantically very close to the Lankan languages, and the sentential and NP structures are also syntactically very Lankan. Sinhala and English influence affected the language more recently, mostly in the lexical domain, and in discourse strategies, in the form of alignment of pragmatic systems along the line of [Matras \(2008\)](#).

### 3.7 Conclusion

What happened with SLM is fairly unusual on a global scale, and it must have to be a special type of social situation in which this process took place. It is far from unique, though. It is reminiscent of processes elsewhere, which have been labeled convergence, metatypy and fusion, and such phenomena can be encountered in all sprachbund situations (linguistic areas). SLM (with SLP) may have attracted more attention since it contradicts an assumption that a linguistic area must consist of more than two languages, as for instance in [Thomason \(2000\)](#). Of course many languages are

spoken in the Sri Lanka sprachbund (Tamil, Sinhala, Malay, Portuguese), but there is no evidence of widespread mutual multilingualism, sometimes considered a precondition for convergent areas.

Where do we find similar processes, in which a language preserves all of its lexicon, but models its structure on other languages, at the same time using native morphemes to develop semantic equivalents of categories in the other language? In all linguistic areas we find such processes, but their exact nature differs from place to place. Furthermore, linguistic areas rarely if ever show identical structures of two languages, and specific languages in a sprachbund typically share only a subset of the features deemed typical for the area.

Where do we find the more radical and overarching process as in SLM (and SLP)? One case that has been discussed in connection with creole studies is the famous Kupwar scenario, in which local Dravidian and Indic languages diverged from related varieties spoken elsewhere, and in which the languages involved came to share a large proportion of their respective grammars. [Nadkarni \(1975\)](#) describes the influence of Dravidian on the Indo-Aryan language Konkani.

A case much more reminiscent of the SLM case is that of the Wutun language ([Li 1983, 1984, Lee-Smith & Wurm 1996](#)), and also some of the other languages spoken in the same region that have converged towards an agglutinative and verb-final profile ([Wurm 1996](#)). Wutun is lexically Chinese, but as is true of the surrounding varieties of Tibetan and Mongolian, it is verb-final, has case markers and an agglutinative structure – a development very reminiscent of what happened with SLM.

Also in regions in which Austronesian and Papuan languages border on one another similar things have been reported, such as the Papuanized Takia languages (Ross 1996; other examples are given there as well, not all equally convincing). The Non-thaburi Malay variety in Thailand ([Tadmor 1992, 1995, 2004](#)) converged on Thai, and in so doing, became an isolating language. The Tariana Arawak language in Brazil converged in the direction of local Tucanoan structures ([Aikhenvald 2002, 2003](#)). Such processes are not uncommon – in fact, the whole principle of linguistic areality is based on convergent changes. [Matras \(2008\)](#) provides an attractive framework for language contact, embedded in a functional and pragmatic perspective, in which bilingual speakers reduce their cognitive burden by replicating discourse patterns from a pragmatically dominant language into the home language.

A process affecting the entire grammatical system is rare, perhaps even nonexistent. Even the landmark examples such as Sri Lanka Malay, Kupwar languages and Takia retain a number of their original structures – Sri Lanka Malay for instance retained the preverbal position of tense, mood and aspect, despite Tamil and Sinhala suffixation. A study comparing the more radical changes cross-linguistically is a desideratum for contact linguistics.



# Bibliography

- Aikhenvald, Alexandra Y., 2002. *Language contact in Amazonia*. Oxford: Oxford University Press.
- Aikhenvald, Alexandra Y., 2003. "Mechanisms of change in areal diffusion: new morphology and language contact." *Journal of Linguistics* 39:1–29.
- Ansaldi, Umberto, 2005. "Typological admixture in Sri Lanka Malay. The case of Kirinda Java." Unpublished manuscript.
- Ansaldi, Umberto, 2008. "Sri Lanka Malay revisited: Genesis and classification." In *A world of many voices: Lessons from documented endangered languages*, edited by Arienne Dwyer, D. Harrison, & David Rood. Amsterdam, Philadelphia: John Benjamins., 13–42.
- Bakker, Peter, 2006. "The Sri Lanka *Sprachbund*: The newcomers Portuguese and Malay." In *Linguistic Areas – Convergence in Historical and Typological Perspective*, edited by Yaron Matras, April McMahon, & Nigel Vincent. Hounds-mills, New York: Palgrave Macmillan, 135–159.
- Bakker, Peter, Daval-Markussen, Aymeric, Parkvall, Mikael, & Plag, Ingo, 2011. "Creoles are typologically distinct from noncreoles." *Journal of Pidgin and Creole Languages* 26:5–42.
- Baldaeus, Phillipus, 1958–1959. *A True and exact description of the great Island of Ceylon (...) 1672. Translated into English by Pieter Brohier*. Special issue of *The Ceylon Historical Journal* 8 (1-4). (originally 1672: Beschrijving der Oost-Indische Kusten Malabar en Coromandel der selver aengrenzende Koninkrijcken en Vortendomme als oock het Keijserrijck Ceylon. Amsterdam: Janssonius van Waasberge en van Someren).
- Becker, Angelika & Veenstra, Tonjes, 2003. "Creole prototypes as basic varieties and inflectional morphology." In *Information structure, linguistic structure and the dynamics of language acquisition*, edited by Christine Dimroth & Marianne Starren. Amsterdam: John Benjamins, 235–266.
- Bichsel-Stettler, Anne, 1989. *Aspects of the Sri Lanka Malay Community and their Language*. Master's thesis, University of Bern.
- Bruyas, Émile, 1898. *Deux Mois à Ceylan, Colombo, Kandy, Nurrelyya, Badulla, Ratnapoura, le musée de Colombo, l'île Ramescheram, Anuradhpoura, chronique et statistique*. Lyon: A. Rey.
- Coperahewa, Sandagomi, 2009. "The language planning situation in Sri Lanka." *Current Issues in Language Planning* 10:69–150.
- Cotteau, Edmond, 1889. *Promenade dans l'Inde et à Ceylan*. Paris: E Plon.
- Cysouw, Michael, 2009. "APiCS, WALS, and the creole typological profile (if any)." Paper presented at the APiCS workshop 2009.

- URL [http://web.mac.com/cysouw/presentations/invited\\_files/cysouwAPICS\\_slides.pdf](http://web.mac.com/cysouw/presentations/invited_files/cysouwAPICS_slides.pdf)
- De Silva Jayasuriya, Shihan, 2002. “A Unique Malay: Sri Lankan Malay Creole.” *NUSA* 50:43–57.
- Emerson Tennent, James, 1859. *Ceylon, an Account of the Island: Physical, Historical and Topographical*. London: Longman, Green, Longman, and Roberts.
- Haspelmath, Martin, Dryer, Matthew S., Gil, David, & Comrie, Bernard (eds.), 2005. *The World Atlas of Language Structures*. Oxford: Oxford University Press. Updated online version: <http://wals.info>.
- Holm, John & Patrick, Peter L (eds.), 2007. *Comparative Creole Syntax Parallel Outlines of 18 Creole Grammars*. London: Battlebridge.
- Hussainmiya, B. A., 1990. *Orang Rejimen: The Malays of Sri Lanka*. Bangi: Universiti Kebangsaan Malaysia, 1st edition.
- Lee-Smith, Mei W. & Wurm, Stephen A, 1996. “The Wutun language.” In *Atlas of languages of intercultural communication in the Pacific, Asia, and the Americas. Vol 2.1*, edited by Stephen A Wurm. Walter de Gruyter, 883–898.
- Li, Charles N., 1983. “Languages in contact in western China.” *Journal of East Asian Languages* 1:31–51.
- Li, Charles N, 1984. “From verb-medial analytic language to verb-final synthetic language: a case of typological change.” In *Proceedings of the Tenth Meeting of the Berkeley Linguistics Society*. 173–185.
- Madrolle, Claudio, 1926. *De Marseille à Saigon, les escales de la traversée de France en Indochine : Djibouti, Éthiopie, Ceylan, Malaisie*. Paris: Hachette.
- Matras, Yaron, 2008. *Language Contact*. Cambridge: Cambridge University Press.
- Meuron, Guy de, 1982. *Le Régiment Meuron 1781-1816*. Lausanne: Editions d’En bas.
- Nadkarni, Mangesh V., 1975. “Bilingualism and syntactic change in Konkani.” *Language* 51:672–83.
- Nordhoff, Sebastian, 2008. “Directions of language change in a trilingual environment.” Ms. University of Amsterdam.
- Nordhoff, Sebastian, 2009. *A Grammar of Upcountry Sri Lanka Malay*. Ph.D. thesis, University of Amsterdam, Utrecht.
- URL <http://www.lotpublications.nl/publish/articles/003744/bookpart.pdf>
- Nordhoff, Sebastian, this volume. “Introduction.”
- Paauw, Scott, this volume. “The Lexical Sources of Sri Lanka Malay Revisited.”
- Papiha, S. S., Mastana, S. S., & Jayasekara, R, 1996. “Genetic variation in Sri Lanka.” *Human Biology* 68:707–737.
- Parkvall, Mikael, 2008. “The simplicity of creoles in a cross-linguistic perspective.” In *Language Complexity. Typology, Contact, Change*, edited by Matti Miestamo, Kaius Sinnemäki, & Fred Karlsson. Amsterdam: John Benjamins, 265–285.
- Percival, Robert, 1805. *An account of the island of Ceylon, containing its history, geography, natural history, with the manners and customs of its various inhabitants*. London: C & R Baldwin, 2nd edition.
- Powell, Geoffrey, 1973. *The Kandyan Wars - The British Army in Ceylon 1803-1818*. London: Leo Cooper.
- Robuchon, Gérard, 2003. “Malayo Language in Sri Lanka.” Paper presented at the 7th

- International Symposium on the Malay/Indonesian Linguistics in Berg en Dal.
- Saltikov (Soltykoff), Alexis, 1840. *Lettres sur l'Inde*. Paris: Amyot.
- Schiffman, Harold F., 1999. *A Reference Grammar of Spoken Tamil*. Cambridge: CUP.
- Schweitzer, Christoph, 1931[1680]. *Reise nach Java und Ceylon, 1675–1682*. Den Haag: Nijhoff. Originally published by Nijhoff in Haag in 1680.
- Seuren, Pieter A. M., 1998. *Western Linguistics – An historical introduction*. Oxford: Basil Blackwell.
- Slomanson, P., 2011. “Dravidian features in the Sri Lankan Malay verb.” In *Creoles, their Substrates, and Language Typology*, edited by C. Lefebvre. Amsterdam: John Benjamins.
- Slomanson, Peter, this volume. “Known, inferable, and discoverable in Sri Lankan Malay research.”
- Smith, Ian R., 2003. “The provenance and timing of substrate influence in Sri Lankan Malay; Definiteness, animacy and accusative case marking.” Paper presented at the South Asian Language Analysis Roundtable XXIII, Austin, Texas.
- Smith, Ian R. & Paauw, S., 2006. “Sri Lanka Malay: Creole or convert.” In *Structure and Variation in Language Contact*, edited by Ana Deumert & Stephanie Durrellman. Amsterdam: John Benjamins, 159–182.
- Smith, Ian R., Paauw, S., & Hussainmiya, B. A., 2004. “Sri Lanka Malay: the state of the art.” *Yearbook of South Asian Languages* 2004:197–215.
- Soltykoff, le Prince Alexis, 1853. *Voyage dans l'Inde et en Perse*. Paris: V. Lecou.
- Stoudt, Howard W, 1961. *The Physical Anthropology of Ceylon*. Colombo: The Colombo Museum.
- Szemerécsanyi, Benedikt & Kortmann, Bernd, 2009. “The morphosyntax of varieties of English worldwide: a quantitative perspective.” *Lingua* 119:1643–1663.
- Tadmor, Uri, 1992. “The Malay Dialects of Central Thailand: a Preliminary Survey.” In *The Third International Symposium on Language and Linguistics*. Bangkok: Chulalongkorn University, 1367–1378.
- Tadmor, Uri, 1995. *Language contact and systemic restructuring: the Malay dialect of Central Thailand*. Ph.D. thesis, University of Hawaii.
- Tadmor, Uri, 2004. “Dialect endangerment: The case of Nonthaburi Malay.” *Bijdragen tot de Taal-, Land- en Volkenkunde* 160(4):511–531.
- Thomason, Sara Grey, 2000. “Linguistic areas and language history.” In *Languages in Contact*, edited by D. Gilbers, J. Nerbonne, & J. Schaecken. Amsterdam: Rodopi, 311–327.
- Valentia, George, 1813. *Voyages dans l'Hindoustan, à Ceylan, sur les deux côtes de la Mer Rouge, en Abyssinie et en Egypte, durant les années 1802, 1803, 1804, 1805 et 1806. Traduits de l'anglais par P.-F. Henry*. Paris: Veuve Lepetit.
- Veth, P. J., 1889. *Uit oost en west, Verklaring van eenige uitheemsche woorden*. Arnhem: P. Gouda Quint.
- Wickremesekera, Channa, 2004. “Military organisation in pre-modern Sri Lanka: the army of the Kandyan kings.” *South Asia: Journal of South Asian Studies* 27(2):133–151.
- Winterbottom, Anna, 2009. “Producing and using the Historical Relation of Ceylon: Robert Knox, the East India Company and the Royal Society.” *British Journal for the History of Science* 42:515–538.
- Wurm, Stephen A, 1996. “North China: Intercultural communications involving lan-

guages other than Chinese.” In *Atlas of languages of intercultural communication in the Pacific, Asia, and the Americas, Volume 2, Part 1*, edited by Stephen A Wurm. Berlin: Walter de Gruyter, 815–826.

## Chapter 4

# Known, inferable, and discoverable in Sri Lankan Malay research

Peter Slomanson

*Aarhus University*

### 4.1 Introduction

A considerable amount of investigation has been carried out over the course of the preceding decade on the history of Sri Lankan Malay, its speakers, the grammar and lexicon of their language, and on its current prospects for survival (Ansaldi (2008, 2011a), Lim & Ansaldi (2006, 2007), Ansaldi & Nordhoff (2009), Nordhoff (2009), Slomanson (2006, 2008, 2009, 2011), Smith et al. (2004), Smith & Paauw (2006) *inter alia*).<sup>1</sup> In my own work, I have focused on morphosyntactic description and explanation, with special emphasis on comparison and contrast with the other languages of Sri Lanka. I have also worked on plausible explanations of various aspects of the language's development, based on fieldwork data collected in different communities over the course of several years. In order to explain how Sri Lankan Malay assumed its modern form, both grammatical and sociohistorical investigation are indispensable. We continue to make progress on a number of important questions, not least of which on the matter of glottogenesis and subsequent development. This is important because the grammar of the language differs so profoundly from the range of Malay dialects spoken in the ancestral areas of the Indonesian archipelago, and because the answers

---

<sup>1</sup>I am indebted to many people for the content of this paper, but in particular to B.A. Hussainmiya, who hosted me on a research visit to Brunei in the summer of 2008, providing me with extensive historical documentation and immeasurably increasing my understanding of Sri Lankan Malay history. I am also extremely grateful to the honourable Al Haj Muzni Ameer for permitting me to consult the marriage records discussed in this paper. I am naturally indebted to all the native speakers from whom I have learned over the years, but my understanding of the grammar of Sri Lankan Malay owes most to Mohamed Jaffar of Colombo and Toronto, and to Mohamed Thawfeek Mohamed Rihan of Kirinda. Mohamed and Rihan are not just excellent linguistic and cultural informants, but true friends. For discussion of the material in this paper, I would also like to thank Sebastian Nordhoff, Scott Paauw, and Romola Rassool, who are indispensable.

have the potential to increase the understanding in linguistic science of the nature of grammatical change under contact conditions.<sup>2</sup>

In section two, I will focus primarily on what is known and knowable, presenting points that bear on a lingering controversy over the external context that gave rise to the language, emphasizing this matter because misconceptions about its external history can give rise to misguided approaches to its internal development. In section three, dealing with what is inferable, I will discuss the need to construct plausible diachronic scenarios in the absence of actual attestations. In section four, I will discuss what is discoverable, including new synchronic information, lexical information, as well as a broad range of sociolinguistic phenomena, including information on patterns of variation and change, as well as macrosociolinguistic information about when and how the language is used, and by whom.

## 4.2 What is known?

### 4.2.1 External facts

I will briefly and selectively discuss the history of Malay settlement in Sri Lanka. I will then discuss contact with Sri Lankan Muslim Tamil speakers (also known as Moors), a matter which has become a highly resolvable bone of contention. I will frequently use the term Shonam, the native glottonym for Sri Lankan Muslim Tamil, a variety which is historically distinct from other varieties of Tamil spoken in Sri Lanka. The controversy in question concerns the relative extent to which Shonam and colloquial Sinhala, respectively, have influenced the development of Sri Lankan Malay.

#### Settlement history

In some of my historical comments, I will use the term “Indonesian”, meant to be understood here as “coming from the region that includes the territories of what are now Indonesia, Malaysia, Singapore, and Brunei”, in order to avoid the ethnic connotations of the term “Malay”. The Dutch colonial period in Sri Lankan history began in the mid-seventeenth century. We know that prior to that period there had been contacts between the Indonesian world and Sri Lanka, but that the actual modern Sri Lankan Malay community came into existence at the beginning of the Dutch period in Sri Lanka’s history. We know that there were different social categories of migrants, and that their status under the Dutch administration was not the same. In spite of the fact that the ethnic Javanese among the migrants were unwilling political exiles, they represented a privileged segment of the community. Far more of the migrants were or would become soldiers, however, and throughout the period preceding the twentieth century, the Malay presence in Sri Lanka had an active military character. The decline in the economic stability and cultural autonomy of the Malay community in the late British period are attributed to the loss of this military role (Hussainmiya 1987, 1990).

The military character of Malay history in Sri Lanka is part of what helps us to know more about the community’s history than we would otherwise be able to know.

---

<sup>2</sup>Unlike my other publications on Sri Lankan Malay, this is not a paper whose goal is grammatical description, analysis, and explanation, but rather it is a paper on the field of Sri Lankan Malay studies itself, emphasizing linguistics and other research that contributes to our understanding of linguistic matters. The content of the paper is intended to be viewed as corrective, with respect to misconceptions, and programmatic with respect to future work in the field. Since it is not a grammar paper, I have included only one, hopefully illuminating, data example.

If there is anything that colonial powers leave records of, it is their military goals and policies, if not the negative consequences of those goals and policies. The Netherlands East Indies Company (VOC), in addition to using coastal Sri Lanka and the South African Cape of Good Hope as places of exile for politically subversive individuals in Indonesia, also strategically used Indonesia as a source of military and other labor, including slave labor. This was based on a policy of maintaining relatively cordial relationships with indigenous elites and local populations in the colonial territories themselves. The Dutch refrained from enslaving and indenturing the Sinhalese in Sri Lanka, the Khoikhoi in South Africa, and the Javanese and Sundanese on Java in Indonesia ([Ward 2009](#)). The local indigenous population were not to see local people being subjected to undesirable treatment at home. This policy resulted from political calculation, and is the colonial policy that is most responsible for the Malay presence in Sri Lanka.

### History with respect to the so-called Tamil bias

Now we come to what is known, but which has recently been rendered controversial. This is the part of the historical picture that has to do with the relationship of the Malay community to the Moorish community, the much larger community of Sri Lankan Muslims, whose historic ethnic language, Shonam, is a contact variety of Tamil, which was written until the early 20th century in modified Arabic orthography, known as Arwi (also a term for Shonam, when written using this script, as is Arabu-Thamul or Arabic-Tamil, in which “Arabic” refers primarily to the orthography).

In the linguistic literature on Sri Lankan Malay, the authors Ansaldo and Nordhoff, respectively, have explicitly objected to an ostensible Tamil bias in Sri Lankan Malay linguistic studies. The phrase “the Tamil bias” is associated with [Ansaldo \(2008\)](#), but Nordhoff, in his 2009 University of Amsterdam Ph.D. dissertation investigated the claim in detail, examining the external arguments for a closer connection to Tamil (Shonam) than to Sinhala. Ansaldo places great emphasis on the significance of Malay-Moor intermarriage in Smith, Paauw & Hussainmiya (2004), going so far as to claim that the existence of this type of intermarriage is little attested and that the number of such marriages that there have been is negligible. This claim, essentially the criticism that the existing literature lacks evidence for one of its points of departure, relies on support from Ansaldo’s own investigative work. I will review this approach and the findings that it yielded. I will subsequently address what I find to be missing from the discussion, as well as further points made in [Ansaldo \(2008, 2011a\)](#) with which I disagree, based on evidence that I have found in the course of my research, and in that of the historian B.A. Hussainmiya.

[Nordhoff \(2009\)](#) also takes issue with the idea that Malay-Moor intermarriage provided the external context for major grammatical change in Sri Lankan Malay. However in contrast with Ansaldo, he does not contest the fact that Malay-Moor intermarriage is a frequent occurrence and has been in the past. While Nordhoff has done part of the work of challenging Ansaldo’s claim about intermarriage, by discussing such sources as native genealogical investigation that demonstrate the frequency of Malay-Moor intermarriage, I will continue to pursue the matter with additional evidence. I am doing this because the intermarriage issue is a major part of the claim that there is a “Tamil bias” that is unmotivated by available evidence, and because this claim continues to be repeated in print ([Ansaldo 2011a](#)). Nordhoff concurs with Ansaldo in rejecting the view that Malay-Moor intermarriage could have created the conditions for glottogenesis, and concludes that there is little room, based on external evidence, for

treating Shonam as of greater importance than colloquial Sinhala in the divergence of Sri Lankan Malay from vehicular Malay.<sup>3</sup> I will present new external arguments for the case that a Malay-Moor symbiosis is the context in which linguistic change took place prior to Sri Lankan independence in the mid-twentieth century. With respect to Ansaldo's position, the best counter-evidence is nineteenth and twentieth century Muslim marriage records. With respect to Nordhoff's position, the existence of bilingual religious literature, as well as Shonam literature written by ethnic Malays, suggests a close relationship between ordinary Malays and Moors, but also between Malays and the Shonam language itself. This should demonstrate that, rather than an a priori bias, the case for the Shonam role in SLM glottogenesis is based on evidence.

I will divide this discussion into three categories: (a) intermarriage, (b) residential patterns, and (c) religious and cultural life.

**Intercultural marriage and the “Tamil bias”** Living in close proximity and sharing a demanding devotional culture was conducive to intermarriage between the Moors and the Malays, regardless of frequently divergent occupational paths. In the earliest period, a poor ratio of male to female migrants must have led to considerable intermarriage, although we do not have statistics for this. Mixed marriages do not necessarily lead to a non-Malay identity or a loss of language loyalty. In the modern period, we find many examples of a pattern in which individuals raised in Malay/Moorish homes have been raised bilingually. The one parent, one language approach that happens to be favored by modern linguists seems to have been adhered to as a matter of course. The position adopted by Smith, Paauw and Hussainmiya is that Moorish mothers spoke L2 Malay which assumed target variety status for children.

The comments in this paragraph might well be categorized as “inferable”, however I will include them here as a follow-up to the topic addressed in the preceding paragraph. I take Smith, Paauw, and Hussainmiya’s position to be plausible. However it is more likely, if bilingual homes in an earlier period were actively bilingual, that there was convergence in these homes, so that after the earliest stage, the Malay that was spoken natively began to be more and more strongly influenced by L2 varieties. Children will have heard a considerable amount of L2-influenced Malay. Given the status of Shonam as a vehicular language for the dissemination of Islamic knowledge, the L2-influenced Malay of native Shonam speakers is unlikely to have been stigmatized, particularly in the absence of a normative Malay standard. What we now think of as standard Malay is based on a literary variety that had little general currency throughout the Indonesian archipelago, whereas vehicular Malay varieties have been widely spoken as auxiliary languages for centuries. Migrants to Sri Lanka are consequently more likely to have conformed to (new) local vernacular norms, beginning with pragmatic norms, in the absence of a normative Malay variety.

As early as the Dutch period, 1658-1798, members of socially prominent Malay families married into elite Moorish families (B.A. Hussainmiya, p.c.). As a rule, however, the Malays preferred endogamous marriages within their own community, and this is not particularly surprising. Marriages to Moors were far from taboo, as we can see from evidence in the *thombos*<sup>4</sup> maintained by the Dutch colonial authorities, *pace*

---

<sup>3</sup>This term, which was coined by Ian Smith, was introduced into the literature on Sri Lankan Malay by Smith and Paauw, to refer collectively to the vernacular intercommunal Malay varieties of Indonesia. These were the varieties that were introduced to Sri Lanka in the seventeenth century. Contrasts between varieties spoken are likely to have been leveled through koineization, although we have no direct evidence of this process.

<sup>4</sup>These are collection of household records, listing all household members. The practice predates the

claims made in [Ansaldo \(2008\)](#) and [Ansaldo \(2011a\)](#), and still farther from rare, as can be seen from voluminous evidence from the British period. It must be borne in mind that as compared to the number of such intermarriages between these two minority communities, marriages between the Malays and the majority Sinhalese were rare events. It is the available documentation from marriage registries maintained by both communities that reveals the frequency of Malay-Moor intermarriage.<sup>5</sup>

What follows is a series of citations from [Ansaldo \(2008\)](#) with respect to the so-called Tamil bias.

Bakker's claim of 'heavy Tamil pressure' may perhaps have been present in the S(ri) L(ankan) P(ortuguese) community, but there is absolutely no historical evidence that this occurred in the Sri Lankan Malay communities.

The most specific claim regarding the creolization of Sri Lankan Malay is that it developed as a result of intermarriage between Malay men and (Tamil) Moor women (e.g. [Smith et al. 2004](#)); this view is based primarily on the historical observations of [Hussainmiya \(1987, 1990\)](#) regarding the records of marriage under the Dutch (thombos) which, according to him, show several cases of intermarriage between SLM and Tamil Moors.

Unless the Tombo [sic] in Hussaimiya's [sic] possession reveal completely different data from the ones of the National Archives, it is safest to discount Hussaimiya's observations.

Hussainmiya has not claimed to own Dutch period records and manuscripts that are not in the possession of the Sri Lankan National Archives. His comments are based on what is found in the archives. It may be the case that Ansaldo saw different records from those that Hussainmiya, a professional historian, examined.

While sharing a common religion may have played a role in individual marriages between SLM and Tamil Moors, there is no historical evidence to lend support to a claim of diffuse intermarriage between these two communities, especially of such a magnitude that could conceivably lead to restructuring of the vernacular.

If by "diffuse", the author means *extensive* intermarriage, there is in fact ample evidence of extensive intermarriage. While the number of records for the Dutch period is not great, those we have found provide evidence for Malay-Moor intermarriages, and no evidence of Malay-Sinhala intermarriages, although it was noted in an account by VOC employee [Schweitzer \(1931\[1680\]\)](#) that these did take place ([Nordhoff 2009:11](#)).

---

Dutch colonial administration. The records of greatest interest, all of which are written in Dutch, are numbered 1/3990 and 1/3991.

<sup>5</sup>Naturally, census materials and marriage records include no information on physical characteristics. It is worth mentioning though that, depending on the community, a significant proportion of Malay people appear far more classically Sri Lankan than Indonesian. This is particularly true in Kirinda, a community that is somewhat isolated from other Malay communities, and was founded at the beginning of the nineteenth century. This suggests that extensive intermarriage is not a recent development, although it may have occurred less in some communities than in others.

At the beginning of the 19th century, Percival, a British military officer, remarked that

"Although they (Malays) intermarry with the Moors and other castes (sic) particularly in Ceylon and by this means acquire a much darker colour than is natural to a Malay; still their characteristic features are strikingly predominant." ([Percival 1803:115](#)).

Nordhoff makes a very cogent point, that “the Dutch had just pushed the Moors away from the colonial cities” (Nordhoff 2009:37). The extent of Dutch antipathy to the coastal Moors was great, because they represented highly skilled commercial competition, consequently the colonial authorities would eventually attempt to ban Moors from living in Colombo altogether. Portuguese antipathy to the coastal Moors had also been great, primarily for theological reasons, and nevertheless Colombo was largely a Muslim city at the end of the Portuguese period. The Dutch did not begin attempts to exclude the Moors from Colombo as early as they did from Galle however, and given the extent of Moorish presence in Colombo, it would have been difficult to exclude them entirely. I agree with Nordhoff that this detail weakens the case for extensive contact with Moors in at least part of the Dutch period, and that this contributes to a kind of historical quandary. It is quite clear that the history of Malay-Moor interaction is a complex one about which more needs to be learned. Ansaldo (2008, see quotation on page 8 of this paper), in his investigation of stated attitudes to intermarriage in certain circles, has come across what one can call an antipathy to a Moorish connection in certain urban Malay circles, although there is no evidence that that antipathy has any historical depth.<sup>6</sup> <sup>7</sup>

The Malays were a very small Muslim population living in a foreign environment, maintaining an ethnic language in the absence of normative authority. With respect to external criteria alone, given simultaneous language maintenance, apparent domestic bilingualism in the periods for which we have evidence, and most significantly, the status of Shonam as a language of Islamic process and practice, it would be surprising if the grammatical organization of Shonam had not been replicated in Sri Lankan Malay, although the *extent* of this replication continues to surprise observers. It is ultimately the interaction of external factors, rather than a single factor, that has lent itself to the outcome. The competing diachronic narrative proposed by Ansaldo is based on what he views as feature competition, of a type in which all definable surface properties of grammars are regarded as features. My objection to this rests on the implausibility of the external context (ecology, in Ansaldo’s terms) and on the fact that the success of one feature or another has no particular trigger, either linguistic or sociolinguistic. This detracts from the potential explanatory value of Ansaldo’s approach. Feature competition seems to be a restatement of the incontestable observation that Sri Lankan Malay owes some of its properties to Shonam and some to Sinhala. (I take what little unambiguous Sinhala influence that there is to be adstratal, i.e late, however I will put this matter aside for the moment.) What motivates success? If this is not stated, then this approach has no explanatory value.

As Hussainmiya (1987) is the only work to directly, though briefly, address this issue, it is important to verify the claim in a precise manner. My investigation of the Dutch *tombos* referred to in that work, for the period 1678-1919, in the National Archives at The Hague (microfilm copies of the Colombo archives) and the National Archives of Sri Lanka in Colombo yielded the following results:

- 1.The records for the period up to 1796 are damaged by water, making

---

<sup>6</sup>See further discussion of this antipathy later in this paper.

<sup>7</sup>In my own work, I have naturally dealt far more extensively with the linguistic evidence for a Shonam model than with the historical details. I think this follows from the linguist’s desire for straightforward historical answers, in order to get on with the business of linguistic research. Be that as it may, history in the real world is not always as cooperative as we might like, and this forces us, as contact linguists interested in diachrony, to get involved in complex questions that we would normally leave to historians.

parts of the entries illegible. The most revealing information for identification here are the signatures of the parties. There is, however, hardly any information of ethnic group, which makes it difficult to identify Malay/Indonesian and Moors given that both groups share the practice of adopting Arabic names. In a particularly interesting section in the *tombos* dedicated to mixed marriages (cf. [Hussainmiya 1987](#)), only five of 238 entries clearly refer to individuals of Javanese origin: of these, two records refer to Javanese-Moor marriage, one to a Javanese-Javanese marriage, and the remaining two are unclear.

2. The following period until 1919, albeit under British rule and therefore less interesting for our claim, shows a more structured archiving system where indication of race is given. Where legible, this reveals still a majority of Western marriages, a growing number of marriages between Eurasians and Burghers (locally born of Dutch/Western heritage), and between Burghers. There are two clear entries involving Malays, one married to a Eurasian (between 1867-1897), and one to a Burgher (1885-1897). From 1897 onwards, race is clearly specified; of 196 entries, only one is Malay.<sup>8</sup>

Of course Ansaldo is correct that historical claims require evidence, although his evidence ought to be retrievable, and this includes negative evidence (i.e. that documentation is *not* available). In fact, the documentation in question is available.

The implicit assumption that religious identity could have been a peripheral consideration in Malay marriage arrangements is difficult to reconcile with knowledge of the way other Muslim communities in the region are organized. A strong tendency for Malays to marry Moors can be demonstrated for at least the nineteenth century onwards, due to the preservation of *kaduthams*,<sup>9</sup> which provide clear evidence that Sri Lankan Malays who remained within the Malay community (i.e. who remained Muslims) have married Moors in great numbers, and that they have not married Sinhala converts to Islam in similar numbers. Since the imperative that Muslims marry Muslims has been well-documented in the ancestral Indonesian communities as well as in Sri Lankan Muslim communities, it would seem that the burden of proof rests with Ansaldo to demonstrate that communal religious culture and identity has *not* been a primary consideration in Malay marriage arrangements, and has not been sufficient to privilege Malay-Moor marriages over Malay-Sinhala marriages. (The fact that Malays have married non-Malays is clear from the documentary record, so there is no question of complete ethnolinguistic endogamy.)

The evidence presented in statement 2 (see above) in [Ansaldo \(2008\)](#) follows from a misconception. There were in fact no *thombos* in “the period until 1919”, subsequent to 1796, the last year of the Dutch colonial administration. Under British colonial administration, Muslim marriage records were kept in mosques. Let us suppose that the content of statement 2 were accurate. It is not clear to me what could be meant by “this reveals still a majority of Western marriages”. A potential explanation for the

---

<sup>8</sup>The preceding sections 1 and 2 from [Ansaldo \(2008\)](#) are cited uncritically in ([Nordhoff 2009:42-43](#)).

<sup>9</sup>The term *kadutham* is a Tamil word literally meaning “letter”. It has specialized meaning in Sri Lanka however, where it refers to a Muslim marriage certificate that is issued at the time of marriage. It was originally kept by the issuer, known as a *khatib*. The *khatib* is affiliated with a particular mosque, and responsible for its catchment area. These records were kept well into the twentieth century, and the text continued to be written in an orthography derived from that used for writing Arabic. Eventually, *khatibs* began to keep marriage records in English that had (and continue to have) the character of logs, with a row for each new marriage.

fact that Ansaldo only found “two clear entries involving Malays, one to a “Eurasian”, likely a Burgher, and the other to a Burgher (Burghers are Christian), is that the records that Ansaldo is describing are church records, such as those kept in Wolvendahl church in Colombo, and the “two clear entries” there do not involve Muslim Malays. The usefulness of this information is questionable, in contrast with the data to be found in mosques, where records of marriages involving Muslims is ordinarily kept, and has been since the beginning of the British period at the end of the eighteenth century.

With respect to demography and its effects, [Ansaldo \(2008\)](#) assumes that the language of the majority was necessarily acquired or influential, and that all communicative interaction is of equal significance and intensity.

Even if we were to accept the claim that Tamil might have been more closely involved in the evolution of SLM due to the fact that the language of the religious texts and practices would have been Tamil ([Hussainmiya 1987](#)), *the sheer numerical and social predominance of Sinhala in Sri Lankan society cannot be ignored* (emphasis mine, PS), and for this reason we should consider the two languages as, at least, adstrates of similar significance in the evolution of SLM.

Historians and historical sociolinguists will find evidence for the relative density and multiplexity of Malay-Moor networks in documentation collected by Hussainmiya. As it stands, the evidence for close Malay-Sinhala interaction is largely limited to military service in the Kandyan kingdom in the Dutch period and similar episodes. This involved a minority of Malays, since the original colonial function of the many Malay soldiers brought by the Dutch to Sri Lanka was to serve as non-native *adversaries* of the Kandyan Sinhalese (see Bakker, this volume), so those who aligned themselves with the Sinhalese were those who managed to flee Dutch control. In the British period, by which time the Kandyan kingdom had come under colonial control, Malays were deliberately organized into all-Malay regiments ([Hussainmiya 1990, 2008](#)), the purpose of which, again, was essentially to insure the submission of the Sinhalese.

Returning to the matter of intermarriage, Malays, as we have seen, frequently married Moors during the British colonial period and subsequent to it. In purely quantitative terms, the available evidence for Malay-Moor intermarriage during the preceding Dutch period, found in the *thombos*, is less striking. It is quite clear however, based as it is on instances of explicit ethnic labeling (with the terms *Javaan* (“Javanese”) and *Moorman* (“Moor”), as well as on onomastic evidence. For example, some of the Moorish family names end in *-poelle*, from Tamil *-pillai*. By contrast, there are no Javanese or Malay marriages to Sinhalese in the *thombos*. This is not to say that they never occurred, but the evidence for them has not been found in those records. It is for this reason that the burden of evidence rests with anyone wanting to show that there was extensive Malay-Sinhala intermarriage or that Malay-Sinhala marriage was not much less likely than Malay-Moor intermarriage. By contrast, documentation of Malay-Moor intermarriage (*kaduthams* and Muslim marriage registers) is extremely extensive, whereas there is comparatively little evidence for any significant number of Malay-Sinhala intermarriages.<sup>10</sup> Marriage records in mosques include the original name of Sinhalese converts to Islam, which is one reason that we can see in the records

---

<sup>10</sup>Members of non-Muslim ethnic groups have converted to Islam to marry Malays, however I have found no evidence that this was ever a frequent occurrence. The *kadutham* records to which I have had access distinguish between Malay-Malay marriages and Malay-Moor marriages. The officiating *khātib* would be aware of the fact that one Muslim partner was of Sinhala origin, or at least that s/he was neither Malay nor Moor, unless the partner was attempting to conceal this fact.

when a Malay married a Sinhalese person. The original name will always be unmistakable, though such marriages are far less frequent than marriages to Moors. If marriage to a Sinhalese convert was not viewed positively, marriage to a non-convert was tantamount to apostasy. Even if evidence were found that such unions were extensive, those were not unions that would have been able to contribute to accumulating changes in the form of SLM, since the new family would have been forced to weaken or sever existing network ties, with little potential for renewing ties with a Malay community. It is necessary to bear the cultural-ideological “ecology” in mind, in addition to the demography. While Hindus do not believe in one god, Buddhists technically do not believe in a god at all. It follows that though the extent of Malay-Sinhala interaction may have varied across periods and locations, it is most unlikely that this weakening of boundaries would extend to marriage, which would be the last boundary. None of this applies to Moors, irrespective of any cultural and occupational differences between the Malays and the Moors.

Ansaldo takes for granted that since Sri Lanka is a multi-ethnic society, this means that social, commercial, and other networks are *necessarily* characterized by comparable relative levels of interaction across ethnic lines, so the presence of a greater number of Sinhala speakers means that Malays must have interacted with more Sinhala speakers or “at least” with as many Sinhala speakers as they did with Moorish Tamil speakers, and outside of religious matters, they must have done so in much the same ways. This is a post-industrial western view of social organization which bears little resemblance to the historical ecology of Sri Lankan communities.

It is ultimately inconclusive to introduce anecdotal evidence into such a discussion, however Ansaldo’s sample of fifty families in which there is almost no intermarriage between Malays and Moors is problematic for a number of reasons. In the first place, the socioeconomic class of the correspondents is a potentially relevant variable. Ansaldo has claimed that his oral history evidence was collected among the “older generations”. What is the significance of this fact? In order for a claim with respect to intermarriage to be conclusive, we would expect to see a systematic survey carried out in a range of communities, certainly including substantially more than fifty families, in order to ascertain what Malay attitudes toward Malay-Moor intermarriage might be in the present period, and to find out what proportion of families interviewed have at least one Moorish member. The paragraph referred to above reads as follows,

In addition to the historical record, clear evidence against a solid Tamil influence in the development of the SLM community comes from oral history recorded in three different SLM communities: in Kirinda, Colombo and Kandy, of approximately 50 families interviewed in total, only two revealed genealogies including Moor-Malay intermarriage. Most families report that marrying outside the SLM community was considered taboo and only allowed in extreme cases. It is only in the present generations that weddings outside the community start being allowed. Moreover, the Moors appear to have had very low status in the eye of the SLM communities and, their low status may be seen as a counteracting force to the hypothetical appeal of religious affinity. Moreover, in at least one community - Kirinda - intermarriage with Sinhalese is well attested in the history of several families.

In order for a pattern of ethnolinguistic endogamy to tell us anything significant about a general tendency or reluctance to marry members of another Muslim ethnic

group (in this case Moors), members of the first ethnolinguistic group (Malays, in this case) must have access to potential out-group partners. Given the demography of the village and of the adjacent rural district, results from Kirinda can only skew the outcome of an investigation of this kind, and not in a way that can contribute to answering the question Ansaldo asks (i.e. to what extent Malays are prepared to marry Moors) which logically *presupposes* that the Malays in question have access to potential Moorish partners. In fact virtually the only Moors present in the village are teachers in the village school who reside elsewhere, and members of the community are not aware that this was ever different. Consequently Ansaldo's Kirinda evidence cannot have the same status as genealogical evidence from elsewhere on the island. The Kirinda community was founded over two hundred years ago, in a then unpopulated rural coastal area, and as Ansaldo himself has written, the community has been rather isolated.<sup>11</sup> What opportunity would its members have to marry Moors, without either settling far from the village or bringing in spouses from far afield? Most other Malay areas, by contrast, have historically been mixed Malay-Moor areas. In the present, the surrounding area is far less sparsely populated than in previous generations, however the present-day surrounding communities are entirely Sinhala-speaking and Buddhist. If Ansaldo had asked the same question in the heavily Muslim town of Hambantota, he would have found a substantial number of Malay-Moor marriages.<sup>12</sup>

Only a large-scale study to obtain quantitative data would provide complete closure in this matter, however social history, and for that matter anthropology, do not require quantitative data, but rather intensive engagement with communities and plausible interpretation of qualitative evidence. Almost every Malay family I have met in Sri Lanka, in a range of urban and rural communities over the course of eight years of active involvement, one calendar year of field research, and many additional research visits, contains at least one Moorish member. This is minimally the basis for a hypothesis that if there is any Malay-Moor marriage taboo at all, it is an exceedingly weak one. The families of Kirinda are the exception to this otherwise ubiquitous pattern, for the reasons stated above. Aside from recent marriage patterns, the core of Ansaldo's claim that there has been minimal or no intermarriage with Moors is also not born out by the *historical* evidence, which is the variety of evidence at the core of his claim (because this is most relevant to the claim that a Tamil variety was the primary model language for the grammatical changes that took place in Malay as it came to be spoken in Sri Lanka). I have already referred to the *kadutham* evidence. Ansaldo examined two sources. The first source examined by Ansaldo consists of evidence from the *thombos* (as discussed earlier). The second source, as we have seen, is British, according to Ansaldo's description. With respect to evidence in the *thombos*, Ansaldo (2008) states

The first occurrence of a claim about Malay-Moor intermarriage occurs in Hussainmiya (1987): “a number of” these marriages are reported, next to Malay-Ambonese/Malabarese/Sinhalese unions recorded in the Dutch Tombos. The same work however suggests that SLM may be influenced by Sinhala, Tamil or both.

---

<sup>11</sup>The area is of high symbolic importance in Sinhala Buddhist history, and there is a Buddhist temple on the water. Sinhala village life in the immediately surrounding area is historically rather recent however.

<sup>12</sup>One of my consultants in the village, a young woman who is a schoolteacher in the village, is engaged to marry a Moor. She went to school in Hambantota, and in that sense, her personal network extends beyond the village. This seems to be completely unproblematic. The village is quite pious, and increasingly traditional, due to the influence of itinerant Moorish preachers whose Friday message in Shonam is spread by loudspeaker from the village mosque. It is difficult to imagine how marriage to a Buddhist would be received, and asking how it would be received would be quite awkward.

I have seen no evidence of Malay-Ambonese/Malabarese/Sinhalese unions in the Dutch *thombos* that I have examined myself in Colombo. This claim by Ansaldo may have been precipitated by the following well-known historical comment by the Dutch East Indies Company employee [Schweitzer \(1931\[1680\]\)](#), as discussed in [Nordhoff \(2009\)](#).

The wives [of the Malays], which in part are Amboinese, in part Sin-gulayans [Sinhalese], and Malabarians [South Indians] may say nothing against [the stripping of their ornaments].

With respect to another claim from [Ansaldo \(2008\)](#), cited above (1), there is no section of the *thombos* "dedicated to mixed marriages". The Dutch did often add the designations *Javaan* ("Javanese") and *Moorman* ("Moor") when listing the members of households in particular areas in Colombo. These entries need to be read carefully because they listed unrelated household members such as servants as well. This is part of the argument for bilingual Malay-Shonam homes, although it was not made in Smith et al. The concept of a modern nuclear family is anachronistic in this discussion. In my own investigation, I found no water damage, although the florid Dutch script is easier to read in some records than in others. For records in which there is no explicit ethnic designation, it is still fairly easy to discern ethnicity. During the relevant period (1) there were more Javanese than there would be later, (2) the Javanese were likely to appear in the *thombos*, because of their status relative to mere soldiers, (3) Javanese family names were not Arabic names and are in fact quite distinctive, and (4) Moor and Malay names are distinguishable in a number of ways. For example, *Lebbe* ("priest") appears again and again, as does *Marikar*.<sup>13</sup> These are associated with Moors only. Most usefully, in marital records from the Dutch period to the present, Malay titles appear again and again, particularly *Thuan*. *Thuan* was never born by Moors, and the female title *Gnei* was never born by Moors either. The words themselves are Malay words (*tuan* and *nyi*, respectively in vehicular Malay). While it is true that there are Arabic names that can obscure ethnic origins, these are relatively infrequent, the list of family names exclusively born by Malays is quite long, and there are plenty of onomastic markers of Moorish identity as well.<sup>14 15</sup>

Ansaldo comments further that

While necessarily brief, the report above of the contents of the Tombos shows that there is hardly any reason to comment on the nature of inter-marriage of the Malays in general, and less even to make specific claims about the origins of the parties. While sharing a common religion may have played a role in individual marriages between SLM and Tamil Moors, there is no historical evidence to lend support to a claim of diffuse inter-marriage between these two communities, especially of such a magnitude that could conceivably lead to restructuring of the vernacular.

---

<sup>13</sup>Nordhoff (2009:43) states "I am not aware of any names exclusively borne by Moors."

<sup>14</sup>The word is *nyi* in Indonesian spelling, and was previously a title for any woman, although the word has since taken on a number of narrower meanings.

<sup>15</sup>Ansaldo may have been influenced by his field experiences in Kirinda, where the names born by Malays are Arabic ones. However, against the general Sri Lankan Malay pattern, the names are concatenated in a specifically Moorish way. For example, Mohamed Thawfeek Mohamed Rihan, a consultant and close friend, bears the given name Rihan. The preceding names are the name of Rihan's father. Rihan's brothers and sisters were all named in accordance with this pattern. This is the Moorish pattern and most certainly does not attest to any aversion to Moorish cultural traits.

Recall the following comments by Ansaldo, repeated here for convenience.

My investigation of the Dutch Tombos referred to in that work ([Hus-sainmiya 1987](#)), for the period 1678-1919, in the National Archives at The Hague (microfilm copies of the Colombo archives) and the National Archives of Sri Lanka in Colombo yielded the following results:

There were no *thombos* kept in 1919, since that was over a century after the British took control of Sri Lanka from the Dutch, and as is well known, the *thombos* ended with the Dutch. The relevant microfilmed *thombo* documents are not accessible at the National Archives at the Hague. The function of the microfilming project is to store back-up copies in the Netherlands, and there are no plans to make these accessible to researchers.

[Ansaldo \(2008\)](#) states with respect to ostensible documentation from the British colonial period,

The following period until 1919, albeit under British rule and therefore less interesting for our claim, shows a more structured archiving system where indication of race is given. Where legible, this reveals still a majority of Western marriages, a growing number of marriages between Eurasians and Burghers (locally born of Dutch/Western heritage), and between Burghers. There are two clear entries involving Malays, one married to a Eurasian (between 1867-1897), and one to a Burgher (1885-1897). From 1897 onwards, race is clearly specified; of 196 entries only one is Malay.

In the first place, had there been British period *thombos*, it is not clear why the British period should be less interesting.<sup>16</sup> In the second place, it is surprising that Ansaldo implies that only people of whole or partial European origin were married in the relevant period, or that only records of the marriages of such people were kept. In fact, the keeping of marriage records during the British period was communalized. This means that in order to find out who was marrying whom in Muslim Sri Lanka, it is necessary to consult the *khatib* for records. These were typically kept in Malay or Shonam in modified Arabic script, as I have described. The often extreme extent of Malay-Moor intermarriage in the nineteenth and twentieth centuries is of great relevance in this discussion, because of the high incidence of bilingualism and language maintenance in Malay-Moor homes, yielding ethnically-mixed offspring, bilingual in Moorish Tamil and Sri Lankan Malay. With respect to both to the *kaduthams* and the registers, there is explicit evidence of ethnicity. In the registers (as opposed to the older *kaduthams*) that are kept at the Wekanda Jumma mosque in Slave Island, the practice has been to maintain separate records for Malay-Malay marriages and for marriages involving non-Malays. The fact that Malay-Malay marriages are a category of its own reflects the fact that the mosque is a Malay mosque in a Malay area in Colombo. The records of mixed marriages are most interesting, since they contain many marriages that are indisputably marriages to Malays, based on the onomastic evidence, and this is true regardless of the year. A representative Moorish register page from 1990 (Figure 3) contains eighteen marriages.<sup>17 18</sup> Of those eighteen marriages, eight are to Malays.

---

<sup>16</sup>[Ansaldo \(2011b\)](#) refers to [Slomanson \(2011\)](#) as “criticizing archival data in support of a Tamil bias”. In fact, I stated that British period *thombo* data on marriages do not exist, and therefore are not found in archives at all.

<sup>17</sup>Although in several cases, the ethnic identities are unmistakable, as in the entries in which the bride’s name includes the SLM title *gnei*, the onomastic evidence was reviewed by Mohamed Jaffar, who grew up bilingual in SLM and Shonam, in an ethnically mixed neighborhood in Colombo.

<sup>18</sup>This does not represent all the Moorish marriages for that year.

This is only a representative sample, as there are several pages of records for that year. There are two marriages of Moors to converted Sinhalese.<sup>19</sup> I am assembling statistics for a range of years representing several periods.

The function of the above reference to a subpart of one year is that it demonstrates that there is no lack of evidence for Malay-Moor marriages or for a cultural taboo that would have prevented them from occurring. Had there been such a taboo, that would have weakened the external (socio-cultural) basis for investigating a variety of linguistic change. That is the variety specifically motivated by Malay-Shonam bilingualism and replication of a range of structural characteristics found in the Shonam language. What the evidence compiled thus far suggests is that there was (and continues to be) a *preference* for marriages to Malays within the Malay community, but this preference has not precluded marriages to other Muslims. The intra-ethnic preference is most likely to be evident in the records maintained in a Malay mosque in a Malay neighborhood. However even at this Malay mosque, we find evidence for a large number of marriages of Malays with Moors. With respect to the ostensible taboo, it is likely that class needs to be considered as a variable. Questionnaires demonstrating the presence of a taboo were distributed at the Colombo Malay Club. This renders the results questionable because the club's members are socioeconomically removed from the general Malay community. It is the professional middle class of Colombo Malays that increasingly sought to distinguish itself from the Moors in the previous century, partly due to the goal of retaining an independent voice in political life (Hussainmiya 1987:16-18). The professional class constitute an elite with collective interests that are to some extent removed from those of the general Sri Lankan Malay community, and its views therefore not be treated *a priori* as representative of views held by the Sri Lankan Malay community as a whole. In addition, antipathy to the Moorish community in this group is the product of twentieth century communal politics, rather than events predating that period.

**Residential patterns** The concept of the ethnically-specified settlement within and adjacent to an urban area was familiar to Malays prior to migration to Sri Lanka, as it characterized the spatial organization of Batavian (i.e. Jakartan) districts (*kampungs*) on Java. Even the settlement of Malays in urban areas within Sinhala-majority regions does not guarantee that some degree of spatial proximity to Sinhala people could play a glottogenetic role. According to Shahul Hasbullah (p.c.), a demographer affiliated with the University of Peradeniya, a comparable pattern was in existence historically even in predominantly Moorish communities, where the designation Muslim, not an official category as it became in the twentieth century, was nevertheless a defining category for Muslims themselves, and this contributed heavily to where Moors chose to reside, in a way that language did not. Malay residential patterns in most areas have been characterized historically by spatial proximity to Moors and extensive social and cultural interaction with them. After the disbanding of the Malay regiment in the late 19th century, some Malay families fanned out to new areas especially in the hill country, to work as overseers and security personnel in the newly opened plantations of rubber, coffee and later tea estates. This fact in itself is significant, with respect to the Malay relationship to varieties of Tamil. Malays easily performed an indispensable middle-man function between essentially monolingual English-speaking plantation owners and monolingual Tamil-speaking plantation workers, since the Malays spoke a Tamil variety (Shonam), although they were not ethnically Tamil themselves, and since they were

<sup>19</sup>There is no evidence for marriages of Malays to Sinhalese, although this does occur infrequently.

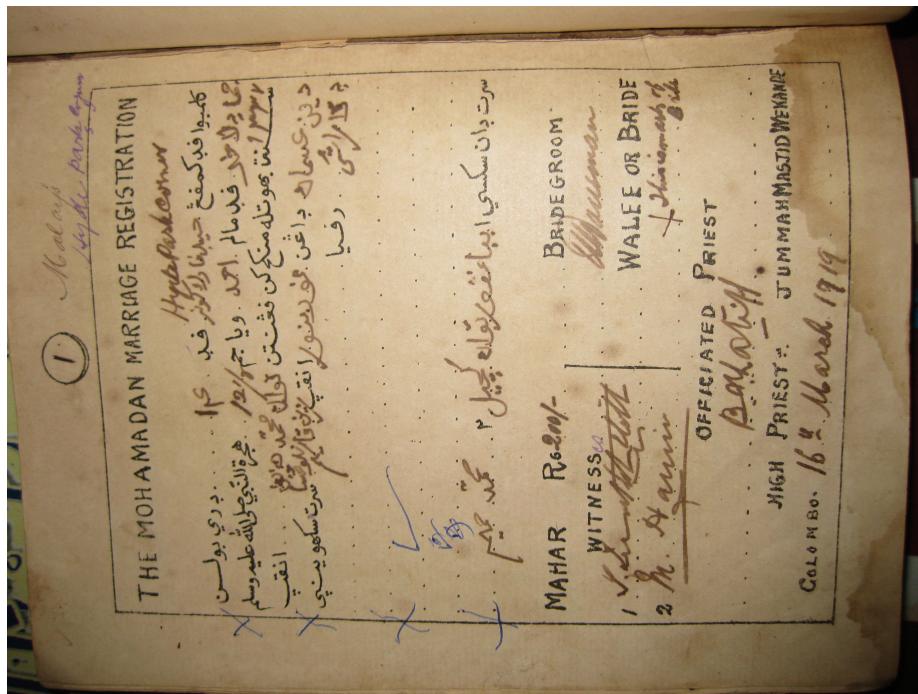


Figure 4.1: A Malay *kadutham* from 1919.

Figure 4.2: A page from a Malay marriage register.

Ref.	Male Partner	Female Partner	Wife's Name	Date
830	Moor	Moor	Moor	2/10
831	Moor	Moor	Moor	2/10
832	Moor	Moor	Moor	2/10
833	Moor	Moor	Moor	2/10
834	Moor	Moor	Moor	2/10
835	Moor	Moor	Moor	2/10
836	Moor	Moor	Moor	2/10
837	Moor	Moor	Moor	2/10
838	Moor	Moor	Moor	2/10
839	Moor	Moor	Moor	2/10
840	Moor	Moor	Moor	2/10
841	Moor	Moor	Moor	2/10
842	Moor	Moor	Moor	2/10
843	Moor	Moor	Moor	2/10
844	Moor	Moor	Moor	2/10
845	Moor	Moor	Moor	2/10
846	Moor	Moor	Moor	2/10
847	Moor	Moor	Moor	2/10
848	Moor	Moor	Moor	2/10
849	Moor	Moor	Moor	2/10
850	Moor	Moor	Moor	2/10
851	Moor	Moor	Moor	2/10
852	Moor	Moor	Moor	2/10
853	Moor	Moor	Moor	2/10
854	Moor	Moor	Moor	2/10
855	Moor	Moor	Moor	2/10
856	Moor	Moor	Moor	2/10
857	Moor	Moor	Moor	2/10
858	Moor	Moor	Moor	2/10
859	Moor	Moor	Moor	2/10
860	Moor	Moor	Moor	2/10
861	Moor	Moor	Moor	2/10
862	Moor	Moor	Moor	2/10
863	Moor	Moor	Moor	2/10
864	Moor	Moor	Moor	2/10
865	Moor	Moor	Moor	2/10
866	Moor	Moor	Moor	2/10
867	Moor	Moor	Moor	2/10
868	Moor	Moor	Moor	2/10
869	Moor	Moor	Moor	2/10
870	Moor	Moor	Moor	2/10
871	Moor	Moor	Moor	2/10
872	Moor	Moor	Moor	2/10
873	Moor	Moor	Moor	2/10
874	Moor	Moor	Moor	2/10
875	Moor	Moor	Moor	2/10
876	Moor	Moor	Moor	2/10
877	Moor	Moor	Moor	2/10
878	Moor	Moor	Moor	2/10
879	Moor	Moor	Moor	2/10
880	Moor	Moor	Moor	2/10
881	Moor	Moor	Moor	2/10
882	Moor	Moor	Moor	2/10
883	Moor	Moor	Moor	2/10
884	Moor	Moor	Moor	2/10
885	Moor	Moor	Moor	2/10
886	Moor	Moor	Moor	2/10
887	Moor	Moor	Moor	2/10
888	Moor	Moor	Moor	2/10
889	Moor	Moor	Moor	2/10
890	Moor	Moor	Moor	2/10
891	Moor	Moor	Moor	2/10
892	Moor	Moor	Moor	2/10
893	Moor	Moor	Moor	2/10
894	Moor	Moor	Moor	2/10
895	Moor	Moor	Moor	2/10
896	Moor	Moor	Moor	2/10
897	Moor	Moor	Moor	2/10
898	Moor	Moor	Moor	2/10
899	Moor	Moor	Moor	2/10
900	Moor	Moor	Moor	2/10
901	Moor	Moor	Moor	2/10
902	Moor	Moor	Moor	2/10
903	Moor	Moor	Moor	2/10
904	Moor	Moor	Moor	2/10
905	Moor	Moor	Moor	2/10
906	Moor	Moor	Moor	2/10
907	Moor	Moor	Moor	2/10
908	Moor	Moor	Moor	2/10
909	Moor	Moor	Moor	2/10
910	Moor	Moor	Moor	2/10
911	Moor	Moor	Moor	2/10
912	Moor	Moor	Moor	2/10
913	Moor	Moor	Moor	2/10
914	Moor	Moor	Moor	2/10
915	Moor	Moor	Moor	2/10
916	Moor	Moor	Moor	2/10
917	Moor	Moor	Moor	2/10
918	Moor	Moor	Moor	2/10
919	Moor	Moor	Moor	2/10
920	Moor	Moor	Moor	2/10
921	Moor	Moor	Moor	2/10
922	Moor	Moor	Moor	2/10
923	Moor	Moor	Moor	2/10
924	Moor	Moor	Moor	2/10
925	Moor	Moor	Moor	2/10
926	Moor	Moor	Moor	2/10
927	Moor	Moor	Moor	2/10
928	Moor	Moor	Moor	2/10
929	Moor	Moor	Moor	2/10
930	Moor	Moor	Moor	2/10
931	Moor	Moor	Moor	2/10
932	Moor	Moor	Moor	2/10
933	Moor	Moor	Moor	2/10
934	Moor	Moor	Moor	2/10
935	Moor	Moor	Moor	2/10
936	Moor	Moor	Moor	2/10
937	Moor	Moor	Moor	2/10
938	Moor	Moor	Moor	2/10
939	Moor	Moor	Moor	2/10
940	Moor	Moor	Moor	2/10
941	Moor	Moor	Moor	2/10
942	Moor	Moor	Moor	2/10
943	Moor	Moor	Moor	2/10
944	Moor	Moor	Moor	2/10
945	Moor	Moor	Moor	2/10
946	Moor	Moor	Moor	2/10
947	Moor	Moor	Moor	2/10
948	Moor	Moor	Moor	2/10
949	Moor	Moor	Moor	2/10
950	Moor	Moor	Moor	2/10
951	Moor	Moor	Moor	2/10
952	Moor	Moor	Moor	2/10
953	Moor	Moor	Moor	2/10
954	Moor	Moor	Moor	2/10
955	Moor	Moor	Moor	2/10
956	Moor	Moor	Moor	2/10
957	Moor	Moor	Moor	2/10
958	Moor	Moor	Moor	2/10
959	Moor	Moor	Moor	2/10
960	Moor	Moor	Moor	2/10
961	Moor	Moor	Moor	2/10
962	Moor	Moor	Moor	2/10
963	Moor	Moor	Moor	2/10
964	Moor	Moor	Moor	2/10
965	Moor	Moor	Moor	2/10
966	Moor	Moor	Moor	2/10
967	Moor	Moor	Moor	2/10
968	Moor	Moor	Moor	2/10
969	Moor	Moor	Moor	2/10
970	Moor	Moor	Moor	2/10
971	Moor	Moor	Moor	2/10
972	Moor	Moor	Moor	2/10
973	Moor	Moor	Moor	2/10
974	Moor	Moor	Moor	2/10
975	Moor	Moor	Moor	2/10
976	Moor	Moor	Moor	2/10
977	Moor	Moor	Moor	2/10
978	Moor	Moor	Moor	2/10
979	Moor	Moor	Moor	2/10
980	Moor	Moor	Moor	2/10
981	Moor	Moor	Moor	2/10
982	Moor	Moor	Moor	2/10
983	Moor	Moor	Moor	2/10
984	Moor	Moor	Moor	2/10
985	Moor	Moor	Moor	2/10
986	Moor	Moor	Moor	2/10
987	Moor	Moor	Moor	2/10
988	Moor	Moor	Moor	2/10
989	Moor	Moor	Moor	2/10
990	Moor	Moor	Moor	2/10
991	Moor	Moor	Moor	2/10
992	Moor	Moor	Moor	2/10
993	Moor	Moor	Moor	2/10
994	Moor	Moor	Moor	2/10
995	Moor	Moor	Moor	2/10
996	Moor	Moor	Moor	2/10
997	Moor	Moor	Moor	2/10
998	Moor	Moor	Moor	2/10
999	Moor	Moor	Moor	2/10
1000	Moor	Moor	Moor	2/10

Figure 4.3: A page from a Moorish marriage register in a Malay mosque, in which almost half of the marriages are Moor-Malay marriages.

frequently fluent in English as well.

**Religious and cultural life** In a number of communities, Malays founded their own mosques, in which Malay-speaking religious officials were to officiate, so that Malay could be used in administering religious rites. This fact is cited in Nordhoff (2009). However Malay mosques did not exist in Sri Lanka prior to the end of the Dutch colonial period (almost a century and a half into the Malay settlement in Sri Lanka), and these have never been exclusively Malay institutions, but rather were shared (B.A. Hussainmiya, p.c., Al Haj Muzni Amer, p.c.). Malays and Moors developed a kind of symbiotic relationship, which only began to break down in the last century, under the pressures of ethnopolitical communalism. Despite the fact that most Moors did not speak Malay, the two groups shared a common religious culture, a fact that held greater cultural significance than the fact that their occupational lives often diverged.<sup>20</sup> This is very different from following a single confessional tradition in modern western terms, since the extent to which daily life revolves around the organization of spiritual life and its strictures was far greater than what is familiar to relatively secularized people nowadays. Both groups belonged to the Shāfi'i sect and their written religious literature used similar Arabic-based orthographic systems. The Malay religious texts were written in Jawi and the Moorish texts were written in Arabic-Tamil, also referred to as Arwi (Shu'ayb Alim 1993). Even the content of the religious texts was frequently

<sup>20</sup>On the east coast and in other Sinhala-majority areas, Moors have been associated historically with commerce, whereas Malays were most closely associated with military activity, plantation supervision, and police work.

identical, some being direct translations from Arabic-Tamil to Malay written in Jawi orthography, and vice versa.

It is significant that the Tamil Muslim-Indonesian religious and cultural symbiosis that we find evidence for in Sri Lanka extends beyond Sri Lanka and it may be that Muslim Tamil culture was familiar to the Indonesian immigrants prior to their arrival in Sri Lanka. The culture that became known as “Moorish” in Sri Lanka was not restricted to that island, but flourished and spread as the seafaring mercantile culture of Tamil-speaking Muslims of Arab and South Indian descent. Coastal southern India (“Malabar”), coastal Sri Lanka, and coastal area in western Indonesia and Malaya were its focal areas. In the Indonesian context, evidence exists supporting Muslim-Tamil and Indonesian (*inter alia* Malay and Javanese) interactions, gleaned from close readings of textual sources in Javanese, Tamil and Malay. In the Sri Lankan context, Malays and Moors shared religious and literary texts. This is a matter for which we have hard evidence, as many of the texts still exist.<sup>21</sup> A recent dissertation by the literary historian Ronit Ricci (Ricci 2006) describes the importance of literary links between the Javanese and Malay world on the one hand and the Tamil-speaking Muslim world in southern India and Sri Lanka on the other, based on common religious texts belonging to Tamil-speaking Muslims, the Malays and the Javanese in the Indonesian archipelago.<sup>22</sup>

The coasts of Southeast India and Indonesia were part of the Indian Ocean’s commercial network that was permeated – beginning in the fifteenth century – by an Islamic ethos, where goods and shared texts and values crossed the seas carried by Muslim merchants, pilgrims, soldiers and scholars, and where coastal towns, which functioned as important trade centers and ports, developed into major centers of Islamic learning and culture.<sup>23</sup>

The Muslims of South India and the archipelago shared a variety of relationships: from at least as early as the seventeenth century they had a shared set of pilgrimage sites, some of which are still popular today. Well known in South India is the lineage of the seventeenth century sufi mystic sheikh Sadaqatullah of Kayalpatnam, whose tomb continues to attract devotees from Malaysia and Indonesia; Muslims on both shores maintained mutual trade contacts, with the Nagore-Acheh route becoming one of the most profitable in the eighteenth century network. The Marakkaiyar trading clans of the Coromandel coast, claiming Arab seafarers and traders as their ancestors, had well established ties with the Muslim ports of the archipelago, exporting tobacco, cotton textiles, gems and pearls; men of the Tamil coastal towns traveled to Southeast Asia, sometimes remaining there for months at a time; members of the two communities even intermarried, the Marakkaiyars preferring intermarriage with the Muslims of the archipelago over marriage with the lower strata of Tamil Muslim society.

The *madhab* (school of Islamic law) followed by Javanese and South In-

---

<sup>21</sup>A not insignificant number of these are in the possession of B.A. Hussainmiya, who may compile and publish an annotated collection of these texts.

<sup>22</sup>The footnotes on the following page are Ricci’s.

<sup>23</sup>On the relationship between trade and Islam in these regions see, for example, Andre Wink, “*Al-Hind*”. India and Indonesia in the Islamic World Economy, c. 700-1800 A.D.” in *India and Indonesia During the Ancien Regime* (Leiden: Brill, 1989) 48-49, and Kenneth McPherson, *The Indian Ocean: A History of People and the Sea* (New Delhi: Oxford University Press, 1993) 76-78.

dian Muslims living along the coast is one and the same (Shafi'i) contacts in the sphere of Islamic education appear to have been strong, with similar institutions emerging in Tamil Nadu, Sumatra and Java; Indonesian pilgrims on their way to Arabia used to stop in the Maldives; in the eighteenth century a Coromandel mosque existed in Batavia, while in the early nineteenth century an approximate 2% of Batavia's population were "Moormen", natives of the Coromandel and Malabar coasts; under colonial auspices contacts – whether through trade or the deployment, employment or exile of subjects. (Ricci 2006:12-13)

A Tamil tradition credits Umar Wali of Kayalpattnam with establishing Islamic schools in Sumatra, where he spent fourteen years in the mid eighteenth century. When Shu'ayb visited the region in 1978 he noted the similarities in curriculum between the contemporary *pesantren* of northern Sumatra and the earlier *madrasah* schools of Tamil Nadu and Sri Lanka, some of which no longer taught certain previously shared texts on account of North Indian influences on Muslim education in the south.

Religious teachers often traveled in a quest to disseminate their knowledge and religious convictions to others, expanding the geographical and cultural limits of the cosmopolis. It is known from the *Sejarah Melayu* ("Malay Annals") that Tamil Muslim teachers were influential in the Malay regions in the fifteenth century. The "Annals" also claim – as does the *Hikayat Raja-Raja Pasai* ("Book of the Kings of Pasai") - that the apostles of Islam reached Malay shores from the Coromandel coast.<sup>24</sup> Shuayb discusses at length the deeds of the above-mentioned Umar Wali, a Tamil 'saint' who spent years in the forests of Sumatra, propagating Islam<sup>25</sup>; Bayly mentions a Tamil *pir* from Vethalai who, while meditating in a Sumatran jungle, encountered and overcame a fierce elephant. In gratitude the sultan granted him his daughter in marriage and nominated him as successor to the Achehnese sultanate<sup>26</sup>; Javanese nobles and their retinues, exiled to Sri Lanka in the seventeenth and eighteenth centuries, brought with them – if not in written certainly in oral form – stories and traditions which were eventually shared with other Muslims on the island.<sup>27</sup> Although the historical accuracy of some of these mentions cannot always be determined with certainty such traditions attest to a sustained memory of participation in promoting Arabicized networks of language, literature and learning that connected Muslims across the region. (Ricci 2006:388-389)

The best evidence for a close religious and cultural relationship, a symbiosis between the Malay and Moorish communities in Sri Lanka, is the texts that exemplify the literary life of the relevant communities, just as the work of Ricci demonstrates connections between the Javanese Muslim world and the Tamil-speaking Muslim networks generally. While we may note to our consternation that there is too little direct evidence in literature of what spoken Sri Lankan Malay looked like prior to the twentieth

---

<sup>24</sup>Robson, *Java* 262.

<sup>25</sup>Shu'ayb 502.

<sup>26</sup>Bayly, *Islam* 155.

<sup>27</sup>Hussainmiya, *Orang Regimen* 38-42.

century, there is nevertheless extensive direct literary evidence for the Malay-Moorish symbiosis.<sup>28</sup> This is discussed briefly in the introduction to [Hussainmiya \(2008\)](#).

Still more important is the fact that the texts [ i.e. those associated with the Malay literary tradition ] revealed the strong links between the Malay community and the Tamil-speaking Sri Lankan Moors. Both shared many features of a literary tradition. They read and exchanged kitabs (in the sense of Islamic religious works) and other literary manuscripts. The Moorish texts were written in Arabic-Tamil, employing an adapted Arabic orthography similar to the one in which Malay was written. So Malays easily read Arabic-Tamil, as they knew Tamil, and were familiar with the script. Interestingly, the Malays in Sri Lanka pioneered the publication of trilingual instruction booklets in Arabic, Malay and Arabic-Tamil. They even published Arabic-Tamil newspapers such as *Ajaib as-Sailan* and *Ummai*.<sup>29</sup> The Malay Jawi newspaper *Alamat Langkapuri* (1869-1870) carried letters and notices in Arabic-Tamil. Moreover, it was not uncommon to find works of Malay and Arabu-Tamil literature bound in the same codex of manuscripts owned by the Malays. They used texts on medicine, magic, and religious formulas written in Arabic-Tamil. Apart from the need to use Arabic liturgical works such as *Subhana Maulud* (eulogy on the Prophet Muhammad) a Malay intellectual such as Baba Ounus Saldin translated Arabic-Tamil works such as *Gnanamani Malai* into Malay for wide use among his colleagues. There were erudite Tamil poets among the Malays who wrote their works in Tamil. Poets such as Jumaron Tungku Usmand, who lived close to the Indian Tamil labourers in the hill country plantation districts, composed Malay songs in romanised Malay by employing folk Tamil literary forms of Kummi and Temmangu.

According to the Sri Lankan Malay linguist Mohamed Jaffar (p.c.), the deep cultural and spiritual connection between Sri Lankan Malay and Shonam in the oldest generation of Malays in Sri Lanka, even in Malay-Malay marriages, needs to be understood by linguists who may assume that Shonam is merely one equal component in the linguistic experience of Malay Muslims in Sri Lanka, as opposed to an intimate part of domestic life. Jaffar speaks of “the prevalence of Shonam in their lives, all of it connected with religious practices: the recitation of *doa 'aashura*, a supplication in the first month of the Islamic calendar, *muharram*, in memory of the martyrdom of the Prophet’s grandson Husayn ibn ’Ali at the battle of Karbala the recitation of the *thali faatiha*, a song in praise of the Prophet’s daughter Fáthima, in both Arabic and *Arwi* (*Arabu-Thamul*) authored by Sayyid Muhammad ibn Ahmad Lebbai ‘álím-al-Qáhirí-al Kirkári, great-grandfather of Shu’ayb Alim; and several other publications, usually Arabic-*Shonam* bilingual, numerous to bear mention here, however, easily obtainable. As recently as in 1963 to be precise, the late Mr Saifuddin J Aniff-Doray — a Sri Lanka Malay notable and one-time Principal of Zahira College, Colombo — translated into English from *Arabu-Thamul* the voluminous “Fat-aud Dayyān” (Fata al-Dayyaan) of Sayyid Muhammad ibn Ahmad Lebbai mentioned above. The evidence and arguments that I have presented in this section is meant to demonstrate that exposure to

---

<sup>28</sup>For a methodological critique of recent claims denying Malay-Moor intermarriage, see [Rassool \(this volume\)](#).

<sup>29</sup>Ummai (“The Truth”) is the first Arabic-Tamil (i.e. Shonam/Arwi) newspaper in Sri Lanka. It was edited and published by a Sri Lankan Malay.

4

عرب	ملاي	عرب
عَرْبَ قَلْ	دَارَةٌ	دَمٌ
أَرَاثَمُ	أَوْنَقٌ	سَمَاعٌ
ثَلَاثَةٌ مُوقِبَيٌ	كُوْمَقٌ	شَحْمٌ
كَلْفٌ	سُخْنَسْخَنٌ	تَخَانٌ
مُوضَبَيٌ	لُودٌ	رِيقٌ
أَمْسِيفٌ	أَيْمَاتٌ	دُمُوعٌ
كَرْتِيٌّ	مَصَارِبَتٌ	فَلَكَاسَنٌ
كُلْلَحَضَنٌ	سُوازَرَةٌ	صَوْلَةٌ
كُرْلٌ	بَلْكَسَنٌ	عَطَشٌ
مُوكَلٌ	جَهَاكَةٌ	عَطَشَنٌ
تَاكَمٌ		

Figure 4.4: Scan of a single glossary page from a nineteenth century text used to teach Arabic in Qur’anic schools attended by Malays. From right to left, the first column is in Arabic, the second is in Malay, and the third is in “Arabic-Tamil” (an alternative name, as is Arwi, for Shonam written in Arabic script). Courtesy of B.A. Hussainmiya.

involvement of Malays in the religious and literary culture of Tamil-speaking Muslims is not restricted to Sri Lanka, but was an Indian Ocean regional phenomenon. Not surprisingly, this involvement was perpetuated in Sri Lanka, a context that needs to be understood in order to see that the linguistic behavior of Moors is likely to have been prestigious within Malay networks. It is also necessary to appreciate this in order to see that intensive interaction between Malays and Moors was greater than a husband, wife, and child domestic bilingualism configuration, but rather extended outward from the home to involve those Malays who did not marry Moors. Shonam was in fact a process language for religious practice among Malay in Sri Lanka, as it continues to be in the most conservative communities, including the all-Malay village of Kirinda.

### 4.3 What is inferable in SLM research?

Much of the diachronic analysis of the grammar of Sri Lankan Malay is inferable, and it is fortunate that this is so. In Malay vernaculars in general, including contact Malay varieties, not just in Sri Lanka, investigating the internal history of a variety is rendered more difficult by a highly diglossic culture. We have few attestations for Malay in Sri Lanka. What we do have is associated with the nineteenth and early twentieth centuries, and the Jawi texts from that period are written in what is essentially literary Malay. The British colonial government encouraged Malay literary activities in order to create a culturally favorable climate for the military personnel that were being recruited in Malaysia. The diglossia that characterizes Sri Lankan Malay literary culture during this period is also characteristic of Tamil and Sinhala literary culture. The written varieties of those languages have since been disseminated across the island as a result of the universal vernacular education policies of successive post-independence Sri Lankan governments. While literacy in at least one of the two official languages has been rendered nearly universal, a tremendous achievement for the country, neither of the major Muslim languages, Shonam and SLM, are written any longer, and the number of people able to read the surviving texts is small.<sup>30</sup> Since we do not have corpora or even isolated attestations of early periods in the history of vernacular Sri Lankan Malay, we have to depend on plausible argumentation, in order to reconstruct linguistic events in the language's development, whether in morphosyntax or another grammatical component. Linguists are familiar with this problem from other unwritten vernaculars, certainly including contact varieties. It is only a minority of the world's languages whose internal history can be examined on the basis of extensive corpora, but this fact ought not to dissuade us from investigating diachrony. I have worked, among other topics, on the development of tense and finiteness contrasts in Sri Lankan Malay and the extension of the finiteness contrast to the negation system. The presence of a finiteness contrast is at first glance a surprising development in the grammar of Sri Lankan Malay, since there is no such contrast in vehicular Malay varieties. Moreover, although the function of these contrasts varies cross-linguistically, they do not encode semantic contrasts that are ordinarily likely to be restored in contact languages that do not have them, or that have lost them. My argumentation follows from the principle that grammatical organization that would ordinarily change only very slowly, may change rapidly where a bilingual minority has experienced a radical shift in discourse

---

<sup>30</sup>Note that Shonam or *Arwi* texts were not simply literary Tamil texts in Arabic-derived script. This means that those texts have the potential to tell us more about the characteristics of Shonam, whereas Malay texts were written in literary Malay that was certainly not simply the formal register for the forms of Malay spoken in Sri Lanka, but rather a variety that was quite far removed from ordinary usage.

culture, and considerable interaction with second language acquirers speaking a typologically discordant language. An important function of finiteness in the Sri Lankan sprachbund is to demonstrate the relative status of the most recent predicate in a temporally asymmetrical sequence (representing a sequence of events that do not occur simultaneously). This is the conjunctive participle construction, that conveys the sense of “Having done A<sup>x</sup>, having (subsequently) done B<sup>y</sup>, I did C<sup>z</sup>”, where x and y are participles, and z is a finite matrix verb. As we can see from the conjunctive participle construction and its subordinate (adjoined) relationship to finite matrix predicates, Sri Lankan Malay has developed cross-clausal assymmetries that are not transparently functional, and that consequently require explanation. In the shift from vernacular Malay culture to Lankan culture, the change in information structure conventions happens to align with profound typological difference. We may ask rhetorically why, given continued access to L1 Malay, would speakers restructure their vernacular to such an extent that it began to take on the grammatical characteristics of a Dravidian language? Syntax under conditions of stable bilingualism changes very slowly in response to contact, although surface configurations that do not violate the constraints of the existing grammar may become increasingly frequent in response to pragmatic changes, leading to reanalysis. This is not qualitatively different from a process we find taking place to a greater or lesser extent in linguistic change generally, but the process is accelerated substantially. This is not the same as creolization in the older and more specific technical sense of pidginization with massive deflection, followed by replacement through grammaticalization. Some of the changes in Sri Lankan Malay, such as the post-verbal distribution of aspect markers in finite clauses and the postponing of prepositions can in no sense be characterized as processes of replacement or accretion of missing grammatical phenomena. It is somewhat unexpected that a radical contact language would develop new phonologically-dependent morphology (whether affixes or clitics) to mark case contrasts, given the fact that syntactic and prosodic processes can overtly mark the thematic relationships that case morphology typically marks. However the function of the morphology is clear, since it encodes relationships between arguments of predicates that are abstractly present in all natural languages, although these relationships are encoded differently in the vehicular Malay lexifier varieties.<sup>31</sup> The sociolinguistic scenario in Sri Lanka, and the acquisition of Malay by speakers of a language with robust case morphology, provides the conditions for what took place.<sup>32</sup> The sentence in (1) exemplifies functional and morphological change in Sri Lankan Malay, since none of the functional affixes and clitics we see in bold have immediate formal analogues in vehicular Malay, although they are all derived from free-standing closed class morphology in that language (and in Javanese, in the case of *na*, based on *nang*).

- (1) *Buk=yang e-baca, Miflal=na məsigit=ring tumman attu=na ittu=yang*  
 book=ACC PTCP-give Miflal=DAT mosque=ABL friend DET=DAT DET=ACC  
*mə-kasi=na si-kəmauan.*  
 INF-give=DAT PAST-want  
 “Having read the book, Miflal wanted to give it to a friend from the mosque.”

<sup>31</sup>Descriptive linguists and others who object to this view will not object to the view that the logical relationships between arguments of a predicate are perceptually present in a predictable feature of human cognition, whether or not the formal linguistic system operationalizes the cognitive universal.

<sup>32</sup>It is important to note that a strikingly similar pattern of nominal complexification took place in the development of Sri Lankan Portuguese. It would be implausible to claim that second language acquisition did not contribute to grammatical change in that case, since not only does Portuguese, like vehicular Malay, not feature non-pronominal case morphology, but the speakers of Sri Lankan Portuguese, still a cohesive community in parts of Tamil-speaking eastern Sri Lanka, are of southern South Asian appearance.

The presence of a finiteness contrast, straightforwardly visible in the presence of an infinitival prefix, constitutes complexification that requires plausible diachronic modeling. With respect to finiteness, the term complexification refers not so much to the phonological dependence of functional material on a lexical host, a rather subjective criterion for defining a gradable and ambiguous term, but to the accretion of functional contrasts not present in the language previously.

#### 4.4 What is discoverable in Sri Lankan Malay research?

The most detailed investigation to date of the grammar of Sri Lankan Malay is [Nordhoff \(2009\)](#), which is quite extensive and broad in its scope. The author sensibly focused on one variety (the “upcountry” variety spoken in Kandy) rather than attempting to accommodate all of the dialect variation in one grammar. Much remains to be said about the other varieties, although they are mutually intelligible, and have much in common. Much remains to be said about the variety investigated by Nordhoff as well, as we would expect with any language whose grammar has only recently been extensively investigated. Although Nordhoff was not the first to write on the phonology of SLM (see also Tapovanaye 1986, 1995 and Bichsel-Stettler 1989), he has asked significant questions, for example on pre-nasalization and segmental boundary perception, basing his investigation on acoustic analysis. His conclusions have in turn been challenged in recent work by Ian Smith. The segmental phonology of SLM is an area in need of greater investigation, but there is also evidence for changes in the language’s intonational phonology during the course of the twentieth century, although this topic has not yet been systematically investigated. There is also a range of open questions and controversies concerning the present grammar (particularly in syntax and morphosyntax) and concerning the present lexicon of the language, which is strongly in flux. In Kirinda, for example, in spite of the cross-generational vitality of the language, generalized trilingualism, and frequent contact with monolingual Sinhala speakers, Shonam and Tamil vocabulary is replacing older Malay lexical items. The same source lexical borrowing source is now less likely in communities outside the southeast, where much of the younger generation is being educated in Sinhala, rather than in (any variety of) Tamil. There has so far been no systematic investigation of sociolinguistic variation in the language, although variation is extensive and apparent even to casual observers (cf [Rassool this volume](#)). This variation should be investigated quantitatively based on social categories that are meaningful within the SLM communities, and the sociology of the language. Getting at those social categories requires detailed ethnographic research. Given the extent of anthropological interest in Sri Lanka, it is surprising that no anthropologist, to my knowledge, has taken an interest in its Malay communities. Returning to the matter of grammatical investigation, it would be an obvious mistake to assume that there is any less to be discovered in synchronic linguistic research on SLM than there is in any other language. There are two related factors, which ought to stimulate further linguistic research. Informally, linguist observers who acknowledged the profound divergence of SLM from vehicular Malay varieties in the Indonesian archipelago had assumed that SLM consisted of a bare lexical inventory superimposed upon Tamil (U. Tadmor, p.c.). Recent research on the grammar of SLM has implicitly and explicitly demonstrated that SLM is highly divergent as a Malay variety, and that it remains an autonomous linguistic system, rather than an ethnically emblematic lexical inventory, analogous with, for example, Anglo-Romani. Given that this autonomous system is by all accounts much more Lankan than it is Malay, in all

but its lexical content, it follows that we have many more questions to investigate with respect to microparametric contrasts with the Lankan languages that have influenced its development. Why has replication brought this much convergence, but no more? Published research on Sinhala has been particularly rich in investigations of information structure phenomena (Gair 1986, 1998, Kariyakarawana 1998 et al), partly because focus is a complex process in Sinhala grammar with morphological as well as syntactic reflexes. For those linguists who believe that there is as much reason to investigate diachronic influence from Sinhala as there is to investigate diachronic influence from Shonam (Ansaldo and Nordhoff), it is worth noting that SLM has not grammaticalized the focus of (non-clausal) constituents under the influence of Sinhala, although focus morphology on verbs is one of the most striking and well-investigated features of Sinhala grammar. To what extent have Sinhala focus phenomena been replicated without recourse to specialized morphology in the grammar of SLM? In my own case, I have rejected the external case for Sinhala influence prior to Sri Lankan independence and I have also shown how SLM aligns more closely with Shonam than with Sinhala. This conclusion applies to glottogenesis and subsequent development. However changing social conditions have led to adstratal influence from Sinhala from the mid-twentieth century onwards, and the effects of this influence are likely to increase over time. To what extent can we identify Sinhala influence in the SLM focus system? How does this compare with Shonam focus specifically? Note that far too little is known about differences between the grammar of Shonam and that of Tamil varieties spoken by Sri Lankan Hindus, Christians, and others. References to “Tamil” are not helpful, since general Sri Lankan Tamil is of little relevance in the study of SLM. While there is still a (very) small number of SLM-speaking Malays living in the Eastern Province, which is predominantly Tamil-speaking, most Malays are in greater contact with speakers of Sinhala than ever before, and due to the educational system, the mass media, and the breakdown of ethnic boundaries, Sinhala is rapidly becoming the dominant vernacular in the majority of youth peer networks, including those of Moors, Malays, Burghers and others. The current picture is quite close to Ansaldo’s scenario [of Sinhala dominance], even if the historical picture deviates from it. To what extent is this affecting the grammar of the language? To what extent is SLM in urban areas losing sociolinguistic domains, and what effect has this had on its lexical inventory? Is this accompanied by the loss of grammatical contrasts as we see for endangered languages spoken elsewhere, or is it the case that because the dominant language, in this case Sinhala, has a range of contrasts encoded with inflectional morphology, that we find *new* contrasts? Ansaldo (2005) claimed that Kirinda Malay had developed an accusative-dative split where there was previously none, but he had unfortunately been misled by glosses provided by B.A. Hussainmiya (a historian, not a linguist) in field recordings transcribed for Ian Smith.<sup>33</sup> Although this was not a genuine example of case split, since all modern Sri Lankan Malay varieties investigated have a robust accusative-dative split, Ansaldo was right to assume that changes in the language’s case system, including possible

---

<sup>33</sup>While Ansaldo may have been misled in that case, he also repeatedly (a) misrepresents the research that I have published on the grammar of Sri Lankan Malay, and (b) makes misleading statements that mask its existence. In a list of researchers judged guilty of some measure of “Tamil bias”, Ansaldo Ansaldo (2011a:377-378) lists me as a weak offender. (“Slomanson 2006 argues for convergence between SLM, Sinhala and Tamil, though Malay-Tamil bilingualism is invoked to explain some aspects of tense-marking.”) I do not recognize my work in that description. He later summarizes his list of biased researchers with a reference to Smith Smith (2003). “The most serious attempt to date to argue for Tamil as a primary “substrate” (i.e. Smith 2003) is a claim which, crucially, failed to find any evidence.” Not only have I published research providing grammatical evidence for the primary role of Shonam, but Ansaldo was present at Slomanson (2007) on the Dravidian (i.e. Shonam) character of the Sri Lankan Malay negation system.

elaboration, are still possible. Given changes in the sociocultural configurations in which SLM speakers now find themselves and given the absence of a written norm, the whole matter of variation and change merits much more extensive investigation. The rate of change may be more rapid than the rate of change in demographically dominant languages whose development is mediated by a written tradition and its associated conventions. One of the standard strategies for investigating variation is to employ apparent time methodology, and to do so before too much time passes, since the passing of time leads to the loss of potential informants, with predictable information loss as its result. There are many elderly members of the community who are still alive, but it is important to start working with them now. There are SLM speakers in their nineties in some communities. In others, such as Kirinda, there are no individuals who have reached the age of eighty (M.T.M. Rihan, p.c.). This also leads to a loss of culture and historical memory, since not everything is being passed on. If we look at a community such as Kirinda, we see the effects of cultural change quite starkly. This is because none of what we see there is attributable to language loss as such. The entire younger generation is completely fluent in SLM. Nevertheless, the lexical inventory is changing rapidly, with Shonam and Tamil words, and to a much lesser extent, Sinhala and English ones, replacing Malay ones. The most common informant response to a question about the acceptability of a particular etymologically Malay word is “That is what old people say” and the counterpart used by young people is generally Tamil or Shonam. This is not surprising, since contacts with the SLM world beyond Hambantota are few, education is Tamil-medium, half of the teachers are Moors, Shonam/Tamil is the medium of Friday religious instruction that is broadcast over loudspeakers, and Tamil-medium television blares in every home. Although Moors do not reside in the village, some village young people attend Hambantota schools, where they become integrated into Moorish peer networks, in which the ordinary language is Shonam. Discovering what is happening linguistically in this age cohort (late teens and early twenties) is a matter of some importance, because this is the cohort that will be raising the next generation. Another area of sociolinguistic inquiry that is indispensable for investigating variation and change is network analysis. My historical claim with respect to the Shonam versus Sinhala question is that Malay-Moorish networks were densest, however there is no way to demonstrate this with mathematical precision, because the network members are no longer alive, and because there has been too little documentary evidence discovered. For present-day youth networks, there is no reason for us not to investigate.

## 4.5 Conclusion

I have provided external arguments in support of the view that Shonam was the primary Lankan influence on Sri Lankan Malay in its development since the beginning of the Dutch period in the island’s history. I have reviewed comments on external argumentation in the literature, and have challenged the comments that I found to detract from our understanding of the social processes that contributed to grammatical change. I have also discussed the importance of inferable diachronic information that we obtain through plausible hypothesis formation. Lastly, I have made suggestions for future research that will advance our understanding of the trajectory of linguistic and sociolinguistic change in Sri Lankan Malay.

# Bibliography

- Ansaldo, U., 2011a. “Sri Lanka Malay and its Lankan adstrates.” In *Creoles, their Substrates, and Language Typology*, edited by C. Lefebvre. Amsterdam: John Benjamins, 367–382.
- Ansaldo, U. & Nordhoff, S., 2009. “Complexity and the Age of Languages.” In *Complex Processes in New Languages*, edited by E. Aboh & N. Smith. Amsterdam: John Benjamins Publishers, 345–363.
- Ansaldo, Umberto, 2005. “Typological admixture in Sri Lanka Malay. The case of Kirinda Java.” Unpublished manuscript.
- Ansaldo, Umberto, 2008. “Sri Lanka Malay revisited: Genesis and classification.” In *A world of many voices: Lessons from documented endangered languages*, edited by Arienne Dwyer, D. Harrison, & David Rood. Amsterdam, Philadelphia: John Benjamins., 13–42.
- Ansaldo, Umberto, 2011b. “Metatypy in Sri Lanka Malay.” In *Annual Review of South Asian Languages and Linguistics*, edited by Rajendra Singh & Ghanshyam Sharma. Berlin: Mouton De Gruyter, 3–15.
- Deumert, Ana & Durrleman, Stephanie (eds.), 2006. *Structure and Variation in Language Contact*. Amsterdam: John Benjamins.
- Hussainmiya, B. A., 1987. *Lost cousins: The Malays of Sri Lanka*. Bangi: Institut Bahasa, Kesusasteraan dan Kebudayaan Melayu.
- Hussainmiya, B. A., 1990. *Orang Rejimen: The Malays of Sri Lanka*. Bangi: Universiti Kebangsaan Malaysia, 1st edition.
- Hussainmiya, B. A., 2008. *Orang Rejimen: The Malays of Sri Lanka*. Bangi: Universiti Kebangsaan Malaysia, 2nd edition.
- Lim, Lisa & Ansaldo, Umberto, 2006. “Keeping Kirinda vital: The endangerment-empowerment dilemma in the documentation of Sri Lanka Malay.” *Amsterdam Centre for Language & Cognition Working Papers* 1:51–66.
- Lim, Lisa & Ansaldo, Umberto, 2007. “Identity alignment in the multilingual space: The Malays of Sri Lanka.” In *Linguistic identity in multilingual postcolonial spaces*, edited by E. Anchimbe. Cambridge: Cambridge Scholars Press, 218–243.
- Nordhoff, Sebastian, 2009. *A Grammar of Upcountry Sri Lanka Malay*. Ph.D. thesis, University of Amsterdam, Utrecht.  
URL <http://www.lotpublications.nl/publish/articles/003744/bookpart.pdf>
- Percival, Robert, 1803. *An account of the island of Ceylon*. London.
- Rassool, Romola, this volume. “Issues of power and privilege in the maintenance of Sri Lanka Malay: A sociolinguistic analysis.”
- Ricci, R., 2006. *Translating Conversion in South and Southeast Asia: The Islamic Book*

- of one thousand questions in Javanese, Tamil and Malay.* Ph.D. thesis, University of Michigan.
- Schweitzer, C, 1931[1680]. *Reise nach Java und Ceylon*. The Hague: Nijhoff.
- Shu'ayb Alim, T., 1993. *Arabic, Arwi and Persian in Sarandib and Tamil Nadu*. Chennai: Imamul Arus Trust.
- Slomanson, P., 2009. “Morphological finiteness as increased complexity in negation systems.” In *Complex Processes in New Languages*, edited by E. Aboh & N. Smith. Amsterdam: John Benjamins Publishers.
- Slomanson, P., 2011. “Dravidian features in the Sri Lankan Malay verb.” In *Creoles, their Substrates, and Language Typology*, edited by C. Lefebvre. Amsterdam: John Benjamins.
- Slomanson, Peter, 2006. “Sri Lankan Malay morphosyntax: Lankan or Malay?” In [Deumert & Durrelman \(2006\)](#), 135–158.
- Slomanson, Peter, 2007. “On the Areal Character of Sri Lankan Malay Negation.” Paper presented at the summer meeting of the Society for Pidgin and Creole Linguistics. University of Amsterdam.
- Slomanson, Peter, 2008. “The perfect construction and complexity drift in Sri Lankan Malay.” *Lingua* 118(10):1640–1655.
- Smith, Ian R., 2003. “The provenance and timing of substrate influence in Sri Lankan Malay; Definiteness, animacy and accusative case marking.” Paper presented at the South Asian Language Analysis Roundtable XXIII, Austin, Texas.
- Smith, Ian R. & Paauw, S., 2006. “Sri Lanka Malay: Creole or convert.” In [Deumert & Durrelman \(2006\)](#), 159–182.
- Smith, Ian R., Paauw, S., & Hussainmiya, B. A., 2004. “Sri Lanka Malay: the state of the art.” *Yearbook of South Asian Languages* 2004:197–215.
- Ward, K., 2009. “Networks of Empire: Forced Migration in the Dutch East India Company.”

## **Chapter 5**

# **Issues of power and privilege in the maintenance of Sri Lanka Malay: A sociolinguistic analysis**

Romola Rassool

*University of Melbourne & University of Kelaniya*

### **5.1 Introduction**

The Sri Lanka Malays are a minority community who form 0.3% of the total population of Sri Lanka.<sup>1</sup> At present there are 55,353<sup>2</sup> Malays in Sri Lanka and they are spread mostly in the Western, Central, and Southern provinces of the country. They trace their origins to the Malays who migrated to Sri Lanka (then Ceylon) primarily during the period of Dutch colonial expansion. The lexifier of their language, Sri Lanka Malay (SLM), is Indonesian in nature, while its syntax, morphology and phonology have been considerably influenced by Sinhala and Tamil, the major languages of Sri Lanka. Since the present state of the SLM language is seen by many individuals in the community

<sup>1</sup>I would like to express my gratitude to Michael Ewing, Sander Adelaar, the editor of this volume, Kaushalya Perera and Dinali Fernando for their comments on earlier versions of this paper. I am also thankful to Eileen Dane and Naushad Rassool for their comments on various aspects of the Sri Lanka Malay community and to Harsha Vithaanaarachchi for helping me with formatting and citation issues.

<sup>2</sup>This number is arrived at by adding the 54,782 quoted in the census of 2001 (<http://www.statistics.gov.lk/PopHouSat/PDF/Population/p9p8Ethnicity.pdf>) and the 327 in the Trincomalee district, 163 in the Ampara district, and 81 in the Batticaloa district reported in the preliminary reports of the census conducted by the Department of Census and Statistics of Sri Lanka in 2007. The report on the census of 2001 did not contain information regarding the population of these districts due to the fact that travel to those districts was not possible because of the inter-ethnic conflict that prevailed in the north and east of Sri Lanka during that period. However, in my own visit to Trincomalee I observed that the number of Malays there is probably well over the figure stated in the census and my observation was confirmed by many Malay residents of the district. It is possible that the numbers in the other districts are similarly understated. Also, these are the figures pertaining to the section of the Sri Lankan population that has identified themselves as Malay by ethnicity and is not a reflection of the number of people claiming proficiency in SLM.

as a consequence of Sri Lanka's language policy of the mid 20<sup>th</sup> century, I provide a brief summary of the political changes of that period and how they affected the SLM language.

Though Sri Lanka was colonised by the Portuguese (1505 – 1656), the Dutch (1656 – 1796) and the British (1796 – 1948), it was not until British rule that inter-ethnic tensions between the majority Sinhalese population and the minority populations emerged to the foreground. Much of the tension centred round the Sinhalese and the Tamil population, the largest minority ethnic group. Feeling that the Tamils were advantaged during British rule, in the lead-up to and after the declaration of independence in 1948, the Sinhalese government introduced policies that were perceived as privileging the Sinhalese people. Much of the political discourse of the time surrounded the issue of language. In 1956 the Official Language Act (popularly known as the 'Sinhala-only' Act) was introduced, declaring Sinhala the sole official language. Though this act was subsequently amended through various provisions and Sinhala and Tamil are both declared national languages with English as the 'link' language, it is commonly believed that this unequal language policy led to the inter-ethnic rivalry that is still in evidence between the two largest ethnic groups in Sri Lanka. These tensions climaxed in a protracted conflict between the Liberation Tigers of Tamil Eelam (LTTE) and the security forces of Sri Lanka over demands by the Tamil Tigers for a separate homeland in the north of Sri Lanka, which ended in 2009 when the LTTE was militarily defeated by the armed forces.

The Sri Lanka Malays felt that they suffered a setback due to the new language policy. Prior to the 1956 Act, they had been assured of jobs due to their proficiency in English. The 1956 Act has also been seen as the indirect cause of the decline of the SLM language. According to [Saldin \(2001\)](#), when English was the medium of instruction in schools, Malay was the language spoken in the home. When the medium of instruction became Sinhala or Tamil, Malay parents felt they needed to speak to their children in English, in order to ensure that the children knew English, which was (and still is) seen as crucial for employment and upward mobility. Therefore, in a sense, the Official Languages Act of 1956 is considered to be responsible for Malay ceasing to be the language of the home ([Ansaldo 2006, Bichsel-Stettler 1989, Lim & Ansaldo 2007, Saldin 2001](#)).

The notion that the Sri Lanka Malay language is 'endangered' has been unquestioned since it was first introduced in the late 1980s and early 1990s ([Hussainmiya 1987, Saldin 2001](#)) and is now taken as a given by members of the community and linguists alike. While there are aspects of a language endangerment situation, the basis of this classification has not been adequately critiqued in the literature. The language shift witnessed in urban centres such as Colombo and Kandy is cited as evidence for this language loss ([Ansaldo 2008, 2006, Lim & Ansaldo 2007](#)). This perception of language loss has led many individuals from within the community and from the linguistic community overseas to take an interest in revitalising the language. One of the proposed remedies has been to teach Malay to members of the community ([Saldin 2001, 2000](#)). But the community is in a quandary as to which variety should be taught: Sri Lanka Malay or one of the 'standard' forms of Malay as used in Malaysia and Indonesia?

Within the context of the above discussion, this paper aims to discuss the implications of using the notion of 'endangerment' by members of the community as well other interested parties as a means of justifying the teaching of 'standard' Malay, and to highlight the ideologies that underlie such a move. It also examines the views of a group of five urban Malay youth on which variety of Malay they think should be taught.

While there is a growing body of literature regarding the SLM language, a quick survey of the writings reveals that the sociolinguistic aspects of the language remain seriously understudied. The first attempt to examine the phonology of the language was undertaken by [Tapovanaye \(1986\)](#) who later examined the segmental phonology and vowel length of SLM (1995). [Bichsel-Stettler \(1989\)](#) undertook a historical analysis of the linguistic development of SLM and [Kekulawala \(1979\)](#) examined the kinship terms used in SLM in relation to language universals. Adelaar traced the roots of SLM in [Adelaar \(1991\)](#) and [Adelaar & Prentice \(1996\)](#) classified the language spoken by the ancestors of the SLM community as a Pidgin-Malay Derived (PMD) variety. The dialect situation of SLM is discussed by [Robuchon \(2003\)](#). [Smith \(2003\)](#) presented evidence from nominal inflection for Tamil as the major substrate language in SLM. [Slomanson \(2003\)](#) examined the verb system of SLM and concluded that the verb in SLM has not converged with Tamil and Sinhala the same way the nominal system has. [Smith et al. \(2004\)](#) concluded that the lexifier of SLM is Indonesian in nature while the influence of the local languages is strongest in syntax and weakest in phonology. In 2006, Smith and Paauw argued that Tamil is the main source language for the TMA categories of SLM, citing linguistic as well as sociohistoric evidence. [Slomanson \(2008\)](#) discussed the morphosyntactic complexity of the SLM verb. Recently, the various aspects of the grammar of Upcountry Malay have received detailed attention from [Nordhoff \(2009\)](#). The only work of a sociolinguistic nature has been undertaken by [Ansaldo \(2006\)](#), [Lim & Ansaldo \(2006, 2007\)](#), and [Ansaldo \(2008\)](#), and will be discussed in detail in this paper. In addition, [Ansaldo \(2009\)](#) has discussed the multi-faceted nature of the linguistic repertoire as well as the identity of the Sri Lanka Malays and suggested that notions such as language ‘shift’ and ‘death’ may not apply well to multilingual communities such as the Sri Lanka Malays. Given that in present-day Sri Lanka SLM arguably attracts the most academic interest and is in the most dynamic state of all languages, it is important that scholarly attention be paid to the sociolinguistic aspects of the language. This is especially important since inter-communal and intra-communal power differentials affect the ideologies that underpin language maintenance initiatives, as will be discussed in this paper.

The nexus between language ideologies, identities, and power has been examined by many linguists ([Blackledge & Pavlenko 2004](#), [Heller 1995, 1999](#), [Hornberger 1998](#), [Irvine & Gal 2000](#), [Skutnabb-Kangas & Cummins 1998](#)). [Bourdieu \(1977, 1982, 1991\)](#) views linguistic diversity as inextricably linked to the production and reproduction of social inequality. He also views linguistic practices as a form of symbolic capital, convertible into economic and social capital, and distributed unequally within any given speech community. Influenced largely by Bourdieu’s notions, poststructuralist linguistic anthropologists have examined the complex interplay between language ideologies and power structures. Approaching the field with the consensus that “the exercise of power, in modern society, is increasingly achieved through ideology, and more particularly through the ideological workings of language” ([Fairclough 1989:2](#)), linguists have focused on studying the “contestable, socially positioned, and [political]” nature of language ideologies ([Hill & Mannheim 1992:382](#)). Informed by these views, I examine the current situation of the Sri Lanka Malay language and the political underpinnings of various language revitalisation initiatives.

In the next section, I discuss the notion of language endangerment, how it has been applied to the SL Malay context to justify various political stances, and the ideologies behind such justifications.

## 5.2 The formulation of SLM as an ‘endangered’ language

Over the last two decades, scholarly attention has been paid to the alarming loss of languages (Abley 2003, Crystal 2000, Dalby 2003, Florey 2005, Kinkade 1991, Krauss 1992, Nettle & Romaine 2000, Robins & Uhlenbeck 1991, Wurm 1996). Language shift in its simplest terms can be viewed as loss of speakers and domains of use (Romaine 2006). Language maintenance occurs when the community collectively decides to continue using the language or languages it has traditionally used (Fasold 1984). Language revitalisation involves efforts to impart new vigour in a language already in use through increasing the language’s domains, often entailing increased institutional power (Paulston 1994).

Highlighting the political nature of language loss, (May 2004:37) states that “[l]anguage loss is not only, perhaps not even primarily, a linguistic issue – it has much more to do with power, prejudice, (unequal) competition and, in many cases, overt discrimination and subordination”. So too is the case with language revitalisation, as the case of Sri Lanka Malay, discussed below, illustrates.

A few members of the SLM community (Saldin 2000, 2001, 2007b, Thaliph 2005) and non-Malay linguists (Ansaldo 2008, 2006, Lim & Ansaldo 2006, 2007) have focused on the issue of language vitality with regard to the SLM community. These discussions have been based on the assumption that Sri Lanka Malay is ‘endangered’ (Saldin 2001, Ansaldo 2006, Lim & Ansaldo 2007). The political nature of this formulation of endangerment and its consequences deserves further analysis.

In the discussion on what form language revitalization efforts should take, the voice from the community that is most frequently heard is that of B. D. K. Saldin (2001, 2000, 2007). A member of the educated urban middle-class, Saldin has for many years advocated an alignment with ‘standard’ Malay. For instance, in Saldin (2007:xii) he states:

Assuming that both the expertise and funding are available, it would perhaps be ideal to revive the literary Malay, which is not being spoken in the Malay world. However, if we do this, then the vast majority of Malays are bound to feel alienated because ‘standard’ Malay is unintelligible to most Sri Lankan Malays...

This statement illustrates Saldin’s awareness that the ‘standard’ variety of Malay is ‘unintelligible’ to the majority of the Sri Lanka Malays. It then becomes necessary to ask why a respected elder of the community advocates the teaching of ‘standard’ Malay, when, by his own admission, it is “a foreign tongue” to most Sri Lanka Malays (2007: xii).

A close examination of the ideologies that underpin Saldin’s writings reveals his negative stereotyping of the members of the community. He has positioned the Sri Lankan Malays as ‘indifferent’ to the plight of their language. He implies that it is this perceived ‘indifference’ that is responsible for the endangerment of the language and the erosion of their identity. Wishing to draw the readers’ attention to the fact that Malays should speak their language in order to be distinguished from their co-religionists, the Moors, Saldin (2001:vii) states:

The Malays in Sri Lanka have been lamenting the fact that their identity is being glossed over and being submerged by that of the Moors, and that

they are being referred to by the all encompassing term of Muslim. My personal view is that the Malays have only themselves to blame for this.

Saldin has also stated that the Sri Lanka Malay community “may not be bothered” about the future of the language:

Since the majority of Malays are more concerned with the problems of survival, they may not be bothered about doing something that would not bring them any economic gains. Therefore if one leaves it to the silent majority to revive a language nothing will come out of it. Some sort of guidance will not be out of place for those interested in preventing our language from dying altogether. (Saldin 2000:vii).

This quote reflects Saldin’s attitude to many salient issues:

1. While acknowledging Saldin’s view that the majority of the Malays have economic burdens which prevent them from focusing on issues such as language, it is necessary to point out that there exists an attitude of ‘talking down’ in Saldin’s concern for the future of the language. In the above quote, Saldin is also giving his judgment that, if left to the community to decide, “nothing will come out of it.” Yet, it is important that the community be consulted in the matter of the future of their language.
2. His view that the rural Malay community may not want to engage with their urban counterparts in working towards a solution to language loss is impressionistic and not backed by empirical research.

The above quotes illustrate instances of what Skutnabb-Kangas (1998) discusses as means by which groups are rendered invisible and are indeed invalidated through the labels used to characterise them. The practice of blaming the members of the community for what Saldin perceives as lack of commitment cannot have positive effects on the way the problem is perceived by the community and cannot be seen as encouraging – it is more likely to be disempowering. This opinion and its politics of disempowerment can have grave implications on a minority community of speakers of whom many believe they speak a ‘broken’ or ‘sub-standard’ variety of Malay.

This review of Saldin’s writings reveals the “partial” and “interest-laden” (Hill & Mannheim 1992) nature of his language ideology. It is also evidence for Kroskrity’s (2000) observation that language ideologies are constructed in the interests of a particular group. By promoting an alignment with Malaysian Malay, Saldin is serving the interests of the urban SL Malays who are most privileged by such a move. Ideologically, the members of the urban minority favour this move because it enables them to feel linked to the larger “*rumpun Melayu*” (Malay stock) and more concretely, it means they can communicate in Malaysian Malay when they travel to Malaysia and Indonesia, as they frequently do.

The community’s efforts to teach Standard Malay to its members must also be viewed in terms of the ideologies of the Malaysian government through its high commission in Colombo. The high commission conducted two courses in ‘standard’ Malay for members of the Malay community (in 2002 and 2003) and the six students who fared best in the courses were trained as language teachers in Malaysia. They were then expected to teach ‘standard’ Malay to their respective sub-communities in Sri Lanka. The next project to teach ‘standard’ Malay was in 2008 when a three-month

course was conducted by the Malaysian high commission in Colombo. These initiatives were welcomed and appreciated by the community without any critical inquiry into the possible ideological motives behind such moves and without questioning the possible consequences to the Sri Lanka Malay language. The language programmes and trips to Malaysia were made possible through the offices of the *Gabungan Persatuan Penulis Nasional Malaysia* (GAPENA): the Federation of National Writers' Association of Malaysia, which was led by Tan Sri Prof Ismail Hussein. Ismail Hussein's views on the *Dunia Melayu*, which, in his own words, was considered his 'hobby', illustrate his ideologies:

My obsession is to build a Global Malay Tribe (*Suku Melayu Dunia*). To unify Bangsa Melayu throughout the world. I have a dream that the Melayu people whose total population counts as many as 350 million at present and who reside in hundred thousands of islands, could build a single association and fraternity... (Irwan Kelana, Warta GAPENA, January 2001, cited in [Hisao \(2010:34-35\)](#))

This political view, with its goal of uniting the Malays under a common banner, is an indicator of the motives behind the teaching initiatives – to create a sense of identity with and loyalty to the Malaysian state. When he was criticised for investing resources that were better spent in developing Malaysian arts and literature, his response reveals his long-term goals:

When our country is hit by a crisis, then all the relationships and cooperations we have constructed and conducted with the international Malay societies will prove to be very relevant and important... We have seen how the relationships and cooperations with the wide *Dunia Melayu* [Malay world] could be a key for the survival of the Malay language and literature in this country itself." (Warta GAPENA, July 2003 cited in [Hisao \(2010:34\)](#).)

These views suggest a possible agenda behind the seemingly 'helpful' and 'benevolent' decision to teach Standard Malay to the Malay community of Sri Lanka. The Malays of Sri Lanka are wooed by the Malaysian government, which chose to conduct the second GAPENA conference in Colombo (1985), arranges periodic visits by representatives of the Malaysian government to the Malay Club in Colombo where grants are given for various community projects, finances occasional trips to Malaysia for members of the urban segment of the community to attend conferences and seminars at the expense of the Malaysian state etc. The fact that the SL Malay community's vulnerable minority status and its weak political position are possibly being exploited to propagate a Pan-Malay ideology is something the community is either ignorant of or chooses to overlook. Highly relevant to this argument is the fact that the beneficiaries of the 'generosity' of the Malaysians are the most forceful advocates of teaching Standard Malay. Some members of the community, when they have an occasion to address the Malaysian or Indonesian officials, have appealed to them to help their Malay brethren, thereby probably justifying the 'reaching out' the Malaysians do through their linguistic and cultural activities. For instance, in his address to the Malaysians at the International Symposium on Malayo-Polynesians in the Commonwealth held in Malaysia in 1998, one member of the SLM community asked

[our] motherlands – Indonesia and Malaysia – to direct their *benevolent* thoughts (faidat-ul-fikr) towards their *hapless protégés* – the Sri Lanka

Malays – in initiating *beneficial actions* (amal-ul-khair) for maintaining *good feelings of humanity* (barakat-ul-insan) and effective cultural rapport . . .” (Thaliph 1998, p. 5), (italics mine)

Such ideas, which are expressed in a language appealing to the Malaysians’ religious sense (through the use of Arabic terminology) as well as emotional sense (as highlighted in the italicised terms), place the Malaysians unquestioningly in a position of power and cast the SL Malays as their ‘powerless’ brethren.

In my interviews with the Malays from various parts of the country, I only heard one opinion that was openly critical of the Malaysian initiatives:

I understand the reason why the Malaysian High Commission would be keen on conducting these classes and trying to teach us Malays to study their language: because of [their idea of making] Bahasa Melayu, Bahasa Dunia (INT/WP/12).

This opinion, also from a member of the urban elite, stands in stark contrast to the views of the rest of the community, although it is quite possible that there may be similar opinions which are not stated publicly. While I acknowledge my own subjective reading of the events and comments that are described in this section, what needs to be emphasised is that there is little critical analysis of the power dynamics that underlie what form language revitalization efforts should take. From a critical ideological perspective, teaching ‘standard’ Malay to the community is a tacit acknowledgment and affirmation of the unequal power distribution between the two communities. Interestingly, I was informed by the member of the community who was liaising between the Malaysian officials and the SLM community when the classes were being organised that all offers to help ceased when the Malaysians understood that the SLM community was at a crossroads and wondering whether it should continue to teach ‘standard’ Malay or promote SLM in its stead.

The Indonesian government does not seem to have similar motives when it offers to help maintain the Malay language in Sri Lanka through its representatives in its embassy in Colombo. The Indonesian language classes that are conducted by the embassy are open to all Sri Lankans, regardless of ethnic origin, unlike the classes held by their Malaysian counterparts, which are exclusively for Sri Lanka Malays. The difference in ideologies has been commented on by [Tirtosudarmo \(2004\)](#) who states: “[w]hen it comes to the notion of ‘Malayness’ . . . the Malaysians are much more assertive than the Indonesians” (2004, p. 2).

Having explored the ideologies behind the views espoused by Saldin and the initiatives of the Malaysian high commission, I turn to an examination of how researchers into the SLM language have perceived the language loss.

[Ansaldo \(2008\)](#) states that SLM varieties “are currently endangered as they are no longer spoken by the younger generation, with one exception, the community in Kirinda” (2008, p. 14) and he states that this is due to the fact that the younger generation is abandoning the language because of their desire to converge with the dominant languages of the country and thus avoid being disadvantaged. With particular reference to the situation in Colombo, [Lim & Ansaldi \(2006\)](#) state that “SLM in [the Colombo Malay] community is no longer a home language for the younger generation and is thus considered an endangered language” (2006, p. 3). As is evident from these statements, the criterion seems to be the maintenance of the language by the younger generation. However, [Himmelmann \(2005\)](#) points out the need to differentiate between the symptoms and causes of endangerment, and it seems that in this context, the reduction in

Community	Characteristics	Vitality
Colombo	Middle-upper class; often bi- or trilingual (Tamil/Sinhala) standardising in Malay; restricted use of SLM; English fairly fluent.	Endangered; no SLM in younger generation.
Slave Island (Colombo)	Lower class; most Tamil influenced; bi- or trilingual; no English.	Very endangered; use of SLM discouraged.
Kandy (and Hill Country)	Similar to Colombo community; weak standardisation forces.	Endangered.
Hambantota	Traditionally heavy Sinhalese-speaking area; low-middle class, often trilingual; limited English.	Mildly endangered.
Kirinda	Lower class; good trilingual competence in middle-younger generations; English limited to a few individuals.	Fully vital; mother tongue even in present generation.

Table 5.1: SLM speech communities (from [Ansaldo 2008](#)).

the number of children speaking SLM could be a symptom for which the causes are to be found elsewhere. Himmelmann states that language endangerment may be defined as “a rapid decline in the number and quality of domains in which a given language is used” (2005, p. 3). Therefore, I argue that the notion of ‘endangerment’ should be problematised to include an examination of domains of use, attitudes to the language, and possibly more criteria.

Ansaldo presents his assessment of the vitality of the language in an overview of the SLM speech communities, reproduced in Table 5.1.

Generalisations based on locations and age groups are a necessary and useful point of departure for studies of this nature. However, they need to be questioned in order to avoid essentialism and to move towards a more nuanced view of the current situation. While acknowledging the potential helpfulness of social class as a category along which endangerment is determined, it is useful to delve deeper into what actually constitutes social class. Similarly, the employment of labels such as ‘endangered’ and ‘very endangered’ can be problematic when they are used without adequate definition. In summary, it is felt that the categorisation of the SLM speech communities and their features can benefit from a more critical and detailed examination.

Having reviewed the opinions of the urban minority, the Malaysian officials, and the researchers, it is necessary to reflect on how the ideologies of these ‘powerful’ agents potentially affect the larger SLM community.

The majority of the Sri Lankan Malays possess little socioeconomic power and live away from the urban centres of the country. Generally speaking, they are also less proficient in English and, in a country where English has huge social and economic capital, have lower social standing as a result. Using Bourdieu’s (1977, 1982, 1991) notion that linguistic practices are a form of symbolic capital, which is convertible into

economic and social capital, it can be pointed out that the majority of the SLM population are doubly disadvantaged: on the one hand, they are part of a minority community whose linguistic (and other) interests are largely overlooked by the central government, and on the other, they are dominated by a more socioeconomically powerful minority within their own community.

Bourdieu's model of symbolic domination rests on his notion that the dominated group is complicit in the misrecognition, or valorisation, of one language or variety as an inherently better form than another. It is possible that, after many years of having their variety of Malay devalued by the dominant segment of the Sri Lanka Malay society, the dominated group has misrecognised their form of Malay as inferior to 'standard' Malay.

The nexus between language and power becomes crucial to the question of the language ideologies of the SLM community. Heller (1982) links language and power in two significant ways. On the one hand, language is seen as part of the processes of social action and interaction and in particular, as a way in which people influence others. On the other, it is a symbolic resource which may be tied to the ability to gain access to, and exercise, power. In the case of the SL Malay community, the dominant minority exert power over the rest of the community through two languages: the first is English, the knowledge of which is seen as symbolically opening the doors to privilege and wealth, and the second is 'standard' Malay, the knowledge of which connects the users to the larger Malay world of the East.

It has been observed that "in the most restrictive formulations of this connection, ideology is always the tool, property, or practice of dominant social groups; the cultural conceptions and practices of subordinate groups are, by definition, nonideological" (Woolard 1998:7). I believe that the field of SLM studies can benefit from a questioning of this position as it is possible that subordinate groups might have an alternative ideology, but it does not get voiced due to the entrenched dominance of the more powerful minority. Bucholtz & Hall (2004) state that "speakers who resist, subvert, or otherwise challenge existing linguistic and social norms are vital to the theoretical understanding of identity as the outcome of agency..." (p. 373).

### 5.3 The language–identity nexus in the SLM context

Recent work in language and identity in the fields of sociolinguistics and linguistic anthropology has emphasised the need for the examination of identity as a social and linguistic construction (Bucholtz & Hall 2004, Joseph 2004). There has also been emphasis on the multiple nature of identity – for instance, Joseph (2004) posits that one's identity shifts according to the context one finds oneself in and also that it there are many versions of one's identity based on other people's construction of it.

While some linguists view language and ethnic identity as inextricably linked (Fishman 1991, Joseph 2004, Pozzetta 1991), more recent work has questioned the assumed relationship between the two. Within sociolinguistics, sociology, and the anthropology of ethnicity, there is a growing consensus that language is at most only a contingent factor of one's ethnicity. Edwards (1985, 1994, 2001) and Eastman (1984) assert that language is often only a surface or secondary characteristic of ethnicity. Barth (1969) cautions us against assuming a simple one-to-one relationship between ethnic units and cultural similarities or differences such as language. In other words, the specific linguistic community decides how salient language is as a marker of ethnicity: "it is the *perceived* usefulness of these cultural attributes in maintaining ethnic boundaries

which is central” ([May 2004](#):40). The above notion seems to suggest that there is nothing intrinsic to one’s ethnicity and thus specifically rejects any automatic link between ethnicity and language. This view is also in keeping with the postmodern/poststructuralist trend of viewing all forms of identity as multiple, shifting, contingent, and hybrid. But that does not mean it will never be a significant or constitutive factor of identity. In theory, language may well be just one of many markers of identity. In practice, it is often much more than that since the language-identity link encompasses significant dimensions that are both cultural and political. One’s individual and social identities, and their complex interconnections, are inevitably mediated in and through particular languages. The political dimension is significant because languages formally (and informally) become associated with particular ethnic and national groups.

My inquiries into the role of language as a marker of ethnic identity within the SLM community have been framed by the academic debates mentioned above. During the course of the fieldwork I conducted for my doctoral thesis from January to July 2010, I learned that all segments of the SLM community perceive that language is intrinsic to one’s identity. This might possibly explain how the language has survived this long. Another reason for the community’s attachment to the language could be the heightened awareness of ethnicity (and language as a marker of ethnic identity) which came about due to the inter-ethnic tensions and conflict in the island state. The reasons for the conflict are also largely though not exclusively language-based: since the implementation of the Official Language Policy in 1956, the language policies of Sri Lanka are perceived by minority communities as privileging the Sinhalese. It has been observed that “language, functioning as a marker of identity and ethnic group interest, has played a major catalytic role in the generation of barriers to national unity and the peaceful development of post-independence Sri Lanka” ([Dharmadasa 2007](#):116).

The present day Sri Lanka Malay community has organised itself into many “Malay Associations” under the Sri Lanka Malay Confederation (SLAMAC), and the maintenance of the SL Malay language is high on the agenda of these organisations. Many associations have conducted language classes in the past (Mabole Malay Association, Sri Lanka Malay Association, Kandy Malay association, to name a few) and they have chosen to teach Malaysian Malay, usually by teachers who were trained by the Malaysian government. However, for some of the newer organisations, it is a pressing issue which variety of Malay should be taught. For instance, the Women’s Association of Sri Lanka Malays (WASLAM), the newest social, cultural and religious organisation formed by Malay women (established on 4th April 2010), has stated in its constitution that one of its goals is “to propagate the use of the Malay language”. The chairperson of the committee for Islam and Malay language states: “it has not been specified in the WASLAM constitution that Malay language shall be ‘Sri Lanka Malay language’ but, as a matter of policy, the Board of Management will go along with Sri Lanka Malay” (p.c.). WASLAM is also in the process of organising a seminar “to arrive at a consensus as to which kind of Malay should be encouraged” (*ibid*). This is evidence of a more questioning attitude to the issue of language maintenance than before.

The issue of how Sri Lankan Malay identity is being (re)negotiated through the community’s linguistic preferences has been the subject of some of Ansaldo and Lim’s academic writing ([Ansaldo 2008, 2006, Lim & Ansaldo 2007](#)). While acknowledging the varied nature of the linguistic repertoire and the communicative practices of the Sri Lanka Malay community as a whole, [Lim & Ansaldo \(2007\)](#) discuss the fact that the Malays of Colombo in particular are facing a situation of language endangerment. This observation has been made previously by many others including [Saldin \(2000, 2001, 2007\)](#), [Hussainmiya \(1987, 1990, 2008\)](#) and [Smith et al. \(2004\)](#). It is an accurate reflec-

Perceived language loss (among urban Malays)  $\Rightarrow$  Desire to revitalise language  $\Rightarrow$  Desire to learn/teach Malaysian Malay  $\Rightarrow$  Identity re-alignment (for all Malays)

Figure 5.1: Path from perceived language loss to identity re-alignment.

tion of only one segment of the Colombo Malay population, albeit the most powerful one. My research has shown that in suburban areas around Colombo – for instance, Wattala, Hunupitiya, Mabole, Makola, Kolonnawa, and Mattegoda – at the very least – there are sub-communities of Malays who *do* speak SLM at home. This observation should lead us to reassess our decision to classify sub-communities and their linguistic range solely by geographical location and move towards an inquiry of social class as at least an additional (if not primary) determiner of language maintenance or shift, a process already begun by [Ansaldo \(2008\)](#). The definition of what constitutes a social class is notoriously problematic, but one that academics need to grapple with if they wish to arrive at a more nuanced understanding of perceived language loss. It must be emphasised that what is being advocated here is not a rejection of geographical location as a determinant of language vitality but a recommendation that geographical location (including the urban-rural divide) and social class (including but not limited to wealth, education, employment, access to cultural capital) be combined when determining the language vitality of specific sub-communities.

This leads me to the issue of who provides the linguists with the data that their writings are based on. The awareness that the ‘gate-keepers’ of a community are usually members of a more powerful segment of the community has been observed. Therefore the choice of community members the linguist has access to can be influenced by the “partial” and “interest-laden” ([Hill & Mannheim 1992](#)) view of the ‘gate-keeper’, leading even the most well-meaning scholars to make generalisations based on the minute sub-section of the community they have access to. As a result, what the linguists might advocate could be in the interests of that sub-section of the community and not necessarily be in the best interests of the community as a whole.

I wish to illustrate the above point using Lim and Ansaldo’s (2007) assessment of the language loss witnessed among the Colombo Malays. Based on their observations of this sub-community, [Lim & Ansaldo \(2007\)](#) state:

it is with Malaysia that the SLM community align themselves, both in terms of language and identity, and the choice in the revitalization process is consequently not for Sri Lanka Malay but for Malaysia’s Standard Malay.” ([Lim & Ansaldo 2007:233](#))

This is a synecdochic situation where a part of the community (the Colombo Malays) is made to represent the whole (all Sri Lanka Malays) and can be illustrated as in Figure 5.1

The academic argument supporting re-alignment with Malaysian Malay is summarised below as it plays a crucial role in intellectualising the shift and thus possibly serving to justify it. [Lim & Ansaldo \(2006\)](#) suggest that the acquisition of a more global identity by the SLM community be viewed as positive agency by researchers. This is done by citing arguments of linguistic citizenship, where “the community is served by its linguistic resources – which comprise negotiable, multiple, diverse and shifting identities – and is not restrained by its language” (2006, p. 6). Lim and Ansaldo state that, viewed from a linguistic human rights paradigm, the case of SLM would be viewed as one of loss of language diversity and forsaking of a unique identity.

But they point out that the linguistic human rights paradigm has been criticized in recent years for “endors[ing] an ethno-linguistic stereotyping in the form of monolingual and uniform identities” (2006, p. 5). They also state that the linguistic human rights paradigm forces groups of speakers to work actively to differentiate themselves by claiming unique linkages of language and identity so as to gain political leverage in the competition for scarce resources. Alternatively, Lim and Ansaldo recommend that the case of SLM be viewed from a linguistic citizenship perspective, where language is viewed at the same time as a semiotic resource for the (re)construction of agency and self-representation, an economic resource, and site of political and economic struggle, a global resource to address local–global concerns, and an intimate resource as the foundation of respect for difference on a global level.

Influenced by current discourses in linguistic citizenship ([Freeland & Patrick 2004](#), [May 2004](#), [Stroud & Heugh 2004](#)), [Lim & Ansaldo \(2007\)](#) propose a new interpretation of the traditional notion of ‘language shift’ – instead of the traditional idea that the shift away from a mother tongue would lead to the loss of a crucial part of the community’s identity, Lim and Ansaldo suggest that this shift be viewed more as positive agency on the part of the community. Arguing from a stance which states that the identity of the SL Malay is defined by being multilingual, [Lim & Ansaldo \(2007\)](#) also state that a potential shift from SLM to Standard Malay “does not make a qualitative difference to the Malays’ multilingual repertoire, nor to the identity they have” (p. 225). They are of the view that Standard Malay still fulfills the same function as that of SLM, of identifying them as Malay (both in SL and in the wider Malay world) and that this can be viewed as identity alignment coming full circle: this is the region that constitutes the SL Malays’ origins and the SL Malays of the previous generations “always used to talk about going home” (Salma Suhood Peiris, 2006, cited in [Lim & Ansaldo \(2007:225\)](#)).

Aside from the observation that this argument is based on what is viewed only in one sub-community (which was discussed earlier), I wish to comment on some of the researchers’ opinions.

Though there is the perception that there could be economic benefits from learning ‘standard’ Malay, this is a largely misplaced notion: there have been no employment opportunities created as a result of ties with the Malaysian government and even if such openings were to exist in the future, they are most likely to benefit only the sub-community ‘closest’ to the Malaysians – the urban, Colombo-based Malays. As stated in Section 5.1, any other opportunities that have been granted (scholarships, attendance at seminars and conferences held in Malaysia etc) have also been enjoyed by the urban Malays. In my interactions with the rural Malays, a few participants showed me that they were aware of the fact that such ‘perks’ were enjoyed by their urban counterparts, and that they were resentful of it.

Considering that community members who are most likely to acquire ‘standard’ Malay will be from the urban minority, Ansaldo and Lim’s recommendation would exacerbate an already existing class division within the community. By making the members of the urban middle class who are proficient in ‘standard’ Malay the ‘authorities’ of the language, such a move can only reinforce the power dynamics that already exist within the community.

Lim and Ansaldo state that a shift to ‘standard’ Malay “does not make a qualitative difference to the Malays’ multilingual repertoire” because the community’s identity is defined being multilingual and adding one more language will not change that. In contrast, I am of the opinion that such an alignment *will* make a difference to “the identity they have” by adopting a language that is not used in speech or in writing by anyone in Sri Lanka, save the proponents of such alignment (who tend to identify with

Malaysian Malays). In order to illustrate this point, I draw on the findings of a focus group discussion conducted in 2010.

## 5.4 Alternative views

In the next section I provide excerpts from a focus group discussion that was conducted as part of my fieldwork amongst five Malay youth from the suburbs of Colombo whose ages ranged from 19 to 23. They can be considered as representing an urban middle class youth segment of the Malay population in terms of the access they have to economic, social and cultural capital (Bourdieu 1977). While three of the participants are working at junior management level in the corporate sector, two are studying in private university colleges which are affiliated to universities overseas. The discussion was conducted in English as all five participants speak English as their home language. The purpose of highlighting the views of these members of the community is to examine the range of opinions that exists within one supposedly similar demographic. Further, it sheds light on how young people of the community view the link between their language and their identity and I feel that the discussion on the future of the language will benefit from including the opinions of youth members of the community.

I am very comfortable with Sri Lankan Malay and that ought to be the language we learn first, mainly because we are currently living in Sri Lanka and there is a need to speak Sri Lankan Malay more than any other form of Malay. (FG/ WP 05)

You don't want to learn Malaysian or Indonesian Malay – you want to learn our own Malay – Sri Lankan Malay. Sri Lankan Malay personally for me is a unique language compared to Malaysian Malay and Indonesian Malay. (FG/ WP 01)

These two comments underscore the perception among the youth that their identity as Sri Lankans is what defines them and is opposed to the emphasis on Pan-Malay identity that is witnessed especially among the older generation. This may be because, unlike the older Malays who lived through the change in language policy and have a sense of being increasingly marginalized politically, the Malays of the 18 – 25 age group have grown up with an identification of Sri Lanka as home. I believe that the sense of patriotism that has been advocated implicitly and explicitly in the public sphere in the wake of the conflict between the Liberation Tigers of Tamil Eelam (LTTE) and the Sri Lankan Armed Forces could also be one of the reasons for the importance given to a Sri Lankan identity. In this regard, it is useful to recall the statement made by Sri Lankan president Mahinda Rajapakse after the military defeat of the LTTE in May 2009:

We [have removed] the word minorities from our vocabulary three years ago. No longer are the Tamils, Muslims, Burghers, Malays and any others minorities. There are only two peoples in this country. One is the people that love this country. The other comprises the small groups that have no love for the land of their birth. Those who do not love the country are now a lesser group.”<sup>3</sup> (Rajapakse 2009)

<sup>3</sup>This statement by the president has also been interpreted as a denial of the rights of the minority communities of Sri Lanka. Some critics have voiced concerns that this speech may justify the continued domination of the country by the Sinhalese, and be thus “evidence of a majoritarian mindset” (Ismail 2009).

Influenced by this and similar political rhetoric superficially promoting ‘social cohesion’ and ‘inclusiveness’, it is possible that the present generation of Sri Lankan youth are (re)negotiating their ethnic and linguistic identities in relation to their national identity and their wish to align themselves with a Sri Lankan form of Malay is a reflection of this evolution.

Providing a variation of the same theme is this opinion by another member of the focus group:

We should be proud of the language we speak which is Sri Lankan Malay and not Malaysian or Indonesian Malay, because our grandparents they’ve spoken Sri Lankan Malay why not we continue the same language? (*FG/WP 02*)

In contrast to many members of the community who have harked back to Indonesia and Malaysia as ‘our motherlands’ (*Thaliph 1998:5*), this participant feels his grandparents, and through them his history, is rooted in Sri Lanka. It is possible that to the present generation, the fact that the Sri Lankan Malays originated from what is present day Indonesia and Malaysia may mean less than the fact that all their associations with the past are rooted in Sri Lanka.

These opinions have to be contrasted with those of older urban Malays who state that they feel embarrassed when they speak SLM. But this embarrassment is only when they speak with Malaysian or Indonesian dignitaries they meet in Colombo or Malaysians or Indonesians they meet during their travels to these two countries. For instance, a member of the urban Malay community who has regular contact with Malaysians and Indonesians stated:

I feel a little ashamed to – I’m being frank with you – I feel a little ashamed because I feel the language is not..., even though it is okay for us, we communicate. (*IN/WP/09*)

Later in the interview, the same participant states:

... in the presence of Indonesians and Malaysians or Malays, I feel a little uncomfortable because I think the language is not up to ... maybe it is not grammatical because I have seen people when they hear us speaking, the smiles. You get what I mean? You get this amused look which is not an amusement to us. I’m being very frank. (*IN/WP/09*)

The hesitation to actually verbalise the extent of his embarrassment and the actual reasons for it (as seen by the incomplete sentences) could be due to a perception that it might be ‘disloyal’ to criticize SLM but could also be an unwillingness to engage with me as the interviewer on the perceived ‘incorrectness’ of SLM in relation to Standard Malay. This notion that SLM may not be a ‘proper’ language is commonly held among many segments of the community as it is among minority or marginalized linguistic communities everywhere. Linked to this is the idea that SLM is a ‘creole’ with the layman’s understanding of the word ‘creole’ as signifying a ‘lesser’ language. One opinion which had a significant impact on the way the language is viewed by the SLM community was expressed in 1986:

The present SLM, although perceived by the local Malays as ‘Bahasa Melayu’, is but a heavily creolised language and therefore widely divergent from the standard Malay spoken in the Malay peninsula and the archipelago. (*Hussainmiya 1987:153*)

This classification and the accompanying implication of speaking a ‘corrupt’ or ‘broken’ variety of the language has been detrimental to the way the language has been viewed by the community and “has a significant impact on the type of [language] shift that may occur as well as its speed...” ([Lim & Ansaldi 2007:220](#)). As illustrated in the case cited above, this perception exists in the mind of the community members as a rationale for moving towards a ‘better’ language.

Returning to the opinions of the participants of the focus group, one is made aware of how multilingualism is perceived by the youth.

And even if you do learn Standard Malay I don’t think it would be a hindrance to Sri Lankan Malay – it’s kind of like knowing how to speak English, Sinhala and Malay. (*FG/ WP 05*)

This quote reflects the participant’s attitude towards the acquisition of multiple languages. It possibly highlights the reason that has made the Malays be regarded as one of the most multilingual communities in Sri Lanka ([Lim & Ansaldi 2007](#), [Saldin 2001](#), [Vijaycharya 2004](#)): their actual multilingualism as well as their positive attitude towards learning many languages. This participant’s opinion also reveals that he views Standard Malay and SLM as two separate languages, which is why he speaks of them alongside English and Sinhala.

We don’t really have the materials to learn our own language – I mean there are no text books. No kind of dictionary of any sort. The only way we can learn is from parents, grandparents – people who already know the language... it would be better if we had some kind of materials for Sri Lankan Malay.

*(FG/ WP 01)*

Similar opinions are frequently heard among Malays from all age groups and sub-communities. This is possibly due to the perception that SLM is a ‘lesser’ variety because it does not exist in written form. Regarding this view, it is useful to cite Kloss’ (1978) discussion of the factors that determine if a given idiom qualifies as a ‘language’. He states that a variety can be termed a ‘language’ if it demonstrates considerable ‘Abstand’ and ‘Ausbau’. Abstand or structural difference is the amount of disparity the variety shows in relation to other ‘languages’ under which it may be subsumed. Kloss defines Ausbau as “language by development” and points out that the ‘language’ in question needs to have developed into “tools for qualified purposes or spheres of application” ([Kloss & McConnell 1978:25](#)). Joseph has included “publication, education and other functions associated with the more prestigious realms of Western culture” (2004, p. 3) as some of the factors that help create Ausbau. Applying this principle to the Sri Lanka Malay context, it is clear that SLM can be considered a language different from ‘standard’ Malay by the Abstand factor because it has now been established that in terms of structure SLM has diverged considerably from the Malay varieties to which it traces its origins ([Adelaar 1991](#)). However, SLM does not demonstrate sufficient Ausbau to qualify as a language in its own right because it lacks the ‘official’ status of having developed tools for education and communication. This is why it is heartening to note that members of the community and linguists are uniting to work towards developing teaching and literary material in SLM. One example of such an effort was the organisation of a seminar entitled “Symposium on the

Mother Language of the Sri Lanka Malays” in Colombo, Sri Lanka in May 2011.<sup>4</sup> At this forum, linguists working on various aspects of the SLM language were invited by the Women’s Association of Sri Lanka Malays (WASLAM) and the Confederation of Sri Lanka Malays (COSLAM) to present their opinions on the preservation of the spoken and written varieties of SLM. As a follow-up to this initiative, WASLAM is currently in the process of working with other Malay organisations throughout the island towards formulating means of developing teaching and literary resources in SLM. Further, a short children’s story has been translated into SLM and its narration has been uploaded onto YouTube.<sup>5</sup> It is hoped that these and similar efforts will be successful in revitalizing a variety of Malay which is viewed as a crucial marker of identity by the members of the community and is of value to academic discussions surrounding extreme language contact.

## 5.5 Summary

The Sri Lanka Malay community has managed to retain its language for the last 350 years despite facing many challenges. The fact that SLM has survived these pressures from other larger, more economically strong, more socially acceptable languages points to the crucial role played by language in defining identity for this community. This was confirmed by most of the participants of my study who felt that language was a crucial marker of their ethnic identity.

However, the members of the community seem to be divided as to which identity should be given priority – the Pan-Malay identity or the Sri Lanka Malay identity, and this difference of opinion drives much of the current debate surrounding SLM. While some members of the urban majority feel that ‘standard’ Malay should be taught, I have pointed out the power dynamics that possibly underpin such a discourse. Using data from a focus group discussion with young Malay participants, I also discussed the views of the younger generation regarding language revitalization and point out that the youth seem to desire a more Sri Lankan identity and hence show signs of wanting to preserve Sri Lanka Malay.

---

<sup>4</sup>More details of this seminar can be viewed at <https://sites.google.com/site/symposiumonslmalay/>

<sup>5</sup> <http://www.youtube.com/watch?v=VuVBjrIVjk8>

# Bibliography

- Abley, M., 2003. *Spoken here. Travels among threatened languages.* Toronto: Random House of Canada.
- Adelaar, K. A., 1991. "Some Notes on the Origin of Sri Lanka Malay." In *Papers in Austronesian Linguistics*, edited by H. Steinhauer. Canberra: Australian National University, 23–37.
- Adelaar, K. A. & Prentice, D. J., 1996. "Malay: its History, Role and Spread." In *Atlas of Languages of Intercultural Communication in the Pacific, Asia and the Americas*, edited by Stephen A. Wurm, Mühlhäusler Peter, & Darrell T. Tryon. Berlin: Mouton de Gruyter, 673–693.
- Ansaldo, L., U.and Lim, 2006. "Globalization, empowerment and the periphery: the Malays of Sri Lanka."
- Ansaldo, Umberto, 2008. "Sri Lanka Malay revisited: Genesis and classification." In *A world of many voices: Lessons from documented endangered languages*, edited by Arienne Dwyer, D. Harrison, & David Rood. Amsterdam, Philadelphia: John Benjamins., 13–42.
- Ansaldo, Umberto, 2009. *Contact languages, Ecology and Evolution in Asia.* Cambridge: Cambridge University Press.
- Barth, F., 1969. "Introduction." In *The Social Organisation of Culture Difference.* Boston,MA: Little Brown & Co., 9–38.
- Bichsel-Stettler, Anne, 1989. *Aspects of the Sri Lanka Malay Community and their Language.* Master's thesis, University of Bern.
- Blackledge, A. & Pavlenko, A. (eds.), 2004. *Negotiation of identities in multilingual contexts.* Clevedon; Buffalo: Multilingual Matters.
- Bourdieu, P., 1977. "Outline of a theory of practice."
- Bourdieu, P., 1982. *Ce que parler veut dire : l'économie des échanges linguistiques.* Paris: Fayard.
- Bourdieu, P., 1991. *Language and symbolic power.* Cambridge: Polity Press in association with Basil Blackwell.
- Bucholtz, M. & Hall, K., 2004. "Language and Identity." In *A Companion to Linguistic Anthropology*, edited by A. Duranti. Maddem, MA;Oxford,UK: Basil Blackwell, 369–394.
- Crystal, D., 2000. *Language death.* Cambridge: Cambridge University Press.
- Dalby, A., 2003. *Language in danger. How language loss threatens our future.* Harmondsworth: Penguin.
- Deumert, Ana & Durrelman, Stephanie (eds.), 2006. *Structure and Variation in Language Contact.* Amsterdam: John Benjamins.
- Dharmadasa, K. N. O., 2007. "Sri Lanka." In *Language and National Identity in Asia*,

- edited by A. Simpson. Oxford, New York: Oxford University Press, 116–138.
- Eastman, C., 1984. “Language, Ethnic Identity and Change.” In *Linguistic Minorities, Policies and Pluralism*, edited by J. Edwards. London: Academic Press, 259–276.
- Edwards, J., 1985. *Language, Society and Identity*. Oxford: Basil Blackwell.
- Edwards, J., 2001. “The Ecology of Language Revival.” *Current Issues in Language Planning* 2:231–241.
- Edwards, John, 1994. *Multilingualism*. London and New York: Routledge.
- Fairclough, Norman, 1989. *Language and Power*. London: Longman.
- Fasold, R., 1984. *The sociolinguistics of society*. Oxford, New York: Blackwell.
- Fishman, J., 1991. *Reversing language shift*. Clevdon, Avon: Multilingual Matters.
- Florey, M., 2005. “Language shift and endangerment.” In *The Austronesian Languages of Asia and Madagascar*, edited by K. Alexander Adelaar & Nikolaus P. Himmelmann. London, New York: Routledge, 43–64.
- Freeland, J. & Patrick, D., 2004. “Language Rights and Language Survival: Sociolinguistic and sociocultural perspectives.” In *Language Rights and Language Survival*, edited by J. Freeland & D. Patrick. Manchester; Northampton: St. Jerome Publishing, 1–33.
- Heller, M., 1982. “Negotiations of language choice in Montreal.” In *Language and social identity*, edited by J. Gumperz. Cambridge, New York: Cambridge University Press, 108–118.
- Heller, M., 1995. “Choice, Social Institutions and Symbolic Domination.” *Language in Society* 24:373–405.
- Heller, M., 1999. *Linguistic Minorities and Modernity: A Sociolinguistic Ethnography*. London, New York: Longman.
- Hill, J. H. & Mannheim, B., 1992. “Language and World View.” *Annual Review of Anthropology* 21:381–406.
- Himmelmann, N. P., 2005. “Language endangerment scenarios: A case study from Northern Central Sulawesi.” In *Endangered languages of Austronesia*, edited by Margaret Florey. Oxford: OUP, 45–72.
- Hisao, T., 2010. “Old and new aspects of Malayness in the contemporary Dunia Melayu movement.” In *Tinta di Dada Naskhah*, edited by H. Ismail. Kuala Lumpur: Universiti Melayu, 29–44.
- Hornberger, N. H., 1998. “Language policy, language education, language rights: Indigenous, immigrant and international perspectives.” *Language in Society* 27:439–458.
- Hussainmiya, B. A., 1987. *Lost cousins: The Malays of Sri Lanka*. Bangi: Institut Bahasa, Kesusastraan dan Kebudayaan Melayu.
- Hussainmiya, B. A., 1990. *Orang Rejimen: The Malays of Sri Lanka*. Bangi: Universiti Kebangsaan Malaysia, 1st edition.
- Hussainmiya, B. A., 2008. *Orang Rejimen: The Malays of Sri Lanka*. Bangi: Universiti Kebangsaan Malaysia, 2nd edition.
- Irvine, J. T. & Gal, S., 2000. “Language ideology and linguistic differentiation.” In *Regimes of Language*, edited by P. V. Krook. Santa Fe, NM: School of American Research Press, 35–83.
- Ismail, Q., 2009. “Critiquing the President’s Victory Speech.” Retrieved November 24, 2009.  
URL <http://groundviews.org/2009/08/20/>

- [critiquing-the-presidents-victory-speech-evidence-of-an-majoritarian-mindset/](#)
- Joseph, J. E., 2004. *Language and identity: national, ethnic, religious*. New York, NY: Palgrave Macmillan.
- Kekulawala, S. L., 1979. "Kinship terminology in Sri Lankan Malay - a contribution to the study of language universals." *Kalyani* 1:207 – 224.
- Kinkade, M. D., 1991. "The decline of native languages in Canada." In *Endangered Languages*, edited by R. H. Robins & E. M. Uhlenbeck. London: Berg Publishers Ltd., 157–176.
- Kloss, H. & McConnell, G. D., 1978. *The written languages of the world: A survey of the degree and modes of use - 1: The Americas; Les langues érites du monde: Relevé du degré et des modes d'utilisation*. Quebec: Les Presses de l'Université Laval.
- Krauss, M., 1992. "The world's languages in crisis." *Language* 68:4–10.
- Kroskrity, P. V., 2000. "Introduction: Regimenting Languages: Language Ideological Perspectives." In *Regimes of Language: Ideologies, Polities and Identities*, edited by P. V. Kroskrity. Santa Fe, NM: School of American Research Press, 1–34.
- Lim, Lisa & Ansaldi, Umberto, 2006. "Keeping Kirinda vital: The endangerment-empowerment dilemma in the documentation of Sri Lanka Malay." *Amsterdam Centre for Language & Cognition Working Papers* 1:51–66.
- Lim, Lisa & Ansaldi, Umberto, 2007. "Identity alignment in the multilingual space: The Malays of Sri Lanka." In *Linguistic identity in multilingual postcolonial spaces*, edited by E. Anchimbe. Cambridge: Cambridge Scholars Press, 218–243.
- May, S., 2004. "Rethinking Linguistic Human Rights: Answering questions of identity, essentialism and mobility." In *Language Rights and Language Survival*, edited by J. Freeland & D. Patrick. Manchester;Northampton: St. Jerome Publishing, 35–53.
- Nettle, D. & Romaine, S., 2000. *Vanishing voices. The extinction of the world's languages*. Oxford: Oxford University Press.
- Nordhoff, Sebastian, 2009. *A Grammar of Upcountry Sri Lanka Malay*. Ph.D. thesis, University of Amsterdam, Utrecht.
- URL <http://www.lotpublications.nl/publish/articles/003744/bookpart.pdf>
- Paulston, C. B., 1994. *Linguistic minorities in multilingual settings: implications for language policies*. Amsterdam, Philadelphia: J. Benjamins Pub. Co.
- Pozzetta, G. E., 1991. *Ethnicity, ethnic identity and language maintenance*. New York, NY: Garland Publishers.
- Rajapakse, M., 2009. "Address by HE President Mahinda Rajapakse at the ceremonial opening of Parliament." Retrieved March 03, 2011.
- URL [http://www.president.gov.lk/speech\\_New.php?Id=74](http://www.president.gov.lk/speech_New.php?Id=74)
- Robins, R. H. & Uhlenbeck, E. M., 1991. *Endangered languages*. Oxford: Berg.
- Robuchon, Gérard, 2003. "Malayo Language in Sri Lanka." Paper presented at the 7th International Symposium on the Malay/Indonesian Linguistics in Berg en Dal.
- Romaine, S., 2006. "Planning for the Survival of Linguistic Diversity." *Language Policy* 5:441–473.
- Saldin, B. D. K., 2000. *A guide to Malay*. Colombo: Nihon Printers.
- Saldin, BDK, 2001. *The Sri Lankan Malays and Their Language*. Kurunegala: Nihon, 2nd edition.
- Saldin, BDK, 2007a. *A Concise Sri Lankan Malay Dictionary*. Kurunegala: Nihon

- Printers.
- Saldin, BDK, 2007b. *A concise Sri Lankan Malay Dictionary – Kamus ringkas bahasa melayu Sri Lanka*. Kurunegala: Nihon.
- Skutnabb-Kangas, T., 1998. “Human Rights and Language Wrongs - A future for diversity?” *Language Sciences* 20:401–426.
- Skutnabb-Kangas, T. & Cummins, J., 1998. *Minority education: from shame to struggle*. Clevedon, Avon, England ; Philadelphia, PA: USA Multilingual matters.
- Slomanson, Peter, 2003. “Sri Lankan Malay Grammars: Lankan or Malay?” Revised Version of a paper presented at The 7th International Symposium on Malay/Indonesian Linguistics, Nijmegen June 2003 and the SPCL Summer meeting 2003 University of Hawaii, August 2003.
- Slomanson, Peter, 2008. “The perfect construction and complexity drift in Sri Lankan Malay.” *Lingua* 118(10):1640–1655.
- Smith, Ian R., 2003. “The provenance and timing of substrate influence in Sri Lankan Malay; Definiteness, animacy and accusative case marking.” Paper presented at the South Asian Language Analysis Roundtable XXIII, Austin, Texas.
- Smith, Ian R. & Paauw, S., 2006. “Sri Lanka Malay: Creole or convert.” In *Deumert & Durleman (2006)*, 159–182.
- Smith, Ian R., Paauw, S., & Hussainmiya, B. A., 2004. “Sri Lanka Malay: the state of the art.” *Yearbook of South Asian Languages* 2004:197–215.
- Stroud, C. & Heugh, K., 2004. “Language Rights and Linguistic Citizenship.” In *Language Rights and Language Survival: Sociolinguistic and Sociocultural Perspectives*, edited by J. Freeland & D. Patrick. Manchester;Northampton: St. Jerome Publishing, 191–218.
- Tapovanaye, Suthadara, 1986. *Analysis of the segmental phonemes of the Sri Lanka Malay language*. Master’s thesis, Univeristy of Kelaniya. Written in Sinhala.
- Tapovanaye, Suthadara, 1995. *Vowel Length in Sri Lankan Malay*. Master’s thesis, University of Hawaii.
- Thaliph, F., 1998. “Experience in documenting Malay society and culture in Sri Lanka - Terang Newsletter.” Paper presented at the International Symposium on Malayo-Polynesians in the Commonwealth, Kuala Lumpur, Malaysia.
- Thaliph, M Farook, 2005. *Book on Malay Grammar (Nahu) – A guide to write and speak grammatical Malay*. Wattala: Terang.
- Tirtosudarmo, R., 2004. “The lands below the winds that is called the Malay world: Notes on some crossing points and beyond.” Paper presented at the Thinking Malayness, Fuchu, Tokyo.
- Vijaycharya, S., 2004. “Ethnic consciousness of the Malays of Sri Lanka: some observations.” Paper presented at the Symposium on “Thinking Malayness”, Tokyo University of Foreign Studies.
- Woolard, K. A., 1998. “Language Ideologies as a Field of Inquiry.” In *Language Ideologies: Practice and Theory*, edited by In K. A. Woolard, B. B. Schieffelin, & P. V. Kroskrity. Oxford, New York: Oxford University Press, 3–47.
- Wurm, S. A., 1996. *Atlas of the world's languages in danger of disappearing*. Paris: Unesco Pub./Pacific Linguistics.

## **Part III**

# **Language contact**



## Chapter 6

# The Lexical Sources of Sri Lanka Malay Revisited

Scott Paauw

*University of Rochester*

In 2004, I wrote a paper entitled “A Historical Analysis of the Lexical Sources of Sri Lanka Malay” in which I looked at the source languages which have contributed to the lexicon of Sri Lanka Malay (SLM), with special attention paid to the Malay varieties which have provided lexical items.<sup>1</sup> Since that time, our knowledge of the Sri Lanka Malay language has increased significantly,<sup>2</sup> although there has been no further work done on the lexical origins of the language, and, sadly, no comprehensive dictionary or word list is yet available. During the past several years as well, our understanding of colloquial Malay varieties has also been improved.<sup>3</sup> The time is ripe, therefore, to reexamine the lexical sources of Sri Lanka Malay.

A brief summary of the socio-historical setting in which Sri Lanka Malay developed is helpful to give a background to the analysis presented in this paper. Most of what we know of the history of the Sri Lanka Malay community comes from the work of the historian (and community member) B.A. Hussainmiya. The following summary is based largely on Hussainmiya’s work, and in particular [Hussainmiya \(1987, 2008\)](#). The Dutch colonial administration began bringing Indonesians to Sri Lanka in 1656, following the expulsion of the Portuguese. These immigrants consisted of political exiles, criminal deportees, soldiers and slaves. Political exiles consisted chiefly of Javanese aristocrats, but also included members of the elite from Sumatra, Maluku (the Moluccas), Makassar, and Timor. Soldiers made up a much larger contingent of immigrants, and came from Ambon, Banda, Bali, Java, Madura and from the Buginese and

<sup>1</sup>This paper is based on research funded in part by a grant from the United States National Science Foundation (NSF). I would like to thank the students who worked with me under this grant, Cihangir Okuyan and Christine Wingrove, as well as colleagues who have advised me on matters relating to Sri Lanka Malay: Ian Smith, Peter Slomanson, Sebastian Nordhoff, Peter Bakker, B.A. Hussainmiya, and David Gil. I am especially grateful to the two linguists from the SLM community, Mohamed Jaffar and Romola Rasool, who have generously shared their knowledge.

<sup>2</sup>The work of Peter Slomanson and Ian Smith, as well as Sebastian Nordhoff’s grammar of up-country Sri Lanka Malay ([Nordhoff 2009](#)) have added a great deal to our knowledge of SLM.

<sup>3</sup>David Gil’s Malay/Indonesian Dialect Mapping Project has yielded some very informative results about colloquial Malay varieties. It is hoped that my dissertation ([Paauw 2008](#)) has been a valuable contribution to the field as well.

Malay areas ([Adelaar 1991:24](#)). A form of Vehicular Malay became the lingua franca of the community, and led to the community identifying itself as ‘Malay’ despite its diverse origins. This identity was reinforced by later waves of immigration under the British administration beginning in 1796, which were recruited mainly from the Malay Peninsula.

Most of the immigrants were Muslim, and the community soon became identified with the Muslim religion. Immigrants who were not Muslims likely integrated into other communities and lost their Malay identity. Due to their shared religion, the Malays associated most closely with the Muslim Tamils (known in Sri Lanka as Moors). [Adelaar \(1991:24\)](#) explains the importance of this connection:

Although from a cultural point of view they lost many of their traditional customs and practices due to their integration with the Moors, (with whom they have often intermarried), it is to them that the Malays owe the maintenance of their religious identity and possibly even their identity as a separate ethnic group.

Although the Sri Lanka Malay language was frequently written during the 19<sup>th</sup> century, and newspapers in the language were published, it ceased to be used as a written language early in the 20<sup>th</sup> century. During the 19<sup>th</sup> century, the SLM community was still in contact with the larger Malay-speaking world, and Standard Malay was used as a model for the written language, creating a diglossic situation. When Sri Lanka Malay was no longer written, and contact with other Malay speakers was lost, the colloquial variety of the language became the only variety which speakers knew.

We now have a much clearer picture of the nature of the SLM language today, thanks to the work of Nordhoff, Slomanson, Smith and others, although the picture of how it got to this point is largely unknown. It is clear that SLM is the result of extreme language contact between Malay and the other languages of Sri Lanka. The degree to which phonological, morphological and syntactic features have been influenced by the languages of Sri Lanka has been described quite well. It is useful, however, to have some idea of the language which was originally brought to Sri Lanka by the Indonesian immigrants, and for this purpose, an understanding of the diversity of Malay varieties is helpful.

The Malay language has been a lingua franca in the Indonesian archipelago and beyond for at least two thousand years. From its origin in western Borneo as a member of the Malayo-Polynesian languages within the Austronesian family, it has spread to Sumatra, and to the Malay Peninsula, as well as to communities in Java and the islands of Eastern Indonesia. In addition to the traditional classification of Malay into “literary Malay, lingua franca Malay, and ‘inherited Malay’” ([Adelaar & Prentice 1996:673](#)), it is essential to account for the divide between colloquial and formal varieties. In many areas of Indonesia, there exist two or three varieties which interact in a complicated diglossia. Table 6.1 shows the complicated nature of these sometimes very divergent varieties which exist under the label of ‘Malay’.

Throughout Indonesia and Malaysia, the standard varieties exist alongside the colloquial versions of the national language. The colloquial varieties of the national languages are very different from the standard languages, featuring “undressed” morphology, different vocabulary, highly variable syntax, and largely context-dependent semantic interpretations. In addition, in areas which are traditionally Malay-speaking (the Malay Peninsula, parts of Sumatra and Borneo, and communities in Java, Bali and Eastern Indonesia), these two national varieties exist alongside “inherited” varieties of Malay in a complex “triglossia”.

Type	Sub-Type	Languages
National Languages	Standard Varieties	Bahasa Indonesia Bahasa Malaysia
	Colloquial Varieties	Modern Colloquial Indonesian, Nusatenggara Jauh Indonesian, Riau Indonesian, many others
“Inherited” Varieties	Malayic varieties	<i>Borneo</i> : “Dayak”, Banjar, Iban, Salako, others <i>Sumatra</i> : Minangkabau, Kerinci
	Group I (Malay varieties)	<i>Sumatra</i> : Middle Malay (Seraway, Besemah), Riau Malay, Medan (Deli) Malay, others <i>Malaya</i> : Trengganu Malay, Kedah Malay, Kelantan Malay, Penang Malay, others
	Group II (Malay varieties)	<i>Borneo</i> : Brunei Malay, Sarawak Malay <i>Maluku</i> : Bacan <i>Sumatra</i> : Palembang, Bangka, Beliton <i>Java</i> : Jakarta
Contact Varieties	East IndonesiaWest Indonesia	<i>Established Communities</i> : Ambon, North Maluku, Banda, Manado, Kupang, Larantuka, Papuan Malay <i>Regional Lingua Francas</i> : Makassar, Alor, Tual, Tenggara Timur Jauh, Aru, Tanimbar Peranakan Malay (Java), Loloan (Bali)
	Malaysia	Baba Malay, Malaysian Bazaar Malay
	Outside Indonesia	Sri Lanka Malay, Nonthaburi Malay (Thailand), Melajoe Sini (Netherlands), Lugger Malay (Australia), Cocos Malay

Table 6.1: Malay(ic) Varieties.

To complicate matters further, there are three major dialect regions in the Malay world, roughly equating to patterns of settlement. Group I varieties are those which left the homeland first, and are spoken in Central and Northern Sumatra and the Malay Peninsula (as well as in the Malaysian areas of Borneo). Group II varieties represent a later wave of emigration and are spoken in Southern Sumatra, Jakarta, Bacan Island in Maluku and in the area of the original Malay homeland in Western Borneo. The third group of Malay varieties are those which developed through trade in Eastern Indonesia, including the Malay varieties spoken in North Maluku, Manado, Ambon, Banda, Kupang and Indonesian Papua. These varieties share many features, and descend from a single language, which has been given the name Eastern Indonesia Trade Malay (EITM) ([Paauw 2008:298](#)).

Apart from the shared features of EITM described in [Paauw \(2008\)](#), there has been little information available on the three major dialect regions of Malay. Happily, that situation is improving. For the past 17 years, David Gil has been working on the Malay/Indonesian Dialect Mapping Project, cataloguing 300 linguistic features (lexical, phonological, and grammatical) used in the colloquial varieties of the national language in locations throughout the Malay world. He has given names to the three varieties, and his terms will be used here for these varieties. The “Group I” languages are referred to as “Malaka”, after the trading entrepôt in the region where those language varieties are spoken, which had significant historical importance. The “Group II” languages are “Java”, representing the cultural and political center of those language varieties, while the varieties descended from EITM are known as “Maluku”, for the spice islands which were the center of the Eastern Indonesia trade.

All three of these regions, Malaka, Java and Maluku, were intensively involved in trade, both within and outside of the archipelago, for hundreds of years. All three regions, therefore, employed a variety of Malay which could be termed Vehicular Malay. The question then is what form the Vehicular Malay took which brought the Malay language to Sri Lanka.

Three previous studies have attempted to address this question. The first observer to look at the question was K.A. Adelaar (1991), who collected a limited sample of data from two informants in Colombo, including a word list of 262 words. Based on this data, Adelaar identified 19 features in SLM which differed from Standard Malay (SM), some of which lined up with other varieties of colloquial Malay. From these features, Adelaar determined that the data he collected “agree with Hussainmiya’s (1987) account in-so-far as they reflect the Moluccan Malay, Javanese, Jakartanese (or Batavian) and Tamil components that made up the Sri Lanka Malay community. But they disagree as to Hussainmiya’s implication that the basis of SLM was Jakartanese, and they also do not show a strong influence from Javanese.” Adelaar found that the varieties which had the most in common with SLM were Moluccan Malay varieties (North Moluccan Malay and Ambon Malay), Baba Malay (a creole spoken by Chinese in Malaysia and Singapore), Bazaar Malay (a pidgin used for interethnic communication in Malaysia and Singapore), and Jakartanese (in this case, the Betawi language of Jakarta, distinct from colloquial Jakarta Indonesian).

Adelaar compiled the following list of “the most striking differences between SM<sup>4</sup> and SLM” ([Adelaar 1991:25-26](#)):

In my own study ([Paauw 2004](#)), although I addressed Adelaar’s conclusions, I was

---

<sup>4</sup> Adelaar does not distinguish between the dialects of Malay in his analysis, and his use of the term “Standard Malay” refers to a variety with features largely shared between the Malaka and Java regions (although not necessarily with the Maluku region), roughly comparable to the national language varieties of Indonesia and Malaysia.

Standard Malay	Sri Lanka Malay	Malay varieties following SLM
1)*-h > -h	*-h > Ø	Moluccan, Bazaar, Baba, Jakartanese
2)---	retroflex series	(Tamil and Sinhalese influence)
3)---	(contrastive) vowel length	(Tamil and Sinhalese influence)
4)---	(contrastive) consonant gemination	(Tamil and Sinhalese influence)
5)-m/-n/-ŋ	-ŋ	Moluccan
6)ə	i/u or ə varying with i/u	Moluccan
7) retention of most of the inherited morphology	loss of most of the inherited morphology	Moluccan, Bazaar, Baba, Jakartanese
8) locative preposition + noun phrase	noun phrase + linker + locative preposition	Bazaar
9) noun + determiner	determiner + noun	Moluccan, Bazaar, Baba Jakartanese
10) possessed + possessor	possessor + linker + possessed	Moluccan, Bazaar, Baba, elsewhere
11) noun + adjective	adjective + noun	(Tamil and Sinhalese influence)
12) prepositions	postpositions	(Tamil and Sinhalese influence)
13) subject-verb-object	subject-object-verb	(Tamil and Sinhalese influence)
14) <i>ada</i> denoting existence of a noun	<i>adə/arə</i> : progressive aspect of a verb	Moluccan, Bazaar, Baba
15)---	negators: <i>tər-/tra</i>	Moluccan, Baba
16) full tense-mood-aspect adverbials	full and reduced tense- mood-aspect adverbials	Moluccan, Baba
17) plural personal pronouns are independent lexemes	plural personal pronouns are historically compound forms with * <i>oraiŋ</i>	Moluccan, Bazaar, Baba
18)---	1st and 2nd personal pronouns borrowed from Hokkien Chinese	Bazaar, Baba, Jakartanese
19)---	plural marker <i>-pada</i>	Jakartanese

Table 6.2: Features collected by [Adelaar \(1991\)](#)

Feature	Maluku	SLM	Corresp.
Word for ‘cockroach’	kakarlak	kakarlath <sup>7</sup>	strong
Word for ‘laugh’	tharthaava	tertawa	strong
Coda of <i>bodoh</i> ‘stupid’	-k	-k, Ø	strong
Different suprasegmental pattern for words with historical penultimate ə	yes	yes	strong
Reduced form of POSSESSIVE <i>punya</i>	pe, pu, pung	pe, ppe	strong
Short ~ long pronoun alternation	yes	yes	strong
PROGRESSIVE with <i>ada</i>	ada	ara-	strong

Table 6.3: Gil’s feature sets for strong correlations for SLM with Maluku.

looking primarily at lexical data. Drawing from three word lists (including Adelaar’s) with about 1300 discrete words, I found that 88.4% of the lexicon consisted of words originating in Malay, 10% were words originating in Sri Lanka, and 1.6% were of unknown origin.<sup>5</sup> Looking more closely at the words of Malay origin, I found that 63% were words used in both Indonesian Malay (the Java and Maluku regions) and Peninsular Malay (the Malaka region), 11.7% were words of Indonesian origin, and 0.8% were words of Peninsular Malay origin (the remaining 21.7% were loan words). This indicated that the lexicon was much more heavily influenced by Indonesian Malay than by Peninsular Malay. I did not look at the Indonesian data in terms of origin in the Java or Maluku dialects.

The most recent study of the origins of the Vehicular Malay which lexified SLM is Gil (2010). Through analysis of the 300+ lexical, phonological and grammatical features included in the Malay/Indonesian Dialect Mapping Project, Gil found that there was a significant eastern (Maluku) component, plus a smaller Java-centered component.<sup>6</sup> Gil found that when comparing features shared by SLM and Maluku, but not found in Java or Malaka, there were eight features which he counted as showing a “strong” likelihood of showing a shared origin, ten features which he labeled as “possible” and six features marked as “weak”. When compared with Java, SLM had two strong features, two possible features, and five weak features. Finally, when compared with Malaka, SLM had no strong features, two possible features, and five weak features. The comparisons suffered from a lack of data from SLM for many members of the feature set, yet the conclusions show a marked correspondence.

Gil’s feature sets for strong correlations for SLM with Maluku and Java are given in tables 6.3-6.4 (Gil’s transcription of SLM data is kept, and may differ from transcriptions elsewhere in this paper):

The results of the Adelaar and Gil studies indicate a clear connection between SLM and the Maluku varieties through primarily structural and phonological features. This is a promising area for future comparative work to determine the depth of the correspondence between SLM and Maluku.

The current study concentrates on the lexical inventory of SLM, and is not intended

<sup>5</sup>The total for words of Malay origin includes words borrowed into Malay from a variety of languages before the Malay settlement in Sri Lanka. The total for words of Sri Lankan origin includes words from Tamil and Sinhala and also words borrowed from other languages after the Malay settlement of Sri Lanka.

<sup>6</sup>It should be noted that Gil does not refer to these regions as “dialects”, and his data is only drawn from colloquial varieties of the standard language (and not “inherited” varieties of Malay).

<sup>7</sup>The Maluku and SLM forms are a loan word from Dutch. The Java form is *kecoa* while the Malaka form is *lipas*

Feature	Java	SLM	Corresp.
Word for ‘ear’ is <i>kuping</i>	yes	yes	strong
Prenasalization in <i>tangis</i> ‘cry’	yes	yes	strong

Table 6.4: Gil’s feature sets for strong correlations for SLM with Java.

to discount in any way the importance of work on phonological and structural comparison of SLM with other Malay varieties. Important evidence can be gathered by looking at each area of a language. Examining a language’s word inventory can inform us about its origins and changes in the lexicon of a language can tell us a great deal about language contact.

For this study, a word list of 1710 discrete headwords was used, drawn from four sources:

- A word list of approximately 1000 discrete entries collected by Bichsel-Stettler as part of her M.A. thesis in 1989 ([Bichsel-Stettler 1989](#)).
- A word list of over 500 items drawn from recorded conversations collected by B.A. Hussainmiya in the late 1970s.
- A word list of 262 items collected by K.A. Adelaar in 1987 from two informants and included in [Adelaar \(1991\)](#).
- Saldin’s *Kamus Bahasa Melayu Sri Lanka 2007*, provided to our project by Sebastian Nordhoff, containing about 2000 entries, many of which are duplicates, and containing many of the items in Bichsel-Stettler’s word list.

The lists were combined, and duplicate entries were removed, as well as place names and ethnonyms, and instances of obvious code-switching (in the transcribed conversations). The original transcriptions used in each document were kept, illustrating the wide variation used in transcribing SLM.

The words in the combined list were tagged for origin, as follows.<sup>8</sup>

*Malay origin:*

- Words shared between the three regions
- Words exclusive to the Malaka region
- Words exclusive to the Maluku region
- Words either exclusive to the Java region, or shared by the Java and Maluku varieties (but not found in the Malaka region); This category was necessitated by the lack of dictionaries for Maluku region varieties
- Loan words which were part of the Malay language before it was brought to Sri Lanka (identified by language of origin)

*Sri Lankan origin:*

- Words borrowed from Tamil

<sup>8</sup>Dictionaries of Bahasa Indonesia, Bahasa Malaysia, and Tamil were used to identify word origins. Much valuable information on loan words was provided by Coope’s (1976) etymological dictionary of Malay.

	2004 Study	2012 Study	
<b>Malay words</b>	1179	88.4%	1526
<b>Sri Lankan words</b>	133	10.0%	116
<b>Unknown origin</b>	21	1.6%	68
<b>Total</b>	1333		1710

Table 6.5: Origin of SLM lexemes in two studies.

	Number of words	Percentage of all words
Shared Malay origin	918	53.7%
Java region	260	15.2%
Maluku region	12	0.7%
Malaka region	9	0.5%
Loan words <sup>11</sup>	327	19.1%
Total	1526	

Table 6.6: Importance of the three possible regions of origin for the lexicon.

- Words borrowed from Sinhala
- Words borrowed from other languages after settlement in Sri Lanka (not found in other Malay varieties)
- Words of unknown origin (not found in other Malay varieties)<sup>9</sup>

The results of the study were remarkably similar to the 2004 study, despite a larger word list and an improved study methodology. The lexicon of Sri Lanka Malay is overwhelmingly of Malay origin (Table 6.5 totals include loan words).

The maintenance of such a high percentage of words of Malay origin illustrates that the lexicon is the chief linguistic marker of Sri Lanka Malay identity, even as the morphology and syntax of the language have undergone significant changes through convergence with the other languages of Sri Lanka. It is notable that SLM has a higher percentage of Malay-derived vocabulary than some Malay varieties in Indonesia, such as Ambon Malay (85.9%) and Betawi (85%).<sup>10</sup>

Analysis of the 1526 words of Malay origin show that the majority are shared by all Malay dialects, but that the Java region has contributed a larger share of the vocabulary than the other varieties (note that the totals for Java include words which are shared between the Java and Maluku regions as well as words exclusive to the Java region).

These results show that there is a very strong Java region component, even if the fact that words counted as Java region include an unknown number of words shared with the Maluku region. This points to a strong influence of the Java region in the formation of Sri Lanka Malay. This evidence, along with the evidence presented by Adelaar and Gil emphasizing the influence of the Maluku region, leads to the obvious

<sup>9</sup>There were 68 words of unknown origin. A number of these may originate in colloquial or regional varieties of Tamil or Sinhala, as the author's knowledge of these languages is not extensive (but does include a reading ability in Standard Tamil). Other words of unknown origin appear to be Austronesian in form, though not from any of the three regions of Malay. These may be borrowings from other languages contributed through the SLM community's diverse origins in Indonesia.

<sup>10</sup>Figures are from Blust (1988:3). The figures represent cognates with Standard Indonesian for a 200-word list.

<sup>11</sup>Loan words are from non-Austronesian languages. Eighteen identified loans from Austronesian languages (all are from Javanese and Sundanese) are included under Java region.

conclusion that these two regions each had a significant role in the formation of SLM. The historical facts support this conclusion, in that many of the immigrants who formed the Sri Lanka Malay community were from Eastern Indonesia, and thus probably spoke the Maluku dialect. Others, including most of the political exiles, were from Java and spoke the Java dialect (although almost certainly as a second language). Furthermore, all of the soldiers were recruited in Batavia, and thus had resided in a Java Malay-speaking milieu before leaving for Sri Lanka. These facts explain the role of both Java Malay and Maluku Malay in the formation of SLM. The SLM language was already established, and the Malay community had resided in Sri Lanka for nearly 150 years before the British colonial administration began bringing immigrants from the Malay Peninsula. Because these later immigrants encountered an established language in Sri Lanka, their Malaka dialect had little effect on SLM.

David Gil's Malay/Indonesian Dialect Mapping Project referred to earlier includes a number of lexical items which can be associated with the patterns of the three regions of Malay. Although Gil (2010) does not address many of these in relation to Sri Lanka Malay, presumably because most items show a convergence with two varieties while Gil was concerned with items converging with a single variety, it is worthwhile to examine these lexical items and how SLM fits into the picture. This is particularly illuminating because these are lexical items selected specifically because they are associated with particular varieties and thus can provide information about the origins of SLM. Table 6.7 presents these items.

Java and Maluku each have 18 correspondences, while Malaka only has 7, further underscoring the closer relationship between Java and Maluku with SLM. Three further items from Gil's feature set, the words for 'car', 'bicycle' and 'bus' show correspondences between Malaka and SLM, but these can be explained by the fact that these concepts entered the language in the 19<sup>th</sup> or early 20<sup>th</sup> centuries, during a period when SLM and Malaka shared a colonial master, the British, and there was intensive contact between the two languages.

One area of particular interest when considering the origins of SLM is the pronoun set. Pronouns are usually quite stable in languages, and are often used in comparisons between languages to show genetic relatedness. However, the first and second person singular pronouns in many Indonesian languages are often unstable, and varieties of Malay are particularly notorious for variation in pronouns and, indeed, avoidance of first and second person singular pronouns altogether. Among the seven Eastern Indonesian Malay varieties described in Paauw (2008), each variety (with one exception) has different second person pronouns from each of the other varieties, although all are varieties of the Maluku region.

In addition to this pronominal variation, Manado Malay, North Moluccan Malay, Ambon Malay, Kupang Malay and Papua Malay have short forms of the pronouns in alternation with the long forms, as SLM does. This was one of the features Gil labeled as a strong correspondence between Maluku and SLM.

The plural pronouns in Maluku consist of the singular pronouns + *orang* 'person', a feature which also occurs in SLM. This method of plural pronoun formation is also found in contact varieties of the Malaka region (specifically the pidgin used for inter-ethnic communication known as Bazaar Malay and the creole Baba Malay, spoken by Chinese in Singapore and Malacca). In addition, the form in SLM which one would expect to exist as a 2P form, *lorang* (derived from *lu* + *orang*), is a singular pronoun in

---

<sup>12</sup>It is notable that the second person singular pronouns are borrowed in several cases: *ngana* (from Terate, a Papuan language), *ose* (from Portuguese *voce*), and *lu* (from Hokkien).

Word	Malaka	Java	Maluku	SLM	SLM aligns with
'large'	kena	kena	dapat	kena	Malaka, Java
'card'	besar	gede	besar	besar	Malaka, Maluku
'ear'	kad	karto	karto	karto	Java, Maluku
'easy'	tingga	kuping	telinga	kuping	Java
'shrimp paste'	senang/muda	gampang	gampang	gampang	Java, Maluku
'give'	belacan	terasi	belacan/terasi	belacang	Malaka, Maluku
'rose-apple'	bagi	kasi	kasi	kasi	Java, Maluku
'meet'	jambu	jambu	jawas	jambu	Malaka, Java
'pink'	jumpa	ketemu	ketemu	kutumung	Java, Maluku
'put'	merah jambu	pink	merah muda	merah jambu	Malaka
'look'	letak	taro	taro	taro	Java, Maluku
'shoe'	tengok	lihat	lihat/jia	liyat	Java, Maluku
'delicious'	kasut	sepatu	sepatu	supatu	Java, Maluku
'tree'	sedap	enak	sedap/enak	enak	Java, Maluku
'understand'	pokok	pohon	pohon	pohong	Java, Maluku
'village'	faham	ngerti	mengerti	mengerti/nerti	Malaku
'wall'	kampung	desa	kampung/negri	kampung/nigiri	Malaka, Maluku
1s and 2s pronouns from Chinese	dinding	tembok	dinding/tembok	tembok	Java, Maluku
'when'	bilâ	gua, lu	go, lu	go, lu	Java
'navel'	kapan	kapan	kapan	kapan	Java, Maluku
'ginger'	pusat	pusar	pusar	pusar	Java, Maluku
'cassava'	halia	jahe	halia/jahe	jaye	Java, Maluku
'cold'	ubi kayu	singkong	kasbi	ubi kayu	Malaka
	sejuk	dingin	dingin	dingin	Java, Maluku

Table 6.7: Comparison of the vocabulary of the different varieties.

Variety	1s Pronoun	2s Pronoun <sup>12</sup>
Manado Malay, North Moluccan Malay	kita	ngana
Ambon Malay	beta	ose
Banda Malay	beta	pane
Kupang Malay	beta	lu
Larantuka Malay	kita	eNko, no (MASC), oa (FEM)
Papua Malay	say, kita	kow, ko

Table 6.8: 1st and 2nd person pronouns in a number of varieties.

SLM	Java	Maluku	Origin of SLM form
1s <i>go saya/se</i>	<i>gua saya</i>	<i>kita beta</i>	Java
2s <i>lu lorang</i>	<i>lu kamu, kau</i>	(various)	Java
3s <i>de</i>	<i>dia</i>	<i>de</i>	Maluku
1P <i>kitang, serang</i>	<i>kita</i>	<i>katong (ki)torang</i>	Maluku ( <i>kita + orang</i> )
2P <i>lorang pada</i>	(none)	(various)	Java and Maluku
3P <i>derang</i>	<i>mereka</i>	<i>dorang</i>	Maluku ( <i>de + orang</i> )

Table 6.9: Comparison of pronouns between SLM, Java, and Maluku.

SLM, alternating with *lu*.

The alternating short and long pronoun forms, as well as the plural pronoun formation, point to an origin in Maluku for the SLM pronouns. However, the first and second person singular pronouns in SLM are from the Java region, including the Hokkien loans *go* ‘1s’ and *lu* ‘2s’, as well as another 1s pronoun, *saya* (with the short form *se* in SLM). This somewhat confusing picture points once again to an origin in both Maluku and Java for the SLM pronouns

Another interesting area of convergence between SLM and Maluku is in the negator *ter/tra-*, variants are which are found in most varieties of the Maluku region, but are largely unknown outside of the Maluku region.<sup>13</sup>

The words which are categorized in the previous analysis of SLM word origins as of Sri Lankan origin are those which have been added to the language since the Malay settlement in Sri Lanka, and are not found in other Malay varieties. These include loan words from Tamil and Sinhala, as well as loans from other languages which have occurred since SLM was formed. These words of Sri Lankan origin are described in the list below:

<sup>13</sup>There is some evidence that a negator with a form similar to *ter-* exists in the Malaysian pidgin known as Bazaar Malay, and even evidence that a similar form was used in Java Malay in the 19<sup>th</sup> century. However, today, negation with *ter-* is only commonly found in Maluku and SLM.

	Number	Percentage of all words
Tamil	71	4.2%
Sinhala	19	1.1%
Other <sup>14</sup>	26	1.5%
Unknown origin <sup>15</sup>	68	4.0%
Total	184	10.8%

Table 6.10: Loan words from Sri Lankan languages in SLM.

Language of Origin	Number of Loan Words
Arabic	150
Sanskrit	79
Portuguese	28
Dutch	21
Persian	16
Hindi/Urdu	15
Tamil	11
Hokkien	4
English	3
	327

Table 6.11: Loanwords from other languages in SLM.

Although SLM has been notably resistant to borrowing, and retains a very large percentage of Malay words, some borrowing has occurred over the centuries, most often relating to the culture, flora and fauna of the Malays' new home. As the SLM community has historically maintained the closest ties with the Muslim Tamil community, it is not surprising that the largest number of loan words is from Tamil.

Although it may not directly relate to the development of SLM, it is interesting to note the language of origin of loan words in Malay which have become a part of SLM, as shown below.

The preponderance of Arabic loan words is probably a reflection of the SLM community's strong degree of religious devotion. In other varieties of Malay, the percentage of loan words from Arabic is much lower.

In conclusion, the lexical evidence points to Java Malay as the most important source of the lexicon of Sri Lanka Malay. Other evidence, provided by Adelaar and Gil, uses phonological and structural data to identify the Maluku region as the primary source. However, there are problems with identifying Maluku as the source for SLM. While there are certain features (the pronoun set, the negator *ter-*, a few lexical items) which could only have had their origin in Maluku, there are other features (word final nasal coalescence, schwa replacement) which only occur in some words in SLM and are not regular processes. In addition, if we consider both these features and another features of Maluku which doesn't occur in SLM, final stop deletion, and we posit Maluku as the source for SLM we are left with the impossible task of explaining how a variety of nasals appear in word final position if they had coalesced in the source, and how a variety of word final stops occur if they had been deleted in the source. Once nasals have merged or consonants have been deleted, they cannot somehow be replaced. Furthermore, there are features in SLM which can only have come from Java Malay, such as the plural marker *pada*, which is only found in Java Malay among Malay varieties of the Indonesian archipelago. These features and the preponderance of vocabulary associated with Java Malay indicate that Java Malay *must* have had a role in the formation of SLM.

A much more reasoned approach, and one which is supported by the historical record, is that both the Java and Maluku regions had an important role in the formation

<sup>14</sup>These consist of words borrowed from English (18), Portuguese (4), Hindi (2), Dutch (1) and Arabic (1).

<sup>15</sup>These are words which are not of Malay origin. It is likely that many of them are loans from Tamil or Sinhala dialects.

of SLM.



# Bibliography

- Adelaar, K. A., 1991. "Some Notes on the Origin of Sri Lanka Malay." In *Papers in Austronesian Linguistics*, edited by H. Steinhauer. Canberra: Australian National University, 23–37.
- Adelaar, K. A. & Prentice, D. J., 1996. "Malay: its History, Role and Spread." In *Atlas of Languages of Intercultural Communication in the Pacific, Asia and the Americas*, edited by Stephen A. Wurm, Mühlhäusler Peter, & Darrell T. Tryon. Berlin: Mouton de Gruyter, 673–693.
- Bichsel-Stettler, Anne, 1989. *Aspects of the Sri Lanka Malay Community and their Language*. Master's thesis, University of Bern.
- Blust, Robert, 1988. "Malay historical linguistics: A progress report." In *Rekonstruksi dan Cabang-Cabang Bahasa Melayu Induk [Reconstruction and the branches of Malay]*, edited by Mohd. Thani Ahmad & Zaini Mohamed Zain. Kuala Lumpur: Dewan Bahasa dan Pustaka, 1–33.
- Coope, A. E., 1976. *A Malay-English Dictionary*. London: Macmillan.
- Gil, David, 2010. "Malay/Indonesian Dialect Geography and the Sources of Sri Lankan Malay." Presentation at The Workshop on Sri Lanka Malay, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany.
- Hussainmiya, B. A., 1987. *Lost cousins: The Malays of Sri Lanka*. Bangi: Institut Bahasa, Kesusasteraan dan Kebudayaan Melayu.
- Hussainmiya, B. A., 2008. *Orang Rejimen: The Malays of Sri Lanka*. Bangi: Universiti Kebangsaan Malaysia, 2nd edition.
- Nordhoff, Sebastian, 2009. *A Grammar of Upcountry Sri Lanka Malay*. Ph.D. thesis, University of Amsterdam, Utrecht.  
URL <http://www.lotpublications.nl/publish/articles/003744/bookpart.pdf>
- Paauw, Scott, 2004. *A Historical Analysis of the Lexicon of Sri Lanka Malay*. Master's thesis, York University (Toronto). Presented as a paper at the meeting of the Linguistic Society of America/Society for Pidgin and Creole Linguistics, Oakland, CA, January 2005.
- Paauw, Scott, 2008. *The Malay Contact Varieties of Eastern Indonesia: A Typological Comparison*. Ph.D. thesis, University at Buffalo.
- Saldin, BDK, 2007. *A Concise Sri Lankan Malay Dictionary*. Kurunegala: Nihon Printers.



## Chapter 7

# Sri Lankan Languages in the South-South Asia Linguistic Area: Sinhala and Sri Lanka Malay

James W. Gair

*Cornell University*

### 7.1 Introduction and Background

Murray Emeneau defined a linguistic area as ‘an area which includes languages of more than one family but showing traits in common which are found not to belong to the other members of (at least) one of the families’, and South Asia has long been recognized as such an area, beginning with the pioneering work of Jules Bloch.<sup>1</sup> As Emeneau characterized that area in his influential 1956 paper: “The end result of the borrowings is that the languages of the two families, Indo-Aryan and Dravidian, seem in more respects more akin to one another than Indo-Aryan does to the other Indo-European languages.” (Emeneau 1956:16), and its characteristics were further defined in the work of a number of scholars such as Colin Masica and others. Within a linguistic area, there always exists the possibility for sub-areas, and within the South Asia one there clearly appears to be a southern subarea (SSLA), which I have attempted elsewhere to begin to enumerate (Gair 1994) and will repeat here while proposing some extensions. Geographically, this area coincides essentially with the South and South-Central Dravidian Language area, extended to bordering languages, including some island languages. Within it there exist a number of Non-Dravidian languages, Indo-Aryan and otherwise. Most if not all of its defining features are characteristic of both

---

<sup>1</sup>This paper is based on a presentation given at the Workshop on Language contact in and around the Indian Ocean, University of Amsterdam, November 26, 2009. I would like to express my appreciation the participants and especially to Sebastian Nordhoff, Kees Hengeveld, and Umberto Ansaldi for encouragement and comments.

South and South-Central Dravidian and thus their appearance in the non-Dravidian languages in the sub-area may confidently be claimed to have Southern Dravidian as its ultimate source.<sup>2</sup> The island Indo-Aryan languages are Sinhala and Dhivehi (Maldivian), the national languages of Sri Lanka and the Republic of the Maldives respectively. The group also includes the language of the island of Minicoy, where it is known as Mahl (or Mahal), but that is linguistically a dialect of Dhivehi. Together, they form a southern branch of Indo-Aryan, which can be referred to as “Southern Insular Indo-Aryan” (SIIA).<sup>3</sup> Sinhala and Dhivehi are now not mutually intelligible, but in addition to sharing the features characteristic of the general South-South-Asian area, they exhibit others specific to SIIA, such as the development of prenasalized stops (“half nasals”) and indefinite affixes derived from the number “one”, and thus clearly form a sub-family within IA. Within the southern subarea there exist other Indo-Aryan isolates from their northern kin. These other Indo-Aryan isolates are later comers to the area, by migration or political-cultural importation. Among them are Sourashtra, dealt with by Ian Smith ([Smith this volume](#)), centered in Madurai, resulting from migration in the 16<sup>th</sup> century CE, and Dakkhini Hindi/Urdu, chiefly in Hyderabad, dating from a somewhat earlier time and the result of Islamic rule in areas of the Deccan.<sup>4</sup> Marathi, bordering the area on the northwest, also requires some mention. Typologically as well as geographically, it is a “border language”, sharing some features with its Dravidian neighbors sometimes as alternates, but also remaining fundamentally akin to its north Indian Indo-Aryan relatives. Sri Lankan Portuguese, described by Ian Smith ([Smith this volume](#)), is another Indo-European isolate, though not, of course, Indo-Aryan. Languages of Sri Lanka, by and large, reflect the characteristics, largely syntactic, that characterize the area. The most prominent of these languages is Sinhala, the others being Sri Lankan Tamil, with several dialects, Sri Lankan Portuguese, and Sri Lankan Malay. For Sri Lankan Malay, we now have extensive available information, thanks to a flourishing of research and description by several scholars, including the recent dissertation of Sebastian Nordhoff. In relation to language contact, SL Portuguese and SL Malay are of special interest for at least two reasons. Unlike Sinhala and Dhivehi, as non Indo-Aryan languages, they did not, at the time of importation, share the typological characteristics common to the wider South Asian area, setting aside any prior creolization involving the Indian subcontinent. Also, the time depth of contact is relatively short, compared to Sinhala/Dhivehi, for which contact extended for millennia, and this has led to interesting work concerning the necessary time span for creolization, in Nordhoff as well as in previous work by Ansaldi, Bakker, Paauw, Smith, Slomanson, and others. This paper is devoted primarily to Sinhala and SL Malay, with the aim of demonstrating the degree to which the latter has incorporated the features defining the SSLA subarea. Sinhala examples are thus presented essentially for purposes of comparison so as to provide a reference point for the subarea features as they occur in SL Malay. As stated earlier, the defining features of the subarea are essentially if not completely South Dravidian in origin. I have exemplified those features in Dravidian languages elsewhere (esp. 1994 and 2009) and will not repeat that material here. They can, however, be taken as a background assumption throughout to be present as in those

<sup>2</sup>In what follows, I will use the term “Southern Dravidian” for both the South and South-Central languages, unless further specified, since the difference is irrelevant here.

<sup>3</sup>This term is an amplification of the useful term “Insular Indo-Aryan” originated by Sonia Fritz ([Fritz 2002](#)), extended slightly to indicate more fully the geographical location and to emphasize the separation from the mainland IA languages.

<sup>4</sup>There is at least one other Indo-Aryan isolate in Vaagri Boli, (among other names) spoken by generally semi-nomadic groups in Tamil Nadu, Maharashtra and Karnataka, but I have too little data on it to include it here.

languages as well as those exemplified here. Other than Sinhala, the other major contact language for Malay was some variety or varieties of Sri Lankan Tamil<sup>5</sup> and it has been argued to be the primary one. My not including examples from Sri Lankan Tamil and restricting myself here to Sinhala does not imply that the SL Malay features are the result of direct Sinhala influence, and should emphatically not be taken as reflecting a conclusion to that question. It does however, seem clear from SL Malay evidence in the works cited here that it has absorbed features from both Tamil and Sinhala, whatever the relative weighting. I do not include SL Portuguese here partly for reasons of time and space, but also because I do not have extensive enough information at his time. Peter Bakker has, in fact, raised the question of a Sri Lankan *Sprachbund*, and made a number of interesting important observations particularly regarding these relatively recent comers. However, there is a terminological and empirical question in regard to the term if it implies a sprachbund territorially limited to the island. If we look only at languages resident in the island and extend this to the Maldives and Minicoy, ignoring the subcontinent, the term is apt, but some of the features cited are characteristic of the languages of Southern South Asia generally and others are shared with the wider South Asian area. Thus the proposed sprachbund would in fact constitute at best a sub-area of the South-South Asia one, sketched out below. Also, if Tamil is included, that automatically invokes an extension to South India, unless one restricted it to features of Sri Lankan Tamil not found on the mainland, and that would involve several local varieties distributed geographically and culturally. Thus, in order to justify the existence of such an island-limited sprachbund it would be necessary to find features limited to Sri Lanka, including Sri Lankan Tamil as well as the other languages. This is, of course, a real, if so far unaddressed possibility. Bakker is to be thanked for raising it, and Nordoff, for one (p.c.), is investigating it further.

## 7.2 The Defining Features

Earlier, I referred to the proposed set of nine features that characterized the South-South Asian area. These are given in Table 7.1 below.<sup>6</sup> It should be recalled that these are as a set defining features of the South-South Asia Linguistic area, which in turn forms a part of a larger South Asian one, and thus shares those features as well. In the case of Sinhala and Dhivehi, they have not surprisingly been inherited from their mainland source. This is not of course the case for the non-SA imports SL Malay,

---

<sup>5</sup>Sri Lankan Tamil itself includes several varieties, and they differ in various respects from the Tamil of the mainland (Suseendirarajah 2008). However, on the basis of available evidence, they all share the areal features here. For the best studied Jaffna variety see especially Gair et al. (1978) and Suseendirarajah (1993).

<sup>6</sup>One reviewer pointed out that many of these features are not limited to SSLA languages but are commonly found in head final languages. This is indeed true, and not limited to South Asian languages. The point within the universe of discourse here, however, is that they form a constellation that distinguishes them within the larger South Asian environment, while also sharing the common features of that larger area of which they form a part, while noting that the main initial source of the traits here was the southern Dravidian languages. The same reviewer also raised the interesting observation that many of them are found in languages of the northeast, especially some in Nepal as well as in some Munda languages, stating that (s)he found and it “quite intriguing that these traits seem to continue throughout the eastern part of the subcontinent right up to Nepal”. This raises the question as to whether these are separate developments or are in some way connected with SSLA. Given that the traits at issue are indeed not uncommon in head final languages, the former is plausible, and there is the added possibility of a different subarea. It might be added that some of the features, such as the ‘say’ complementizer, are found in Bengali, and in fact Dravidian influence has been posited there (Klaiman (1977), and see Bayer (2001)). This invites further research involving the cultural and linguistic history of the area, but that is clearly beyond the confines of the present paper.

or SL Portuguese, which as stated at the beginning were typologically very different from the South Asian ones. My concern in this paper is with the extent to which SL Malay participated in the SSLA sprachbund, and features of that language not specific to that subarea have not been dealt with. In investigating contact induced typological change in SL Malay, however, it is obvious that those more general features are of equal importance to those specified here, since they were all part of the donor languages. As one crucial example of such a vitally important feature, we may note basic SOV and right-headedness, and in fact the very strong right-headedness that is characteristic of SS Asia and underlies several of the features here. In relation to Sri Lankan Malay, one interesting aspect is that this did not extend to the morphological level, though it does pervade the syntax. This also appears to be the case with SL Portuguese, as noted by Ian Smith, who states in relation to SL Portuguese and Sourashtra “Structurally, Sourashtra was closer to Tamil, again because of a closer starting position; SLP showed more accommodation to Tamil, but had not modified the structure of some closely bound elements such as verbal prefixes” (Smith 2001:408). This raises some interesting questions as to variable resistance of aspects of the grammar to contact induced change, but that is beyond our concerns here.

A sentence-final reportative or hearsay particle is found in all of the SSLA languages, but it is not, as far as I have been able to determine, found outside that region in the northern Indo-Aryan languages, with the exception of Marathi, which has the form *mhaNe*, (Bashir 2006:19) and Marathi is a kind of “border” South-South language, sharing some of the SSLA features, but retaining its general northern IA character.

The distributions of these reportative forms with regard to their use on constituents as well as in sentence final position varies across the languages, and in the case of Sinhala at least, it is intimately integrated with the system of focus and clefting (Gair 1986, 1994, 1997, 2009). Another likely possibility is the existence of coordinating or additive particles or affixes cited by Bakker (2006:143) such as Tamil *-um*, Sinhala *da*, SL Malay *le(y)*. SL Portuguese *ta:m* has also been cited. However, it remains to be established that non-southern Indian languages do not have this feature. Emphatic particles have also been mentioned for Sri Lanka (as in Bakker 2006:142-3), but in this case it appears to be characteristic of the wider South Asia area, with such forms as Hindi *bhii*.

It should be noted that the all of the features that I have listed are syntactic. The literature on Malay and SL Portuguese, however, presents a number of relevant phonological and morphological features that invite wider investigation. For example, to cite one new contribution, Nordhoff notes the development of prenasalized stops in upcountry SL Malay, an uncommon change that puts it in a set with Sinhala and Dhivehi and clearly indicates Sinhala influence since it is lacking in the other languages. I will not attempt to pursue the non-syntactic features further here, since it represents a separate and in part new investigation, though one that is much to be desired.

### **7.3 The diagnostic features as evidenced in the Sri Lankan languages, especially Sinhala and SL Malay**

In what follows (7.3.1 through 7.3.8), I will go through the features individually, illustrating with Sinhala and SL Malay. As stated earlier, my not including Tamil in no way represents a position on the discussion prominent in the literature on the relative extent to which Sinhala and Tamil influence affected the changes in SL Malay that brought

7.3.1	Question marker appears at the end of sentence (postverbal) as the unmarked location, but may also occur on questioned sentence-internal constituents)
7.3.2	Subordinate Clauses marked at the end, by a verbal affix or a conjunctive form of some kind, rather than by initial conjunctions (which are rare or missing altogether except for sentence adverbs)
7.3.3	Preposed Relative Clauses (Adjectival Sentences) as the only or main alternative.
7.3.4	Correlatives use a WH rather than a correlative form of the Indo-Aryan type and are generally restricted to indefinite or conditional contexts and commonly employ a sentence particle (dubitative or question) on the subordinate clause.
7.3.5	Sentence-Final quotative from ‘say’
7.3.6	Sentences may be nominalized without genitivization (or deletion) of subject, by employing a sentence-final form or verbal affix.
7.3.7	Focused (Nominal Cleft) Sentences, including those with rightward focus.
7.3.8	Negatives:
•	Negative varies with type of main clause (Verbal, Equational, Existential).
•	Negative verbs in subordinate clauses.
•	Cleft sentences negated like nominal equational ones.
7.3.9	Conjunctive participles may occur with overt lexical subjects, not co-indexed with main subject (or agent). [Extent yet to be determined]
7.3.10	A sentence-final reportative or hearsay particle.

Table 7.1: The nine features typical of the SSA, plus one additional feature, with references to the sections where they are discussed.

it strikingly into alignment with the other SSLA languages. SL Malay language is of special interest to me since I had not considered it earlier as a member of the set, a consideration now made possible by the available scholarship on the language and greatly facilitated by Sebastian Nordhoff's detailed and comprehensive grammar. In what follows, Malay examples can be assumed to be taken from that grammar unless otherwise noted. Their sources are cited by page and example number.

### 7.3.1 Question marker at end of sentence (postverbal) as the unmarked location, but may also occur on questioned sentence-internal constituents

This is found in both Sinhala and Sri Lanka Malay, as illustrated in (1) through (4).

- (1) Sinhala “Simple” Sentence Question

*ee minihaa iiye gunapaala-ta sali dunnaa=da?*  
that man yesterday Gunapala-DAT money gave=INTERR

‘Did that man give Gunapala money yesterday?’

- (2) Sinhala Constituent (Focused) Sentence Question

*ee minihaa=da iiye gunapaala-ta sali dunn-e?*  
that man=INTERR yesterday Gunapaala-DAT money gave-foc

‘Was it that man who gave Gunapala money yesterday?’

- (3) SL Malay “Simple” Sentence Question

*Sebastian pùddas arà-maakang=si*  
Sebastian spicy NONPAST-eat=INTERR

‘Do you eat spicy food, Sebastian?’ (361, ex.339)

- (4) SL Malay Questioned Constituent

*Daging baabi=si anà-billi?*  
meat pork=INTERR PAST-buy

‘Did you buy PORK?’ (274, ex. 16)

Sinhala has a special property here, in that the questioned constituent with *da* requires the focusing affix on the verb, as part of a general system of focus and clefting. This is also the case in some SSLA Dravidian languages, but not in Tamil. SL Malay does not have a focusing system employing a focusing form of the verb, so it is not surprising that it lacks this feature ([Slomanson this volume](#)). However, there is an intriguing pattern that bears a resemblance to Sinhala in relation to the verb affixes that co-occur with constituent questions. This will be dealt with in section 7.3.7 on focus structures.

### 7.3.2 Subordinate Clauses marked at the end, by a verbal affix or a conjunctive form of some kind, rather than by initial conjunctions (which are rare or missing altogether except for sentence adverbs)

One example of this is the conjunctive participle, illustrated in (5) for Sinhala.

- (5) *siri [kæema kaalaa] gedara giyaa*  
 Siri food eat-CP home went  
 ‘Siri ate and went home.’

SL Malay has a form of this as well, as in (6):

- (6) *Samma oorang school=nang asà-pii arà-blaajar cingalaa*  
 all man school=DAT CP-go NONPAST-learn Sinhala  
 ‘Everybody goes to school and learns Sinhala’ (242, ex.26)

The conjunctive participle as such is a feature shared by the SSLA languages, but it is not a feature marking that subarea. Rather, as has been noted by numerous scholars, it is an areal feature of South Asia as a whole (Emeneau 1956, Masica 1976).<sup>7</sup> That does not, of course diminish its importance as a striking and important development in SL Malay, and like other features aligned with South Asia in general, it was a product of contact with languages within the subarea. Though the conjunctive particle as such is not a defining characteristic of the South-South Asia area, it is in those languages one exponent of a general SSLA pattern by which all markers of subordination, including the conjunctive participle, are at the right margin, and may be affixal, clitics, or other subordinating forms. This distinguishes these languages from the North Indian ones, which though SOV, are ‘mixed’ in that regard, having left-marginal conjunctions and complementizers as well as conjunctive participles.<sup>8</sup> Also, conjunctive participles in SSLA languages are commonly not restricted to a single function, but occur in a variety of roles. In Sinhala, for example, they occur also as a participle in periphrastic perfect constructions, and may even occur as a main verb (Gair 2005). SL Malay is in this respect in general accord with the overall SSLA pattern. However, it differs from the other languages in the morphological alignment of the subordinating element(s), employing a prefix *asà-* rather than a suffix. As only one other example, of this parallelism in function with a difference in morphological realization, there is an SL Malay temporal prefix *kapang* (also appearing as *kaN*, *kal* and *ka*), illustrated in (7):

- (7) *Mosque=nang kapang-pii samma ooran=nang go athi-kaasi*  
 mosque=DAT when-go all man=DAT 1s.FAMILIAR IRR-give  
 ‘When I go to the mosque, I give to everybody’. (386, ex.451)

This invites comparison to the Sinhala form *kota* ‘when’, which follows the verb, but has a similar subordinating function and much the same sense:

- (8) *pansala-ta yana-kotा mama hæma-ken-eku-ta kæema de-nawaa*  
 temple-DAT go-when I every-one-INDEF-DAT food give-PRS  
 ‘When I go to the temple, I give food to everyone.’

<sup>7</sup>See Subbarao (in press), section 7.2 for a general account covering four major families.

<sup>8</sup>See Bayer (2001) for a description and analysis of “mixed”, i.e., double complementizer languages in South Asia.

### 7.3.3 Preposed Relative Clauses (Adjectival Sentences or Participial Relatives) as the main or only alternative.

Sinhala and Tamil relative expressions are formed by preposing a clause with a relativizing, or adjectival, tensed verb form to the head, and the pattern is essentially the same in the other SSLA languages. SL Malay relative clauses, like Sinhala and the other South-South Asian Languages also precede the head, and this is the exclusive pattern (Nordhoff 2009:370). Thus (9):

- (9) *[Seelong=nang dhaataang aada Ø] Mlaayu oorang ikkang*  
 Ceylon-DAT come exist Malay man fish  
 ‘The Malays who came to Sri Lanka were fishermen.’ (518, ex.61)

Sentence (9) also includes a perfect formed with an existential verb, another parallel to in Sinhala and Tamil (but also found in the North Indian languages). In Sinhala and Tamil, as well as in in the other SSLA languages, there are no TAM restrictions. Thus for Sinhala, parallel to (9):

- (10) *lankaawa-ṭa æwi-lla hita-pu minissu govitaŋ kora-pu aya.*  
 Lanka-DAT come-CP exist\PAST-REL men cultivation do-PAST.REL people.  
 ‘The people who came to Lanka were folk engaged in agriculture.’

The nominal head in SL Malay can have a wide range of semantic roles within the clause, and there is no restriction to specific grammatical relations, such as subject or object (Nordhoff 523ff). This is also the case in Sinhala, Tamil, and the other SSLA languages (Subbarao in press:Section 9.1.2))<sup>9</sup> in contradistinction to restrictions on preposed participial modifiers in many other languages including the North Indian IA ones such as Hindi. Thus (11)-(13) in Sinhala illustrate heads linked to subject, object and indirect object:

- (11) *[siri gunapaala-ṭa dunn-a] pota*  
 Siri Gunapala-DAT give.PAST-ADJ book  
 ‘The book that Siri gave Gunapala’
- (12) *[gunapaala-ṭa pota dunna] siri*  
 '(the) Siri who gave the book to Gunapala'
- (13) *[siri pota dunna] gunapaala*  
 '(the) Gunapala to whom Siri gave the book'

Note also that there is no change of case on constituents in the relative clause in either language, which occur in the same case that they would have in an independent sentence, and that this includes the subject. Thus there is no genitivization like that in northern IA languages like Hindi, or in some familiar European ones. For SL Malay, (9) illustrated a subject/agent. Theme/object and recipient are shown in (14) and (15):

<sup>9</sup> As (Subbarao in press:section 9.5.1.2) asserts, any subcategorized element can be relativized in all Dravidian languages [also in Sinhala], though some elements such as non-subcategorized comitatives cannot be. Such seems to be the case in Sinhala regarding some forms with postpositions. This might be expected from the fact that the head will naturally take the case required by the matrix sentence but it also seems to be a pragmatic function of interpretability.

- (14) [Kirras pinthu=nang arà-thatti      hathu svaara] su-dinngar  
 strong door=DAT SIMULT-hammer INDEF noise PAST-hear  
 'They heard a noise of hard hammering at the door.' (441, ex.105)
- (15) [Se      duvith anà-kaasi oorang] su-iilang  
 1S=DAT money PAST-give man PAST-disappeared  
 'The man I gave money to disappeared.' (525, ex. 93)

As Nordhoff notes, (528), SL Malay differs from Sinhala, Tamil, and other SSLA languages, in that it has not developed a relative participle and there is no overt marking for relativization. Nevertheless, it is clear that the shift of relative clauses to a right-headed pattern is a major typological one, and a striking result of contact with Sri Lankan languages. One striking feature of SL Malay not found in Sinhala, Tamil, or the other languages as far as I am aware, is the existence of genuinely headless relatives, i.e., zero or null headed ones, in which there is no overt form in any position representing a head. Nordhoff gives a number of undoubted examples, of which (16) is one:

- (16) [Lorang anà-maasak Ø] eenak  
 2PL PAST-COOK tasty  
 'What you cooked is tasty.' (453, ex. 163)

Sinhala does not have this structure, nor as far as I am aware, do any of the Dravidian languages. Rather, the equivalent would have a pronominal form expressed, as in (17):

- (17) oyaa iww-a      eewa      raha-yi  
 you cook\PAST-REL 3PL.PRON INANIM tasty-PRED  
 'What ('Those) you cooked is tasty.'

The plural pronominal in (17), *eewa*, can have indefinite or mass reference, but could also be specifically referential, so that the meaning could be 'Those that you cooked are tasty.' In Sinhala, the pronominal form is also marked for number, and if one thing is specifically referred to, a singular pronoun such as *eeka* 'that one' or *eka*, the inanimate number 'one', can occur as head, as can *deka* 'two', or any other number under the appropriate circumstances:

- (18) oyaa iww-a      eeka      rahayi  
 you cook\PAST-REL 3SG.PRON tasty-PRED  
 'That one you cooked is/was tasty.'
- (19) oyaa iww-a      eka      rahayi  
 you cook\PAST-REL 3SG.PRON tasty-PRED  
 'The one you cooked is/was tasty.'<sup>10</sup>

Number can also be expressed with SL Malay headless clauses. As Nordhoff remarks, "NPs based on headless relative clauses can only be modified by the plural marker *pada...*, but by nothing else." (396). This is illustrated in (19):

---

<sup>10</sup>Clauses with *eka* can lead to ambiguity, since action nominalized clauses of this form are of the same shape, but do not necessarily have a gap co-indexed with *eka*. Thus (18) could also have the meaning 'your cooking (the activity) is tasty' which is ruled out pragmatically in this case. Cf. Exx. (32-33).

- (20) [Seelon=nang anà-dhaatang Ø] pada mlaayu pada.  
 Ceylon=DAT PAST-come PL Malay PL  
 ‘Those who had come to Ceylon were the Malays’ (454, ex. 166)

The parallel here is interesting, and especially the occurrence of *pada* attached to the entire clause, but apparently relating to the null element.

### 7.3.4 Correlatives use a WH rather than a correlative form of the Indo-Aryan type and are generally restricted to indefinite or conditional contexts and commonly occur with a sentence particle (dubitative or question) on the subordinate clause.

This is apparently not a feature of SLM, which is not surprising, since the model is not found in either of the Tamil or Sinhala adstrates in the varieties to which the Malay speakers would have been exposed. Sinhala does have a correlative structure, but it is characteristic of the Literary variety and has features like the Dravidian languages. It uses a correlative pronoun derived from the Old (and Middle) Indo Aryan *ya-* forms, rather than an interrogative as in Dravidian, but like Southern Dravidian it is generally confined to the “who-ever, what-ever” sense, and the correlative clause occurs with the question particle *da* or with the conditional (“if”) form *nam*.

### 7.3.5 Sentence-Final quotative from “say”

The SSLA languages employ a sentence/clause final quotative affix or particle to mark indirect or direct speech as well as related functions such as thought or intention. It is characteristically derived from ‘say’ but in a frozen form or one from a ‘say’ verb no longer in active use, as with Tamil *enRu/(e)NNa*. In Sinhala, the quotative is *kiyalaa*, which is homonymous with the conjunctive particle of *kiyanawaa*, but it has a separate range of distribution specific to the quotative use. That range is partially exemplified in (21), (22), and (23):

- (21) [siri iiye aawaa kiyalaa] gunapaala kiwwa  
 Siri yesterday came QUOT Gunapala say\PAST  
 ‘Gunapala said that Siri came yesterday.’
- (22) ma-ta siri kiyalaa kiya-nawaa  
 I-DAT Siri QUOT call-PRS  
 ‘They call me “Siri”.’
- (23) raksawa hoyaa-ganna kiyalaa kolañba-ta giyaa.  
 job seek-find QUOT Colombo-DAT go-PAST  
 (I) went to Colombo in order to find a job.’ (i.e., with the intention)

SL Malay has an equivalent form, with much the same range of functions, partially illustrated in (24), (25), and (26). It is not derived from a “say” verb as such, but rather from a Malay word for “word” but it is clearly modeled on the Sinhala/Tamil form (Nordhoff 396). (24) through (26) are examples:

- (24) Se=pe oorang thuuya pada anà-biilang [kitham pada Malaysia=dering  
 Is=POSS man old PL PAST-say 1PL PL Malaysia=ABL  
 anà-dhaathang] katha  
 PAST-come QUOT

'My ancestors told me that we had come from Malaysia.' (401, ex 465)

- (25) [Aashik=nang hathu soldier mà-jaadi suuka]=si katha arà-caanya.  
 Ashik=DAT INDEF soldier INF-become like=INTERR QUOT NONPAST-ask  
 'He asks if you want to become a soldier, Ashik.' (393, ex..480)
- (26) See=yang Tony katha arà-panggel  
 I=ACC Tony QUOT NONPAST-call  
 'I am called "Tony".' (390, ex. 466)

### 7.3.6 Sentences may be nominalized without genitivization (or deletion) of subject, by employing a sentence-final form or verbal affix.

In the Southern Dravidian languages, the general pattern has a relative clause of the type discussed in in 7.3.6 headed by a pronominal form, characteristically 3<sup>rd</sup>. person neuter or some other non-agreeing form, as in (27) from Tamil, where the pronominal form appears as affixed to the relativizing form of the verb.

- (27) *avan va-nt-atu nallatu*  
 he came-REL-NMLZ good  
 'It's good that he came.'

In essence, the nominalizing form nominalizes the entire sentence, producing an action nominal. These differ from relative clauses modifying a pronominal head in that there is no necessary gap within the sentence; i.e., they are complete, aside from the ever-present possibility of ellipsis in these languages. However, if the sentence does include a gap (i.e., a null pro-form), it will not be co-indexed with the head, but will generally have external reference. In present day Sinhala, the pronominal form is *eka*, the inanimate numeral 'one', as in (28) and (29). *Eka* is also used to adapt foreign, especially English, loans, as in *kaar-eka* 'the car'.

- (28) [Silva mahattaya ma-ta eeka kiww-a-eka] ætta  
 Silva gentleman I-DAT that say\PAST-ADJ-EKA truth  
 'It is true that Mister Silva said that to me.'

- (29) eyaa aaw-a-eka hoñda-yi.  
 he come\past-ADJ-eka good-PRED  
 'It is good that he came.'

In all of these, the entire sentence is intact and in the same form as an independent equivalent except for the relativizing (adjectival) verbal affix and the pronominal form or *eka*. There is no change in internal case marking. Thus in (28) and (29), the subjects are in the nominative, as in an independent sentence equivalent. Thus also, in (30) the dative case marking of the subject, required by the verb *teerenawaa*, is retained as is that of the direct object (required by the verb) in (31):

- (30) lamayaa-ṭa teer-un-a-eka hoñda-yi.  
 child-DAT understand-PAST-ADJ-eka good-PRED  
 'It is good that the child understood.'

- (31) *[noonaa lamayaa-ta banin-a-eka] puduma-yi*  
lady child-DAT scold-ADJ-eka surprising-PRED  
‘It is surprising that the lady scolded the child.’

*Eka* can also function as a co-indexed pronoun, so that given the possibility of null pronouns in these languages there is the possibility of homonymous sentences, as in (32), in which *eka* is the co-indexed neuter singular relative clause head co-indexed with a gap, but which can also have the structure and reading in (33), in which *eka* is the sentence nominalizer and the direct object is represented by a null indefinite pronoun.

- (32) *gunapaala ammaa-ta (Ø)<sub>i</sub> dunn-a-eka*                  *maŋ dækkaa*  
Gunapala mother-DAT give\PAST-ADJ-ONE.INANIM I saw  
‘I saw what (the one) Gunapala gave mother.’

or:

- (33) *gunapaala ammaa-ta (Ø)<sub>indef</sub> dunn-a-eka*                  *maŋ dækkaa*  
gunapaala mother-DAT give\PAST-ADJ-eka I saw  
‘I saw that Gunapala gave (something) to mother.’

Malay nominalized sentences exist, but they are simpler than those in Sinhala and the other SSLA languages, since they lack any overt nominalizing element. As Nordhoff reports: “in SLM, clauses can be used as noun phrases as they are. No further morphological flagging of this use is necessary” (450). They may also occur with case postpositions, as in (34):

- (34) *Suda butthul suuka [[asà-dhaatang]<sub>CLS</sub>]<sub>NP</sub>=nang.*  
this very like CP-COME=DAT  
‘So, I am very pleased that you have come.’ (452, ex.158)

In (34) above, the verb was in the conjunctive participle form, but an uninflected verb is possible (35):

- (35) *[Manis-an maakang]<sub>CLS</sub>]<sub>NP</sub>=nang go*                  *suuka bannyak.*  
Sweet-NMLZ eat=DAT                  1S.FAMILIAR like much  
‘I like very much to eat sweets.’ (452, ex.159)

In (34) and (35), there is no overt subject in the nominalized clause, but SL Malay, as in the other SSLA languages, the subject may be overt, with no change of case, as in (36):

- (36) *kitham=pe baapa su-biilang [[lorang suurath=yang mlaayu=dering*  
1PL=POSS father PAST-SAY 2PL letter=ACC Malay=ABL  
*anà-thuulis]<sub>CLAUSE</sub>=nang bannyak arà-suuka].*  
PAST-WRITE=DAT much SIMULT-LIKE  
‘Daddy said that he liked very much that you wrote the letter in Malay.’ (450, ex.150)

SL Malay does have a nominalizing suffix *-an*, but this appears primarily to serve to create nominal forms from verbs, such as the noun *aajar-an* ‘teach(ing)’ from the verb *aajar* ‘teach’, and not to nominalize sentences (37).

- (37) *Lorang=nang see=yang ingath-an=si?*  
 2PL=DAT 1S=ACC think-NMLZ=INTERR  
 ‘Do you have thoughts on me/are you thinking of me? (512, ex. 36)

As pointed out earlier in Section 7.3.3, SL Malay has true headless relatives, in which there is no form expressing a head, and these could be considered, from their distribution and ability to occur with case, to be a kind of nominalized sentence.

### 7.3.7 Focused (Nominal Cleft) Sentences, including those with rightward focus

The basic pattern for the Southern Dravidian languages is for the presupposition to be expressed in a nominalized clause headed by a pronominal form, characteristically 3rd person neuter.<sup>11</sup> It thus resembles or is identical to a nominalized sentence of the kind described in Section 7.3.6, but will contain a gap linked to the focused element. The focused element commonly follows the presupposition, but different languages may also allow focus *in situ* or in other orders.<sup>12</sup> The basic pattern for rightward focus is thus as in (38), where the content of NMLZ varies from language to language and where  $XP_i$  represents the focused element. It is exemplified for Sri Lankan (Jaffna) Tamil in (39):

- (38) [s[s...Ø<sub>i</sub> ... v-TNS-REL-NMLZ]  $XP_i$  ]
- (39) *naan poonatu yaasppaanat-ukku*  
 I go-PAST-NMLZ Jaffna-DAT  
 ‘It was to Jaffna that I went.’ (Sri Lankan Tamil)

The pattern was borrowed into Sinhala at least by the eighth or ninth century, and was subsequently elaborated to become an integral feature that intersects with other patterns such as negation and both WH and yes-no interrogatives, and it is of very frequent occurrence in discourse.<sup>13</sup> In Sinhala, the verb is not marked by a pronominal form, but by a focusing (sometimes called “Emphatic”) affix *e(e)*. That affix, however, does indeed derive historically from a 3<sup>rd</sup> person masculine/neuter pronominal affix, although it is now specialized for focus/clefts and one kind of negation. Colloquial Sinhala examples are given in (40) and (41). As (42) shows, Sinhala allows items to be focused *in situ*, especially when marked by one of a set of focus-inducing forms, including the question particle and the reportative clitic *lu* described in Section 7.3.10 below among others. This characteristic is shared to a varying degree by some of the SSLA languages, and was referred to in Section 7.2 above as such.

- (40) *mama giy-ee gama-ta*  
 1S go\PAST-FOC village-DAT  
 ‘It is to the village that I went.’

<sup>11</sup>While the general pattern is clear, there are variations across languages. Thus Tulu and Dhivehi utilize focus affixes on the verb, but they are not clearly pronominal (Somashekar 1999, Cain & Gair 2000). Such focused sentences are also found in Dakhini Hindi, and Sri Lanka Portuguese shows a form of this structure as well.

<sup>12</sup>Where in some of the languages including Sinhala, the focused element occurs internally, as *in situ*, there will of course not be a gap co-indexed with a rightward form, but the focused element is marked in some fashion, commonly a clitic such as an interrogative or emphatic one.

<sup>13</sup>I have dealt with this at length in several places, beginning with Gair (1970[1998]). For the history, see especially Gair (1986), and I will not repeat that here. A similar elaboration of the interaction with structures such as WH questions occurred apparently independently in Malayalam, but not in Tamil.

- (41) *mama kiyeww-e ee pota*  
 I read\PAST-FOC that book  
 ‘It was that book that I read.’
- (42) *eyaa heṭa=da kolañba yan-ne?*  
 (s)he tomorrow=INTERR Colombo go.PRS-FOC  
 ‘Is it tomorrow that (s)he is going to Colombo?’

Despite the extensive information provided by Nordhoff, and his inclusion of work specifically on focus, it is not yet clear to me whether focusing sentences of this general type or a variant are present in SL Malay, but there are some interesting and suggestive hints as to the possibility that this might be the case. SL Malay constituents can be focused by attaching the emphatic clitic *jo*, as in (43) (Nordhoff 379).

- (43) *[itthu katha]\_UTT=jo Mahinda arà-biilang.*  
 DIST QUOT=EMPH Mahinda NONPAST-say  
 ‘That’s what Mahinda [Rajapaksa, President of Sri Lanka] is saying.’ (379, ex.420)

A Sinhala equivalent could be formed using the emphatic form *tamayi*, which is one of several forms requiring clefting, i.e., the focusing affix *e(e)* on the verb:<sup>14</sup>

- (44) *eeka tamayi mahinda kiyan-ne*  
 that EMPH Mahinda say.PRS-FOC  
 ‘That’s what Mahinda is saying.’

One would naturally not expect similar verb marking in SL Malay, since it lacks the verb marking form. Tamil would not mark the verb for cleft function in a parallel sentence, but would simply use the emphatic/reflexive clitic *taan* unless special focusing was required.<sup>15</sup> Although the focusing verb form is not found in SL Malay, as Nordhoff points out (p.275) there is an intriguing pattern of co-occurrence of past affixes with focus that is reminiscent of the Sinhala situation. There are two past affixes *su-* and *anà-*. The differences between them are subtle, and they are interchangeable in many contexts (*ibid*). However, in past questions where a constituent is emphasized by the question particle, (Nordhoff’s polar questions) *su-* is ruled out, as in (45):

- (45) *Daging baabi=si \*su/anà-billi???*  
 meat pork=INTERR \*su/anà-buy  
 ‘Did you buy PORK? (274, ex.16)

As Nordhoff also points out, the occurrence of the question particle *si* here rules out the possibility of *su-*, but *si* can occur in simple past questions:

- (46) *Itthu binaathan lorang=yang anà-/su-giigith=si*  
 DIST animal 2PL=ACC PAST/PAST-bite=INTERR  
 ‘Did that animal bite you?’ (275, ex.19)

<sup>14</sup>*tamayi* itself is a complex form composed of an emphatic/reflexive form *tama* and the predicate/emphatic marker *-yi*. The former does not induce the focus affix on the verb, the latter does when occurring on a constituent other than the verb.

<sup>15</sup>Sinhala has another emphatic clitic *-ma*, which does not induce the verb marking, but would have a different sense, emphasizing the identity of Mahinda, rather than his participant function in the action. SL Malay *-jo* seems to encompass that function as well.

This non-occurrence of *su-* also to be the case with Content (Wh) questions (3.7.10), which require the focusing affix in Sinhala:

- (47) *mana binaathang lorang=yang anà/\*su-giigith?*  
 which animal 2PL=ACC past-bite  
 ‘Which animal bit you?’ (274, ex.17)<sup>16</sup>

Interestingly, unlike Sinhala, the question particle apparently does not occur on WH forms in SL Malay, and this as also the case in Tamil. In the pluperfect, however, *su-* is the only form that can be used, and *anà* is ruled out. (*Ibid.*) Nordhoff notes also that it is possible that *anà* could have a specification for clauses with argument focus, since it often occurs with the emphatic marker *jo*, as in (48): (*Ibid.*)

- (48) *TV=ka=jo anà-kuthumung*  
 TV=LOC=EMPH PAST-see  
 ‘It was on TV that we saw it.’ (275, ex.21)

Nordhoff states:

This distinction according to information structure would also make sense from a contact language perspective, since Sinhala has an ‘emphatic’ verb form used in focal contexts (Gair 1985)

However, as he goes on to say:

Things seem to be more complex in SLM than in Sinhala, though. While in Sinhala, the use of the emphatic [our ‘focusing’] form [of the verb] is obligatory in focal contexts, this is not the case with *anà-* in SLM. True, most argument focus constructions with past reference have *anà-*, but there are some examples where we find *su-* as well”. (p.240)

At the very least, this invites further investigation and explanation. Nordhoff also states that focus can also be indicated by rightwards extraposition, as in (49):

- (49) *Itthu=nang blaakang su-dhaatang [Hambanthota mlaayu pada]*  
 DIST-DAT after PAST-come Hambantota Malay PL  
 ‘After that came the Hambantota Malays.’ (691, ex. 29)

(50) and (51) seem to be similar, though they were not singled out as examples of focus:

- (50) *[Seelon=nang anà-dhaatang Ø] pada mlaayu pada.*  
 Ceylon=DAT PAST-come PL Malay PL  
 ‘Those who had come to Ceylon were the Malays’ (454, ex. 166)

- (51) *Itthu vakthu [kithang=nang nya-aada Ø] asàdhaathang ini JVP katha*  
 DIST Time 1PL=DAT PAST-exist ] COPULA PROX JVP QUOT  
*hathu problem*  
 INDEF problem  
 ‘What we had at that time was the so-called JVP problem.’ (454, ex. 168)

<sup>16</sup>The difference in the final consonant in *binaathan(g)* between this and other examples (cf. (47)) may be due to a typo.

As mentioned in Section 7.3.6 above, the form of the presupposition in Dravidian clefts is essentially that of an action nominalized sentence of the kind described in 7.3.6, but it includes a gap (null item) co-indexed with the focused element. SL Malay nominalized sentences and relative expressions do not, however, have overt morphological marking.<sup>17</sup> Thus if there were to be developed a cleft/focus construction parallel to that of Sinhala and the other languages as a result of contact, we could not expect it to have the same morphological characteristics, given their absence in the language. SL Malay does, however have true headless relatives. Thus if we were to speculate, the existence and nature of these headless relatives would seem to make them a likely candidate for expressing presuppositions, since they have an empty slot, parallel to the gap in the Sinhala and Tamil presuppositions (the Ø in examples (50) and (51)). Indeed, example (49), presented by Nordhoff as an example of focus, does seem to be analyzable as containing such a clause, and correspondingly, (50) and (51), which exemplified headless relatives, seem to involve something like the focus in (49). Thus we might very tentatively conclude that cleft / focused structures on the SSLA pattern do exist in that language, even if in inchoate form. Establishing this would require investigation of their role in discourse and relation to other sentence types, but in any case, SL Malay seems to have all of the building blocks for such a development to take place. In a footnote (p.690, fn.1), Nordhoff points to the existence of one speaker who has a kind of pseudo-cleft construction involving the copula *a(bbi)sdhaatang*, which is shown in the following two examples (52) and (53).

- (52) *See anà-pass.out] abbisdhaathang University of Peradeniya=ka.*  
 1s PAST-graduate COPULA University of Peradeniya=LOC  
 Where I graduated was the University of Peradeniya.' (690, fn. 1)
- (53) a. *[Ithu arà-kirja] abbisdhaatang,*  
 DIST NONPAST-make COPULA  
 'How you make it is'  
 b. *thullor asà-ambel=apa baaye=nang asà-puukul=apa*  
 egg cp-take=after good-DAT cp-hit=after  
 'You take eggs and beat them well, and...' (690, fn. 2)

These really do resemble Sinhala clefts, though the role of *abbisdhaathang*, not clearly having a counterpart there, is puzzling. Nordhoff remarks that the construction seems to be idiolectal, but in any case it shows that the development of a SSLA type cleft/focus structure under apparently obvious contact is a real possibility, and in part realized, at least in this one case. Clearly this is another inviting topic for further exploration.

---

<sup>17</sup>See sections 7.3.3 and 7.3.6. Slomanson (2006:147) states that Colombo SL Malay has a suffix *-nya*, which nominalizes sentences, as in (his 21b)

- (1) *[Ali ara-pi-nya] sE ara-liyat*  
 Ali PRS-go-NMLZ 1s PRS-see  
 'I see Ali going.'

This does in fact resemble Sinhala and Tamil. I cannot find this form in Nordhoff, except as a dative, as in ex. 110, p.531, and as an accusative, in the same example, however, so that it may represent a difference in dialect. It may also be an allomorph of a different form. See Nordhoff (2009:529) on this form.

### 7.3.8 Negation

There are a number of points in which SL Malay negative distribution bears resemblance to Sinhala, as well as to Tamil. The details are complex, given that each of the languages has a complex system of negation involving several forms and their interaction with other factors such as sentence type, aspect, and, tense, as well as with independent versus subordinate status (for SL Malay, see the helpful survey and table given by Nordhoff ([this volume](#))). A full investigation would also require a consideration of Tamil and the other Dravidian languages that time considerations did not allow me to undertake, so I shall not take this up in detail here, but save it for later and separate treatment. We might note, however a few points in passing. In both Sinhala and SL Malay, negation varies with the type of clause, and there are different negators for verbal, equational, attributive and existential predicates. In Sinhala, NP predicates negate with *nemeyi* (also *nevi* and *neveyi*), as in (54), and in SL Malay with *bukang*, as in (55).

- (54) *mee pota puskola potak nemeyi.*  
 this book ola.leaf book NEG.NONV  
 ‘This is not an ola leaf manuscript.’
- (55) *Deram Islam oorang bukang (SL Malay)*  
 3PL Islam man NEG.NONV  
 ‘They were not Muslims.’ (259, ex.406)

Adjectival predicates in Sinhala negate with *nææ*, as in (56), and in SL Malay with *thraa*, as in (57), or the prefix *thàrà-*, as in (58).

- (56) *mee pota hoñda nææ (Sinhala)*  
 This book.DEF good NEG  
 ‘This book is not good.’
- (57) *Itthu muusing gampang thraa. (SL Malay)*  
 DIST time easy NEG  
 ‘It was not easy back then.’ (258, ex.302)
- (58) *Itthu thàrà-baae*  
 this NEG-good.  
 ‘This is not good.’ (297, ex.99)

Nordhoff (p.224) remarks that “The exact conditions that trigger one or the other adjectival negation [in SL Malay] are unclear.” Nordhoff correctly notes, in a footnote (p.221, fn 27), that Sinhala has two negators with similar uses to SL Malay *thraa* and *bukang*, further noting that Indian Tamil has lost the distinction though it is retained in Jaffna Tamil. It would be interesting to find whether the distinction holds in any other variety of Sri Lanka Tamil that might have been in contact with SL Malay. Independent non-focused Sinhala verbal sentences in all tenses, negate with *nææ*:

- (59) *mama ada wæda kəranne nææ*  
 I today work do NEG  
 ‘I am not working today’,

In SL Malay, the negation of verbal sentences varies with tense and aspect. In the non-past, the form is *thuma-* or *thumau-*, (60) and in the past, *thàrà-* (61):

- (60) *Go kaapang=le saala thamau-gijja.*  
 1S.FAMILIAR when-ADDIT WRONG NEG.NONPAST-do  
 ‘I never do wrong.’ (299, ex. 107)
- (61) *Mister Yussuf thàrà-siggar=le*  
 Mister Yussuf NEG-well=ADDIT  
 ‘Mister Yussuf was also unwell.’ (297, ex.100)

This is not the case in Sinhala. There are other points of resemblance, such as that existential sentences negate by replacing the existential verb with a negative form, and that the perfect also negates by replacing the existential light verb, but I will not pursue those here, except for one set to be dealt with below in connection with focus and constituent negation. The list of SSLA features also included the appearance of negative verbs in subordinate clauses. Sinhala has a negative verbal prefix *no-*, which may be used with non-finite or dependent verbs, including the focusing verb form and the finite verb if it is in an embedded sentence. Comparison with SL Tamil or Malayalam in regard to the negative verb or prefix is especially difficult owing to the complexity of all of the systems, and I will not pursue it here, but the occurrence of the *no-* prefix on focusing verbs in Sinhala relates to the negation of cleft/focusing sentences, and it is illustrated in (62). The effect is essentially to exclude the focused element from the positive equivalent of the presupposition. Sinhala focused sentences can also negate using *nætte*, a focusing form of *nææ*, as in (63).<sup>18</sup>

- (62) *mama no-ka-nn-e harakmas witara-yi*  
 I NEG-eat-pres-foc beef only-PRED  
 ‘It’s only beef that I don’t eat.’
- (63) *minihaa ya-nn-e næ-tte gamə-tə*  
 man go-FOC not-FOC village-FOC  
 ‘It’s to the village that the man does not go.’ (DeAbrew 1981)

Sentences negating the focus i.e, clefted sentences with constituent negation, negate with with *neveyi/nemeyi*, as in (64).

- (64) *iiye gunapaala-ta salli dunne ee minihaa nemeyi*  
 yesterday Gunapala-DAT money gave-foc that man NEG.NONV  
 ‘It was not that man who gave Gunapala money yesterday.’

Nominal predicate sentences also negate with *nemeyi*, as stated earlier:

- (65) *mee pota puskola potak nemeyi*  
 this book olaleaf book-INDEF NEG.NONV  
 ‘This is not an ola leaf manuscript.’ (=54)

Here, there is a striking similarity with SL Malay. Nordhoff states (p.588) that *bukang* is used to negate both nominal predicates and constituents, as in (66) and (67).

- (66) *Deram Islam oorang bukang (SL Malay)*  
 3PL Islam man NEG.NONV  
 ‘They were not Muslims.’ (=55)

---

<sup>18</sup>Sinhala negation is a rich subsystem involving several forms with complex distributional characteristics. An account of the forms and semantics is given in DeAbrew (1981), and some elaboration is presented in Foley & Gair (1993).

- (67) *Thaangang=dering bukang kaaki=dering masà-maayeng*  
 hand=INSTR NEG.NONV foot=INSTR must-play  
 ‘You must not play with the hands, but with the feet.’ (299, ex.408)

The list of features relevant to the SSLA area included one saying that cleft sentences negated like nominal equational ones (2.1.8.3),<sup>19</sup> and despite the possible absence of marked cleft sentences of the Dravidian-Sinhala type in SL Malay, the parallel just mentioned and illustrated in (67) is suggestive, and appears worthy of investigation in relation to a previously unnoted effect of contact on negation.

### 7.3.9 Conjunctive Participles

Conjunctive participles may occur with overt lexical subjects, not co-indexed with main subject (or agent). In Sinhala example (68), the conjunctive participle clause has an overt subject not coreferential with that of the main clause, and it is in the same case that it would have in an independent sentence.<sup>20</sup>

- (68) *ammaa gihil-laa ma-ta seerama gedara-wæða kéra-nna oona*  
 mother go-CP I-DAT all house-work do-INF necessary  
*unaa*  
 become.PAST  
 ‘With Mother gone, I had to do all the housework.’

In SL Malay, the same possibility holds, as noted by Nordhoff (p.464), who provides the following example:

- (69) a. *Go asà-niiŋgal*,  
 1s.FAMILIAR CP-die  
 ‘I having died’  
 b. *Alla go=nya asà-dhaathang*,  
 Allah 1s.FAMILIAR=DAT CP-come  
 ‘Allah having come towards me.’  
 c. *kuburan asà-gaali*,  
 grave CP-dig  
 ‘The grave having been dug’  
 d. *Go=nya kubur-king!*  
 1s.FAMILIAR=ACC bury-CAUS  
 ‘Bury me!’ ‘I die and Allah comes for me and the grave will be dug and they will have me buried.’ (539, ex110a)

### 7.3.10 Reportative/hearsay particle

A form commonly referred to as a reportative or hearsay particle or affix is a feature of the SSLA languages, including SL Malay.<sup>21</sup> It may also have evidential force indicating knowledge not directly known. Nordhoff refers to it an evidential clitic (pp. 337ff),

<sup>19</sup>This actually referred to the negation of the focus, since there are other forms of negation in Sinhala clefts, as illustrated earlier in (62) and (63).

<sup>20</sup>See especially Gair (2005) and McFadden & Sundaresan (2010) for case and coreference in relation to finiteness.

<sup>21</sup>Malayalam may be an exception here, Bashir (2006:14) remarks that “in Malayalam evidentiality distinctions are not morphologically encoded”, but she does give an example with a verbal noun with the

and that is certainly an important aspect of its character. For Sinhala speakers, is commonly taken as having the English translation equivalent ‘it seems’, though the hearsay implication is stronger in Sinhala, which generally carries the sense that someone has said/heard it. The unmarked position for its occurrence is sentence final, but in some languages, as in Sinhala, it may occur on internal constituents as well. Other languages within the general South Asia area may, of course have forms with similar functions, but they are generally of a different form and location. Marathi and Dakkhini Hindi, however, do have such forms, clearly as a result of Dravidian influence. A Sinhala example is given in (70).

- (70) *poliisiy-ej gunapaala-ta hariyata gæhuwaa=lu.*  
 police-INSTR Gunapala-DAT really hit\PAST=EVID  
 ‘The police really beat Gunapala, they say/it seems.’

(This example also illustrates a SLM feature shared with Sinhala that is not in the list; the use of the instrumental form in “corporate” subjects.) The equivalent in SL Malay is given by Nordhoff as *kiyang*, or *keyang*, and as he notes, others have given *kanyang* (p. 337). Like Sinhala *lu*, *kiyang* can mark hearsay or information for which the speaker does not take responsibility. Nordhoff gives several examples with somewhat different implications, including (71):

- (71) *Seelong Airport=yang duva-pulu-umpath vakhtu=le asà-bukka arà-simpang kiyang*  
 Ceylon Airport=ACC two-ty-four hour=ADDIT CP-open NONPAST-stay EVID  
 ‘The Ceylon Airport is open 24h, it seems.’ (388 ex. 458)

SL Malay *kiyang*, or its variants, have generally been described as a sentence final form that must attach to the predicate, as is said to be the case with Tamil –*aam*, on which it is said to be modeled, as in Bakker (2006:142) or Smith & Paauw (2006:175).<sup>22</sup> In Sinhala, *lu* commonly appears sentence final, with scope over the entire sentence, but it may also occur on internal constituents. It is a focus-marking or more precisely, perhaps, a ‘Cleft-linked’ or ‘Cleft-requiring’ form, along with a number of others. That is, its appearance on an internal constituent requires the cleft/focus

---

evidential function:

- (1) *Rāman-re acchan -i vīTunirmmiccu kēTTu*  
 Raman-GEN father(NMLZ) this house build(VERBAL NOUN)  
 ‘Raman’s father built this house’ (Speaker has learned this from a third party).

<sup>22</sup>Whether it is the case that –*aam* must occur on the last constituent or the predicate, is not entirely clear. The account in Gair et al. (1978:214), which calls it a ‘quotative marker’ also has “usually” Suseendirajah (1993:148) for Jaffna Tamil, simply says that it occurs “finally”. Schiffman (1999:151) however, says that –*aam* can be added to various constituents to indicate that the speaker does not claim responsibility for the veracity of the statement, but merely reports something. He states that It is usually added to the last constituent of the sentence, but adds that “it can also occur somewhat idiomatically (or ironically) in other places in a sentence, e.g. with reduplicated noun phrases”;

- (1) *periya ivar-aam ivaru*  
 big this.man.HON-EVID this.man.HON  
 ‘Well, la-de-da, get a load of him.’ (i.e., he thinks he’s hot stuff.)

construction described in 7.3.7 as requiring a special verb form. As such, it is part of a complex system that is very much integrated into Sinhala structure, and which I have described elsewhere (cf. footnote 12). Nordhoff notes that it is difficult to elicit *kiyang*, and that is quite understandable, given its semantics, discourse function, and specificity. It is, as he points out optional, to which he might have added that it is triggered by the discourse situation and the wider context, which are difficult to induce. He does provide, however, examples of sentences elicited using Sinhala (Exx. 455-457, p. 338) that show a distribution like Sinhala. They are given here as (72)-(74), and I have provided a separate item-by item gloss for Sinhala. Note that (73b) does not accord with the characterization referred to above of *kiyang* as necessarily sentence final.<sup>23</sup>

- (72) a. *Haturaa balahatkaaray-eij lamun bañdavaagannavaa=lu* (*Sinhala*)  
enemy force-INSTR children recruit=EVID  
'The enemy is recruiting children, it seems.'<sup>24</sup> (387, ex. 455)
  - b. *Satthuru paksa aanak pada arà-kumpulkang kiyang.* (*SL Malay*)  
enemy force children PL NONPAST-recruit EVID  
'The enemy is recruiting children, it seems.' (387, ex. 455)
- (73) a. *Haturaa=lu balahatkaaray-eij lamun bañdavaaga-n-ne.* (*Sinhala*)  
enemy-EVID force-INSTR children recruit-PRS-FOC  
'The enemy, it seems, is recruiting children.' (387, ex 456)
  - b. *Satthuru=kiyang paksa aanak pada arà-kumpulkang.* (*SL Malay*)  
enemy=EVID force child PL NONPAST-recruit  
'The enemy, it seems, is recruiting children.' (387, ex. 456)
- (74) a. *Haturaa balahatkaaray-eij lamun=lu bañdavaaga-n-ne.* (*Sinhala*)  
enemy force-INSTR children=EVID recruit-PRS-FOC  
'The enemy is recruiting CHILDREN, it seems.' (387, ex. 457)
  - b. *Satthuru paksa aanak pada=kiyang arà-kumpulkang.* (*SL Malay*)  
enemy force child PL=EVID NONPAST-recruit  
'The enemy is recruiting CHILDREN, it seems.' (387, ex. 477)

Nordhoff presents these with a due cautionary statement since they were elicited using Sinhala *lu*, but if they do in fact represent sentences that could occur in natural SL Malay discourse, they are further striking evidence of detailed and complex assimilation to Sinhala.<sup>25</sup>

## 7.4 Summary

It is clear that SL Malay, in the time since its arrival on the island of Lanka, has adopted through contact a number of features characteristic of the languages of that island,

<sup>23</sup>This was pointed out by one reviewer, for which I am grateful. In this case, of course, the placement could simply be calquing the eliciting form (72a), but there is a possibility that there is a parallel with the characterization of the placement of *-aam* as in the preceding footnote. In any case this is beyond the scope of this paper and invites further research.

<sup>24</sup>I would take this as 'The enemy is recruiting children by (illegal or improper) force.'

<sup>25</sup>Nordhoff notes that unlike in Sinhala *kiyang* does not trigger a special verb form, but appears with different past forms. However, the Sinhala verb form he refers to is not specific to *lu*, but is a central part of the cleft/focus system mentioned earlier, and it does not appear to have a SL Malay equivalent, so that its non-appearance here is not surprising.

whether Sinhala or Tamil, most likely both. These Sri Lankan languages, along with Dhivehi, in turn are part of a larger South-South Asia linguistic area, the features of which are essentially South Dravidian in origin. However, It should be recalled that the South-South Asia linguistic area, in turn, forms a part of a larger South Asian one, and thus shares the relevant general features of that larger area as well.<sup>26</sup> In the case of Sinhala and Dhivehi, of course, those features were carried over as inheritances from their mainland source. This is not the case for the non-SA imports SL Malay, or SL Portuguese, which, as stated at the beginning, were typologically very different from the South Asian ones, and are especially interesting for that reason. My concern here was with the extent to which SL Malay participated in the SSLA sprachbund, and features not specific to that subarea have not been dealt with directly. In investigating contact induced typological change in SL Malay however, it is obvious that those more general features are of equal importance with those specified here, since they were all characteristic of the donor languages. As one crucial example of such a vitally important feature, we may note basic SOV and right-headedness, and in fact the strong right-headedness that is characteristic of SS Asia and underlies several of the features here. One interesting aspect here is that this did not extend to the morphological level in SL Malay, though it does pervade the syntax. This was noted here earlier in relation to the conjunctive participle and temporal prefixes developed under influence from post-element forms. Something similar also appears to be the case with SL Portuguese, as noted by Ian Smith, who states in relation to SL Portuguese and Sourashtra (Smith 2001:408): “Structurally, Sourashtra was closer to Tamil, again because of a closer starting position; SLP showed more accommodation to Tamil, but had not modified the structure of some closely bound elements such as verbal prefixes.” This raises some interesting questions as to the variable resistance of aspects of the grammar to contact induced change, but that is beyond our concern here, which was simply to illustrate the extent to which SL Malay conforms to the SSLA pattern. In the case of Sinhala, and most likely Dhivehi, the Dravidian contact spanned millennia. For SL Malay, as for SL Portuguese, the time span was shorter, encompassing a few centuries at most (see papers in this volume for views on this for Malay). Thanks to the work of Sebastian Nordhoff and others represented here, SL Malay is now an unusually well documented language in regard to both its structure and its history and can serve as an unusually clear and important example of contact induced change resulting in a major typological shift in branching direction from a strongly right branching language to an essentially left branching one as part of a linguistic subarea in a relatively short span of time.

---

<sup>26</sup>An extensive presentation of syntactic typological characteristics of languages of India, and some extent wider South Asia, from four major families, covering both similarities and differences, is given in Subbarao (in press) and in part in his contribution in Kachru et al. (2008).

# Bibliography

- Bakker, Peter, 2006. “The Sri Lanka *Sprachbund*: The newcomers Portuguese and Malay.” In *Linguistic Areas – Convergence in Historical and Typological Perspective*, edited by Yaron Matras, April McMahon, & Nigel Vincent. Hounds-mills, New York: Palgrave Macmillan, 135–159.
- Bashir, Elena, 2006. “Evidentiality in South Asian Languages.” In *Proceedings of the LFG06 Conference, Workshop on South Asian Languages Konstanz Universität*. Stanford, 30–50.
- Bayer, J., 2001. “Two Grammars in One: Sentential Complements and Complementizers in Bengali and Other South Asian Languages.” In *The Yearbook of South Asian Languages: Tokyo Symposium on South Asian Languages - Contact, Convergence and Typology*, edited by P. Bhaskararao & K. V. Subbarao. New Delhi: Sage Publications, 11–36.
- Cain, Bruce D & Gair, James W., 2000. *Dhivehi (Maldivian)*. München: Lincom Europa.
- DeAbrew, K. Kamal, 1981. *The Syntax and Semantics of Negation in Sinhala*. Ph.D. thesis, Cornell University.
- Deumert, Ana & Durrleman, Stephanie (eds.), 2006. *Structure and Variation in Language Contact*. Amsterdam: John Benjamins.
- Emeneau, Murray B., 1956. “India as a linguistic area.” *Language* 32:3–16.
- Foley, Claire & Gair, James W., 1993. “The Distribution of *no-* in Sinhala.” Paper given at SALA, Iowa.
- Fritz, Sonia, 2002. *The Dhivehi Language: A Descriptive and Historical Grammar of Dhivehi and its Dialects*. Würzburg: Ergon Verlag.
- Gair, James W., 1970[1998]. “About Colloquial Sinhalese Clause Structures.” In *Studies in South Asian linguistics; Sinhala and other South Asian Languages*, edited by Barbara C. Lust. New York; Oxford: Oxford University Press.
- Gair, James W., 1985. “Sinhala Focused Sentences: Naturalization of a Calque.” In *South Asian Languages: Structure, Convergence and Diglossia*, edited by Bh. Krishnamurti, Colin P. Masica, & Anjani Sinha. Delhi: Motilal Banarsi-dass, 147–164. Reprinted in Gair (1998), pp. 153–169.
- Gair, James W., 1986. “Sinhala Focused Sentences: Naturalization of a Calque.” In *South Asian languages: Structure, Convergence and Diglossia*, edited by Bh. Krishnamurti, Colin P. Masica, & Anjani K. Sinha. Delhi: Motilal Banarsi-dass, 147–164. Reprinted, with additions, in Gair (1998), pp. 153–169, 323–327.
- Gair, James W., 1994. “Universals and The South-South Asian Language Area.” Keynote Address SALA (South Asia Language Analysis). University of Pennsylvania, May 20 1994.

- Gair, James W., 1997. "La Focalisation en Singhala." In *Faits de Langues, Revue de Linguistique: Les Langues d'Asie du Sud*, edited by Annie Montaut. Paris: Orphys, 155–162.
- Gair, James W., 2005. "Some Aspects of Finiteness and Control in Sinhala." In *The Yearbook of South Asian Languages and Linguistics*, edited by Rajendra Singh & Tanmoy Bhattacharya. Berlin and New York: Mouton-De Gruyter, 116–143.
- Gair, James W., 2009. "Focus, Clefts and Interrogatives in South-South Asia and Southern Island Indo- Aryan." Paper presented at the École dété (INALCO) Agay, June 8.
- Gair, James W., Suseendirarajah, S., & Karunatilaka, W. S., 1978. *An Introduction to Spoken [Sri Lankan] Tamil*. External Services Agency, University of Sri Lanka. Reprinted 2005, S. Godage and Brothers, Colombo.
- Kachru, Braj B., Kachru, Yamuna, & Sridhar, S. N., 2008. *Language in South Asia*. Cambridge: Cambridge University Press.
- Klaiman, M. H., 1977. "Bengali Syntax: Possible Dravidian Influence." *IJDL* 3:135–153.
- Masica, Colin P., 1976. *Defining a Linguistic Area: South Asia*. Chicago: University of Chicago Press.
- McFadden, Thomas & Sundaresan, Sandya, 2010. "DP Distribution and Finiteness in Tamil and Other Languages: Selection vs. Case." *JSAL* 4:1–25.
- Nordhoff, Sebastian, 2009. *A Grammar of Upcountry Sri Lanka Malay*. Ph.D. thesis, University of Amsterdam, Utrecht.  
URL <http://www.lotpublications.nl/publish/articles/003744/bookpart.pdf>
- Nordhoff, Sebastian, this volume. "Synchronic grammar of Sri Lanka Malay."
- Schiffman, Harold F., 1999. *A Reference Grammar of Spoken Tamil*. Cambridge: CUP.
- Slomanson, Peter, 2006. "Sri Lankan Malay morphosyntax: Lankan or Malay?" In [Deumert & Durrellman \(2006\)](#), 135–158.
- Slomanson, Peter, this volume. "Known, inferable, and discoverable in Sri Lankan Malay research."
- Smith, Ian, 2001. "Creolization and Convergence in Morphosyntax: Sri Lanka Portuguese and Sourashtra Nominal Marking Typology." In *The Yearbook of South Asian Languages and Linguistics*, edited by Rajendra Singh, Peri Bhaskararao, & K. V. Subbarao. New Delhi, Thousand Oaks, and London: Sage Publications, 391–409.
- Smith, Ian, this volume. "Hijacked constructions in Second Language Acquisition: Implications for Sri Lanka Malay."
- Smith, Ian R. & Paauw, S., 2006. "Sri Lanka Malay: Creole or convert." In [Deumert & Durrellman \(2006\)](#), 159–182.
- Somashekar, Shamitha, 1999. *Developmental Trends in the Acquisition of Relative Clauses: Cross-Linguistic Experimental Study of Tulu*. Ph.D. thesis, Cornell University.
- Subbarao, K. V., in press. *South Asian Languages: a Syntactic Typology*. Cambridge: Cambridge University Press.
- Suseendirarajah, S., 1993. *Jaffna Tamil: Phonology and Morphology*. Thirunelveliy, Sri Lanka: University of Jaffna Publication.
- Suseendirarajah, S., 2008. "Tamil Language in Sri Lanka and Some of its Archaisms

and Peculiarities.” In *Dakkhina: Papers on Buddhist thought, Linguistics, and Sri Lankan Culture (James Wells Gair Felicitation Volume)*. Colombo: Godage International publishers, LTD, 67–85.



## Chapter 8

# Hijacked constructions in Second Language Acquisition: Implications for Sri Lanka Malay

Ian Smith

*York University, Toronto*

### 8.1 Introduction

Most creolists accept that there is a cline between creole and non-creole. Categories of contact-languages are, like most human categories, fuzzy-edged and centred on abstract *prototypes*.<sup>1</sup> (Rosch 1975, 1978, Thomason 1997, Smith 2005b). The content of prototypical contact-language types is still a matter of negotiation and debate. Small wonder, then, that the position of Sri Lanka Malay in the typology of contact varieties is a contentious issue and the main *raison d'être* of this volume.

One of the problems with categorizing contact varieties in general is that the processes of grammar development in contact-induced change and in change without contact may lead to similar results (cf. the debate over Dravidian influence in Old Indo-

<sup>1</sup>I would like to acknowledge the support of the Social Sciences and Humanities Research Council for fieldwork on Sri Lanka Portuguese (1974-5) and Sourashtra (1989-90; 1992). I am grateful to Mohamed Jaffar, B.A. Hussainmiya, and Romola Rassool for their assistance with Sri Lanka Malay, to the late Ronald Rosario and Richard Starack for their assistance with Sri Lanka Portuguese, to Richard Starack, Jothy Jeeveratnam for assistance with Sri Lanka Tamil, to Mohamed Jaffar for assistance with Shonam, to Ranjhani Raghunathan for assistance with Indian Tamil, to K. S. Shanti, A. R. Kumar, K. Pasumpon, O. K. Ramanandam, V. V. Ramlakshmanji, R. R. Parimalam, and N. M. Omprakash for assistance with Sourashtra, and to Romola Rassool for assistance with Sinhala. Thanks to Bill Foley for comments on an oral presentation of a related paper. Standard abbreviations are used as prescribed by the Leipzig glossing rules and the list of abbreviations provided in this book. Additional abbreviations are as follows: Ind. Indian; OIA Old Indo-Aryan; Ptg. Portuguese; SL Sri Lanka; SLP Sri Lanka Portuguese. In transcriptions of Tamil allophonic voicing of stops is shown for the convenience of the reader, following Schiffman (1999) and others; word-final *a* and *e* are distinguished even though most speakers lack a surface contrast. (The distinction does appear in sandhi with a following vowel.)

Aryan – see [Munkwitz-Smith \(1995\)](#) for an overview). Within the category of languages that have developed through contact, differences are even harder to discern. Indeed, there may be no sharp division in terms of process between certain types of convergence<sup>2</sup> and creolization. Creoles are now recognized to involve processes of second language acquisition, in which influence flows from the L1 of the (untutored) learner to the L2 being learnt/created. One (complex) process involved in contact-induced language developments is the mapping of categories (or bundles of categories) from one language onto structures abduced from the other language ([Smith 1985a](#)). Although the sociolinguistic context influences the outcome of this process, an equally significant factor is the distinction between L1 > L2 influence and L1 < L2 influence. It is this distinction that I want to highlight in this paper. In Section 8.3 below I will elaborate on this distinction and its significance in distinguishing between convergence and creolization. First, however, it is necessary to lay the groundwork by explaining the relevance of the abduction process to untutored second language acquisition. After these preliminaries are covered, the paper will explore the Dravidian verbal noun in Tamil and four languages in contact with Tamil: Sinhala, Sourashtra, Sri Lanka Portuguese and Sri Lanka Malay.

## 8.2 Abduction and untutored second language acquisition

The relevance of Pierce’s term *abduction* to linguistics was first pointed out by Henning Andersen in a paper on morphophonological change ([Andersen 1973](#)). As Andersen explains, *abduction* is a third type of syllogistic reasoning, in addition to *deduction* and *induction*:

“Abduction proceeds from an observed result, invokes a law, and infers that something may be the case. E.g., given the fact that Socrates is dead, we may relate this fact to the general law that all men are mortal and guess that Socrates was a man. This inference differs essentially from the conclusions reached by inductive and deductive reasoning. Although it, too, is based strictly on its premises, it is not necessarily true, even though its premises are: if we have matched the given result with the wrong law, our conclusion may be false.” ([Andersen 1973:775](#))

Andersen applied the term to the intergenerational transmission of language, in which new learners must abduce rules of grammar from linguistic facts. Similarly, abduction is one of the processes that applies in second language acquisition. The primary role of learners in establishing their own grammar of the target has long been recognized and is the premise behind the concept of *interlanguage* ([Selinker 1972](#)).

In untutored second language learning, speakers of the target provide the initial data, but the analysis comes from learners who are guided by their L1. Since learners have no direct access to the grammars of target language speakers, the L2 grammars which they create need not resemble those of target language speakers. Specifically, learners need not ascribe the same grammatical status to the target language items they isolate as do target language speakers. Abduction underlies one of the forms of *transfer*, or the influence of a learner’s L1 on their version of L2, namely that in

---

<sup>2</sup>I prefer the traditional term *convergence*, to the more recent (and more narrow) *metatypy*.

which learners map categories (or bundles of categories) onto target structures that they isolate. The abduction inference is as follows (cf. [Smith 1985a:292-6](#)):

- (1) Fact: L1 expression and L2 expression are more or less pragmatically equivalent in certain contexts.  
“Law”: The L1 expression encodes specific semantics and has specific usage.  
Abduction: Therefore, the L2 structure encodes the semantics of the L1 expression.

Lefebvre and Lumsden's *relexification* (1994) encapsulates this process in the framework of a lexicon-based grammar.

The results of abduction can be indistinguishable from perfect acquisition when the L1 semantics matches that of L2 and learners successfully analyze the corresponding L2 structures. I do not claim, however, that abduction accounts on its own for second language acquisition, only that it is one of a number of processes involved. The results of abduction are observable when the L1 semantics do not match the L2 semantics of the analyzed form. In extreme cases, a grammatically and semantically incongruent L2 structure is *hijacked* - chosen by an abductive leap of faith to represent the L1 semantics. This phenomenon is rarely reported in the second language acquisition literature, probably because in a tutored setting it is usually quickly rectified. An example of such an abductive leap of faith comes from my own experience as a learner of French. In the first few months of study I isolated the holophrase [kiaj'krisa] as meaning ‘That’s wrong’; the source /ki a ekri 'sa/ in fact means ‘Who wrote that?’ (Qui a écrit ça?). The context in which my analysis developed was the teacher’s habit of reviewing homework students had written on the blackboard by pointing to answers that were wrong, asking, ‘Qui a écrit ça?’ Hijacking in a tutored second language acquisition context does not usually persist (I began to question my analysis following the teacher’s reaction when I tried to use the phrase to point out an error.) When a language is influenced by mass second language acquisition, however, the results of hijacking can persist. For example, the Singapore English clause-final objection particle *what*, seen in (2), encodes the semantics of Hokkien *ma* ([Smith 1985b](#)).

- (2) [Singapore English. ([Smith 1985b:111](#))] Context: a student wants to post a notice on the notice board  
 Student [to department secretary]: *Can I have some pins ah?*  
 Secretary: *Notice board got pins what.*  
 [*What* serves to deny the perceived presupposition that pins are not already available.]

A final example of hijacking is instantiation of the West-African focus marker or ‘highlighter’ function in Atlantic creoles. Holm’s summary of the situation is an eloquent characterization of hijacking:

“The creole highlighters represent a syntactic category in the substrate languages that does not correspond very closely to anything in the superstrate languages. The variety of forms that the highlighter has taken in the creoles suggests a ghost-like syntactic function rummaging through the European lexicons in search of some suitable corporeal form.” ([Holm 2000:203](#))

Hijacked constructions may not be limited to second language acquisition. I have observed one case in a child’s acquisition of English as a first language. Around the

age of four the child apparently needed to differentiate subordinate clauses from main clauses, but abduced *is* as performing the function of subordination, rather than one of the English subordinators (*that*, *which*, *who* etc.). Examples such as (3)-(5) were observed. The child was observed regularly over the course of development of this structure. The earliest examples recorded are of relative clauses, but the usage was soon also seen in complement clauses. Hijacked constructions in first language acquisition do not persist, however: by age 5.3 the child displayed normal relative and complement clauses, and the use of subordinating *is* had disappeared.

- (3) This is a game is we play in the house. [age 4.0, author's observation]
- (4) Don't go without telling me is you're going, OK? [age 4.3, author's observation]
- (5) I wish is Papa putted me in my bed. [age 4.10, author's observation]

### 8.3 Convergence and creolization: L1 > L2 influence and L1 < L2 influence

The above discussion has focussed on the influence of a group's L1 when learning an L2, symbolically L1 > L2 influence. In cases of stable bilingualism, by contrast, it is often the case that a group's first language is influenced by their knowledge, and frequent use, of another language, their L2. In this case the influence flows in the reverse direction, i.e. *on* the L1 *from* the L2, symbolized here as L1 < L2. L2 categories (or category bundles) are imported into the L1, through direct borrowing of the morpheme/structure in question, by calquing, or by mapping the categories onto existing L1 structures. In the latter case, because speakers are more familiar with their L1, the repurposed structures influenced by an L2 are likely to display grammatical and semantic consistency with their L2 models and thus less likely to be hijacked.

Thus, given the association between untutored second language acquisition and creole-formation, hijacked constructions are more likely to be found in creoles than in converged languages, which are the product of mother-tongue maintenance in the context of extensive and intensive bilingualism (Nadkarni 1975). As a simple illustration of this thesis, consider the sources of tense, mood and aspect morphemes of the converged language, Sourashtra (see 4.2 for a brief description of the language) and the creole, Sri Lanka Portuguese (see 4.3), displayed in Tables 1 and 2. While the Sourashtra morphemes are derived from repurposed Gujarati tense-mood-aspect morphemes and one possible borrowing, four of the morphemes of Sri Lanka Portuguese come from further afield: adverbs, nouns etc. Often these more radically repurposed forms are semantically consistent with the SLP function, but the reflexive imperfective morpheme lacks even this, and is thus hijacked.

The distinction between L1 > L2 influence and L1 < L2 influence may thus shed light on the role of untutored second language acquisition in the formation of Sri Lanka Malay and answer the question of whether it is more creole-like or more like a converged language. As an illustration, this paper looks at a single construction in Sri Lanka Malay and three other languages in contact with Tamil.

Sourashtra morpheme	Probable source
PRS <i>-ares</i> etc.	AUX <i>rah</i> (DUR) + ch-
PAST <i>-es</i> etc.	AUX <i>ch-</i>
(PRS IPFV) ??FUT <i>-u</i> etc.	“Old Present” <i>-ū</i> etc.
PFV <i>-ud/d/t/ł/-</i>	auxiliary <i>de/di</i> (COMPL/ALTER-BEN) or ?Telugu <i>-di</i> (PFV)
DEB <i>-no</i>	<i>-vaano</i> (DES)
CAUS <i>-ad</i>	<i>-ad</i> (CAUS)
REFL/IPFV <i>-ul/ł/-</i>	auxiliary <i>le/li</i> (COMPL/SELF-BEN)
CP <i>-i</i>	<i>-i</i> (CP)

Table 8.1: The sources of Sourashtra tense-mood-aspect markers.

SLP morpheme	Probable source
PRS <i>ta-</i>	aux <i>está</i>
(PRS PROG) PAST <i>jaa-</i>	adverb <i>ja</i> ‘already’
FUT <i>lo-</i>	adverb <i>logo</i> ‘soon’
PFV <i>kaa-</i>	verb <i>acabar de</i> + V ‘to finish V-ing’
DEB <i>mes-/mesta-</i>	noun <i>mister</i> ‘need’: <i>ser mister</i> + V ‘to be necessary to V’
CAUS <i>faya</i>	verb <i>fazer</i> + complement V (CAUS)
REFL/IPFV <i>taam</i>	<i>tambem</i> ‘also’
CP <i>=tu</i>	Tamil <i>-tu</i> (CP)

Table 8.2: The sources of Sri Lanka Portuguese tense-mood-aspect markers.

## 8.4 Tamil and four adjacent languages

Tamil, as a major regional language of Southern India and Sri Lanka and historically a language of political and cultural power, has influenced a number of nearby languages (some indirectly through a third language). These provide a valuable natural laboratory in which experiments in language contact under different social circumstances have taken place. The four languages compared here are Sinhala, Sourashtra, Sri Lanka Portuguese and Sri Lanka Malay.

### 8.4.1 Sinhala

It seems likely that Sinhala speakers arrived in Sri Lanka in the 6<sup>th</sup> C. B.C. (Gair 1976:200), a date that accords with their own tradition. Nothing is known about the sociolinguistics of the early contact period, but the influence of Tamil-Malayalam (before these languages split between the 9<sup>th</sup> and 13<sup>th</sup> centuries (Krishnamurti 2003:502)) was strong enough that Pali texts of the 5<sup>th</sup> and 6<sup>th</sup> centuries produced by Sinhalese writers already show Dravidian influence (Rhys-Davids & Stede (1921-1925/1966:vi), cited in Gair (1976:200)). On structural grounds, there is no suggestion that Sinhala was initially creole-like, since it has maintained a considerable amount of morphology that is derivable from Indo-Aryan inputs, including such ‘frills’ as subject-verb agreement (now surviving only in the literary variety). Rather, the heavy Dravidian influence comes about through contact over two and a half millennia, with varying degrees of bilingualism at different historical stages, likely involving both L1>L2 influence from L1 speakers of Tamil, as well as L1<L2 influence.



Figure 8.1: Cities and towns having significant numbers of Sourashtra speakers (following Učida 1979:3). Ammapettai, listed by Učida, is not indicated on the map as its reference is ambiguous (Ammapettai, Erode District; Ammapettai, Thanjavur district; or possibly Ammapet, a neighbourhood of Salem). It should also be noted that Učida adds 'etc.' to his list of locations in Tamilnadu.

### 8.4.2 Sourashtra

Sourashtra is a distant relative of Gujarati spoken by a traditionally weaving community in Tamil Nadu and adjacent areas of Andhra Pradesh, predominantly in cities such as Madurai and Tanjavur. According to their own traditions, the Sourashtra-speaking community left their home in what is modern Gujarat in the 11th century, and after stops in Marathi and Telugu-speaking regions probably started arriving in the Tamil country in the 16th C. Sourashtra continues as the exclusive home language in most households and is the first language of all Sourashtra speakers. Again, there is no documentation of the early sociolinguistic context, but as far as one can tell, Sourashtra has continuously been the spoken language of a sizeable community, and at no point has its transmission from one generation to the next been disrupted. Despite heavy Dravidianization (e.g. Pandit 1972), the language, like Sinhala, retains some Indo-Aryan derived morphology, including the remains of subject-verb agreement. The Sourashtra community spent extended periods among Marathi and Telugu speakers before coming into contact with Tamil and has many lexical items of Telugu and Marathi origin. Long periods of contact may well have left their mark on the grammar as well. Sourashtra has always been a minority language and there is no evidence that Tamils have ever spoken the language. Thus Dravidian influence in the language is due exclusively to Sourashtra speakers' bilingualism in Tamil/Telugu (L1 < L2).

### 8.4.3 Sri Lanka Portuguese

Sri Lanka Portuguese was born from Portuguese colonization of the island's littoral region (1515-1658). Created largely by Tamil and Sinhala speakers who learned simplified Portuguese as a second language, it became the language of identity for the

mixed-race community that was engendered by unions of local women and Portuguese men and their camp-followers. It thrived during the subsequent period of Dutch rule (1656–58), but began a long decline after the British take-over in 1795. There remains a small (but unknown) number of speakers, and the language was on the verge of extinction even before the twin disasters of civil war and tsunami. See [Nordhoff \(forthcoming\)](#) for a recent report. Nevertheless, the Dravidian features found in Sri Lanka Portuguese are *not* due to language death – most, if not all, were already in place by the mid 19<sup>th</sup> C. ([Smith \(1979:231–214\)](#), [Smith \(2005a\)](#)). Sri Lanka Portuguese is uncontroversially a creole, arising at about the same time, and in the same socio-political context as its sister varieties of Indo-Portuguese, thus the Dravidian influence is attributable in the initial stages to the L1 of local learner/creators (L1 > L2); once the community was established, influence could also flow in the reverse direction (L1 < L2).

There is evidence of ongoing substrate (acting as adstrate) influence on word order in Korlai Portuguese ([Clements 1991](#)), and more general evidence that both lexifier and sub/adstrate exert an influence on an established creole ([Clements 2009](#), [Smith forthcoming](#)). The ouster of the Portuguese from Sri Lanka by the Dutch by 1658 left free rein to the L1 < L2 influence of the substrates as adstrates. For most of the structures of the language, it is unclear whether the source is Sinhala or Tamil ([Smith 1979:216](#)); from a historical point of view, since the Portuguese first established themselves in the southwest, Sinhala-speaking, area of the island, it is likely that Sinhala was the most important early influence (cf. [De Silva Jayasuriya 1999:254–5](#)). Once the language had been spread to Tamil speaking areas, Tamil could be the source of both L1 > L2 (particularly through intermarriage) and L1 < L2 influence. However, the possibility must also be entertained that certain South Asian features were already present in the pidgin Portuguese/foreigner talk brought to Sri Lanka; such would seem to be the case for the postnominal genitive marker *-su(a)*, found in Sri Lanka as well as in other Indo-Portuguese varieties.

#### 8.4.4 Sri Lanka Malay

[Smith & Paauw \(2006\)](#) give the following account of the origins of Sri Lanka Malay:

“SLM is spoken by descendants of immigrants from the Indonesian archipelago and the Malay peninsula brought to Sri Lanka by Dutch and British colonial administrations. Most of the immigrants were soldiers, with smaller numbers of political exiles, slaves, and convicts. . . . The first waves of immigrants brought to Sri Lanka by the Dutch colonial administration, beginning in 1656, came from various locations in the Indonesian archipelago, and were native speakers of a number of West Austronesian languages.”  
[\(Smith & Paauw 2006:2006:160–1\)](#)

This much is uncontroversial. Disagreement arises, however, over other statements:

“From their arrival, the immigrants closely associated with the established Tamil-speaking Sri Lanka “Moor” community, with whom they shared the Muslim religion ([Hussainmiya 1987:45](#)). As most of the immigrants were soldiers, and therefore single males for the most part, a high degree of intermarriage with the Moor community was inevitable. Through this interaction between the two communities, a new language arose, strongly

influenced by Tamil structure, with a lexicon almost entirely derived from Malay." (2006:160)

The claim that the Moors played a primary role in providing human and linguistic input to the new community and its language has been challenged, [Ansaldi \(2008\)](#) going so far as to claim that it reflects a "Tamil bias". At issue also is the question of whether there was a disruption in the transmission of Malay serious enough to warrant the application of the term "creole". On the other hand, some claim Sri Lanka Malay is a contact language of a different type: a mixed language ([Meakins to appear](#)).

One important aspect of the contact situation that has largely been ignored is the presence of Standard Malay in the 19<sup>th</sup> century, particularly through the provision of Malay-language education for Malay soldiers and their families. It is known that the presence of its lexifier and its availability through education exerts a significant influence on a creole. Therefore, it is quite likely that some Standard Malay features were introduced into Sri Lanka Malay during this century.

## 8.5 Dravidian verbal noun

The verbal noun is a well-established structure in several southern Dravidian languages (See [Gair 1986:155](#) for descriptive sources) which has spread to all four of the languages in contact with Tamil described here.

### 8.5.1 The Tamil verbal noun

The verbal noun in Tamil is formed by adding a 3<sup>rd</sup> person non-human singular pronoun to a tensed verbal stem. The Sri Lanka Tamil present and past verbal nouns for the verbs *poo* 'go' and *keel* 'hear; ask' are shown in (6).

- (6) [Sri Lanka Tamil, constructed]
  - a. *poo-r-adu*; *poo-n-adu*  
go-PRS-3SG.NHUM; go-PAST-3SG.NHUM
  - b. *kee-kk<sup>3</sup>-r-adu*; *kee-t̪-adu*  
hear-AUG-PRS-3SG.NHUM; hear-AUG-PAST-3SG.NHUM

The stem to which the pronoun is affixed is arguably a verbal adjective (e.g. *poo-r-a*, *poo-n-a*, *kee-kk-r-a*, *kee-t̪-a*) which has undergone apocope, since this is the stem that combines with other pronouns, as in the Sri Lanka Tamil habitual shown in (7). Apocope of the final vowel is seen in (8) when the 3<sup>rd</sup> person nonhuman pronoun is used; this form is homophonous with the verbal noun.

- (7) [Sri Lanka Tamil, elicited]

*naan tyuvishan-ukku poo-r-a-naan.*  
1SG tuition-DAT GO-PRS-ADJZ-1SG

'I go for tuition.'

- (8) [Sri Lanka Tamil, elicited]

---

<sup>3</sup>-*kk-* is a stem augment found in the present stem and infinitive of certain verb classes.

*adu      tyuvishan-ukku poo-r-Ø-adu*  
 3SG.NHUM tuition-DAT      go-PRS-ADJZ-3SG.NHUM

'He/she [child] goes for tuition.'

The functions of the verbal noun differ somewhat from one variety of Tamil to another. The principal varieties mentioned here are Sri Lanka Tamil (especially Jaffna Tamil), Shonam (Sri Lanka Muslim Tamil) and Indian Tamil (in which there is much regional and social variation). Although examples may be specified as one variety or another because of phonological, morphological or lexical specificities, usages are pan-dialectal unless otherwise specified.

### Nominal clause

The verbal noun is used to nominalize a clause so that it may then function as the subject or complement of a verb, complement of a postposition etc. The verbal noun is the only signal of subordinate status; other elements of the clause remain the same as they would be in a main clause. This contrasts with languages like English in which a nominalized verb may take different kinds of arguments from its main clause counterpart. (For example, *The hunters shot the lions.* vs. *The shooting of the lions by the hunters.*)

- (9) [Shonam; elicited]

*sigarat    kudi-kk-r-adu      oqambu-kku nall=aa      ille.*  
 cigarette smoke-AUG-PRS-NMLZ body-DAT      good=ADVZ NEG

'Cigarette smoking is not good for health.'

### (Past) experiential negative

The experiential negative is formed by adding the negative *ille* to the past verbal noun.

- (10) [Tamil: elicited]

*naan ange poo-n-ad      ille.*  
 1SG there go-PAST-NMLZ NEG

'I never went there./I have never been there.'

### (Present) habitual negative

The habitual negative is formed by adding the negative *ille* to the present verbal noun.

- (11) [Tamil: elicited]

*naan ange poo-r-ad      ille*  
 1SG there go-PRS-NMLZ NEG

'I don't go there.'

- (12) [Tamil: elicited]

*naan sigarat    kudikk-r-ad      ille*  
 1SG cigarette smoke-PRS-NMLZ NEG

'I don't smoke.'

### 'Thinking of V-ing'

In Indian Tamil, the equivalent of the English expression ‘thinking of V-ing’ is rendered by present verbal noun with the addition of the adverbializer *=aa* and the verb *iru* ‘be’. This usage is not found in Sri Lankan Tamil varieties.

- (13) [Indian Tamil: elicited]

*adan-aala sigarat̪ kud̪i-kk-r=adu vid̪-r-ad=aa*  
 That-ins cigarette smoke-AUG-PRS-NMLZ leave-PRS-NMLZ=ADVZ  
*iru-kk-r-een*  
 be-AUG-PRS-1SG

‘Therefore I am thinking of giving up cigarette smoking.’

### Focus/cleft (optional)

A non-verbal constituent of a Tamil main clause may be focussed by the addition of the focus particle *=taan/daan*. The verb is optionally nominalized. If the verb is nominalized, the focal element may be placed in post-verbal position, where it carries the major final intonation contour. This contrasts with “afterthought” elements which follow a non-nominalized verb; in this case the major final intonation contour falls on the verb and the “afterthought” is on low pitch. (14) shows an unfocussed sentence; (15) shows various focussing options. Murugaiyan notes that when the verb is nominalized the focus marker is pragmatically obligatory (Murugaiyan 2009:56).

- (14) [Tamil, elicited]

*naan bas-le poo-n-een.*  
 1SG bus-LOC go-PAST-1SG

‘I went by bus.’

- (15) [Tamil, elicited]

- a. *naan=taan bas-le poo-n-een. ~ \*bas-le poo-n-een naan=taan*  
 1SG=FOC bus-LOC go-PAST-1SG
- b. *naan=taan bas-le poo-n-adu. ~ bas-le poo-n-adu naan=taan.*  
 1SG=FOC bus-LOC go-PAST-NMLZ ~ bus-LOC go-PAST-1SG 1SG=FOC  
 ‘I was the one who went by bus; I went by bus; it was I who went by bus.’

Since WH-words are inherently focussed, they behave in the same way as a focussed element, viz. they can occur with a verb showing subject concord or with a nominalized verb and with the latter they may be postponed. These possibilities are illustrated in (16) and (17).

- (16) [Tamil, elicited]

- a. *yaar ange poo-n-anga? ~ \*ange poon-aanga yaar?*  
 who there go-PAST-3PL
- b. *yaar ange poo-n-adu? ~ ange poo-n-adu yaar?*  
 who there go-PAST-NMLZ ~ there go-PAST-NMLZ who  
 ‘Who went there?’

- (17) [Tamil, elicited]

- a. *avan enge poo-n-aan?* ~ \**avan poo-n-aan enge?*  
3SG.NHON where go-PAST-3SG.NHON
- b. *avan enge poo-n-adu?* ~ *avan poo-n-ad enge?*  
3SG.NHON where go-PAST-NMLZ  
'Where did he go?'

Speakers of both Indian and Sri Lanka Tamil have expressed unease with a nominalized verb when the subject is honorific, as in (18). This seems to be because of the use of the nonhuman pronoun as nominalization marker.

- (18) a. [Tamil, elicited]
- niinga enge poo-n-iinga?*  
2HON where go-PAST-2HON
- b. (\**niinga enge poo-n-adu?*  
2HON where go-PAST-NMLZ  
'Where did you go?'

#### Unspecified subject (optional)

When the subject of the main clause is unspecified (and unexpressed), the verb may be optionally nominalized. Often the most appropriate English rendering uses the passive voice. Again, the nominalized verb is dispreferred for some speakers when the nonhuman marking of the nominalized verb clashes with an honorific participant, as in (21).

- (19) [Tamil, Sourashtra:07B<sup>4</sup>]
- onga pa||ikkuudatt-ile enam=oo nad-a-kk-r-adu*  
2HON.GEN school-LOC what-INDEF run-AUG-PRS-NMLZ

'They are putting something on in your school.'

- (20) [Tamil, elicited]
- ad-e cup=nu solr-adu*  
that-ACC cup=QUOT say-PRS-NMLZ
- 'That's called a "cup".'

- (21) [Indian Tamil; Sourashtra:05A]
- avanga]-e atte=nu sollu-v-oom/(\*sol-r-adu*  
3PL-ACC atte=QUOT say-FUT-1PL/say-PRS-NMLZ
- 'They are called "atte" [=father's mother].'

<sup>4</sup> Tamil examples sourced from "SourashtraXXS" are from the Tamil translations of conversational conversations in Sourashtra recorded in Madurai in 1989-90; XX refers to the cassette no. and S is the side (A or B). Tamil translations were initially provided by bilingual Sourashtra speakers and were checked for naturalness by a Tamil speaker. All examples cited here have been checked with at least one additional Tamil speaker.

### Main Clause with future ref: (IT only, dialectal?, optional)

In a main clause with future time reference a present nominalized verb may sometimes be used in place of a future-tensed verb with subject agreement. This usage is not reported in the literature and has not been investigated in detail. It is not found in Sri Lanka and may be geographically restricted within India.

- (22) [Indian Tamil, Sourashtra:05A]

*id        ellaam diipaavalij ka|i-nj-a        peragu*  
 3SG.NHUM all      Divali      finish-PAST-ADJZ after  
*gavani-pp-aanga/gavani-kk-r-adu*  
 look.after-FUT-3PL/look.after-AUG-PRS-NMLZ

‘[They] will look after all this when Divali is over.’

### 8.5.2 Sinhala ‘focus’ form

The Sinhala focus (or ‘emphatic’) form is constructed by adding *ee* to a tensed (prs or pst) stem, as seen in (23) for *bala-n-awaa/ baeluw-aa* ‘see-PRS-fin/see.PAST-fin’.<sup>5</sup> The origin of this form seems to be a verbal adjective to which is added a 3sg.m/n suffix (Gair (1986:156), citing Geiger (1938:134-5); i.e. it is a calque on the Tamil model. This analysis is now opaque, and as Gair points out, the form is “not a clearly nominal form at present” (1998:156). Indeed, Sinhala has two synchronic nominalization strategies (e.g. *balaa-ma, balana-eka* ‘seeing’, the latter a more recent calque on Tamil<sup>6</sup>), which are used for non-focus nominalizations, e.g., the nominalized verb in (24) The focus form is tightly integrated into the grammar of Sinhala; its functions are outlined below. For a detailed description, see Gair (1986).

- (23) [Sinhala: elicited]

*bala-nn-e; baeluw-ee*  
 see-PRS-FOC; see.PAST-FOC

- (24) [Sinhala: elicited]

*sigarat̄ bii-ma        saukya-ṭa ho"da nææ*  
 cigarette smoke-NMLZ health-DAT good NEG

‘Cigarette smoking is not good for the health.’

### Focus constructions (obligatory)

In contrast to Tamil, where the nominalized verb is optional in focus constructions, in Sinhala the focus form is obligatory whether or not the focal element is postposed. (25) gives a non-focus construction and (26) its focussed equivalents.

<sup>5</sup>It is not the general scholarly practice to decompose Sinhala verbs into concatenative morphemes. I do so here only to make clear the structure of the focus form. Some problems associated with the analysis here are as follows. First, the fact that *bala* by virtue of its opposition to *baeluw* also indexes nonpast is not indicated. Second, the single *-n-* found in the present form also indexes “finite”, since the present focus form and the infinitive have *-nn-*. Third, “finite” is an unsatisfactory label; perhaps “unmarked” would be better. Finally, the “finite” affixes, by virtue of their difference, also index tense.

<sup>6</sup>An anonymous reviewer points out a third form, *bæli-lla*.

- (25) [Sinhala; elicited]

*mama bss-ek-ee giyaa*  
1SG bus-SG.DEF-LOC GO.PAST

‘I went by bus.’

- (26) [Sinhala, elicited]

- a. *mama=y bas-ek-ee giy-ee/\*giyaa*  
1SG=FOC bus-SG.DEF-LOC GO.PAST-FOC/\*GO.PAST
- b. *bəs-ekee giy-ee(/\*giyaa) mama=y*  
bus-DEF.LOC go.PAST-FOC/\*go.PAST 1SG=FOC  
‘I was the one who went by bus.’

As expected, the focus form is also obligatory in WH-questions, as seen in (27). Unfocussed polar questions, on the other hand, use the ‘finite’ form, as seen in (28).

- (27) [Sinhala, elicited]

*oyaa kohee=da giy-ee?/\*giyaa ~ oyaa giy-ee(/\*giyaa) kohee=da?*  
2SG where=INTERR go.PAST-FOC/\*GO.PAST

‘Where did you go?’

- (28) *oyaa kolaṁba giyaa=da?*

2SG Colombo go.PAST=INTERR  
‘Did you go to Colombo?’

### Negation (obligatory)

In 5.1.2 and 5.1.3 we saw that the Tamil verbal noun combines with the negative marker to render specialized negative meanings (viz. experiential and habitual). In Sinhala, by contrast, all negative sentences require the focus form. (29) shows a positive sentence and (30) its negative counterpart

- (29) [Sinhala; elicited]

*mama kolaṁba yanawaa/giyaa*  
1SG Colombo go.PRS/go.PAST

‘I’m going to Colombo / I went to Colombo’

- (30) [Sinhala; elicited]

- a. *mama kolaṁba yann-ee nææ*  
1SG Colombo go.PRS-FOC NEG  
‘I’m not going to Colombo; I didn’t go to Colombo.’
- b. *mama kolaṁba giy-ee nææ*  
1SG Colombo go.PAST-FOC
- c. *\*mama kolaṁba giyaa/yanəwaa nææ*  
1SG Colombo go.PAST/go.PRS NEG

### 8.5.3 Sourashtra -atte/-eyo

Sourashtra forms verbal nouns by adding *-atte* (present) and *-eyo* (past) to a verb base, as illustrated for three verbs in (31); neither form is found in Gujarati. The present form originates from an Indo-Aryan imperfective participle in *-at* (<OIA *-anta*) plus a 3sg pronoun *te* (Masica (1991:272), Cardona (1965:93,100); the past form is from an Indo-Aryan perfective participle in *-ey* (<OIA *-ita*) plus a m.sg. marker *-o* (Masica (1991:272), Cardona (1965:100-1). Both formations are now opaque. The functions of Sourashtra *-atte/-eyo* are outlined below.

- (31) a. *ker-atte/ker-eyo*  
do-PRS.NMLZ/do-PAST.NMLZ
- b. *sig-atte/sig-eyo*  
learn-PRS.NMLZ/learn-PAST.NMLZ
- c. *men-atte/men-eyo*  
say-PRS.NMLZ/say-PAST.NMLZ

#### Nominalized clause

This function closely parallels the corresponding Tamil function (5.1.1).

- (32) [Sourashtra:05A<sup>7</sup>]

*tunko puus men-atte pus-unaan*  
2HON.DAT listen.IMP say-PRS.NMLZ listen-FUT.NEG.PL

‘You won’t listen to what anyone tells you.’ (Lit: You won’t listen to [someone] saying “listen!”)

- (33) [Sourashtra:29A]

*avro gaamu thov-eyo konni serkko cal-ani*  
1PL.GEN town put-PAST.NMLZ nothing well go-NEG

‘What they set up in our town [i.e. the community association] did not go well at all.’

#### Focus (optional)

As in Tamil the use of nominalization in focused sentences is optional. (See (38) for a non-nominalized example.) The focussed nominal may be postposed, and the construction is commonly used with WH-interrogatives.

- (34) [Sourashtra:05A]

*atto soo mhad̪a mundad̪ji=nnaa khad̪-eyo*  
now six month before=FOC.TAG buy-PAST.NMLZ

‘Now it was six months ago that they bought it, eh.’

- (35) [Sourashtra:05A]

*teno teed̪uu=s jiy-eyo*  
3SG.C there=FOC go.PAST-PAST.NMLZ

'It was there that he went.'

- (36) [Sourashtra:34A]

*jeendipuramu jaa-riyo basu koot hing-atte?*  
Jeendipuram go-ADJZ bus where get.on-PRS.NMLZ

'Where do you get on the bus that goes to Jeendipuram?'

- (37) [Sourashtra:35A]

*tumi medraasu-m uj-eyo koot?*  
2HON Madras-LOC be.born-PAST.NMLZ where

'Where in Madras were you born?'

- (38) [Sourashtra:35A]

*hindu koot jiy-aasi, tumi?*  
today where go.PAST-PAST.PL 2HON

'Where did you go today?'

### Habitual negative

As in Tamil, the present verbal noun combines with a negative marker to encode habitual negative. The experiential negative uses an alternate formation for the past verbal noun using the suffix *-iriyo*, as seen in (40).

- (39) [Sourashtra:34A]

*tellamaam puuraa-k ami ja-atte nhaa*  
that.LOC all-DAT 2PL go.PRS-PRS.NMLZ NEG

'We don't go there and all.'

- (40) [Sourashtra, elicited]

*teno ji-iriyo nhii*  
3SG go.PAST-PAST.NMLZ NEG

'He never went/He hasn't gone (ever)'

### Unspecified subject: (optional)

As in Tamil, a nominalized verb may be used with an unspecified subject, often as a passive equivalent. Unlike Tamil, Sourashtra has no restriction on using the verbal noun with an honorific participant, as in (42). This may be because the nominalizing suffixes are not synchronically linked with non-human forms as are the corresponding Tamil suffixes.

- (41) [Sourashtra:07B]

*ture poltam-u kaay=ki dham-atte*  
2HON.GEN school-LOC what=INDEF run-NMLZ

'They are putting something on in your school.'

- (42) [Sourashtra:05A]

*tenko ‘phōy’ men-atte*  
 3PL.ACC ‘phōy’ say-PRS.NMLZ  
 ‘They are called “phōy” [father’s brother].’

#### Main Clause with future reference (optional)

As in Tamil, a main clause with future reference may have a verbal noun instead of an agreement-inflected verb. Sourashtra uses this formation more frequently than Tamil.

- (43) [Sourashtra:05A]

*ella aski dival|i ser-et-teer gavunc-atte*  
 this all Divali finish-PAST.ADJZR-after look.after-PRS.mnlz  
 ‘[They] will look after all this when Divali is over.’

#### Relative clause (variant)

Pandit (1972), in his pioneering work on Sourashtra pointed out that its relative clause structures were similar to Tamil rather than to Gujarati. The salient characteristics of this Dravidian-style structure are prenominal position, absence of a relative pronoun and dependent verb form (the verbal adjective) that marks tense but not agreement. Sourashtra has two forms for the present verbal adjective, one in *-atte*, seen in (44), the other in *ariyo*, seen in (45). Neither of these is used exclusively as a dependent form: *-atte* is familiar as our present verbal noun; *-ariyo* forms a main verb unmarked for agreement and is also an alternative (but rarer) verbal noun formation. As verbal adjectives, the two forms are almost in complementation, with *-ariyo* in general use before a noun head. As in Tamil, Sourashtra relative clauses can modify pronouns and time nominals which contract with the verbal adjective. In Sourashtra the present verbal adjective used in these combinations is almost exclusively the *-atte* form (contracted to *-at* before following *te-*). The complementation is spoilt, however, by the occasional occurrence of *-atte* with noun heads as in (44) and of *-ariyo* with a pronoun head.

The past verbal adjective is formed with *-e* (*-et* before *te-*) as seen in (43) above. The past verbal noun in *-eyo* never marks a relative clause.

- (44) [Sourashtra:05A]

*macc-i phil-l-atte savlo pay-et khat-t-uvaay*  
 change-CP wear-BEN-PRS.NMLZ sari want-COND buy-PFV-PER

‘If [you] want a house sari [we] can buy [one]’ (Lit: ‘a sari that [you] wear having changed’)

- (45) [Sourashtra:05A]

*telle phalaar hand-i vikk-ariyo gheer legutto kyaa?*  
 that idli prepare-CP sell-PRS.NMLZ house near INTERR

‘Near that house where [they] prepare and sell idli?’

- (46) [Sourashtra:07B]

*dekk-et-tenko puuraa bovd-ariyo; av-at-tenu av-an*  
 see-PAST.ADJZ-3PL.ACC all invite-PRS.NMLZ-3PL come-FUT.P

‘We’ll invite all those whom we saw. Those who come will come.’

### 8.5.4 Sri Lanka Portuguese *ki-*

Sri Lanka Portuguese forms its nominalized verb with the verbal prefix *ki-*, followed optionally by the tense markers *ta*-(PRS) or *jaa*-(PAST) as illustrated in (47) for *faya* ‘do’. form: *ki*+(TNS)+Verb base.

- (47) [Sri Lanka Portuguese, constructed]

*ki-faya / ki-ta-faya / ki-jaa-faya*  
NMLZ-do / NMLZ-PRS-do / NMLZ-PAST-do

The marker *ki-* derives from the Ptg. relative clause marker *que*, particularly in cleft constructions of the type illustrated in (48), for which the corresponding Sri Lanka Portuguese sentence is given in (49).

- (48) [Portuguese, elicited]

(É) eli que foi lá  
be.3SG 3SG.MASC REL go.PAST.3SG there

‘It was he who went there.’

- (49) [Sri Lanka Portuguese, constructed]

eli mee ala ki-jaa-foi  
3SG FOC there NMLZ-PAST-go

The functions of *ki-* are outlined in the following subsections. Verbs may also be nominalized by use of the suffix *-saam*, as in *midi-saam* ‘measurement’ from *midii* ‘measure’, illustrated in (50), or *arnaga-saam* ‘hatred; cursing’ from *arnagaa* ‘curse’. This formation does not participate in any of the constructions outlined below, however. It is best treated as a lexical derivation, which does not always produce fully predictable meanings.

- (50) [Sri Lanka Portuguese, conversational, Field notes:2235<sup>8</sup>]

tɔɔna aka midi-saam inda uja pa-araa naa-pooy=ski  
later that measure-NMLZ yet one INF-escape FUT.NEG=HBL=REP

‘Afterwards, someone else cannot ignore that [official] measurement, apparently.’

#### Relative clause (optional)

Given its origin, it is not surprising to find that *ki-* is used to mark a relative clause, as illustrated in (51). It is always optional and very rare in actual speech. All relative clauses directly precede the noun they modify, as in the other languages of the area. When *ki-* is absent, there is no formal distinction between the verb of a relative clause and a main verb.

- (51) [Sri Lanka Portuguese, conversational, Field notes:5313]

<sup>8</sup> Numbers refer to the item in the author’s Sri Lanka Portuguese field notes, 1974–5. Examples taken from spontaneous recorded conversations among native speakers are marked ‘conversational’; examples spoken to the author without elicitation are marked ‘spontaneous’.

*nosa yuunyan-ntu (ki-)tija meshiin-su shatlkeriya*  
 1PL.GEN union-LOC NMLZ-be.PAST machine-GEN shuttle.carrier  
*jaa-kaa-kebraa, um padaas*  
 PAST-PFV-break one piece

‘A piece broke off the shuttle carrier of the machine that was in our union.

### Nominalized clause

One of the most common uses of *ki-* is to mark a nominalized clause. The usage parallels that of Tamil (5.1.1).

- (52) [Sri Lanka Portuguese, conversational, Field notes:5173]

*tyuvishan ki-(ta)-daa graandi viraadu*  
 tuition NMLZ-PRS-give big error

‘Giving tuition is a great error.’

- (53) [Sri Lanka Portuguese, conversational, Field notes:5251]

*avara maara boorda ki-t-andaa kum, padam-pa ki-t-andaa kum, kastiiv*  
 now sea beside NMLZ-PRS-go and movie-DAT NMLZ-PRS-go and trouble

‘Now, going to the beach and going for a movie is troublesome.’ [Context: speaking of decorous conduct for teen-aged girls.]

- (54) [Sri Lanka Portuguese, conversational, Field notes:1429]

*avara botus-pa ki-jaa-daa teem, aka graandi uj kaavs!*  
 now 2HON-DAT NMLZ-PAST-give PFV that big one thing

‘Now the fact that they have given [it] to you, that is a big thing!’

### Focus construction (optional)

As in Tamil, the use of the nominalized verb is optional in focus constructions. The focused element may optionally follow the verb, whether the verb is nominalized or not. This differs slightly from Tamil which only allows a post-verbal focus when the verb is nominalized. Example (57) illustrates a post-verbal focus without nominalization of the verb.

- (55) [Sri Lanka Portuguese, conversational, Field notes:4883]

*kii avara-su paav-su faartus-su preesu voonda pruveetu*  
 what now-GEN wood-GEN thing.PL-gen price by profit  
*ki-ta-vii muytu meenis*  
 NMLZ-PRS-come very less

‘What! due to the price of timber products, the income comes to very little.’<sup>9</sup>

- (56) [Sri Lanka Portuguese, conversational, Field notes:2241]

---

<sup>9</sup> A putative alternative reading ‘the income which comes is very little’ is ruled out by the fact that relative clauses are invariably prenominal in Sri Lanka Portuguese.

*seem, midii-pa dispoos mee ki-poy sartifay=ski*  
 yes measure-DAT after FOC NMLZ=HBL certify=REP

‘Yes, it is only after measuring [it] that [he] can certify [it], apparently.

- (57) [Sri Lanka Portuguese, conversational, Field notes:5111]

*alaa menaa bibeer=ley tijna, jaa-paraa sudu akii mee; isti*  
 there of.course livelihood=ADVZ be.PAST; PAST-live all here FOC this  
*podjiyaas jaa-nasa sudu akii mee*  
 child PAST-be.born all here foc

‘We were there of course for employment [but] where we all lived was right here; where the children were born was right here.’

### WH-Questions (optional)

As in Tamil, WH-questions often have a nominalized verb, but do not require one, as (60) shows. In all the examples in the corpus the WH-word occurs before, rather than after the verb. This may be a deficiency in the corpus or a difference with Tamil, which allows post-verbal WH-words.

- (58) [Sri Lanka Portuguese, conversational, Field notes:5502]

*okaloora isti-ntu doos taya daa senaa. sufraa see kii ki-faya?*  
 then this-LOC two tire give then swell COND what NMLZ-do

‘Put two tyres on this, then. If they blow out, what to do?’

- (59) [Sri Lanka Portuguese, conversational, Field notes:2282]

*eli oondi ki-t-andaa, akii benkal laraa-tu?*  
 3SG.NHON where NMLZ-PRS-go here workshop leave-cp

‘Where is he going, leaving his workshop here?’

- (60) [Sri Lanka Portuguese, song lyric, Field notes:5567]

*liimpu bistidu bunitu, oondi boos t-andaa?*  
 clean clothes pretty where 2SG PRS-go

Where are you going, all dressed-up and pretty?

### Habitual negative

As in Tamil the nominalized verb participates in a habitual negative construction, seen here in (61).

- (61) [Sri Lanka Portuguese, conversational, Field notes:4839]

*eev un tεemptu taam ki-nuku-kuma, farijna; aka diiya jaa-kuma-tu*  
 1SG one time even NMLZ-NEG-eat, flour; that day PAST-eat-cp  
*jaa-padiisa*  
 PAST-suffer

‘I never “used to”<sup>10</sup> eat it, flour; one day I ate it and suffered.’

---

<sup>10</sup> In Sri Lankan English, as in other Asian varieties such as Singapore English, *used to* can be used to signal present habitual.

### Habitual

In the first clause of the previous example the nominalized verb signals habitual. Further examples are seen below. Recall that Sri Lanka Tamil has evolved a habitual formation consisting of the verbal adjective followed by a pronoun agreeing with the subject (section 5.1, where the similarities between this formation and the verbal noun were pointed out). Since Sri Lanka Portuguese verbs do not exhibit agreement, its use of the verbal noun is thus inspired by the [Sri Lanka] Tamil construction.

- (62) [Sri Lanka Portuguese, spontaneous, Field notes:1151]

*botus-su kaaza-ntu verduura-s ki-simipaa gardaa?*  
2HON-GEN house-LOC vegetable-PL NMLZ-plant ben?

‘Do you plant vegetables (for yourselves) at home?’

- (63) [Sri Lanka Portuguese, conversational, Field notes:5248]

*isti=ley greeya t-andaa kum ki-ta-vii mee*  
this=like church PRS-go and NMLZ-PRS-come FOC

‘Like this, we go to church and come only.’ (I.e. ‘We only go to and from church.’)

### 8.5.5 Sri Lanka Malay *yang-*

Sri Lanka Malay forms nominalized verbs with the preverbal particle *yang* (often reduced to [jə], [ayn] etc.). The formation is illustrated in (64) for *maakan* ‘eat’ and *pii* ‘go’.

- (64) *yang-maakan; yang-pii*  
NMLZ-eat; NMLZ-go

The language also retains the Malay nominalizing formation pref-V-*an*, as in *pəcari-yan* ‘work’(N) from *caari* ‘work’(V) and (*kə-)**susah-an* ‘sorrow’ from *suusa* ‘be sad’. This is primarily a lexical derivation and yields meanings that are not fully predictable.

The morpheme *yang* derives from the Malay marker *yang* which has a variety of, primarily pragmatic rather than strictly grammatical, functions, which it will be useful to survey before we examine the Sri Lanka Malay data. Englebretson demonstrates that in colloquial Indonesian *yang* marks “a continuum with three general foci... subordinate clauses modifying a head noun, phrasal constituents expressing presupposed information in clefts, and referring expressions serving as direct arguments of predicates or as unattached NPs” (Englebretson 2008:24). The relative clause is illustrated in (65), where the clause introduced by *yan* gmodifies the head noun *hal* ‘things’. A phrasal cleft is seen in (66), where *yang baca gitu/yang baca itu* express presupposed, backgrounded information as opposed to the cleft focus *dia*. In (67) *yang* marks the referring nominal expressions *yang di kita*, *yang di dia*, and *yang emas*; which contain no verbal element.

- (65) [Colloquial Indonesian, Englebretson (2008:9), morph-by-morph glosses amended for conformity]

*Masih ada hal yang perlu kita lakukan lagi*  
 still be thing YANG need 1PL do.APPL again

‘There are still things we need to do yet.’

- (66) [Colloquial Indonesian, Englebretson (2008:13-14)]

*Maksudnya, dia yang baca gitu. Ayatnya dikasih Ari, dia yang baca itu*  
 mean.DEF 3SG YANG read thus verse.DEF p.give Ari 3SG YANG read that

‘So you mean it was she who read. You (Ari) gave her the verse and it was she who read it.’

- (67) [Colloquial Indonesian, Englebretson (2008:16), morph-by-morph glosses amended for conformity]

*Yang di kita itu batu. Yang di dia itu yang emas*  
 YANG at 1PL that stone YANG at 3sg that yang gold

‘The one we’ve got is stone. The one he’s got is the gold one.’

Steinhauer’s description of the core semantics of *yang* again highlights its pragmatic functions:

“*yang* instructs the hearer to arrive at a projection in his mind of the thing meant through at least two stages: (1) a projection of the set of appropriate referents of the overt or implied antecedent, and (2) a projection of that subset which also belongs to the set of appropriate referents of the adjunct. In other words, *yang* triggers the search for an antecedent and subsequently for a specified subset of the set of appropriate referents of this antecedent.” (Steinhauer (1992:430-1), cited by van Minde (2008:1987))

Van Minde, surveying older and regional material, notes additional discourse functions not usually reported in standard accounts of *yang* that rely on material from standardized Bahasa Malaysia and Bahasa Indonesia. These are illustrated below.

In (68) *yang* marks the theme/topic of the second and third sentences.

- (68) [Malay, 1870, (van Minde 2008:1994-5), morph-by-morph gloss supplied]

*Maka kalau boleh paduka sahabat sahaya beri sedikit surat kepada tuan*  
 CNJ COND can excellency friend 1sg give a.little letter to Mr.  
*dokter yang minta sedikit obat demam serta obat cacing*  
 doctor yang ask a.little medicine fever as.well.as medicine worm  
*barang enam bungkus. Yang obat demam sahaya makan sendiri*  
 approximately six pack. yang medicine fever 1sg eat REFL  
*yang obat cacing hendak memberi anak2 kecil*  
 yang medicine worm intend TR.give child.PL small

‘Could you please give me a note for the doctor asking for some medicine to suppress the fever and about six packets of medicine for worms. The fever medicine I will take myself, the vermicide I will give to the children.’

In (69) *yang* marks the clausal complement of *meluluskan*. Van Minde remarks that in the dialectal folk stories collected in Toer (1982) *yang*, rather than standard *bahwa*, is the usual clausal complementizer following verbs of speaking and cognition, adjectives and nouns (2008:1995-6).

- (69) [Malay, 19<sup>th</sup> C., [van Minde \(2008:1992\)](#), morph-by-morph gloss supplied]

*Maka adat itu meluluskan yang boleh sahaya menegurkan kelakuan*  
 CNJ law that permit *yang* can 1SG criticize behaviour  
*raja~raja yang tiada berpatutan, dengan tiada kena hukum bunuh atau*  
 king~PL *yang* NEG propriety, CNJ NEG pass punishment kill or  
*rampas adanya*  
 plunder being

‘The laws allow that I criticize the unseemly behaviour of the rajas [corrected from raja’s –IS] who kill and rob without being punished.’

Finally, van Minde notes that in the pre-20<sup>th</sup> century texts he surveyed, *yang* is often omitted where the modern standard varieties would require it for complex attributes within an NP (such as relative clauses). He finds the same tendency in a modern Indonesian writer, Seno Gumira Ajidarma, whose “language is quite direct, down-to-earth, with a strong admixture of spoken style language, while influence of Javanese and Jakarta Malay is also clearly felt.” (2008:1997).

In contrast to its morphosyntactic fluidity in Malay/Indonesian, *yang* in Sri Lanka Malay is only pre-verbal marker. The closest pragmatic overlap between Malay/Indonesian and Sri Lanka Malay usages is found in cleft/focus constructions, as illustrated in (66) above (70) and (71) below; see also 5.5.2. The functions of *yang* in Sri Lanka Malay are surveyed in the following sections.

- (70) [Malay, constructed]

*Saya yang pergi sana.*  
 1SG *yang* go there

‘It was I who went there.’

- (71) [Sri Lanka Malay, elicited]

*se=jo saana yang-pii.*  
 1SG=FOC there NMLZ-go

‘It was I who went there.’

### Relative clause (optional – past only)

*Yang* retains its function as a relative clause marker, but its usage has become restricted to clauses with past reference where it replaces the past marker *su-* (For further details, see [Bourdin & Jaffar 2010](#)). (72) illustrates this restriction. The standard Dravidian relative clause formation has the widest distribution in Sri Lanka Malay. Because of the obligatory replacement of the past-marker *su-* by *yang*, a past relative clause is distinguished from a past main clause. In the present and future, however, there is no formal distinction between the two.

- (72) [Sri Lanka Malay, [Bourdin & Jaffar \(2010:2-3\);<sup>11</sup>](#) elicited]

- a. *se yang-beli kar merah*  
 1SG NMLZ-buy car red  
 ‘The car I bought is red.’

- b. \**se* (*yang-*)*su-bəli kar merah*  
1SG NMLZ-PAST-buy car red  
'The car I bought is red.'
- c. *se* (\**yang-*)*anti-bəli kar merah*  
1SG NMLZ-FUT-buy car red  
'The car I will buy is red.'

#### Focus constructions: (optional)

As in the other languages treated here, one of the main uses of nominalization is to mark a focus construction.

- (73) [Sri Lanka Malay, Bourdin & Jaffar (2010:5)]

*[derang Kirinda=nang yang-pi] [gaja pada mem-beli=jo]*  
3PL Kirinda=dat NMLZ-go elephant PL INF-buy=FOC

'It was to buy elephants that they went to Kirinda.'

Nominalization is optional, as in Tamil. Bourdin and Jaffar point to the pragmatic contrast between (74) and (75), arguing that "yang is responsible for effecting this partitioning between backgrounded information and focus" (2010:6). But (76) shows that the focus particle can occur independently of *yang*. Nevertheless, Jaffar (p.c.) advises that "yang (or yang + jo) mark the focus better." Sri Lanka Malay allows the focal element to follow the verb even when the latter is not nominalized, (76); this freedom is not generally permitted in unfocussed sentences, (78), unless the verb is not tense-marked (75).

- (74) [Sri Lanka Malay, Bourdin & Jaffar (2010:6)]

*se yang-minta te, bukang kopi.*<sup>12</sup>  
1SG NMLZ-ask tea NEG coffee

'What I asked for was tea, not coffee.'

- (75) [Sri Lanka Malay, Bourdin & Jaffar (2010:6)]

*se minta te, bukang kopi.*<sup>13</sup>  
1SG ask tea NEG coffee

'I asked for tea, not coffee.'

- (76) [Sri Lanka Malay, elicited]

*se minta tee=joo, koopi bukang*  
1SG ask tea=FOC NEG coffee

---

<sup>11</sup> In this and other examples taken from Bourdin & Jaffar (2010), I have retained their original transcription but modified the glosses where necessary to conform to the usage of this paper.

<sup>12</sup> For many speakers, *bukang* must follow rather than precede the element over which it has scope. Such speakers render this sentence: *Se yang-minta tee(=joo), koopi bukang.*

<sup>13</sup> Other speakers would render this *Se minta tee, koopi bukang.*

'I asked for tea, not coffee.'

- (77) [Sri Lanka Malay, elicited]
- a. *se=jo saana yang-pii.*  
1SG=FOC there NMLZ-go  
'It was I who went there.'
  - b. *saana yang-pii se=jo*  
there NMLZ-go 1SG=FOC  
'It was I who went there.'
- (78) [Sri Lanka Malay, elicited]
- a. *se saana su-pii.*  
1SG there PAST-go  
'I went there.'
  - b. *\*saana su-pii se*  
there PAST-go 1SG

Finally, the contrast between (77) and (79) shows that when the focus marker *=jo* is used in sentences referring to the past, the past marker *su-* is not permitted.

- (79) [Sri Lanka Malay, elicited]
- a. *\*se=jo saana su-pii*  
1SG=FOC there PAST-go
  - b. *\*saana su-pii se=jo*

### WH-Questions (?obligatory in past)

As in the other languages, nominalization may occur in WH-questions. It is not obligatory in Sri Lanka Malay, except possibly when time reference is to the past. As with the other languages, the WH element may stay in situ or be placed after the nominalized verb. Generally, the WH element may not follow a verb that is not nominalized, but in (82) this possibility is acceptable in informal speech.

- (80) [Sri Lanka Malay, elicited]
- a. *saapa saana yang-pii?*  
who there NMLZ-go  
'Who went there?'
  - b. *saana yang-pii saapa?*  
there NMLZ-go who  
'Who went there?'
- (81) [Sri Lanka Malay, elicited]
- a. *de maana yang-pii?*  
3SG where NMLZ-go  
'Where did he go?'
  - b. *de yang-pii maana?*  
3SG NMLZ-go where  
'Where did he go?'
- (82) a. [Sri Lanka Malay, elicited]

- de maana ara-pii?*  
 3SG where PRS-go  
 ‘Where is he going?’
- b. *de ara-pii maana?*  
 3SG PRS-go where  
 ‘Where is he going?’

- (83) [Sri Lanka Malay, elicited]

- a. *??de maana su-pii?*  
 3SG where PAST-go  
 ‘Where did he go?’
- b. *\*de su-pii maana?*  
 3SG PAST-go where

In Malay/Indonesian also, *yang* may occur in WH-questions, whether they contain a verb (84) or not (85).

- (84) [Colloquial Indonesian, [Englebretson \(2008:15\)](#)]

*Siapa yang ngambil uangku*  
 who YANG at-take money-1SG  
 ‘Who stole my money?’

- (85) [Colloquial Indonesian, [Englebretson \(2008:15\)](#)]

*Siapa YANG ulang tahun*  
 who yang birthday  
 ‘Who has a birthday’

The functions of the nominalized verb in Sri Lanka Malay are not a perfect match for those in either Tamil or Sinhala. It lacks the Tamil use in nominalized clauses and in experiential and habitual negatives; it lacks the Sinhala use in all negatives. Its use in (past) relative clauses is not found in either language, though as noted the Tamil verbal noun and verbal adjective (the verb form of relative clauses) are closely related morphologically. The ability of a focal NP or WH-word to appear in stressed post-verbal position when the verb is nominalized but not when the verb is not nominalized is common to all three languages. The optionality of nominalization in focus constructions and WH-questions is common to Sri Lanka Malay and Tamil, while in Sinhala the focus form (whose nominalization function is now obscure) is obligatory. Only on this final point does Sri Lanka Malay display greater similarity to one of its adstrates than to the other. Finally, the linking of nominalization with past-tense interpretation in relative clauses (and possibly WH-questions) and the incompatibility between nominalization and the past tense marker (*su-*) is a Sri Lanka Malay development with no parallels in either Tamil or Sinhala.

Is nominalization a hijacked construction in Sri Lanka Malay? Certainly *yang* as a preverbal marker in Sri Lanka Malay differs radically from its morphosyntactic status in Vehicular Malay. But the fact that all the uses of *yang* in Sri Lanka Malay have parallels in Malay/Indonesian is problematic for a hijacking diagnosis, which would require only minor overlap in usage in a restricted pragmatic context. While it is true that *yang* has not been extended to any uses beyond those found in Vehicular Malay, it

is noteworthy that, apart from the past relative clause, all of the Vehicular Malay usages incompatible with the new status of *yang* as a verbal-noun marker failed to cross over into Sri Lanka Malay.

## 8.6 Discussion

The first section of this paper argued that hijacked constructions are more likely in languages shaped by informal second language acquisition, and thus more common in creoles than in language change in non-creoles. A stringent test for this hypothesis is offered by comparing two types of extreme language contact: creolization and convergence. Converged languages, like creoles display massive contact-induced developments, but, at least in the case of maintained minority languages, this comes primarily from the influence of an L2 on an L1 rather than of an L1 on an L2 as in second language acquisition and creolization. The core sections of the paper carried out a survey of the structure and functions of the nominalized verb in Tamil and four neighbouring non-Dravidian languages. Of these, Sourashtra is a clear case of convergence with the exclusive influence of L2 on L1, while Sri Lanka Portuguese is a clear case of a creole, with early L1 on L2 influence and subsequent L2 on L1 influence. Sinhala is clearly a converged language, but the details of its long sociolinguistic history are imprecisely known. Sri Lanka Malay, the youngest of the four, is of disputed typology, with claims for both creolization and convergence, and plausible sociolinguistic scenarios invoked for both possibilities.

Formally, both Sinhala and Sourashtra have mimicked the structure of the Dravidian nominalized verb by adding a third person marker to a participle. Sri Lanka Portuguese on the other hand has hijacked a relative clause marker. The association of a relative clause structure with a nominalized verb function seems to have been facilitated by clefting constructions using a relative clause in Portuguese. It is noteworthy that the two converged languages, Sinhala and Sourashtra, have both employed calquing rather than hijacking while the undisputed creole (Sri Lanka Portuguese) displays a hijacked construction. Although this study focuses on a single construction, hijacking distinguishes a creole innovation from its counterparts in non-creole languages that have resulted from extreme language contact.

We are now in a position to consider the implications of this study for Sri Lanka Malay. Sri Lanka Malay has repurposed a multifunctional pragmatic marker *yang* as a preverbal-particle to mark nominalized verbs. As in Sri Lanka Portuguese, the adaptation was facilitated by clefting constructions. The morphosyntactic mismatch between Malay/Indonesian *yang* and Sri Lanka Malay *yang* is suggestive of untutored second language acquisition, but since all the current functions of *yang-V* in Sri Lanka Malay likely had precedents in Vehicular Malay, not all the conditions for a diagnosis of hijacking are present. Thus the question of whether this makes Sri Lanka Malay a creole or just creole-like-in-some-respects is not a question that can be resolved here.

The functions of the nominalized verb in the various languages are summarized in Table 8.3. The inter-language differences are striking. Clearly each of the four languages that have adopted the structure has integrated it into its grammar in a unique way. As just noted, the conduit for the borrowing of nominalization in Sri Lanka Malay and Sri Lanka Portuguese seems to have been the focus construction. For a creole such as Sri Lanka Portuguese (and, possibly, Sri Lanka Malay) this accords with the fact that focus is probably an earlier communicative need than other functions, such as a nominalized clause. In Sinhala, too, the focus construction looks at first glance like it played

	Ind Tamil	Sourashtra	SLT	Sinh.	SLP	SLM
Nom. Cl.	+	(+)	+	-	+	-
Focus	opt	opt	opt	+	opt	opt
WH-Q	opt	opt	opt	+	opt	opt/+ in PAST
Negation	EXPER, HABIT	HABIT	EXPER, HABIT	all	HABIT	-
[–spec] sbj	opt	opt	opt	-	- ?	-
'think of'	(+)	-	-	-	-	-
Future	opt	opt	-	-	-	-
Habitual	-	-	(related)	-	+	-
Relative	(related)	(+)	(related)	-	opt rare	+ in PAST / - elsewhere

Legend: + obligatory usage; - absence of usage; (+) optional: other nominalization strategies can also be used; opt. optional usage; (related) a structurally related form is used.

Table 8.3: Summary of verbal noun functions in the five languages. This table covers only the forms examined in detail in the preceding sections. All the languages have additional nominalization strategies which may fill in some of the gaps in the table. E.g. the Sinhala V+eka form (a more recent calque on the Dravidian verbal noun) serves to nominalize a clause.

a major role in the borrowing, given the current restriction in usage, but texts from the 12<sup>th</sup> century show that the Sinhala nominalized verb had a wider range of functions more characteristic of its Dravidian model (Gair 1986:166-9). In Sourashtra similarly, the contexts for borrowing were likely diverse. The wider distribution in these two languages follows from the fact that they both innovated calques on the Dravidian model; the initial usage was thus not restricted by the channelling of the borrowing through a particular function.

The details of the spread of the Dravidian nominalized verb in Southern India and Sri Lanka offers a salutary lesson in historical argumentation. A common position concerning syntactic borrowing claims that language X couldn't have borrowed such-and-such a structure from language Y because the details of their usage differ. This investigation gives the lie to such arguments. For, unless we are prepared to argue that the equivalent of the Dravidian nominalized verb developed independently in all the languages involved, and that the striking similarities in usage are purely accidental, we have to agree with Thomason that to prove contact-induced change, common structural features "need not be identical in all respects" (2001:93).

The question of creole structural uniqueness has been approached from several different angles, from Bickerton's (1984) claim that creoles lay bare the human genetic blueprint for language to McWhorter's (2001) proclamation that "The world's simplest grammars are creole grammars." This study argues that creoles are distinct from a historical perspective in that they are more likely to display hijacked constructions because of the strong untutored second-language input to their formation. There are two caveats to this claim. First, hijacked constructions form only a small minority of creole structures, and there is no claim that creoles eschew other means of grammar formation. Second, hijacked constructions are not exclusive to creoles. In particular, non-creole languages with a significant input from untutored second language acquisition should also be expected to exhibit hijacked constructions, but perhaps less commonly than creoles. The view that creolization is gradual (Arends 1993) blurs the line between such languages and creoles. At the other end of the continuum, languages with less

and less untutored second language input fade gradually into languages with no such input where hijacked constructions are claimed to be extremely rare.

Returning one final time to Sri Lanka Malay, we can observe that the difficulties in resolving its classification result from two factors: first, the similarity of the results of extreme convergence on the one hand and creolization/extreme second language input on the other; second the cline between creoles and languages with no second-language input.

# Bibliography

- Andersen, Henning, 1973. “Abductive and deductive change.” *Language* 49:765–793.
- Ansaldi, Umberto, 2008. “Sri Lanka Malay revisited: Genesis and classification.” In *A world of many voices: Lessons from documented endangered languages*, edited by Arienne Dwyer, D. Harrison, & David Rood. Amsterdam, Philadelphia: John Benjamins., 13–42.
- Arends, J., 1993. “Towards a gradualist model of creolization.” In *Atlantic meets Pacific*, edited by F. Byrne & John Holm. Amsterdam: John Benjamins, 371–380.
- Bickerton, Derek, 1984. “The language bioprogram hypothesis.” *Behavioral and Brain Sciences* 7:173–203.
- Bourdin, Philippe & Jaffar, Mohamed, 2010. “SLM yang: the vanishing relativizer?” Paper presented at ISMIL XIV, University of Minnesota, May 2010.
- Cardona, George, 1965. *A Gujarati reference grammar*. Philadelphia: University of Pennsylvania Press.
- Clements, J. C., 1991. “Two Indo-Portuguese creoles: Languages in transition.” *Hispánica* 74:637–646.
- Clements, J. C., 2009. “Accounting for some similarities and differences among the Indo-Portuguese creoles.” *Journal of Portuguese Linguistics* 8:23–47.
- De Silva Jayasuriya, Shihan, 1999. “Portuguese in Sri Lanka: Influence of substratum languages.” *Journal of the Royal Asiatic Society* 3(9):251–270.
- Englebretson, Robert, 2008. “From subordinate clause to noun-phrase: Yang constructions in colloquial Indonesian.” In *Crosslinguistic studies of clause combining: The multifunctionality of conjunctions*, edited by Ritva Laury. Amsterdam: John Benjamins, 1–33.
- Gair, James W., 1976. “The verb in Sinhala, with some preliminary remarks on Dravidianization.” *International Journal of Dravidian Linguistics* 5:259–272. Selections reprinted in Gair 1998, 200–209, to which page reference is made.
- Gair, James W., 1986. “Sinhala Focused Sentences: Naturalization of a Calque.” In *South Asian languages: Structure, Convergence and Diglossia*, edited by Bh. Krishnamurti, Colin P. Masica, & Anjani K. Sinha. Delhi: Motilal BanarsiDass, 147–164. Reprinted, with additions, in Gair (1998), pp. 153–169, 323–327.
- Geiger, Wilhelm, 1938. *A grammar of the Sinhalese language*. Colombo: Royal Asiatic Society. Reprinted 1995; New Delhi, Madras: Asian Educational Services.
- Holm, John, 2000. *An introduction to pidgins and creoles*. Cambridge: Cambridge University Press.
- Hussainmiya, B. A., 1987. *Lost cousins: The Malays of Sri Lanka*. Bangi: Institut Bahasa, Kesusastraan dan Kebudayaan Melayu.
- Krishnamurti, Bh., 2003. *The Dravidian languages*. Cambridge: Cambridge University

- Press.
- Lefebvre, Claire & Lumsden, John S., 1994. "Le rôle central de la relexification dans la genèse des langues créoles." *Plurilinguismes* 8:47–93.
- Masica, Colin P., 1991. *The Indo-Aryan Languages*. Cambridge: CUP.
- McWhorter, John H., 2001. "The world's simplest grammars are creole grammars." *Linguistic Typology* 5:125–166.
- Meakins, Felicity, to appear. "Mixed languages." In *Contact languages*, edited by Yaron Matras & Peter Bakker. Berlin: Mouton de Gruyter.
- Munkwitz-Smith, Jeffry Clark von, 1995. *Substratum influence on Indo-Aryan grammar*. Ph.D. thesis, University of Minnesota.
- Murugaiyan, Appasamy, 2009. "Focus constructions in modern Tamil." *International Journal of Dravidian Linguistics* 38:43–69.
- Nadkarni, Mangesh V., 1975. "Bilingualism and syntactic change in Konkani." *Language* 51:672–83.
- Nordhoff, Sebastian, forthcoming. "The current state of Sri Lanka Portuguese." *Journal of Pidgin and Creole Languages*.
- Pandit, Parodh B., 1972. "Bilingual's grammar: Tamil-Saurashtra grammatical convergence." In *India as a sociolinguistic area*, edited by Parodh B. Pandit. Poona: University of Poona, 1–25.
- Rhys-Davids, T. W. & Stede, William, 1921–1925/1966. *The Pali Text Society's Pali-English Dictionary*. London: Luzac, 5th edition.
- Rosch, Eleanor, 1975. "Cognitive representations of semantic categories." *Journal of Experimental Psychology: General* 104:192–233.
- Rosch, Eleanor, 1978. "Principles of categorization." In *Cognition and Categorization*, edited by E. Rosch & B. B. Lloyd. Hillsdale, NJ: Lawrence Erlbaum, 27–48.
- Schiffman, Harold F., 1999. *A Reference Grammar of Spoken Tamil*. Cambridge: CUP.
- Selinker, Larry, 1972. "Interlanguage." *International Review of Applied Linguistics* 10:209–241.
- Smith, Ian, 1979. "Convergence in Sri Lanka: A Creole Example." *Lingua* 48:193–222.
- Smith, Ian R., 1985a. "The development of morphosyntax in Sri Lanka Portuguese." *York [UK] Papers in Linguistics* 11:291–301.
- Smith, Ian R., 1985b. "Multilingualism and diffusion: A case study from Singapore English." *Indian Journal of Applied Linguistics* 11:105–126.
- Smith, Ian R., 2005a. "Les missionnaires wesleyens et la conversion du créole portugais de Sri Lanka." Paper read at the annual meeting of Associação de Crioulos de Base Lexical Portuguesa e Espanhola. Orléans, June.
- Smith, Ian R., 2005b. "(Re)conceptualizing the Creole (and Pidgin) Prototype." Paper read at the annual meeting of the Society for Pidgin and Creole Linguistics, San Francisco, January.
- Smith, Ian R., forthcoming. "Measuring substrate influence: Word order features in Ibero-Asian creoles." In *Ibero-Asian Creoles: Comparative Perspectives*, edited by Alan N. Baxter, Hugo Cardoso, & Mário Pinharanda Nunes. Amsterdam: John Benjamins.
- Smith, Ian R. & Paauw, S., 2006. "Sri Lanka Malay: Creole or convert." In *Structure and Variation in Language Contact*, edited by Ana Deumert & Stephanie Durrellman. Amsterdam: John Benjamins, 159–182.

- Steinhauer, Hein, 1992. "On the meaning of *yang* in Indonesian." In *The Language Game: Papers in Memory of Donald C. Laycock*, edited by Tom Dutton, Malcolm Ross, & Darell Tryon. Canberra: Research School of Pacific Studies, 427–439.
- Thomason, Sarah G., 1997. "A typology of contact languages." In *Pidgins and creoles: structure and status including selected papers from the meetings of the Society for Pidgin and Creole Linguistics*, edited by Arthur K. Spears & Donald Winford. Amsterdam; Philadelphia: John Benjamins, 71–88.
- Thomason, Sarah G., 2001. *Language Contact: An Introduction*. Washington DC: Georgetown University Press.
- Toer, Pramoedya Ananta (ed.), 1982. *Tempo doeloe. Antologi sastra pra-Indonesia*. Jakarta: Hasta Mitra.
- Učida, Norihiko, 1979. *Oral Literature of the Saurashtrans*. 6. Hirakata; Calcutta: KUFS; Simant.
- van Minde, D., 2008. "The pragmatic function of Malay *yang*." *Journal of Pragmatics* 40:1982–2001.



## Chapter 9

# The Semantics of Serial Verb Constructions in Sri Lanka Malay

Mohamed Jaffar

*York University, Toronto*

### 9.1 Introduction

The aim of this paper is to examine Serial Verbs<sup>1</sup> in Sri Lanka Malay, at least the current many that I, as a native speaker, can readily tap into.<sup>2</sup> I shall refer often to the work of Sebastian Nordhoff (2009, forthcoming), said work being the only one I know that has dealt with the subject of SLM Serial Verb Constructions in any significant detail. I examine the structure of SVCs (which, in SLM, is straightforward) and, above all, the semantic implications of SLM serial verb constructions. Wherever possible I have tried to place them in the framework of an areal typology, in order to see clearly the influences of the adstrate languages: Sinhala (S), Tamil (T), and Colombo Muslim Tamil (CMT), a variety of Shonam, the language of Tamil-speaking Muslims.

Sri Lanka Malay (SLM) is the language spoken by the descendants of the soldiery that came with the Dutch in the seventeenth century, drawn from a number of islands in the archipelago that comprised the Dutch East Indies. There were also a few exiled Javanese rebel princes, and a few convicts of whom very little information has been recorded or very little is known even anecdotally. There was at one time a high level of language maintenance thanks to the efforts of the British regimental schools that

<sup>1</sup>Serial Verbs (SV) and Serial Verb Constructions (SVC) are used interchangeably in this paper.

<sup>2</sup>For helping me find the right idiomatic expression, I am indebted to Anton Fernando, Nirusha Perera and Iris Wickramasinghe (Sinhala); Zaneera Tegal, Sam Tegal, Zreena Jaffar and Zowriya Jayman (Sri Lanka Malay); Rafiya Cassim and Marhooma Samsudeen (Colombo Muslim Tamil), all native speakers of their respective languages. My grateful thanks go to Ian Smith for his advice in connexion with the morphosyntax of Dravidian languages and Sri Lanka Portuguese — which helped me immensely in understanding analogues in my native Sri Lanka Malay —and to Sebastian Nordhoff for engaging me in lively debate from which I learnt some valuable lessons in navigating the murky waters of terminology. Last but not least, I acknowledge a large debt of gratitude to Philippe Bourdin, linguist, mentor and friend, for his advice on matters requiring fine linguistic judgment, and for his suggestions regarding a few aspects of this paper.

imported teaching materials, mainly from Singapore, and encouraged the children of the Malay expatriates to learn to read Standard Malay and to write it in ‘*gundul*’, the name used in Sri Lanka for ‘*huruf Jawi*’, the Arabic script, suitably modified and used for writing Malay. Primers, copy-books to imitate the ‘*gundul*’ script, prayer-guides, Islamic catechisms and prayer-books, all of these, silver-fish-eaten and coming apart in their bindings, were in the collections of local families. Many have discarded them; many more have donated them to the National Museum. It is reasonable to assume therefore that there was a ‘language-in-the-making’ with Malays intermarrying; mingling with the other communities of the Island; and leading the life of the uprooted as best they could. Over time, a spoken Malay would develop that would retain Malay lexis but re-invent itself morphosyntactically along the lines of Tamil and Sinhala. It is my contention that the influence of Tamil,<sup>3</sup> on balance, is likely to have been greater than that of Sinhala. My reasons for making that assumption are grounded in historical and sociolinguistic realities. As was to be expected, with the winding down of the regimental schools, the Malay teachers of language and religion found employment elsewhere. Itinerant Moor teachers rushed to fill the breach and entered ‘contracts’ with householders to teach Qur’anic Arabic (and Arabic-Tamil) to their little children. Interestingly, side-by-side with writings in Malay — the quality of which was suffering a steady decline — the local Malays also produced works<sup>4</sup> in Arwi or Arabu-Thamul: the variety of Tamil used by Tamil-speaking Muslims, a knowledge of which was important for the community’s religious needs. Many families saved the books in Arwi and gave away those written in Jawi for they no longer were able to read them.<sup>5</sup>

## 9.2 Serial Verb Constructions in SLM

The constructions I propose to examine meet the requirements for serial verb status as specified in [Aikhenvald \(2006\)](#):

1. they are a sequence of verbs acting together as a single predicate;
2. they are monoclausal and allow no markers of syntactic dependency on their components;
3. their intonational properties are those of a monoverbal clause; no intonation breaks or pause markers occur between the components of an SVC;
4. they share TAM, modality and polarity values;
5. together the verbs refer to one single event;
6. prototypical SVCs share at least one argument.

[Aikhenvald \(2006\)](#) also mentions a few other parameters defining the structure of serial verbs and, here again, SLM serial verbs fit unimpeded into the four categories given hereunder:

---

<sup>3</sup>Each Tamil-speaking Muslim community in Sri Lanka has a dialect of its own depending on location (Colombo, Beruwala, Galle, Matara, Kandy, Nawalapitiya, Trincomalee, Batticaloa, and Jaffna, to name a few). It would appear that the community as a whole prefers the name ‘Muslim’ to ‘Moor’ in order to invite into its fold all persons professing the faith of Islam. Sri Lanka Malays, Muslims themselves, have stoutly resisted the invitation on the just grounds that they have a sufficiently clear linguistic identity to subsist on their own.

<sup>4</sup>These works were mainly religious in nature but a few devoted themselves to sorcery and witchcraft.

<sup>5</sup>Personal observations and books retrieved from old collections in my family.

1. COMPOSITION: Asymmetrical and Symmetrical SVCs:

*Asymmetrical* SVCs may consist of a verb from a relatively unrestricted class and another from a semantically or grammatically closed class (*thulis* ‘write’ + (*h*)*abis* ‘finish’ = *write finish* (completive aspect))

*Symmetrical* SVs consist of two or more verbs chosen from a semantically and grammatically unrestricted class: (*lari* ‘run’ + *pi* ‘go’ = *lari pi* = ‘go running’)

2. CONTIGUITY versus NON-CONTIGUITY of the verb components: all SLM serial verbs studied here are consistently contiguous: (*thulis ambel* ‘write take’; *masak simpang* ‘cook keep’; *lapas tharo* ‘let-go put’.<sup>6</sup>)

3. WORDHOOD OF COMPONENTS: all SLM serial verb components studied here are ‘stand-alone’ phonological words that, in combinations, form single lexical entities: *masak* ‘cook’; *simpang* ‘keep’; *thulis* ‘write’; *ambel* ‘take’; *lapas* ‘let go’, ‘set free’; *tharo* ‘put’

4. MARKING OF GRAMMATICAL CATEGORIES IN SV CATEGORIES — such as, for instance, person, tense, aspect, modality, negation or valency changing — may be marked just once per construction (‘SINGLE MARKING’) or can be marked on every component (‘CONCORDANT MARKING’).

The SLM serial verbs studied here are consistently of the ‘single marking’ variety:

- (1) *su-masak simpang* vs \**su-masak su-simpang*  
 PAST-cook keep                    \*PAST-cook PAST-keep  
 ‘was cooked and set aside.’

### 9.3 Verbal Compounds

In describing the Verbal Compound Nordhoff has this to state: ‘The string *kasithaau* in [2] is parsed into one phonological word, which can be seen from the absence of a long vowel<sup>7</sup> in *kasi*’ (Nordhoff 2009:171):

- (2) *Badulla Kandy Matale samma association=nang masà-kasi-thaau*  
 Badulla Kandy Matale all association=DAT must-give-know  
 ‘Badulla, Kandy, Matale, we must inform all associations.’

Before we consider verbal compounds we have first to address the somewhat vexatious ‘long-vowel’ story. Tapovanaye Sutadhara (1995) raised the issue of long and short vowels in SLM and posited the existence of a long penultimate vowel in a word as a mark of its phonological completeness, its wordhood, see Tapovanaye Sutadhara (1995:18).<sup>8</sup>

<sup>6</sup>There being just one type of schwa in my dialect, I have chosen to depict it by /ə/ whereas other contributors might have shown the schwa differently, as /ʌ/, /ɪ/, and /ʊ/ respectively, depending on the perceived surface realizations of the schwa in certain idiolects.

<sup>7</sup>Emphasis mine.

<sup>8</sup>

Sutadhara traces a linguistically plausible phonological surface representation path for the word *paha* ‘thigh’, based on the application of some rules, from initial *paa* → laryngeal insertion *paha* → stress assignment *paha* → vowel lengthening ’*pa:ha* → paragogic glottal insertion ’*pa:ha?* terminating in a word *paa.ha’* (long vowel in the penult, ending in glottal stop). Native speakers of SLM regularly pronounce the word as *paha* (short vowels in both syllables) and would have difficulty in readily identifying *paa.ha’*, as described by Sutadhara, with the SLM word for ‘thigh’.

The issue is whether this is in accord with the intuitions of native speakers. As a native speaker myself, and one who is in constant touch with native speakers of several varieties of SLM (Hambantota, Kandy, Badulla, Colombo), I am inclined to side with the theory of a stress-based pronunciation, as opposed to one of long and short vowels, in determining the ‘phonologically autonomous’ status of compound verbs. Even then, the absence of a stressed vowel in the first verb (*kasi* in *kasi thaau* in the example given above) may not be a defining element. How else could one explain a set of serial verbs, in everyday SLM use, that might lay a legitimate claim to lexicalization as a ‘single-word’ despite being articulated clearly as separate phonological words: *aja*<sup>9</sup>

‘invite’ + *ba’wa* ‘bring’ resulting in *a ja bawa*, ‘escort someone toward the speaker’; and, also with *bapi*<sup>10</sup> ‘take away’ in *a ja bapi* ‘escort someone away from the speaker’. The first verb (in diachrony, *ajak*, from Vehicular Malay, ‘to invite, solicit’ is no longer in autonomous use in SLM, and is no longer productive. There is no inkling of the possible meaning of that first verb in SLM collective memory today, and both speaker and hearer consider *a ja bawa* (often contracted to *a ja ba*), and *a ja bapi*, to be single lexical entities. This is all the more evident in the serial verb construction *pi+a ja bawa*:

- (3) *Mina, chàpat bapa=nya pi aja bawa!*  
 Mina, soon father=acc go accompany bring!  
 ‘Mina, go fetch Dad quick!’

Yet, these verbs are pronounced respectively *a ja bawa* and *a ja bapi*, i.e. with equal stress in the penult of each contributing verb. The forms *\*ajabawa*, and *\*ajabapi* — where the stress is on the penult of only the second contributing verb, as it would rightly be had the two verbs been parsed as a single phonological word — are both not attested in any variety of SLM that I am acquainted with.

However, contrary to what I claimed in [Bourdin & Jaffar \(2010\)](#), I now agree with [Nordhoff \(2009\)](#) that both ‘*kasithau*’ and ‘*kasikawing*’ must be regarded as single verbs<sup>11</sup> in their own right. To sum up, SLM possesses five SVCs that have undergone lexicalization, and there appears to be no likelihood of their increasing in number:

1. *kasi thau*, ‘advise’, ‘inform’,
2. *kasi kawing*, ‘give in marriage’,
3. *kasi kànal* ‘make known’, ‘introduce’, ‘present’, as in one person to another, a rare but nonetheless attested variant of *mà-kànakang lìpas*, also *kànakang lìpas*,
4. *a ja bawa*, ‘accompany toward the speaker’,
5. *a ja bapi* ‘accompany away from the speaker’.

This is in accord with what [Aikhenvald \(2006:51\)](#) has to say about one-word serial verbs, ‘[they] tend to be restricted to a more limited set of verb roots. That is, if a language has one-word and multiple-word serial verbs, the former tend to be limited, and the latter productive.’

<sup>9</sup>Sutadhara proposes a faulty etymology, (*h)ajat* (VM from Arabic, meaning ‘wish’- n.) leading to *a ja(t) bapi* [?a:ja’ba:pɪ?] ([Tapovanaye Sutadhara 1995:26](#)). Irrefutably, the better choice for this serial verb pair is *ajak* (VM ‘invite’ –v.) producing *a ja(k) bapi* [a:ja ba:pɪ]. I have never heard this combination pronounced as suggested by Sutadhara, with a short vowel in *a ja* and a long vowel in *bapi* [a:ja.ba:pɪ], among native speakers of SLM.

<sup>10</sup>*Bapi* is undoubtedly a grammaticalization of an older SVC: *bawa* ‘bring’ + *pergi* ‘go’ evolving into →*bawa+pi* → *ba+pi* → today’s *bapi* ‘take away’.

<sup>11</sup>He has put forward other convincing arguments save for those regarding the ‘long vowel’.

## 9.4 Full verb serialization

I would propose, for SVCs proper, the following general characterization:

Two or more verbs operating in a series, mainly contiguously, to provide in a synergistic manner a shade of meaning that each verb is incapable of conveying on its own.

In SLM (as in other languages) the class of SVCs proper is partitioned into two subsets: (a) the two verbs act in tandem and the resulting combination carries the sum of their lexical meanings, in which case we are dealing with *full verb serialization* (FVS), or (b) one of the two verbs undergoes some degree of semantic depletion in the process of serialization, in which case it is appropriate to call it a *vector verb* as it ‘points’, metaphorically, the direction in which the main verb is semantically headed.

As each of the participating verbs contributes its lexical content to the meaning of the whole, the FVS type is more productive in SLM than the ‘vector verb’ type where both the contributing verbs and resulting combinations belong to restricted classes. Mostly full verb serializations combine verbs of motion: *lari* ‘run’, *pi* ‘go’, *dathang* ‘come’.

SVCs of the following type as noted by Nordhoff (2009) are full verb serials<sup>12</sup> with one TAM marking:

1. *jalang pi* ‘walk go’, ‘go walking’
2. *lari kàluling*,<sup>13</sup> ‘run around’
3. *lari dathang* ‘come running’
4. *cari dathang* ‘come looking for’
5. *ambel bapi* ‘take away’

I shall give two examples here:

- (4) *kàcil (k)anak-anak pada lapang=ka adà lari kàluling*  
 little child PL field=LOC NONPAST run around  
 ‘The little children are running around on the green.’
- (5) *polis as=dathang yang-curi barang=pada=nya su-ambel bapi*  
 police CP-come REL-stole(n) goods=PL=ACC PAST-take take-away  
 ‘The police came and took away the goods that were stolen.’

## 9.5 Vector Verbs

I shall conform here to the definition in Nordhoff (2009) which I think captures accurately the semantics of the linguistic operation performed by this type of serial verb.

<sup>12</sup>The SLM words shown here have been re-written in a spelling convention of my own for conformity’s sake and vary from Nordhoff’s in that they do not portray pronunciation, e.g. ‘long vowels’ and doubled consonants.

<sup>13</sup>Nordhoff gives the meaning ‘go astray’ as elicited by him among the up-country Malays. In my dialect, as in many others, *lari kàluling* simply means ‘run about’.

A verb in this construction, *usually*<sup>14</sup> the second in the series, imparts some useful aspectual or other grammatical information, to complement or augment the meaning of the first verb. For SLM, Nordhoff (2009:174) lists a number of these vector verbs, viz. *ambel* ‘take’, for (self)-benefactive and ingressive; *kasi* ‘give’, for benefactive; *(h)abis* ‘finish’, for completive; *tharo* ‘put’, for affective; *simpang* ‘keep’, for continuative, prospective; *duduk* ‘sit’, for progressive; *kàna* ‘strike’, for adversative; and *pukul* ‘hit’, for intensive-aggressive (see Section 9.5.11 of this work a detailed discussion of this last-named).

### 9.5.1 *Ambel* ‘take’

Verb combinations with *ambel* as the vector have variously been described as having BENEFACTIVE, INGRESSIVE-INCEPTIVE-INCHOATIVE meanings. As a rule, though, the versatility of *ambel* lies in its ability to impart to the verb-sequence a notion of ‘seizing or appropriating’ (the general acception of ‘take’), thus rendering more lasting the activity described by the main verb. This is true of most of the *ambel* compounds.

In order to highlight the nuances of meaning, I give below a few examples — together with their equivalents in CMT and S — firstly of an utterance employing the core verb unaided by *ambel*, secondly followed by an SVC using the same core verb with *ambel* as vector, e.g. *thulis* ‘write’. However, with *ambel* ‘take’, the combined serial verb ‘write take’ conveys the BENEFACTIVE meaning ‘note down’, ‘record’.

- (6) a. *Amath telfon nombär=nya su-thulis SLM*  
Amath telephone number=ACC PAST-write  
'Amath wrote the telephone number.'
- b. *Amath telfon nombär=nya su-thulis ambel*  
Amath telephone number=ACC PAST-write take  
'Amath wrote down the telephone number (recorded it)'
- (7) a. *Amath telfon nombər=e e|uzinaan CMT*  
Amath telephone number=ACC PAST-write  
'Amath wrote the telephone number.'
- b. *Amath telfon nombər=e e|uzi konqaan*  
Amath telephone number=ACC CP<sup>15</sup>.write PAST-take  
'Amath wrote down the telephone number (recorded it)'
- (8) a. *Amath tælifon angkay=ə liwwa S*  
Amath telephone number=ACC write.PAST  
'Amath wrote the telephone number.'
- b. *Amath tælifon angkay=ə liya gatta*  
Amath telephone number=ACC write.CP take.PAST  
'Amath wrote down the telephone number (recorded it).'

The same processes produce similar differences in meaning with *dàngar* ‘hear’ and *dàngar ambel* ‘hear take’ = ‘heed, listen, pay attention to’ and their equivalences in CMT *keel*/*keettu ko* and Sinhala *ahannə/aha gannə*.

<sup>14</sup>Cf. (*à*)*mbath* further along which refutes this statement by being the first of the two verbs, contrary to the usual pattern.

<sup>15</sup>Arden (1962[1891]) calls it the Verbal or Adverbial Participle.

- (9) a. *yang-bilang=nya anak thra dàngar SLM*  
          REL-say=ACC child NEG hear
- b. *ſenn-att=e pulle=KKI keekkə ille CMT*  
          say-NMLZ=ACC child=DAT hear NEG
- c. *kiyəpu-de lamaya=tə æhune næhæ S:*  
          say-NMLZ child=DAT hear NEG  
          'The child didn't hear what was said.'
- (10) a. *yang-bilang=nya anak thàma(u) dàngar ambel SLM*  
          REL-say=ACC child NEG want hear take
- b. *ſenn-att=e pulle keet̪tuk kollaadu MT*  
          say-NMLZ=ACC child hear take NEG
- c. *lamaya kiyəpu=de aha gannə næhæ S*  
          Child say.PTCP=INTERR hear take NEG  
          'The child will not listen to what is told him.'

Occasionally there is a RECIPROCAL meaning, for example:

*pukul* 'hit'/*pukul ambel* 'hit take; hit each other, fight'. The CMT rendering of this is *adj/adicci.ko*; the Sinhala forms are *gahanno/gaha ganno*.

- (11) a. *ithu orang anjing=nya arà-pukul SLM*  
          DIST man dog=ACC NONPAST-strike
- b. *andə manufan naaiy=e adjikkiraan CMT*  
          DIST man dog=ACC NONPAST-strike
- c. *ara minihaa balla=tə gahanəvaa S*  
          DIST man dog=DAT NONPAST-strike  
          'That man is striking the dog'
- (12) a. *(k)anak-anak jalang=ka arà-pukul ambel SLM*  
          child~RED 3PL street=LOC NONPAST-hit take
- b. *pullei-əl theru=le adjicci kolraanuvəl CMT*  
          child-3PL street=LOC hit.cp.3p take
- c. *lama-yin paar=e gaha gannəvaa S*  
          child-3PL street=LOC hit take  
          'The children are fighting each other in the street'

It seems perfectly legitimate to give here an account of the function of *ambel* in terms of the TRANSITIVITY HYPOTHESIS propounded by [Hopper & Thompson \(1980\)](#). It appears that one fundamental property of *ambel* would be to carry the semantics of the verb-combination to a higher order of transitivity by reason of some resulting characteristic such as VOLITIONALITY, INDIVIDUATION, AFFECTEDNESS, or AGENTIVITY:

In (6b), *thulis+ambel*: Amath is recording the number (deliberateness of the Agent, high degree of volition) and the number itself now takes on, so to speak, a status it did not possess prior to the act of its being written down. In other words, *ambel* has a transitivity effect both in terms of AGENTIVITY of the Subject and AFFECTEDNESS of the Object;

In (9a), *dàngar*, by itself, refers to a sensory event that is involuntary;

In (10a), *dàngar+ambel*, together the two verbs refer to an act of volition. The referent of the subject is not the passive recipient of a sensory impression but an Agent with a degree of control. So, much as with *thulis*, *ambel* acts as a transitivizer by enhancing the AGENTIVE and VOLITIONAL character of the process.

In (11a), *pukul*, by itself, we have one Agent and one Patient.

In (12a), *pukul+ambel*, however, the addition of *ambel* entails an increase in AGEN-TIVITY (the meaning of reciprocity that it contributes elevates both participants to the status of Agents).

### 9.5.2 Àmbath ‘thrash, lash, whip’

Àmbath<sup>16</sup> must qualify as the vector verb of choice in SLM to express the INTENSITY of an act described by the main ‘action’-verb: (*à*)*mbath* “thrash” + V = intensive, do V with vigour, intensity.

- (13) *meja=ka yang-ada makanan=nya se makang*  
table=LOC REL-be food=ACC 1SG eat  
'I ate the food that was on the table.'
- (14) *meja=ka yang-ada makanan=nya se mbath makang* <sup>17</sup>  
able=LOC REL-be food=ACC 1SG strike eat  
'I ate up<sup>18</sup> the food that was on the table.'
- (15) *polis=dari sini dathang (s)aja, maling de lari*  
police=ABL here arrive only thief 3SG run  
'As soon as (they) from the police came here, the thief ran.'
- (16) *polis=dari sini dathang (s)aja, maling de mbath lari*  
police=ABL here arrive only thief 3SG strike run  
'As soon as (they) from the police came here, the thief sped off (running).'
- (17) *ithu bothol sopi=nya de minung*  
DIST bottle arrack=ACC 3SG drink  
'He drank the bottle of arrack.'
- (18) *ithu bothol sopi=nya de mbath minung*  
DIST bottle arrack=ACC 3SG strike drink  
'He drank up the bottle of arrack.'

CMT and S do not have similar constructions, i.e. with an equivalent vector verb of intensity.

<sup>16</sup>An entry for àmbath /əmbat/ in Stevens & Schmidgall-Tellings (2004) reads as follows: ‘embat mengembat 1 to lash/whip/thrash with a strip of bamboo, a piece of rope, etc.’

<sup>18</sup>Breaking away from the general SLM pattern, the bleached verb is on the left of the main verb it modifies, a phenomenon not unusual in multiple verb constructions (cf. vector verb *kàna*). I cite Nordhoff (2009): ‘4.1.6. Position of the bleached and the unbleached verb...the bleached verb can be on the left side or the right side, and the unbleached verb then occupies the other position.’

<sup>18</sup>An action viewed from its endpoint, i.e. a telic action...in the telic sense. *I ate it up*, the activity is viewed as completed, and is carried out in its entirety...' Hopper & Thompson (1980)

### 9.5.3 *Kàna* ‘be struck by’

*Kàna* appears to be the veritable workhorse in vector-verb constructions. It must be noted that *pukul* and *kàna* are not interchangeable in their usage for, when it occurs in isolation, *pukul*, the transitive verb, has the active meaning of ‘strike’ as opposed to *kàna*, the intransitive verb, which has the inherently passive meaning of ‘be struck’:

- (19) a. *ithu orang anjing=nya arà-pukul SLM*  
DIST man dog=ACC NONPAST-strike
  - b. *andə manufən naaiy=e adikkiraan CMT*  
DIST man dog=ACC strike.NONPAST
  - c. *ara minihaa balla=tə gahanəvaa S*  
DIST man dog=DAT strike.NONPAST  
‘That man is striking the dog’
- (20) a. *\*ithu orang anjing=nya arà-kàna SLM*  
DIST man dog=ACC NONPAST- be.struck
  - b. *\*andə manufən naaiy=e pacukiraan CMT*  
DIST man dog=ACC be.struck
  - c. *\*ee miniha balla=tə vadinaavaa S*  
DIST man dog=DAT be.struck  
(‘That man (be struck) the dog’, ungrammatical because of the verb with an inherently passive meaning)

Also, *kàna* has an extended meaning of ‘come in contact with’:

- (21) a. *se=pe kaki pinthu=ka arà-kàna SLM*  
s1s=POSS foot door=LOC NONPAST- be.struck
  - b. *end=e kaal kadavu=le padudu CMT*  
1SG=POSS foot door=LOC be.struck
  - c. *ma=ge kakula dor=e vadinaavaa S*  
1SG=POSS foot door=LOC be.struck  
‘My foot strikes (comes in contact with) the door.’
- (22) a. *\*se=pe kaki pinthu=nya arà-pukul SLM*  
1SG=POSS foot door=ACC NONPAST- strike
  - b. *\*en=də kaal kadavu=e adikkidu CMT*  
1SG=POSS foot door=ACC strike
  - c. *\*ma=ge kakula dor=tə gahanavaaa S*  
1SG=POSS foot door=DAT strike  
(‘My foot strikes/hits the door’, ungrammatical because verb calls for a human Agent and ‘foot’ does not qualify).

SLM, in common with CMT, has no morphological device for Passive Voice. Muslim Tamil turns to an auxiliary verb *padu* (which, coincidentally, has the same inherently passive meaning of ‘be struck’ or ‘come in contact with’ as does *kàna* in SLM) with the difference that where CMT uses ‘infinitive + *padu*’ SLM employs ‘*kàna* + V1’ (core verb). Sinhala, on the other hand, has no passive formation; rather, it relies on the lexical choice of an intransitive or involitive verb to achieve the semantic equivalent of a passive (Gair & Paolillo 1997:39).

As Nordhoff (2009:188) rightly observes: ‘It is therefore not the case that *kìnna* [=kàna] changes the syntactic status of arguments, as a passive construction would do. Rather, it contributes a semantic shade of meaning, very much in the way other vector verbs do.’

- (23) a. *th(r)a kà-thau-an=nang duith su-kàna bayar* SLM  
NEG NMLZ-know-NMLZ=DAT money PAST-be.struck pay
  - b. *teriyaamə salli ket̪tip pat̪ucci* CMT  
Unknowingly cash pay.INF be,struck.PAST  
‘Unwittingly the money was paid out.’
- (24) a. *anging=nang jànela su-kàna buka* SLM  
wind=DAT window PAST-be.struck open
  - b. *kaattuk=kú jannal torandup pat̪ucci* CMT  
wind=DAT window open.INF PAST-be.struck  
‘The window got opened for (by) the breeze.’

*Kàna* also participates in idiomatic constructions, with some intransitive verbs, that have no MT or S equivalents:

- (25) *inceia sama se=dang m-omong th(r)a dapath*  
3SG.HON COMIT 1SG=DAT INF-speak NEG get  
‘I didn’t get to speak with him’
- (26) *inceia sama se=dang m-omong th(r)a kàna dapath, bukang!*  
3SG.HON COMIT 1SG=DAT INF-speak NEG be.struck get EXPL  
‘I didn’t get to speak with him, y’know !’ (the circumstances just weren’t propitious).
- (27) *kàmareng pagi siang~siang se bangung*  
yesterday morning early 1SG rise  
‘I woke up early yesterday morning.’
- (28) *kàmareng pagi siang~siang se kàna bangung*  
yesterday morning early 1SG be struck rise  
‘I woke up early yesterday morning (involuntarily).’

#### 9.5.4 *Simpang*, ‘keep’, ‘preserve’

*Simpang* usually conveys the meaning ‘set aside (for later) and can be seen as prospective: *Masak* ‘cook’ + *simpang* ‘keep’ = ‘cook keep’ (cook and set aside)

- (29) a. *Àma, nasi sàdikith masak simpang sahar wakthu=nang!* SLM  
Ma rice little cook keep sahar time=DAT
- b. *Ummaa, soor konyəm aakki veyyngə sahar neeratt-ukku!* CMT  
Ma rice little cook keep sahar time-DAT
- c. *Amma, sahar velaawə-tə bat̪ tikak uyaala tiyannə!* S  
Ma sahar time-DAT rice little cook keep  
‘Mum, steam some rice (and set it aside) for the *sahar* (ritual Ramadhan meal)!

### 9.5.5 *Liath* ‘see’

*Liath* conative, V+*liath* mean ‘try to V’, e.g. *pake* ‘wear’ + *liath* ‘see’ = ‘wear see’ (try on some garment):

- (30) a. *ithu kemeja=nya pake liath SLM*  
          DIST shirt=ACC wear see  
   b. *andə sattey=e ucluttu paarungə CMT*  
          DIST shirt=ACC wear see  
   c. *ee kamise æ̃ndəla balannə S*  
          DIST shirt wear see  
          ‘Try on that shirt !’

### 9.5.6 *(H)abis* ‘finish’

*(H)abis*, completive, conveys a meaning of V completed,  
*Makang* ‘eat’ + *(h)abis* ‘finish’ = ‘eat finish’, ‘finish eating’:

- (31) a. *nyari sore kithang siang~siang su-makang habis SLM*  
          today evening 1PL early~RED PAST-eat finish  
   b. *ində andikk=i neerattoodə naangə tindu muclicci<sub>t=t=oom</sub> CMT*  
          this evening=DAT early 1p eat.CP finish.PAST=1p  
   c. *adə hæ̃ndææwe api veelapaing kaala ivarayi S*  
          today evening 1PL early eat.CP finish  
          ‘We finished eating (dinner) very early this evening.’  
          (H)abis ‘finish’

### 9.5.7 *Kasi* ‘give’

*Kasi*, benefactive, conveys the notion of V being benefactive to the patient:

- (32) a. *the athu cangker me-thuang kasi=si? SLM*  
          tea one cup INF-pour give=INTERR  
   b. *tee oru koope uutti tara=vaa? CMT*  
          tea one cup pour give=INTERR  
   c. *tee kooppe=yak hadala dennə=də S*  
          tea cup=3SG make give=INTERR  
          ‘May I pour (you) a cup of tea?’

### 9.5.8 *Tharo* ‘put’

*Thaaro* adds the meaning of detrimentally affective to Patient,

- (33) a. *ithu pohong järok=nya derang su-pothong tharo SLM*  
          DIST tree lime=ACC 3PL PAST-cut put  
   b. *andə deefikka maratt=e vetti poo-<sub>t=t</sub>uaa=<sub>ŋ</sub>go CMT*  
          DIST lime tree=ACC cut put-PAST=3PL  
   c. *ara dehi gaha kapala dæmma S*  
          DIST lime tree cut.CP put.PAST  
          ‘They cut down that lime tree.’

### 9.5.9 *Duduk* ‘sit’

*Duduk*, durative, conveys an ongoing aspect of V:

- (34) *nene=nya arà-masak waktu cucu=pada arà-maeng*  
          grandmother=POSS NONPAST-cook time grandchildren=PL NONPAST-play  
*duduk SLM*  
          sit  
        ‘While their grandmother was cooking, the children were playing.’ [CMT and S do not have an equivalent SVC with ‘sit’ as a vector verb].

### 9.5.10 Verb reduplication

Two more examples of Serial Verb Construction, actually a process of reduplication to indicate an ongoing process, bear reporting and these appear to be calqued on Sinhala:

#### CONTINUATIVE or PROGRESSIVE reduplicated verb

*nangis~nangis*: ‘weep.weep’ to mean ‘keep on weeping’

- (35) a. *Mina kamar=ka nangis~nangis arà-duduk SLM*  
          Mina room=LOC weep~weep NONPAST- sit  
        b. *Mina kaamar=e andə~andəa innəvaa S*  
          Mina room=LOC weep~weep NONPAST- sit  
        ‘Mina is weeping (keeping on weeping) in the room.’

It must be noted here that Bahasa Indonesia has both *bertangis~tangisan* and *menangis~nangis* to convey the continuative-progressive sense of ‘keeping on weeping’ — and it is quite possible that SLM has carried this construction over from its forbear, Vehicular Malay.

#### PROGRESSIVE reduplicated SVC with *ambel*

- (36) a. *lari ambel~lari ambel bas=ka as-lompat de su-naik SLM*  
          run take~run take bus=LOC CP-jump 3SG PAST-climb  
        b. *oɔdʒi konɖu~oɔdʒi konɖu paanji bas=le eerit̪aan CMT*  
          run take~run take jump bus=LOC climb  
        c. *duvaa gənə~duvaa gənə pænala bas=ekee nægga S*  
          Run take~run take jumped bus=LOC climb  
        ‘He ran and (while running) hopped on the bus’

### 9.5.11 The case of *pukul*: vector verb or full verb serialization?

Nordhoff (2009:186) classifies *pukul* as a vector verb, as in (37) *buvang-puukul* and (38) *bale-king puukul*, and states that it is ‘indicative of aggressive or vigorous activity.’ This meaning, however, is not attested in the regular speech of Malays in Colombo and suburbs, Hambantota, and Badulla, which suggests that we may be dealing with a form of idiolectal usage.

- (37) [*Incyang=pe<sup>19</sup> kàpaala=ka anà-aada*] *thoppi=dering moonyeth pada=nang*  
          3SG.POLITE=POSS head=LOC PAST-exist hat=ABL monkey PL=DAT  
        *su-buvang puukul*  
          PAST-throw hit

'He took the hat from his head and violently threw it at the monkeys.'

A careful parsing of (37) will show that the meaning is one of 'striking by means of throwing' (*buang+pukul*, typically a full verb serialization), and not one of violent throwing.

- (38) *Ithu=kapang ithu moonyeth pada=le [anà-maayeng duuduk thoppi]*  
 dist=when DIST monkey PL=ADDIT PAST-play sit hat  
*pada=dering inni oorang=nang su-bale-king puukul*  
 PL=ABL NEG man=DAT PAST-return-CAUS hit

'Then the monkeys threw back the hats with which they had been playing.'

Again, in (38) above, *bale-king puukul* (*balek-king pukul*) is clearly 'strike back', not 'throw back violently'.<sup>20</sup> Both verbs contribute to the semantics of the statements in (37) and (38) with their individual literal meanings, and these examples with *pukul* must therefore be read as full verb serializations (FVS). Note that the productivity of *pukul* as a vector is severely constrained for otherwise we should have expected to encounter *\*tholak pukul* 'push violently', *\*thàndang pukul* 'kick violently'; *\*tharek pukul* 'pull violently'; and even *\*pukul pukul* 'hit violently'.

## 9.6 Conclusion

Sri Lanka Malay serial verb constructions are prototypical in that they are a sequence of verbs acting together as a single predicate without any overt marker of coordination, subordination or syntactic dependency of any sort; they operate in a series, mainly contiguously, to provide in a synergistic manner a shade of meaning that each verb is incapable of conveying on its own; consistent with the conditions suggested in Aikhenvald (2006), they describe what can be conceptualized as a single event; they are monoclausal; their intonational properties are those of a monoverbal clause; they have just one tense, aspect and polarity value; and, additionally, they share all arguments. In multi-word constructions (as opposed to one-word or compound serial verbs) they are of the 'single marking' variety, i.e. they are marked just once and not on every component verb. Verbal combinations involving *ambel*, *(a)mbath* and *kàna* have all undergone a process of lexicalization and this is evident from the fact that their meaning is not compositional. All the other verbal sequences we have examined carry meanings that are essentially compositional in the sense that each of the two verbal elements retains its original semantics. Nearly all of the SVCs examined in this paper seem to be calqued on corresponding constructions in Colombo Muslim Tamil, and sometimes on those of Sinhala, suggesting a notable adstrate influence in the evolution of the language. Nonetheless, there are a few constructions — especially in the 'reduplicating' variety (36b) and in the *kàna* 'pseudo-Passive' variety [cf. (24b) and (26)] — where there is reason to postulate a Vehicular Malay origin.

<sup>19</sup>Items (37) and (38) are taken from Nordhoff (2009:186) where they bear the numbers (83) and (84) respectively.

<sup>20</sup>The verb *buang* 'throw' is nowhere in the example even though the author notes that for *puukul* as a vector verb 'the only instances are related to throwing an item,' (*idem*).



# Bibliography

- Aikhenvald, Alexandra, 2006. “Serial Verb Constructions in Typological Perspective.” In *Serial Verb Constructions: A Crosslinguistic Typology*, edited by A. Aikhenvald & RMW Dixon. Oxford: Oxford University Press, 1–60.
- Arden, A. H., 1962[1891]. *A Progressive Grammar of the Tamil Language*. Madras: Christian Literature Society.
- Bourdin, Philippe & Jaffar, Mohamed, 2010. “SLM yang: the vanishing relativizer?” Paper presented at ISMIL XIV, University of Minnesota, May 2010.
- Gair, James W. & Paolillo, John C, 1997. *Sinhala*. München: Lincom Europa.
- Hopper, Paul & Thompson, Sandra A., 1980. “Transitivity in Grammar and Discourse.” *Language* 56:251–299.
- Nordhoff, Sebastian, 2009. *A Grammar of Upcountry Sri Lanka Malay*. Ph.D. thesis, University of Amsterdam, Utrecht.  
URL <http://www.lotpublications.nl/publish/articles/003744/bookpart.pdf>
- Nordhoff, Sebastian, forthcoming. “Multi-verb constructions in Sri Lanka Malay.” *Journal of Pidgin and Creole Languages* 27(2).
- Stevens, Alan M. & Schmidgall-Tellings, A. Ed. (eds.), 2004. *A Comprehensive Indonesian-English Dictionary*. Athens, Ohio: Ohio University Press.
- Tapovanaye Sutadhara, 1995. *Vowel Length in Sri Lankan Malay*. Master’s thesis, University of Hawaii.



# **Chapter 10**

## **The genesis of Sri Lanka Malay as a multi-layered process**

Sebastian Nordhoff

*Max Planck Institute for Evolutionary Anthropology*

### **10.1 Introduction**

Sri Lanka Malay differs remarkably from all other varieties of Malay. The most notable differences are morphological complexity and word order, but more subtle changes, like paragogic velar nasals are also found. In this paper, I take stock of the features which can be seen as defining for Sri Lanka Malay and try to trace them to the input languages Malay, Tamil, and Sinhala. I furthermore try to establish the most likely point in time where the change took place as well as the process at work.

The set of ‘Sri Lankan’ features includes features which can be considered peculiar by the casual observer familiar with the most prominent Malay varieties, Bahasa Malaysia and Bahasa Indonesia. Some of these features are in fact not innovations on Sri Lankan soil, but predate the arrival of the Malays in Sri Lanka and are also found in basilectal varieties of Malay. Some of these features have occasionally been claimed to be of Tamil or Sinhala influence because Standard Malay does not show them. By including them in the set of features to investigate, we can sort out which of the ‘Sri Lankan’ features are indeed Sri Lankan, and which are retentions from basilectal varieties of Malay.

Language change is the result of speakers’ adaptation to the society which surrounds them. As such, social variables like demographics, prestige, political situation, and legislation influence the speech of speakers, and as a result, language change. There is also a residue of internal change, or drift, which will be discussed where relevant. In the Sri Lankan context, however, non-internal language change is by far more important.

In order to arrive at a useful classification of the ‘Sri Lankan’ features, we must first take a look into the social and demographic history of the Sri Lankan Malays. We can distinguish 5 periods based on the following landmark events:

1. The arrival of the Dutch in Ceylon (1656)

2. The eviction of the Dutch by the British and the creation of the Malay regiment (around 1800)
3. The disbandment of the regiment (1873)
4. The independence of Ceylon from the UK (1948) and the establishment of nationalist language policies favouring Sinhala (1956)

The Dutch had captured parts of the Indonesian archipelago in the early 17<sup>th</sup> century and had set up their capital in Batavia on the major island of Java. In order to continue their conquest, they recruited native soldiers from Java and other islands, who would eventually make it to Sri Lanka in 1656. These recruits came from all over the archipelago, but the Eastern islands of the Moluccas were overrepresented ([Paauw this volume](#)). The recruits came from a variety of ethnic backgrounds and spoke widely diverse languages. Around Batavia, so-called *kampungs* were set up, which were ethnically quite homogeneous. For purposes of inter-ethnic communication, the different groups used Malay as a lingua franca. Malay had had a long history as lingua franca in the archipelago, even in areas where no native Malays were found. Different varieties of this trade language existed, with substrate influence giving rise to diverging structures. This difference is still discernible today in contemporary daughter languages, spoken in different location in the East of Indonesia ([Paauw 2008](#)).

These lingua franca versions of Malay are very different in structure from what is known in most of the academic world about Malay/Indonesian, and some of the peculiar features of Sri Lanka Malay are actually shared with them, so that these features need not necessarily have their origin in Sri Lanka. The structures of SLM which can be attributed to its origin as an offshoot of the trade varieties of Malay are an interesting area of research, but are only marginally relevant to the language's fate on Sri Lanka. I will classify them as 'Stage 0'. Stage 0 covers everything before the 17<sup>th</sup> century.

In Batavia, speakers with different version of the lingua franca met, and eventually mingled. The precise time of the mingling is difficult to establish, but it is sure that as soon as the ships to Sri Lanka set sail, ethnic heterogeneity was the norm. This is comparable to other displaced population with dialectal diversity, e.g. Indians in South Africa ([Mesthrie 1993](#)) or British in New Zealand ([Trudgill 2004](#)). Documentation of the events in South Africa and New Zealand can inform us about the processes likely to have taken place in the Malay context. These processes generally go under the rubric of 'dialect levelling'. Dialect levelling probably already took place back in the Jakartan kampungs, but gained in importance as a process when the recruits reached Sri Lanka. The period of dialect levelling was finished when Sri Lanka Malay had established itself as a focussed variety with comparatively little internal variation. There is agreement that the point of stablization predates the arrival of recruits from the Malay peninsula, starting 1819. Peninsular Malay has had close to no impact on SLM, suggesting that the new arrivals adapted to the local standard, which must have been established at that time (compare the 'Founder Principle' [Mufwene 1996](#)). This period of dialect levelling is Stage 1. It starts with the establishment of kampungs in Batavia, and ends before the Malaysians arrived, giving it the approximate time span of about 1640-1810.

In Sri Lanka, the Malays started to interact with the local population. This gave rise to at least three different processes, which are more difficult to pin down than the ones discussed above. The easiest one is the most recent, convergence towards Sinhala. After the independence of Ceylon, the Sri Lankan government enacted a number of laws favouring Sinhala in public life. This led to a sociolinguistic shift with Sinhala

gaining prestige and domains while the other languages receded. This applies not only to Sri Lanka Malay, but also to English and the Moor variety of Tamil. The timing of this process is easy to link to the establishment of said laws in the 1950s. The processes at work are convergence and attrition, and the period extends to the present. I will catalogue this process as ‘Stage 4’.

The remaining stages 2 and 3 are the most difficult to analyse since they overlap to a large extent. Stage 2 covers the reinforcing of marginal Malay structures by Sinhala and Tamil (‘Substrate reinforcement’); Stage 3 covers the imposition of novel Lankan structures upon Sri Lanka Malay, which were not found in any of the Trade Malay varieties. These two periods basically began with the first contact between Malays and Lankans and continue up to this day. The process at work in Stage 2 was substrate reinforcement; as for Stage 3, various proposals have been made: creolization ([Smith et al. 2004](#)), convergence ([Bakker 2006](#)), or metatypy ([Ansaldi 2008](#)). See Table 10.1. It is one goal of this paper to disentangle various claims about language contact, language change, demographics, sociolinguistics and processes at work by sorting the ‘Sri Lankan’ features into the different stages established. Anything which is a result of stages 0, 1, 2, or 4 cannot be used as an argument for the process at work in Stage 3 (cf. [Smith this volume](#), [Slomanson this volume](#)). Note that I also include a Stage 5 as a category for developments which are internal and are not due to any kind language change, while at the same time not being shared with other varieties of Malay either.

stage	process	influence	place	time
0	Pre-SLM	Malay	Indonesia	<1600 -1800
1	Dialect Leveling/koineization	Malay	(Batavia,) Sri Lanka	1650 - 1800
2	Substrate Reinforcement	Malay+Lankan	Sri Lanka	1650 - today
3	???	Lankan	Sri Lanka	??
4	convergence towards Sinhala	Sinhala	Sri Lanka	1950-today
5	internal development		Sri Lanka	1650-today

Table 10.1: The different stages of language contact in Sri Lanka Malay.

I will now discuss the ‘Sri Lankan’ features and their classification.

## 10.2 What happened

In this paper I will investigate 60 features which distinguish SLM from other varieties of Malay. Most of the features are clearly Sri Lankan, but some of the features are common to spoken varieties of Malay and distinguish them from the written standard language. These features are included for the benefit of the readers familiar with Bahasa Indonesia/Bahasa Malaysia, but not with the wealth of the spoken varieties.

The following examples show the drastic nature of the changes

- (1) a. *bapak saya kasi dia emas* COLLOQUIAL INDONESIAN  
father 1SG give 3SG gold  
'My father gave him gold.' (p.c. David Gil)
- b. *Se=ppe baapa incayang=nang ummas su-kaasi* SLM  
1SG=POSS father 3S.POLITE=DAT gold PAST-give  
'My father gave him gold.' (K070000wrt04)

We note a change in word order, long vowels and consonants, and the development of bound morphology. A list of 60 'Sri Lankan' features is given in Table 10.2.

<b>phonology</b>	<b>stage</b>	<b>syntax</b>	<b>stage</b>
loss of initial <i>h</i> -	1	serial verbs	2
loss of initial <i>s</i> - in <i>satu</i>	1	vector verbs	2
phonemic schwa	1	QUANT N	2
existence of schwa in final syllables	1	DEM N	2
dropping of schwa in initial syllable	1	GEN N	2
lowering of high vowels in final syllables	1	relator nouns	2
velarization of final nasals	1	SOV	3
phonemic prenasalization mb,nd,ng	2	RELC N	3
consonant gemination	2	ADJ N	3
vowel length	2	STD DAT N MORE ADJ	3
dental/retroflex distinction	3	postpositions	3
[v]→[u]	3		
žj	5		
paragogic η	5		
/#t/→/#c/	5		
<b>morphology</b>		<b>semantics</b>	
no <i>məN-</i>	0	EVID	3
no <i>di-</i>	0	QUOT	3
plural pronouns in orang	1	tense rather than aspect	3
chinese pronouns	1	dative subjects	3
TAM markers based on the existential <i>ada</i>	1	non-nominative subjects	3,4
enclitic TAM adverbs	1	arà-	3,4
possessive marker =pe	1		
plural marker pada	1		
negator thraa	1		
morphologicization	3		
rigid word classes	3		
coordinating clitics	3		
indefinite pronouns	3		
discourse markers based on DEIC+X	3		
negation pattern	3		
CASE	3		
INF	3		
CP	3		
PRES PTPL	3		
Impolite imperative	3		
copula	3		
existentials duuduk/aada	2,3,4		
INDEF	3,4		
jamà-	5		
<b>lexicon</b>			
Jakartan words like <i>oomong</i>	1		
Javanese words like <i>kulluth</i>	1		
Tamil loans like <i>kattil</i>	3		

Table 10.2: ‘Sri Lankan’ features in a variety of linguistic domains and the most likely point of their emergence.

### 10.2.1 Phonology

The phonological features which can be taken to be Sri Lankan are given in Table 10.3.

loss of initial <i>h</i> -	shared with other Trade Malay varieties	<a href="#">Adelaar (1991)</a> , <a href="#">Adelaar &amp; Prentice (1996)</a> , <a href="#">Paauw (2004, 2008)</a>	1
loss of initial <i>s</i> - in <i>satu</i>	shared with other Trade Malay varieties	<a href="#">Paauw (2004)</a>	1
phonemic schwa	archaism		1
existence of schwa in final syllables	archaism from Jakarta	<a href="#">Adelaar (1985)</a> , <a href="#">Nordhoff (2009)</a>	1
dropping of schwa in initial syllable	Jakarta	<a href="#">Uri Tadmor, p.c.</a>	1
lowering of high vowels in final syllables	shared with other Trade Malay varieties	<a href="#">Paauw (2004, 2008)</a> , <a href="#">Nordhoff (2009)</a>	1
velarization of final nasals	shared with other Trade Malay varieties		1
phonemic prenasalization ṁb, ḡd, ḡg	substrate reinforcement (Sinhala)	<a href="#">Nordhoff (2009:118)</a>	2
consonant gemination	substrate reinforcement	<a href="#">Nordhoff (2009:119)</a>	2
vowel length	substrate reinforcement	<a href="#">Nordhoff (2009:117)</a>	2
dental/retroflex distinction [v]→[v̪]	<b>candidate</b> <b>candidate</b>	<a href="#">Nordhoff (2009:117-118)</a> <a href="#">Smith et al. (2004)</a>	3
phonemic prenasalization ŋj	independent development	<a href="#">Nordhoff (2009:120)</a>	5
paragogic η	independent development	<a href="#">Nordhoff (2009:120)</a>	5
/#t/→/#c/	independent development	<a href="#">Nordhoff (2009:120)</a>	5

Table 10.3: ‘Sri Lankan’ features in phonology.

Next to a couple of features which are common in other varieties of Malay as well, and which are therefore regarded as preceding the language’s arrival in Sri Lanka, we find some salient phonological differences, namely length distinctions in consonants and vowels and the tautosyllabic analysis of certain NC clusters, giving rise to phonemic prenasalized stops. These phonological features are all also found in the Lankan languages<sup>1</sup> and are among the one most often cited for the peculiar status of Sri Lanka Malay. At a closer look, however, it emerges that these structures are not at all novel in Sri Lanka Malay. Many Malay dialects have a regular subphonemic lengthening of the penultimate syllable, which is the only place where vowel quantity distinctions are found in Sri Lanka Malay. We are thus dealing with a phonemicization of a subphonemic contrast through language contact, not with the development of a completely novel feature. Similar things can be said about gemination. Gemination is found in some Malay dialects following a schwa. This is also the case in Sri Lanka Malay, where the overwhelming majority of geminate consonants follow a schwa.<sup>2</sup> The process of gemination of consonants after schwa already present in some forms of Trade Malay was regularized and expanded in Sri Lanka, certainly under influence from Sinhala and to a lesser degree from Tamil, but it is not a phenomenon completely alien to a Malay language. A similar argument can be made for the prenasalized stops. Some dialects of Malay syllabify NC sequences as .NC, others as N.C. Speakers from both types were

<sup>1</sup>Tamil has no prenasalization.

<sup>2</sup>There are a couple of exceptions like *appi* ‘fire’ or *ikkang* ‘fish’.

brought to Sri Lanka, where lexemes from different dialects made it into the languages. This phonemicization was without doubt helped by Sinhala having a phonemic distinction between tautosyllabic and heterosyllabic NC sequences, but Malay was not a blank slate when Sinhala started to exert its influence in phonology.

Things are different for the dental/retroflex distinction. SLM expanded the original /t,d/ set to a fourfold distinction of /ʈ,ʈ,ɖ,ɖ/. There are no known dialects with ʈ or ɖ, so that this development is one of the few instances in SLM phonology where we are dealing with a clear Stage 3 process. Another clear instance of a Stage 3 process is the change in articulation of the labiodental from fricative [v] to approximant [u].

There are some other phonological features of Sri Lanka Malay, which constitute independent developments and are not due to language contact. These are the development of phonemic /ʃ/, the addition of a paragogic velar nasal to some historically coda-less lexemes (e.g. *buunung* ‘kill’<*bunu(h)*) and the change from initial /t/ to /c/ in some lexemes.

For the main concern of this paper, the features of stage 3, we can only retain two items: the development of a dental/retroflex distinction and the change in articulation of the labiodental consonant.

### 10.2.2 Morphology

The morphological features which can be taken to be Sri Lankan are given in Table 10.4.

In the domain of morphology, we note that the absence of voice morphology (*məN-*, *di-*) is not at all particular to Sri Lanka Malay, but actually the norm for basilectal varieties of Malay and found in all dialects of Trade Malay. The particular pronouns employed, which feature Chinese roots for the impolite 1<sup>st</sup> and 2<sup>nd</sup> person as well as a plural formative *rang* < *orang* ‘man’ are also comparable to what is found elsewhere in the Malay world. There are two cases of morphological elaboration: the grammaticalization of adverbs and the existential *ada* to yield TAM clitics. These also have their origin in Indonesia, although the grammaticalization has spread further in SLM than in other varieties. The possessive marker *punya~pe*, the plural marker *pada* and the negator *thraa* are salient characteristics of Sri Lanka Malay, but these three features are shared with other offshoots of Trade Malay and do not constitute a case of language contact in Sri Lanka.

Where we find language contact is in the morphologization of the language. No other Malay variety comes close to Sri Lanka Malay in the extent of bound morphology. Where many grammatical categories are optional in other varieties of Malay, and realized by particles or adverbs, in Sri Lanka Malay, we find a lot of obligatory categories, like tense or case, realized as bound forms. There is furthermore the formation of nonfinite categories such as an infinitive and two participles. This greater availability of bound material led to the creation of rigid word classes ([Nordhoff forthcominga](#)), which is very untypical of a Malay variety. Beyond the domain of the word, we also find a number of devices which are clearly the result of language contact, e.g. the Coordinating Clitics, which are also used in the formation of indefinite pronouns and some discourse markers. These clitics have clear counterparts in Sinhala and Tamil, which also use them for the formation of pronouns and discourse markers. They are not found elsewhere in the Malay world, so that we are dealing with a clear case of language contact here. A final case of language contact in morphology is the negation pattern, where the past/nonpast-split in negation morphemes is a Tamil influence in Sri Lanka Malay.

no <i>məN-</i>	shared with other Trade Malay varieties	Adelaar (1991), Adelaar & Prentice (1996), Paauw (2004, 2008)	0
no <i>di-</i>	shared with other Trade Malay varieties	Adelaar (1991), Adelaar & Prentice (1996), Paauw (2004, 2008)	0
plural pronouns in <i>orang</i>	shared with other Trade Malay varieties	Adelaar (1991), Paauw (2004)	1
chinese pronouns	shared with other Trade Malay varieties	Adelaar (1991), Paauw (2004)	1
TAM markers based on the existential <i>ada</i>	shared with other Trade Malay varieties, mainly Moluccan	Adelaar (1991)	1
enclitic TAM adverbs	shared with other Trade Malay varieties	Adelaar (1991)	1
possessive marker = <i>pe</i>	shared with other Trade Malay varieties	Adelaar (1991), Paauw (2004, 2008)	1
plural marker <i>pada</i>	innovation from Jakarta	Adelaar (1991), Paauw (2004)	1
negator <i>thraa</i>	innovation from the Moluccas	Adelaar (1991), Paauw (2004)	1
morphologization	<b>candidate</b>		3
rigid word classes	<b>candidate</b>	Nordhoff (forthcominga)	3
coordinating clitics	<b>candidate</b>	Nordhoff (2009:314-330)	3
indefinite pronouns	<b>candidate</b>	Nordhoff (2009:391-392)	3
discourse markers based on DEIC+X	<b>candidate</b>	Nordhoff (2009:359-361)	3
negation pattern	<b>candidate</b>	Nordhoff (2009:588-590)	3
INF	<b>candidate</b>	Nordhoff (2009:244-245)	3
CP	<b>candidate</b>	Nordhoff (2009:241-243)	3
PRES PTPL	<b>candidate</b>	Nordhoff (2009:358-359)	3
CASE	<b>candidate</b>	Nordhoff (2009:285-313)	3
Impolite imperative	<b>candidate</b>	Nordhoff (2009:265-266)	3
existentials <i>duuduk/aada</i>	substrate reinforcement, convergence towards Sinhala	Nordhoff (2009:142-146)	2,3,4
INDEF	<b>candidate</b> , Sinhala influence	Nordhoff (2009:278-282)	3,4
copula <i>jamà-</i>	independent development	Nordhoff (2011)	5
	independent development	Nordhoff (2009:245-247)	5

Table 10.4: ‘Sri Lankan’ features in morphology.

Sri Lanka Malay has two existentials with a split along animacy. This is due to influence from Sinhala, but a root in Malay dialects cannot be ruled out. Here, and in other cases of Sinhala influence, it is difficult to establish whether we are dealing with a recent influence from Sinhala, or whether this change predates the political changes of the 20<sup>th</sup> century. The same is true of the grammaticalization of the indefiniteness marker *hatthu*, which can be recent or an older phenomenon.

In the domain of morphology, finally, we can mention two internal developments which are not due to language contact. The first is the subordinate negator *jamà-* (*jaŋ-* in Slomanson 2008), the second one is the emergence of a copula (Nordhoff 2011).

Neither of these developments has a good model in either Malay, Sinhala, or Tamil.<sup>3</sup>

### 10.2.3 Syntax

The syntactic features which can be taken to be Sri Lankan are given in Table 10.5.

serial verbs	substrate reinforcement	Paauw (2004), Nordhoff (forthcomingb)	2
vector verbs	substrate reinforcement	Paauw (2004), Nordhoff (forthcomingb)	2
QUANT N	substrate reinforcement	Adelaar (1991), Paauw (2004, 2008)	2
DEM N	substrate reinforcement	Adelaar (1991), Paauw (2004, 2008)	2
GEN N	substrate reinforcement		2
relator nouns	substrate reinforcement	Paauw (2004)	2
SOV	<b>candidate</b>	Adelaar (1991)	3
RELC N	<b>candidate</b>	Adelaar (1991)	3
ADJ N	<b>candidate</b>	Adelaar (1991)	3
STD DAT N MORE ADJ	<b>candidate</b>	Nordhoff (2009:582-583)	3
postpositions	<b>candidate</b>	Adelaar (1991)	3

Table 10.5: ‘Sri Lankan’ features in syntax.

In the domain of syntax, we find two different types of serial verb constructions, which are modelled mainly on Tamil structures, but have some resemblances to certain Trade Malay varieties (Nordhoff forthcomingb, Jaffar this volume). We are thus dealing with substrate reinforcement. The bulk of the other features have to do with word order, where Sri Lanka Malay has a clear preference for head-final structures. These features are split as to whether they are also found in some Trade Malay varieties, or whether they are pure influence from Sinhala and/or Tamil. The last feature, relator nouns to specify spatial and temporal relations like *duppang* ‘in front of, before’ and *blaakang* ‘after’ are again a clear case of influence from both Sinhala and Tamil, a Stage 3 process.

---

<sup>3</sup>Tamil *-aama* and to a lesser extent Sinhala *noo* provide some imperfect approximations of *jamā-*.

### 10.2.4 Semantics

The semantic features which can be taken to be Sri Lankan are given in Table 10.6.

QUOT	<b>candidate</b>	<a href="#">Nordhoff (2009:340-346)</a>	3
EVID	<b>candidate</b>	<a href="#">Nordhoff (2009:337-340)</a>	3
tense rather than aspect	<b>candidate</b>		3
dative subjects	<b>candidate</b>	<a href="#">Nordhoff (2009:421-430)</a>	3
accusative and instrumental subjects	<b>candidate</b> , Sinhala Influ- ence	<a href="#">Nordhoff (2009:421-430)</a>	3,4
<i>arà-</i>	substrate reinforcement, Sinhala Influence	<a href="#">Nordhoff (2009:253)</a>	3,4

Table 10.6: ‘Sri Lankan’ features in semantics.

It is in the domain of semantics that we find some clear instances of Stage 3 processes. Sri Lanka Malay has a number of grammatical categories completely absent from other Malay varieties, but found in both Sinhala and Tamil. These include evidentiality and the quotative. In a different domain, Sri Lanka Malay gives preference to the encoding of tense, rather than aspect, what is found in other Malay varieties. This mirrors what we find in Sinhala and Tamil. The possibility to use dative subjects is due to influence from Tamil and Sinhala, while the use of instrumental and accusative subjects is due to influence from Sinhala alone.

Next to the clear Stage 3 processes, there are again a number of features which could be due to early influence from Sinhala (Stage 3), or to late influence (Stage 4). These are the use of instrumental and accusative subjects as well as the semantics of the nonpast marker *arà-* , which can be used in subordinate contexts with past reference next to its more common usage in present or future contexts in both Sinhala and SLM.

### 10.2.5 Discourse

The discourse feature which can be taken to be Sri Lankan are given in Table 10.7.

tail-head linkage	<b>candidate</b>	<a href="#">Nordhoff (2009:474-475)</a>	3
-------------------	------------------	---	---

Table 10.7: ‘Sri Lankan’ feature in discourse.

SLM discourse is organized through tail-head linkage, like in Sinhala and Tamil. This is a feature not found in the relevant varieties of Malay, so that we are dealing with a Stage 3 process.

### 10.2.6 Lexicon

The lexical features which can be taken to be Sri Lankan are given in Table 10.8.

<i>oomong</i> etc	vocabulary from Jakarta	1
<i>kulluth</i> etc	vocabulary from Javanese	1
<i>kattil</i> etc	vocabulary from Tamil	3

Table 10.8: ‘Sri Lankan’ lexical features.

Sri Lanka Malay shows a number of vocabulary items from Jakarta, which are obviously a retention. Some other vocabulary items are of Javanese origin, but without a clear link with Jakarta. These are also retentions. SLM has borrowed from Tamil, mainly animal names, but also some very basic items like *kattil* ‘bed’, *marakari* ‘vegetables’ or *dhaatha* ‘elder sister’. These borrowings belong in Stage 3. There are no noteworthy borrowings from Sinhala.

In the following, I will discuss what happened during the various stages.

## 10.3 When did it happen

Having charted the linguistic domains of ‘Sri Lankan’ features, I now turn to sorting these features into the different stages.

### 10.3.1 Summary of Stage 0 features

no *məN-*    no *di-*

Table 10.9: Stage 0 features.

Stage 0 refers to features of Sri Lanka Malay which distinguish it from the modern standard languages, but are in fact shared with all other offshoots of Trade Malay. These features are limited to the absence of voice morphology. They are of no particular interest to the study of language contact.

### 10.3.2 Summary of Stage 1 features

loss of initial <i>h</i> -	plural pronouns in <i>orang</i>	<i>oomong</i> etc
loss of initial <i>s</i> - in <i>satu</i>	chinese pronouns	<i>kulluth</i> etc
lowering of high vowels in final syllables	TAM markers based on the existential <i>ada</i>	
velarization of final nasals	enclitic TAM adverbs	
phonemic schwa	possessive marker <i>=pe</i>	
existence of schwa in final syllables	plural marker <i>pada</i>	
	negator <i>thraa</i>	

Table 10.10: Stage 1 features.

Stage 1 features stem from different dialects of Trade Malay, which met in Batavia and then Sri Lanka and whose mixture formed the very first instance of a Sri Lankan variety of Malay in the 17<sup>th</sup> century. Some features can be traced to a specific area, while others are found in several areas, but not all ([Paauw this volume](#)). The features we find today are those which made it through the dialect levelling process; an equally important number of features did not make it into Sri Lanka Malay and was lost on the way.

The extent of dialect levelling is quite strong in phonology and morphology, but mainly in pronouns and particles, not so much in bound morphology. The lexicon also shows signs of dialect levelling, but syntax and semantics are completely absent. In

these areas, Sri Lanka Malay either retains a general Trade Malay pattern (Stage 0) or adapts to the Lankan model (Stages 2, 3, and 4).

### 10.3.3 Summary of Stage 2 features

retroflexes	serial verbs	QUANT N
consonant gemination	vector verbs	DEM N
vowel length	relator nouns	GEN N
phonemic prenasalization		
ṁb, ūd, ūg		

Table 10.11: Stage 2 features.

The features of Stage 2 were marginally present in Trade Malay varieties, but received a boost upon arrival in Ceylon because the Lankan languages made frequent use of them. Substrate reinforcement is found in the quantity contrasts of vowels and consonants, and in prenasalization. Furthermore, verb serialization and head-final structures which were present in some Trade Malay varieties gained in importance upon contact with Tamil and Sinhala ([Nordhoff forthcomingb](#)).

### 10.3.4 Summary of Stage 3 features

SOV	EVID	coordinating clitics	morphologicization $v \rightarrow v'$	
RELC N	QUOT	indefinite pro- nouns	rigid word classes	tail-head linkage
ADJ N	CASE	DEIC + X	impolite	impera- tive
STD DAT N	INF	copula <sup>4</sup>		<i>kattil</i> etc
MORE ADJ				
postpositions	CP			
	PRES PTPL			
	DAT SUBJ			
	ACC/INSTR			
	SUBJ			

Table 10.12: Stage 3 features.

Stage 3 contains the features which have the most relevance for theories of language contact and change. These features are neither plain nor modified retentions from earlier varieties, and they cannot be attributed to independent developments either. Some of the features in this stage could also be attributed to Stage 4, recent influence from Sinhala.

Stage 3 contains two large groups of features, one relating to word order, the other to grammatical categories. A smaller number of features pertain to the emergence of

<sup>4</sup>The development of the copula *asādhaatahg* was analyzed as an independent development in [Nordhoff \(2011\)](#). Newer research, however, shows that Tamil (*vantu*) and Sinhala (*āvillaa*) models do exist for this form.

clitics and word formation processes involving them. The last noteworthy category contains processes referring to the development of morphology. The last column is the wastebasket category. The composition and relevance of these categories will be discussed in detail below, but it can already be noted here that these features do not fall within the area of features typically associated with creolization.

### 10.3.5 Summary of Stage 4 features

existentials <i>duuduk/aada</i>	INDEF	ACC/INSTR SUBJ	semantics of the nonpast form <i>arà</i>
------------------------------------	-------	-------------------	---

Table 10.13: Stage 4 features.

Stage 4 includes features which are due to recent influence from Sinhala. For all of those features, an earlier date than the 20<sup>th</sup> century is also possible, in which case we would be dealing with stage 3, or even Stage 2 in the case of the existentials (Nordhoff 2012b).

The features with Sinhala influence all draw upon existing material and only change the use a little bit. They do not involve huge changes. The difference in existentials involves the reinterpretation of one verb. The grammaticalization of the indefinite article renders a formerly optional feature obligatory. The non-nominative subjects are a straight copy of Sinhala verb categorization, facilitated by the fact that ablative and accusative are available as cases anyway. The extension of nonpast to subordinate past context finally is also easy to accomplish since the nonpast form is already available and the new meaning can easily be added to its range of functions.

Given the superficial nature of the changes through Sinhala influence, these could easily be very recent, although an earlier origin cannot be ruled out.

### 10.3.6 Summary of Stage 5 features

ᬁጀ	paragogic ŋ	#t→#c	<i>jamà</i>
----	-------------	-------	-------------

Table 10.14: Stage 5 features.

As for independent developments (Stage 5), we find three phonological features, the development of a new clause pattern, the copula construction, and a new verbal prefix. All in all, this is a mixed bunch which could have occurred like that in any other language without external influence.

## 10.4 Why did it happen

In this section, I will survey in more details the processes which led to the emergence of the typical features of Sri Lanka Malay.

### 10.4.1 Stage 0: general Trade Malay features

The process which led to Stage 0 features in SLM is trivial: we are dealing with retention.

### 10.4.2 Stage 1: dialect levelling

In Stage 1, which started in Batavia and continued in Sri Lanka until the Sri Lankan variety of Malay emerged as a more or less unified language, the different dialects of Trade Malay were in contact. In the beginning, variation was high and coherence low, but as mutual accommodation progressed, divergence diminished and some features crystallized. This process is not unique to Sri Lanka; it is also found in other communities of displaced people. [Trudgill \(1986, 2004\)](#) analyses the general processes which take place when an internally diverse community is displaced to a foreign shore. Trudgill restricts his theory to the formation of new dialects by dialect mixture in a community isolated from other speakers of the same language. His example of choice is New Zealand, but the definition corresponds to the Sri Lankan situation as well. He distinguishes three phases:

I	adult migrants	first generation	rudimentary levelling
II	first native-born speakers	second generation	extreme variability and further levelling
III	subsequent generations		focussing, levelling, and reallocation

Table 10.15: The three phases of dialect levelling according to [Trudgill \(1986\)](#).

In the first phase, adults of linguistically diverse backgrounds find themselves in foreign territory. In order to arrive at successful communication, they avoid the most divergent features of their dialects, which pose the greatest problems to communication. In phase II, children are born to the first generation, who grow up in a very variable linguistic environment where widely diverging forms from a wide variety of dialects are found. The children are exposed to varieties from speakers outside of their immediate family and their own speech eventually integrates some of the non-family members' speech. Through the generations (phase III), redundant constructions are eliminated. The constructions eliminated and retained normally do not all come from the same dialect, and are not predictable on a deterministic basis, although factors like articulatory ease, salience, and integration into other parts of the relevant linguistic subsystem do play a role. Based on these factors, it is possible to make probabilistic predictions about which features are more likely to be retained on a statistic level ([Trudgill 2004](#)). Constructions which are replaced by constructions from other dialects either fall into disuse, or they are reallocated to other functions ([Trudgill 1986](#)). Most of the constructions only have communicative functions, but some become markers of identity ([LePage & Tabouret-Keller 1985](#)) as the displaced community acquires a sense of identity distinct from the mother country population. [Trudgill \(2004\)](#) provides evidence from recordings of speakers born in the late 19<sup>th</sup> century in New Zealand, which show the levelling proceeding through the generations. Trudgill's findings are quantitatively evaluated with a mathematical model in [Baxter et al. \(2009\)](#), who reject Trudgill's reliance on frequency factors alone, but confirm the overall validity of the scenario.

[Mesthrie \(1993\)](#) applies Trudgill's scenario to South African Bhojpuri, where he posits the same sequence of events. The British recruited indentured labourers from India. The Indians all spoke different dialects from the Bhojpuri-Hindi-continuum. In South Africa, the dialects first lost the most extravagant features and were then levelled, similar to what happened in New Zealand. The South African case provides another parallel to what happened in Sri Lanka: initial dialect diversity is reduced and

a common norm emerges.

Sri Lanka Malay shows features from Jakarta (plural marker *pada*, dative/allative *nang*, retention of schwa in final syllables, dropping of schwa in penultimates, lexical features like *oomong* ‘speak’ rather than *becara*) and from the Moluccas (negator *thraa*, velarization of final nasals, lowering of high vowels in final syllables). Gil (this volume) provides further evidence that Sri Lanka Malay is in some respects like Jakartan varieties, in others, like Moluccan varieties, and yet in others, like both.<sup>5</sup> The combination of features from Jakarta and the Moluccas is quite arbitrary and involves features from phonology, morphology and syntax from both areas without any obvious principle. The selection appears to be quite random, reflecting the general ethnical heterogeneity of the immigrant population, where no one subgroup could impose their dialect. An area where we can find phonological reflexes of different dialects is the treatment of schwa. Some varieties of Trade Malay have lost schwa and merged it with /a/ or /i/ in the cases which interest us here. Sri Lanka Malay retains schwa, but there are at least two cases where a lexeme has been taken from a non-schwa variety. This can be seen from the fact that these lexemes have a long vowel, which is never possible in SLM if the underlying vowel is schwa. This shows that the underlying vowel is not schwa in these cases, and that the lexemes must come from a dialect where the distinction between schwa on the one hand and /a/ or /i/ on the other hand was lost. The first case of this is the word *baalai* ‘nonsense’. The cognate word is *bəlai* in schwavarieties, but *balai* in varieties which have lost schwa. Given that the SLM lengthens the vowel, the lexeme must have entered SLM as /balai/, with a full /a/. This contrasts with lexemes which entered SLM with a schwa. These have geminate consonants instead of long vowels, next to raising of schwa. An example is the word for ‘to split’ which is *belah* in Indonesia, but *billa* in Sri Lanka.

The mixing of dialects is followed by regularization. Sri Lanka Malay has generally dropped schwa in antepenultimates, so that the historical form *cəlana* ‘trousers’ is now pronounced *claana*. There is some evidence that this process, which is common in present day Jakarta (Uri Tadmor, p.c.), was not fully completed in all of the varieties which entered Sri Lanka. The word for a certain type of ear jewelry is *kerabu* in Indonesia. It is found in Sinhala as *kerabuva* (Gunasekara 1891), with the schwa intact (and an addition of the suffix -va, which need not concern us here). However in modern SLM, the same lexeme is *kraabu*, without the schwa. Obviously, Sinhala borrowed that lexeme at a point in time where the pronunciation with schwa was still current. The subsequent change in SLM did not affect the Sinhala lexeme, which preserves the former state. This shows that, next to the Jakartan dialects with schwa dropping, there were some dialects without schwa dropping in Sri Lanka, but their treatment of schwa was ironed out in the regularization process taking place within the local-born generations.

### 10.4.3 Stage 2: substrate reinforcement

Jeff Siegel observed in a number of papers that in language contact, some marginal structures of one language become more prominent upon contact with another language. In Melanesia, the pidgin used on plantations in the late 19<sup>th</sup> century by workers of different origin was brought back to the different homelands of the workers (New Guinea, Solomons Islands, Vanuatu). Siegel (1998:349-350) shows that there was initially a good deal of variation in Melanesian Pidgin. This variation was reduced when

---

<sup>5</sup>Malaccan/Malaysian features are not found at all in Sri Lanka Malay.

the workers went back to their home regions and used Melanesian Pidgin as a lingua franca for interethnic communication. In the three different major regions, the erstwhile plantation pidgin changed into different directions. In New Guinea, the features matching the local Western Oceanic languages were retained where variability existed, and gave rise to Tok Pisin. In the Solomon islands, variability was disposed of by retaining features which matched the local languages of the Southeast Solomonic subgroup, giving rise to Solomons Pijin. And in Vanuatu, the features which matched with features of the North-Central Vanuatu subgroup of Oceanic were retained, giving rise to Bislama. The initial variability found in a diverse population was thus reduced ('levelled'), but the levelling we find here is not purely internal, as was the case in New Zealand or South Africa. Rather, it is influenced by the other languages of the linguistic ecology. If there is a choice among alternative expressions for the same content, speakers tend to prefer the alternative which is also found in other languages they know. This is the case for Melanesian Pidgin, and it is also the case for the elimination of the high initial variability of Sri Lanka Malay. The language of the immigrants had variable order for the placement of the possessor. Some dialects preferred prenominal possessors, others postnominal possessors. On Lankan soil, both Tamil and Sinhala only admit prenominal possessors. The initial variability found in the Malay speech community was then reduced and the feature retained was the one which matched the other languages of the area, the 'substrate' so to speak. The reduction of variability to align with a contact language where this variability does not exist is not restricted to Melanesia or Sri Lanka: Siegel (1987, 2000) shows this for Fiji and Hawai'i, Singler (1988) for Liberian English.

For Sri Lanka Malay, this pattern is widely attested. Given the large dialectal variation of the immigrants' languages, there was often at least one variety where a feature was found that 'fit' within the Lankan model. The dialect levelling process is non-deterministic, as discussed above, but a similar structure in another language the Malays had command of (Sinhala or Tamil), facilitated the retention of a feature from a Malay variety.

Syntactic features like serial verbs and relator nouns fall into this category. They are not a very prominent aspect of Malay languages in general, and are not found across the board. Since these features play an important role in Tamil (and in Sinhala, as far as relator nouns are concerned), the Malay varieties which did have them had a greater impact in the dialect levelling process as far as these features are concerned, and the fledgling use of these constructions in some Malay varieties became well-established through contact with Tamil and Sinhala, which made other speakers realize the existence and potential usefulness of these constructions.

Another aspect of substrate reinforcement is the elimination of certain variants from the dialect pool. Trade Malay varieties allow nominal modification to the left and to the right for a number of modifiers. Sinhala and Tamil generally only allow prenominal position.<sup>6</sup> Sri Lanka Malay is not as rigid as Sinhala or Tamil, but is clearly more a right-headed language, even in the nominal domain, than any other variety of Malay. This order of constituents, however, did not come out of the blue, but could build on tendencies already present in Trade Malay varieties, which were reinforced through the speakers' frequent dealing with languages where this order was obligatory.

A clear instance of substrate reinforcement is the velarization of final nasals. This is a feature of Moluccan varieties of Trade Malay, but is also found in Sinhala. The presence of a language in the ecology with a similar phonological rule helped the Moluccan

---

<sup>6</sup>Sinhala (and to a lesser extent Tamil) numerals and quantifiers are an exception to this.

feature to be retained in this domain.

Substrate reinforcement can further be observed in some lexemes, where phonological features from different dialects were retained. The clearest of these cases is prenasalization. NC clusters are syllabified as N.C in some varieties of Malay, while others have .NC (Tapovanaye 1995, Apoussidou & Nordhoff 2008, Nordhoff 2009). In Sri Lanka we find both, e.g. *am.bel* ‘take’ and *gaa.mbar* ‘picture’ or *an.jing* ‘dog’ and *baa.njir* ‘flood’.<sup>7</sup>

Gemination of consonants after schwa as found as a subphonemic contrast in some Malay varieties was phonemicized in Sri Lanka under influence from Sinhala and Tamil. Indonesian words of Indian origin like *topi* ‘hat’ (from Hindi) or *kapal* ‘ship’ (from Tamil), which had lost the original geminate consonant in Indonesia, reinstated the geminate in Sri Lanka under influence from Tamil (*kappal*, *toppi*) or Sinhala (*toppiya*). This leads to near-minimal pairs like *kappal* ‘ship’, *kaapang* ‘when’ or *thoppi* ‘hat’, *soopi* ‘liquor’. Once this phonological contrast was established, the formerly subphonemic contrast between [ku:mis] ‘moustache’ and [kummis] ‘Thursday’ or [t̪i:kam] ‘stab’ and [t̪ikkam] ‘press’ was also analysed as phonemic. At the same time, the above pairs can be used to explain the difference in vowel length in Sri Lanka Malay<sup>8</sup>

The issue of the dental/retroflex distinction is more difficult. Malay varieties generally have a dental /t̪/ and an alveolar /d/. What has to be explained is the emergence of the postalveolar/retroflex /t̪/ and the dental /d/. The voiceless ‘retroflex’ stop /t̪/ is found in loanwords from Tamil and in native words in front of a back vowel e.g. *baa[t̪]lok* ‘coconut shell’ or *paa[t̪]ok* ‘hiss’ or *oo[t̪]ak* ‘brain’. From an articulatory point of view, it is plausible to assume that the backness of the vowel caused the preceding consonant to be articulated further back, yielding a subphonemic contrast. This contrast phonemicized upon the entry of loanwords from Tamil, where /t̪/ and /t̪/ had to be kept distinct, e.g. *ka[t̪]il* ‘bed’ and *a[t̪]e* ‘leech’. It is unclear, however, whether this can be seen as a case of substrate reinforcement, since the subphonemic contrast we hypothesized above had no function in Trade Malay, and its functional usefulness could therefore not be reinforced.

The development of /d/ is more difficult to explain. This phoneme is only found in initial position in a dozen words or so. It is much more restricted than /t̪/. Smith (2003) tries to explain this phoneme through the adoption of a Tamil constraint against retroflex onsets, but fails to notice that Tamil has an equally important constraint against voiced onsets, so that /#d/ is an unlikely outcome of contact with Tamil. Sinhala has words starting with /d/ as well as words starting with /d/, but there seems to be no reasons to assume that the split of Malay /#d/ into /#d/ and /#d/ had anything to do with Sinhala. What we can say is that this distinction was probably already made in the early 19<sup>th</sup> century since there are manuscripts where two different d’s are graphemically distinguished in the Arabic script employed (Hussainmiya 1987). It is tempting to see the split of the voiced stop as a result of language contact, but closer analysis reveals that the explanation is far from obvious (Smith 2012, Nordhoff 2012a).

<sup>7</sup>The long vowel in the prenasalized words is a consequence of the penult losing its coda as a consequence of the different syllabification. See Nordhoff (2009:120ff) for more detailed discussion.

<sup>8</sup>In many cases, a long/short consonant can be explained by a short/long vowel, or the other way round. There are, however, some cases where the long consonant must be specified lexically (*appi* ‘fire’ vs. *baapi* ‘bring’), as well as there are cases where the length of the vowel must be stored in the lexicon (*thurus* ‘straight’ vs. *thuurung* ‘descend’).

#### 10.4.4 Stage 3: where the problems lie

After discarding everything which is either an early development (see above) or possibly a very recent change (see below), we are left with a set of 22 features which are the result of the crucial Stage 3 and can help elucidate the process at work. Was it creolization, convergence, metatypy, or ‘ganging up’?

At this point, it is worthwhile to review the two main theories which have been advanced for the genesis of Sri Lanka Malay. Ian Smith and colleagues have defended the view that Sri Lanka Malay is an instance of creolization. See Smith’s contribution to this volume for an overview of the argument. While in the better known cases of creolization, the displaced population acquires the socially dominant language, in this case, the local population is argued to have acquired the language of the displaced population. More specifically, Tamil speaking women belonging to the Muslim group of the Moors married Malay soldiers and tried to acquire their husband’s language. In doing so, they transferred structures of their native Tamil language to their variety of Malay. This Tamilized Malay was passed on to the children of the union, who nativized this new variety.

This contrasts with the theory of ‘rapid convergence’ advanced by Bakker (1995, 2000a,b, 2006). In his view, Sri Lanka Malay changed within a very short time span, one or two generations. Bakker (2006) also offers ‘metatypy’ as an explanation: similar to what is found in some other heavy contact settings, the grammatical structures of the language of wider communication is imposed on the in-group language, but the in-group language retains its vocabulary as a marker of identity.

This metatypic explanation is also supported by Ansaldi (2008, 2009). Ansaldi additionally adds that Malay converged towards Sinhala and Tamil at the same time since those languages are grammatically very similar. The double weight of this alliance exerts a greater influence on Malay than any of the two would have had individually.

I have argued above that several processes must be kept apart in the genesis of Sri Lanka Malay. The stage which can give us answers about the processes of language change of interest here is stage 3. The earlier stages have no or only disputable contact influence, the later stages do not fall into the formative period. Taking a look at stage 3, we find, a number of word order features, and a number of grammaticalizing semantic domains. Furthermore, we are dealing with some features which have to do with the disconnection of phonological word and morphological word, i.e. the morphologization of the language. Another set is related to Lankan style clitics.

#### Creolization

The term ‘creolization’ has received a lot of interpretations (see Bakker, this volume). The Bickertonian view of Creolization as the breakdown of a communicative system and its successive reinvention clearly does not apply to Sri Lanka Malay, since there are many Malay features which are retained, showing that there was no break in transmission. This shows at the same time that Sri Lanka Malay is not a ‘young’ language (cf Ansaldi & Nordhoff 2009), however defined, but of course the bulk of its morphosyntactic structure is of quite recent origin.

If ‘creolization’ means the transmission of first language features into a target language (see Smith, this volume), it is difficult to see how Sri Lanka Malay fits the bill since the features transferred are not those which would have been expected under this assumption.

None of the features mentioned in Section 10.4.4 looks like anything which has

been proposed for creole formation. Head-final word order is only rarely ever found in creoles, and has occasionally been argued to be an unlikely outcome of creolization, although this is due to the fact that the settings investigated included relatively few head-final input languages to begin with. Languages traditionally analysed as creoles typically involve a reduction of morphology, so that the increase in morphology in SLM and the semantic domains which need not be expressed morphologically do not really fit the bill and would rather suggest an origin different from what theories of creolization suggest. The development of word classes is also not something which is typically associated with creole formation.

Second language learners are known to reduce the grammatical features not necessary for communication. The propositional content of a message does not depend on evidentiality, for instance. If imperfect learning was important for Sri Lanka Malay, we would expect some difficult features of Trade Malay (if there are any) to be imperfectly, or not at all, acquired by Tamilophones. Imperfect learning does generally not suggest that learners invent new categories in the target language, on the other hand. German learners of English, for instance, are not known to invent gender or case when learning English, even if those features are important in their native language. Russian learners of English do not create aspect, Slovenians do not create duals. Germans, Russians and Slovenians simply strip their languages of their morphosyntactic complications instead of transposing them to English.

Nevertheless, it is true that we often do find structures in Creoles which are due to transfer of L1 features. Crucially, this only happens when the target language is used as a medium of interethnic communication, either between populations who do not share another language (e.g. Tok Pisin) or in clear power relations (e.g. between slaves and masters). Neither is found in Sri Lanka. Sri Lanka Malay was never used in communications which involved only non-Malays, and Malays never enjoyed the power and prestige to enforce their dialect upon the people who addressed them in a way similar to the European powers.

The setting where non-Malays spoke Malay is when they married into a Malay family (see Slomanson, this volume). But in this setting, the target language is not used in the way (inter-ethnic communication) which is required in most theories of Creole formation for substrate effects to take place.

### **Metatypy**

As for (rapid) convergence and metatypy in Bakker's sense, I will treat them together here since the morphosyntactic argumentation will be the same. The speed of development relies on sociohistorical information and cannot be evaluated based on morphosyntactic considerations alone. [Smith & Paauw \(2006\)](#) and [Nordhoff \(2009\)](#) have shown that some sociohistorical key assumption for Bakker's theory of rapidity are mistaken, and this argument will not be repeated here.

As for the compatibility of the morphosyntactic features with the theories, other instances of metatypy involve changes in word order (Takia/Waskia in PNG ([Ross 1996, 1997, 2001, 2003, 2007](#)), Cappadocian Greek in Turkey ([Dawkins 1916](#))), so that the SLM facts jibe well with this theory. The development of new obligatory grammatical theories is also often observed in situations of metatypy or convergence. [Aikhenvald \(2002:117ff\)](#) for instance observed that the Arawakan language Tariana developed obligatory evidential marking through contact with Tucano (Tucanoan), a language which was established as a language interethnic communication by missionaries. Crucially, there is no indication that a significant number of Tucano speakers

tried to acquire Tariana in the relevant period. Rather, Tariana speakers were familiar with the language of wider communication Tucano and found the evidentiality status of a proposition an important aspect (Aikhenvald 2004:296), which they expected to be conveyed in their native language Tariana as well.<sup>9</sup>

Similar things can be said about volitionality. In Sinhala, and to a lesser extent Tamil, volition is an important parameter of the grammar. Sinhala sentences typically indicate whether the action was performed with volition or not. This constant exposure to the parameter of volition led to the Malays' expecting this information to be conveyed; failure to do so would result in an infelicitous message and communication problems.

Nordhoff (*forthcominga*) argues that cognitive entrenchment of semantic entity classes with the prototypical acts of reference (for objects, yielding nouns) and predication (for actions, yielding verbs) through contact with Sinhala and Tamil is responsible for the development of rigid word classes in SLM.

Ansaldo (2008, 2009) argues that the structure of SLM can best be explained by metatypy with frequency effects, i.e. a shift away from Malay structures is more likely if Sinhala and Tamil have the same structure (e.g. dative) and less likely if their structure is different (e.g. accusative). For the list of Stage 3 features compiled above, this generalization seems to hold. The only item where SLM did not converge towards Sinhala and Tamil at the same time is the present participle formation by reduplication, which is found in Sinhala, but only marginally in Tamil.

#### 10.4.5 Stage 4: convergence towards Sinhala/attrition

Probably already since 1873, but surely since 1956, Sri Lanka Malay has been losing ground in the Sri Lankan ecology. In 1873, the Malay Regiment, which provided a privileged and sheltered place for Malay interaction and cultural life, was disbanded. The former soldiers found work in police and fire stations across the country, but the concentration of Malays in their new work places was far less than it was before.<sup>10</sup> In 1956, the so-called ‘Sinhala Only’ law changed the linguistic landscape of Sri Lanka. The Malays, who had up to then used Malay at home and English, and to a lesser degree Tamil, outside the home, had to become proficient in that language. In order to keep the economic advantage provided by English, the home language shifted from Malay to English, and Malay started losing domains. This process continues up to today, and there are many Malay children whose first language is Sinhala and who have only limited command of Malay.

The increased bilingualism in Sinhala is likely to have increased Sinhala influence on Malay, and it is probable that present-day SLM is more Sinhala-like than SLM was a hundred years ago. On the other hand, the extent of recent Sinhala influence should not be exaggerated: people born in the 1930s, who had acquired Malay before the nationalist policies were instated, speak pretty much the same way as younger fluent speakers, suggesting that the Sinhala structures in SLM were already present before 1950.

The processes at play here are convergence, but also attrition, as found for instance in Gaelic speaking communities in Scotland (Dorian 1989). Since Malay is used less

---

<sup>9</sup>To be fair, speakers of Tucanoan languages also use evidentials when they speak Portuguese (Aikhenvald 2004:298), so that the transfer of L1>L2 is a possibility as well. It is clear, however, that the status of Sri Lanka Malay with regard to Tamil matches more closely Tariana's relation to Tucanoan than Tariana's relation to Portuguese.

<sup>10</sup>This development actually precedes the disbandment of the Regiment by a dozen years or so.

and less, and the domains of use become fewer and fewer, children do not acquire a full command of the language and do not know their ways in Malay in certain situations. In those situations, they draw on their general linguistic competence to find a way to convey the message. This general linguistic competence is based on Sinhala, so that in ad hoc formations, it is Sinhala, rather than Malay, which provides the communicative strategies.

A good candidate for this process is the indefinite article. While it cannot be excluded that its use in Malay goes back to very early times, it is equally plausible that it is a recent development. Due to loss of competence in Malay *hatthu* was used every time when the speaker would have used *ek~ak* in Sinhala. Using *hatthu* was not ungrammatical before, but it was not required. Through the use of Sinhala on a nearly permanent basis, this pattern became entrenched, and the knowledge about the optionality of *hatthu* disappeared. This change is very local and has no consequences on other parts of the grammar, so that it can very easily be quite recent.

The use of instrumental and accusative subjects is also quite local and could be recent. The accusative and the instrumental are of course much older, but their use with subjects could be due to Sinhala structures gaining importance. It is very easy to add a postposition to an NP, and has no repercussion elsewhere in the grammar, so that this change would not need a lot of time to take place.

Another change which can be explained by attrition is the dominance of the order ADJ N in younger speakers where older speakers also have N ADJ. The continuous exposure to the head final structure in Sinhala led to the dismissal of the historically dominant head-initial structure.

#### **10.4.6 Stage 5: independent developments**

Languages change through contact, but of course internal developments are also common. The fact that Sri Lanka Malay is a contact language does not mean that it is somehow immune to internal change. Since for the features listed for this stage contact is not a viable explanation, they have to be analyzed as independent developments and are of no further interest for the topic of this paper beyond the point made in the beginning of this paragraph.

### **10.5 Conclusion**

In this paper, I have shown that the current shape of Sri Lanka Malay is the result of at least 6 different stages. SLM can thus not be categorized into a handy drawer without being explicit about what particular stage one is talking. In order to come to grips with the structure of this language, a holistic approach must be taken, and the development of a language must be understood as a continuous rather than a punctual process. At different points in time, different processes are at work, and different demographic, social, and political circumstances lead to different behaviour by the speakers, resulting in different types of language change (cf. [Migge 1998, 2003, Migge & Goury 2008](#)).

The stage of most interest for creole studies is stage three. About a third of the ‘Sri Lankan’ features are found in this stage. In this paper I have shown that the phenomena of stage three are probably not the result of creole formation, but of metatypy or convergence.



# Bibliography

- Adelaar, K. A., 1991. "Some Notes on the Origin of Sri Lanka Malay." In *Papers in Austronesian Linguistics*, edited by H. Steinhauer. Canberra: Australian National University, 23–37.
- Adelaar, K. A. & Prentice, D. J., 1996. "Malay: its History, Role and Spread." In *Atlas of Languages of Intercultural Communication in the Pacific, Asia and the Americas*, edited by Stephen A. Wurm, Mühlhäusler Peter, & Darrell T. Tryon. Berlin: Mouton de Gruyter, 673–693.
- Adelaar, K. Alexander, 1985. *Proto-Malayic; The reconstruction of its phonology and parts of its morphology and lexicon*. Ph.D. thesis, Leiden University.
- Aikhenvald, Alexandra Y., 2002. *Language Contact in Amazonia*. Oxford: OUP.
- Aikhenvald, Alexandra Y., 2004. *Evidentiality*. Oxford: OUP.
- Ansaldo, Umberto, 2008. "Sri Lanka Malay revisited: Genesis and classification." In *A world of many voices: Lessons from documented endangered languages*, edited by Arienne Dwyer, D. Harrison, & David Rood. Amsterdam, Philadelphia: John Benjamins., 13–42.
- Ansaldo, Umberto, 2009. *Contact languages, Ecology and Evolution in Asia*. Cambridge: Cambridge University Press.
- Ansaldo, Umberto & Nordhoff, Sebastian, 2009. "Complexity and the age of languages." In *Complex processes in new languages*, edited by Enoch O. Aboh & N. Smith. Amsterdam: John Benjamins, 345–363.
- Apoussidou, Diana & Nordhoff, Sebastian, 2008. "Feet in Sri Lanka Malay – no stress please." Paper presented at OCP 5, January 2008, Toulouse.
- Bakker, Peter, 1995. "Over convergentie in een sprachbund." In *Artikelen van de Tweede Sociolinguïstische conferentie*, edited by Erica Huls & Jetske Klatter-Folmer. Delft: Eburon, 37–51.
- Bakker, Peter, 2000a. "Convergence intertwining: an alternative way towards the genesis of mixed languages." In *Languages in Contact*, edited by Dicky Gilbers, John Nerbonne, & Jos Schaeken. Amsterdam: Rodopi, 29–35.
- Bakker, Peter, 2000b. "Rapid Language Change: creolization, intertwining, convergence." In *Time depth is historical linguistics*, edited by C. Renfrew, A. McMahon, & R. Trask. Cambridge: McDonald Institute for Archeological Research, 586–620.
- Bakker, Peter, 2006. "The Sri Lanka Sprachbund: The newcomers Portuguese and Malay." In *Linguistic Areas – Convergence in Historical and Typological Perspective*, edited by Yaron Matras, April McMahon, & Nigel Vincent. Hounds mills, New York: Palgrave Macmillan, 135–159.
- Baxter, Gareth J., Blythe, Richard A., Croft, William, & McKane, Alan J., 2009. "Modeling language change: An evaluation of Trudgill's theory of the emergence of New

- Zealand English.” *Language Variation and Change* 21:257–296.
- Dawkins, R. M., 1916. *Modern Greek in Asia Minor: a study of the dialect of Silli, Cappadocia and Phárasa with grammars, texts, translations, and glossary*. Cambridge: Cambridge University Press.
- Dorian, Nancy C. (ed.), 1989. *Investigation Obsolescence: Studies in Language Contraction and Death*. Cambridge: CUP.
- Gunasekara, D. Mendis, 1891. *A comprehensive Grammar of the Sinhalese Language*. Sri Lanka Sahitya Mandalaya.
- Hussainmiya, B. A., 1987. *Lost cousins: The Malays of Sri Lanka*. Bangi: Institut Bahasa, Kesusasteraan dan Kebudayaan Melayu.
- Jaffar, Mohamed, this volume. “The Semantics of Serial Verb Constructions in Sri Lanka Malay.”
- LePage, Robert Brock & Tabouret-Keller, Andrée, 1985. *Acts of identity: Creole-based approaches to language and ethnicity*. Cambridge: CUP.
- Mesthrie, Rajend, 1993. “Koineization in the Bhojpuri-Hindi diaspora – with special reference to South Africa.” *International Journal of the Sociology of Language* 99(1):25–44.
- Migge, Bettina, 1998. “Substrate influence in creole formation: the origin of *give*-type serial verb constructions in the Surinamese Plantation Creole.” *Journal of Pidgin and Creole Languages* 13:215–265.
- Migge, Bettina, 2003. *Creole Formation as language contact: the case of the Suriname Creoles*. Amsterdam: John Benjamins.
- Migge, Bettina & Goury, Laurence, 2008. “Between contact and internal development: Towards a multi-layered explanation for the development of the TMA system in the creoles of Suriname.” In *Roots of creole structures: Weighing the contribution of substrates and superstrates*, edited by Susanne Michaelis. Amsterdam: John Benjamins, 301–332.
- Mufwene, Salikoko S., 1996. “The founder principle in creole genesis.” *Diachronica* 13:83–134.
- Nordhoff, Sebastian, 2009. *A Grammar of Upcountry Sri Lanka Malay*. Ph.D. thesis, University of Amsterdam, Utrecht.  
URL <http://www.lotpublications.nl/publish/articles/003744/bookpart.pdf>
- Nordhoff, Sebastian, 2011. “Having come to be a copula in Sri Lanka Malay – an unusual grammaticalization path.” *Folia Linguistica* 45(1):103–126.
- Nordhoff, Sebastian, 2012a. “Commentary on Smith’s papers.” *Journal of Language Contact* 5:73–79.
- Nordhoff, Sebastian, 2012b. “Establishing and dating Sinhala influence in Sri Lanka Malay.” *Journal of Language Contact* 5(1):23–57.
- Nordhoff, Sebastian, forthcominga. “Jack-of-all-trades: The Sri Lankan Malay property word.” In *Flexible word classes: a typological study of underspecified parts-of-speech*, edited by Eva van Lier & Jan Rijkhoff. Oxford: OUP.
- Nordhoff, Sebastian, forthcomingb. “Multi-verb constructions in Sri Lanka Malay.” *Journal of Pidgin and Creole Languages*.
- Paauw, Scott, 2004. *A Historical Analysis of the Lexicon of Sri Lanka Malay*. Master’s thesis, York University (Toronto). Presented as a paper at the meeting of the Linguistic Society of America/Society for Pidgin and Creole Linguistics, Oakland, CA,

- January 2005.
- Paauw, Scott, this volume. "The Lexical Sources of Sri Lanka Malay Revisited."
- Paauw, Scott H., 2008. *The Malay contact varieties of Eastern Indonesia*. Ph.D. thesis, State University of New York at Buffalo.
- Ross, Malcolm, 2007. "Calquing and metatypy." *Journal of Language Contact* 1:117–143.
- Ross, Malcolm D., 1996. "Contact-induced change and the comparative method: Cases from Papua New Guinea." In *The comparative method reviewed: Regularity and irregularity in language change*, edited by Mark Durie & Malcolm D. Ross. New York: Oxford University Press, 180–217.
- Ross, Malcolm D., 1997. "Social networks and kinds of speech-community event." In *Archaeology and language 1: Theoretical and methodological orientations*, edited by R. Blench & Matthew Spriggs. London: Routledge, 209–261.
- Ross, Malcolm D., 2001. "Contact-induced change in Oceanic languages in North-West Melanesia." In *Areal diffusion and genetic inheritance*, edited by Alexandra Y. Aikhenvald & R. M. W. Dixon. Oxford: Oxford University Press, 134–166.
- Ross, Malcolm D., 2003. "Diagnosing prehistoric language contact." In *Motivations for language change*, edited by Raymond Hickey. Cambridge: Cambridge University Press, 174–198.
- Siegel, Jeff, 1987. *Language contact in a plantation environment*. Cambridge: CUP.
- Siegel, Jeff, 1998. "Substrate reinforcement and dialectal differences in Melanesian Pidgin." *Journal of Sociolinguistics* 2/3:347–373.
- Siegel, Jeff, 2000. "Substrate Influence in Hawai'i Creole English." *Language in Society* 29:197–236.
- Singler, John Victor, 1988. "The homogeneity of the substrate as a factor in Pidgin/Creole genesis." *Language* 64:27–51.
- Slomanson, Peter, 2008. "The perfect construction and complexity drift in Sri Lankan Malay." *Lingua* 118(10):1640–1655.
- Slomanson, Peter, this volume. "Known, inferable, and discoverable in Sri Lankan Malay research."
- Smith, Ian, this volume. "Hijacked constructions in Second Language Acquisition: Implications for Sri Lanka Malay."
- Smith, Ian R., 2003. "The provenance and timing of substrate influence in Sri Lankan Malay; Definiteness, animacy and accusative case marking." Paper presented at the South Asian Language Analysis Roundtable XXIII, Austin, Texas.
- Smith, Ian R., 2012. "Adstrate Influence in Sri Lanka Malay: Definiteness, Animacy and Number in Accusative Case Marking." *Journal of Language Contact* :5–22.
- Smith, Ian R. & Paauw, S., 2006. "Sri Lanka Malay: Creole or convert." In *Structure and Variation in Language Contact*, edited by Ana Deumert & Stephanie Durrellman. Amsterdam: John Benjamins, 159–182.
- Smith, Ian R., Paauw, S., & Hussainmiya, B. A., 2004. "Sri Lanka Malay: the state of the art." *Yearbook of South Asian Languages* 2004:197–215.
- Tapovanaye, Suthadara, 1995. *Vowel Length in Sri Lankan Malay*. Master's thesis, University of Hawaii.
- Trudgill, Peter, 1986. *Dialects in contact*. Oxford: Blackwell.
- Trudgill, Peter, 2004. *New-dialect formation. The inevitability of colonial Englishes*. Edinburgh: Edinburgh University Press.

