



BRILL

*Journal of Language Contact* 5 (2012) 1–35JOURNAL OF  
LANGUAGE CONTACT  
  
brill.nl/jlc

# Establishing and Dating Sinhala Influence in Sri Lanka Malay

**Sebastian Nordhoff**

MPI-EVA, Leipzig

[please provide some more details on faculty, department; and contact details, at least email]

## Abstract

The study of Sri Lanka Malay has focussed on the genesis scenario, where theories of creolization (Smith et al., 2004; Smith & Paauw, 2006) with a dominant role of Tamil compete with theories of convergence (Bakker, 2006; Ansaldo, 2008) which allow for a more important role for Sinhala. This paper assesses and reevaluates the empirical data brought forward by both sides and contributes more empirical data on parallels with Sinhala. These parallels are partly due to substrate reinforcement (Siegel, 1998) of marginal structures found in Malay varieties, partly they are clear calques on Sinhala patterns. Some structures must be analysed as the result of Early Sinhala Influence during the colonial period, while for others, a later development following socio-political changes after independence is possible (Late Sinhala Influence). The paper argues that SLM changes towards Sinhala at both periods can be seen as a kind of metatypy comparable to other language contact settings in Eurasia and Papua.

## Keywords

Sri Lanka Malay; Sinhala; Tamil; language contact; metatypy

## 1 Introduction

Sri Lanka Malay is a contact variety of Malay which has undergone heavy grammatical restructuring towards Indian typology (Adelaar, 1991; Paauw, 2004; Smith et al., 2004; Smith & Paauw, 2006; Slomanson, 2006; Bakker, 2006; Ansaldo, 2008; Nordhoff, 2009). This is true for all domains of grammar, from phonology to discourse. Among the features which show that Sri Lanka Malay has become a member of the Indian, more precisely the South Asian, sprachbund, we find phonemic retroflexes as in (1), infinitives as in (2), conjunctive participles as in (3), and preposed relative clauses<sup>1</sup> as in (4), to name but a few of the various instances from different domains of grammar.

<sup>1</sup> These are more typical of the Southern subarea of the South Asian sprachbund (Gair 1994, forthcoming).


- (1) baʔok : baʔok  
cough : coconut shell
- (2) Itthu *cave*=nang kithang=le pii aada *mà*-liyath=nang. SLM  
DIST cave=DAT IPL=ADDIT go exist INF-look=DAT  
'We have also gone to that cave to have a look.' (K051206nar02)
- (3) a. Oorang pada *asà*-pirrang, SLM  
*man* PL CP-wage.war  
'After having waged war'
- b. derang=nang *asà*-banthu,  
*3* PL=DAT CP-help  
'and after having helped them'
- c. siini=jo su-chiggal.  
here=EMPH PAST-settle  
'the people settled down right here.' (K051222nar03)
- (4) [thoppi arà-daagang] oorang SLM  
hat NON.PAST-trade man  
'The man who sells hats; the hat seller' (K070000wrt01)

The features mentioned above are not found anywhere else in the Malay world, but are common among the languages of Sri Lanka, so that language contact is an explanation suggesting itself as the source, a fact that none of the researchers working on Sri Lanka Malay fails to notice. Beyond this broad generalization, things become more difficult. There are two main languages in Sri Lanka, Sinhala (Indo-Aryan) and Tamil (Dravidian), which could have contributed to the formation of these features. While Sinhala is the bigger language of the two, spoken by 4/5 of the population, a variety of Tamil is spoken by the ethnic group of the 'Moors', who share the Malays' Islamic faith and frequently married into Malay families. While the Moors only amount to about 6% of the population, the religious ties to the Malays mean that they had a greater impact than their numbers alone would suggest. One of the most hotly debated topics in the field of Sri Lanka Malay studies is the respective influence of Sinhala and (Moorish) Tamil (see Smith & Paauw (2006) and Ansaldo (2008) for advocacy of the different points of view, Nordhoff (2009) for an overview of the historical development).

The issue is complicated by the fact that Sinhala and Tamil are typologically very similar, and ~~the~~ Sinhalese grammar is indeed closer to the Dravidian type than to its Northern Indo-Aryan cousins.<sup>2</sup> The parallelisms between Sinhala


<sup>2</sup> Actually, the membership of Sinhala in the Indo-Aryan language family (rather than Dravidian) is less obvious than for the ~~Northern language~~ and has occasionally been disputed (Rask 1821, Lassen 1847:199). See Geiger (1938) for a discussion.

and Tamil are indeed so great that more often than not it is not possible to tell whether a given Sri Lanka Malay feature was adopted from one or the other given that both share identical structures. For the illustrative examples (1) to (4) above, we find in Sinhala and Tamil:<sup>3</sup>

- (5) a. kudu : kuḍu SINHALA  
bent powder  
b. kade : kaḍe TAMIL  
story shop
- (6)  kalisang dekaḱ=uyi ṣ **ekak**=uyi madi-**nnā** de-**nnā** tiyēnavāa SINHALA  
trousers two=ADDIT shirt one=ADDIT iron-INF give-INF have  
'I have to give two trousers and one shirt to be ironed (to give to iron)' (Karunatilake, 2004:62)
- b. Kumaar [amerikkaa-v-ukku-p poo-g-**a** virumbu-kiṛ-aṇ TAMIL  
Kumar America=DAT go-INF want-PRES-3SM  
'Kumar wants to go to America' (Lehmann, 1989:257)
- (7) a. Kumaar mehe āvillāa, maṭa kataa.keruvāa. SINHALA  
Kumaar ingee vandū, ennai kuuppittāa TAMIL  
Kumaar here CP.come me called  
'Kumaar came here and called me.' (based on Lehmann, 1989:266)
- (8) a. [kumaar-ai-k kaḍi-tt-a] naay ood-i-p pooy-iṛ-ru TAMIL  
Kumar-ACC bite-PAST-REL dog run-CP go-past 3SN  
'The dog which bit Kumar ran away' (Lehmann, 1989:288)
- b. [mang li-vv-a] liyum vagayak tiyēnavāa SINHALA  
IS write-PAST-REL letters some exist.INANIM  
'There are some letters that I wrote' (Karunatilake 2004:153)

For these features, it is impossible to trace the origin to either of the Lankan languages, since both of them show identical structures. This is already observed by Smith (2003), taken up again in Smith et al. (2004) and Smith & Paauw (2006).

### 1.1 Smith's findings

Smith notes the  logical similarity of Sinhala and Tamil and argues convincingly that only areas where Sinhala and Tamil differ allow insights into the

<sup>3</sup> There is a great deal of variation in the sources with regard to the spelling of Sinhala and Tamil. This concerns the marking of retroflexes *ay* or *ṭ* and the rendering of the labiodental approximant as *v* or *w* in both languages. In Tamil, there is further variation with regard to the presence vs. absence of voiced stops and the use of *ṇ* and *ṟ*. In this paper, I retain the original orthography used by the sources.

respective contributions of these languages. Smith (2003:8) determined three such areas, which will now be discussed in turn: number, the accusative case, and definiteness.<sup>4</sup> Table 1 sums up these generalizations.

1. In Sinhala singular nouns are obligatorily marked  $\square$  definite; in Tamil no specific definiteness markers are found.
2. In Sinhala number is marked obligatorily on all nominals; in Tamil number marking is optional, though it generally occurs more frequently with nouns higher on the animacy hierarchy.
3. Animate direct objects of a verb may optionally be marked accusative in Sinhala; in Tamil object marking is obligatory for definite/specific objects and is variable for indefinite/nonspecific objects, with a stronger tendency to mark animates. (Smith, 2003:6)

According to Smith, SLM exactly mirrors the Tamil distribution in what concerns definiteness, number marking, and accusative marking: there is no grammaticalized marking of definiteness, number marking is optional, and accusative marking is obligatory for animate objects and optional for inanimates. This contrasts with the Sinhala distribution, where definiteness and number are grammaticalized, and accusative is only possible on animates, where it is optional.

**Table 1** Nominal categories in Sinhala and Tamil, based on Smith (2003:6,8).

			man (anim)	book (inanim)	man (anim)	book (inanim)
SG	NOM	def	minih-aa	pot-a	manidan	puttagam
		indef	minih-ek	pot-ak		
	ACC	def	minh-aa(-va)	pot-a	manidan-e	puttagatt-e
		indef	minh-ek(-va)	pot-ak	manidan	puttagam
PL	NOM	def	minissu		manidan-va	puttagam
		indef		pot		
	ACC	def	minissu(ng-va)		manidan-va]-e	puttagatt-e
		indef			manidan-va	puttagam
			Sinhala		Tamil	

<sup>4</sup> While Smith et al. (2004) and Smith & Paauw (2006) are more easily available, both papers ultimately draw and rely on and approve of the findings presented in Smith (2003). I will therefore rely on the presentation of the facts found in Smith (2003).

### 1.1.1 Additional feature: dental onsets

An additional feature, which is given less importance, is the change in Malay coronal stops. Malay varieties generally have two coronal stops, a voiceless dental stop /t̪/ and a voiced alveolar stop /d/. There is thus an asymmetry. In SLM, this asymmetry is levelled out, with two additional phonemes filling the gap, namely /d̪/ and postalveolar/retroflex /ɖ/ (the voiced stop is also postalveolar/retroflex in SLM). What is interesting is that voiced dental stops are only found in onsets. In all other environments, the original articulation further back is preserved. Smith et al. (2004:202) attribute this to influence from Tamil, whose phonology does not permit retroflexes in initial position. This would have led to substitution of the Malay alveolars, which were equated with retroflexes, by the dentals in initial position, but not elsewhere. Sinhala on the other hand has no such phonotactic restrictions, so that this phonological characteristic points to Tamil influence.

### 1.2 Ansaldo's findings

Ansaldo (2008, 2009) deplores the 'Tamil bias' in the study of language contact in SLM and argues that Sinhala was actually the more important contact language. He adduces evidence from the pronominal paradigms and from the semantics of the instrumental and ablative case markers. According to Ansaldo, the dative marker *-dang*, which is only found on first and second person singular in SLM is a result of animacy effects, which can also be observed in Sinhala, although in other areas of grammar. Furthermore, the marker *-ring* can be used for both instrumental and ablative in SLM, similar to what we find in Sinhala, but different from Tamil, where these two cases are expressed by distinct morphemes.

### 1.3 Assessing the evidence

The claim for absence (Smith et al.) or presence (Ansaldo) of Sinhala influence rests on the six features sketched above. I will now analyse the validity of these claims in more detail, taking into account the actual distribution of these features in the three languages. I will also compare them to the grammar of historical varieties of Trade Malay in order to rule out possible inheritance and analyse the plausibility of the path of change proposed.

#### 1.3.1 Number marking

The first domain I will discuss is number marking. In this domain, the facts are clear and I have nothing to add to Smith's description. Smith provides

convincing evidence for the optionality of number marking in Tamil and Sri Lanka Malay. He also shows that Sinhala has obligatory number marking, and that Sinhala influence can thus not be invoked here.

### 1.3.2 *Accusative marking*

The second domain to investigate is accusative marking. In comparison to number marking, the development of case morphology is a very good candidate for language contact. No other well-described variety of Malay has accusative marking, while both Sinhala and Tamil do, so that influence is likely here. As for the exact provenance of accusative marking, Smith argues that the Sri Lanka Malay accusative aligns with Tamil, and not with Sinhala, where the accusative is less developed. Smith bases his argument on the claim that the conditioning parameters for the presence of the accusative are definiteness and animacy in both Tamil and SLM. In Sinhala on the other hand, only animacy plays a role (also cf. Gair, 2003:780,790). Ansaldo (2008) and Nordhoff (2009:329–332) show that the SLM system is actually more complex than what Smith presented, with additional parameters like topicality, affectedness and singular number also positively affect the presence of the accusative marker.<sup>5</sup> This more detailed analysis, however, does not suggest that Sinhala would have had a more important part than what was previously suggested. The additional parameters described by Nordhoff could not possibly be explained by influence from Sinhala, where they play no role. On the contrary, more detailed descriptions of Tamil like Asher (1985:107) or Lehmann (1989:29) show that the Tamil system is actually not as clear-cut as sketched by Smith and that other parameters than definiteness and animacy enter into play. To sum up, a good case for Sinhala influence cannot be made for the accusative, and Tamil influence is more likely in this domain.

### 1.3.3 *Indefiniteness marking*

Smith observes that Tamil does not mark (in)definiteness, while Sinhala has a solid grammaticalized marking of indefiniteness. These findings find support in the literature (Asher (1985); Lehmann (1989) for Tamil; Gair & Paolillo (1997); Gair (2003); Karunatilake (2004) for Sinhala). New data collected in Sri Lanka on SLM point towards a very clearly grammaticalized marking of indefiniteness, just like in Sinhala. The indefiniteness marker (*h*)*atthu* is actually one of the most frequent morphemes and has to be used in exactly the

<sup>5</sup> One could subsume this under the rubric of ‘transitivity’ as done for instance by Hopper & Thompson (1980) or Næss (2007). The higher the transitivity of an event, the more likely the accusative is to surface.

same instances as its Sinhala counterpart (Nordhoff, 2009:319–323), albeit as a clitic rather than a suffix.

- (9) Se **atthu**=aade, se **hatthu**=aade. SLM  
 1S INDEF=younger.sibling 1S INDEF=younger.sibling  
 ‘I am a younger sibling, I am a younger sibling.’ (K061120nar01)
- (10) **Hathu**=oorang=pe muuluth=dering **hathu**=criitha kal-dhaathang. SLM  
 INDEF=man=POSS mouth=ABL INDEF=story when-come  
 ‘When a story comes out of a man’s mouth.’ (B060115prs15)

Compare this to the Sinhala examples:

- (11) mang malli ken-**ek** SINHALA  
 1s younger.brother-CLF.ANIM-INDEF.ANIM  
 ‘I am a younger brother’
- (12) minih-**ek**-gee kaṭ-ing kataav-**ak** ena koṭa SINHALA  
 man-INDEF.ANIM-POSS mouth-ABL story-INDEF.INANIM come when  
 ‘When a story comes out of a man’s mouth’

While the use of Tamil *oru*, ‘one’ would be possible in the examples above, in Tamil it is equally grammatical to leave out *oru*. This is different from Sinhala and SLM, as the following example shows. In these languages, indefiniteness marking is obligatory, albeit on different sides of the noun:

- (13) Kumar vakkil TAMIL  
 Kumar niitignya.varay **-ek** SINHALA  
 Kumar **hatthu** prentha.oo SLM  
 Kumar one lawyer INDEF  
 ‘Kumar is a lawyer.’ based on Lehmann (1989)

Given that the marker surfaces on different sides of the noun, we cannot speak of morphological influence here; what we find is rather a *semantic* influence in the sense that indefiniteness is now obligatory in both languages.

Further evidence for Sinhala influence comes from another use of *hatthu*, namely as a loanword integrator. In Sinhala, English loanwords in the singular are integrated by means of *eka*, cognate to the indefiniteness marker *-k*. So we find *kar eka*, *bas eka*, *pleen eka* etc. Crucially, these words are *definite*, and not indefinite as one could presume. This means that they can co-occur with deictic markers like *mee*, ‘this’, thus *mee kar eka*, *mee bas eka* etc. In order to make them indefinite, the marker *-k* has to be added (*kar eka-k*, *bas eka-k*). As explained above, both *eka* and *-(a)k* are related to the numeral *eka*, ‘one’. Both the indefiniteness marker and the numeral are rendered in Sri Lanka Malay as *hatthu*. This means that it is possible for *hatthu* to occur *twice* in a sentence, once for each function. The following examples show this (cf. Nordhoff, 2009).

- (14) Itthu **atthu**=story=**atthu** SLM

DIST INDEF=story=INDEF

‘This is a story.’ (B060115nar05)

- (15) Kitham=pe **atthu**=-three-tonner=**atthu** aada, duppang=ka SLM

1PL=POSS INDEF=three-tonner=INDEF exist, front=LOC

‘There was a three-tonner of ours at the front.’ (K051206nar16)

The Sinhala markers *eka* and *-k* which were calqued by the SLM construction are shown in the following example.

- (16) apee three-tonner **eka-k** tibuna eliya SINHALA

1PL.POSS three-tonner LOAN-INDEF exist.INANIM.PAST front.LOC

‘There was a three-tonner of ours at the front’

The double occurrence of *oru/onnu*, ‘one’ in Tamil is not grammatical as (17) shows.<sup>6 7</sup>

- (17) \*enggal oru three-tonner *oru* irukku, munnaal-le TAMIL

1PL.POSS one three-tonner one exist-3n front-LOC

In addition to the double use of markers related to the numeral ‘one’, the loan word integrator can co-occur with deictics, as shown in (18).

- (18) **Inni mock wedding =hatthu** mas-gij **SLM**  
**Mee mock wedding eka** karandā oonā **SINHALA**  
**\*inda mock wedding oru** ceyya veerum **TAMIL**

PROX mock wedding=INDEF must- make must

‘I have to do this mock wedding.’ (based on K060116nar10)

There is a difference between Sinhala and Sri Lanka Malay in that in Sinhala the integrator is obligatory while in Sri Lanka Malay, it is optional. When SLM makes use of the integrator, it is a clear calque of the two Sinhala construction and does not have a parallel in the Tamil variety described by Smith.<sup>8</sup>

Contrary to what has been argued by Smith, the indefiniteness marker is thus not an example of absence of Sinhala influence, but, really, one of the best examples for exclusive Sinhala influence one can find in Sri Lanka Malay grammar.

<sup>6</sup> Swapping the first two words to get the possessor adjacent to the possessum does not make the sentence grammatical, either. *Oru* and *oru* are prenominal and postnominal allomorphs, respectively.

<sup>7</sup> An anonymous reviewer objects that this sentence is in fact grammatical in Tamil. I consulted with Prof. Nuhman, emeritus professor for Tamil studies at the University of Peradeniya and a specialist of the relevant dialects of Tamil. He confirmed that the two morphemes cannot co-occur.

<sup>8</sup> It is possible that certain Sri Lankan varieties of Tamil have actually developed a grammaticalized marking of indefiniteness through contact with Sinhala, but this needs further research.



### 1.3.4 Dental onsets

There is consensus that the development of additional stops (voiced dental and unvoiced alveolar/retroflex) is due to language contact. The stop inventory of the relevant languages is given below. For Tamil, voiced and voiceless allophones are both given since they will become important in the course of the argument.

It is immediately obvious that the SLM facts closely mirror the Sinhala and Tamil facts. However, things are different when it comes to phonotactics. There, we find that some of the historically alveolar voiced stops have become dental in SLM. Smith suggests that the occasional rendering of voiced coronal stops in initial position in Sri Lanka Malay as dental is due to Tamil influence. Elsewhere, the historically alveolar /d/ is realized as retroflex, but for some lexemes, it has a dental reflex in initial position. Tamil does not allow retroflexes in initial position, a feature traced to Proto-Dravidian by Smith. Sinhala on the other hand does allow such occurrences. Tamil influence could thus explain why the originally alveolar stops have not developed into the retroflex phoneme, but rather into the dental one. What is puzzling though is that another constraint found in Tamil, and also going back to Proto-Dravidian, namely a ban on initial voiced stops, did not have any influence. The allophones allowed in initial position are given below.

**Table 2** Phonemes and allophones of coronal stops.

Malay	Dental	Alv/retroflex	SLM	Dental	Alv/retroflex
voiced	–	+	voiced	+	+
voiceless	+	–	voiceless	+	+
Tamil	Dental	Alv/retroflex	Sinhala	Dental	Alv/retroflex
voiced	+	+	voiced	+	+
voiceless	+	+	voiceless	+	+

**Table 3** Coronal allophones in initial position.

Malay	Dental	Alv/retroflex	SLM	Dental	Alv/retroflex
voiced	–	+	voiced	+	+
voiceless	+	–	voiceless	+	–
Tamil	Dental	Alv/retroflex	Sinhala	Dental	Alv/retroflex
voiced	–	–	voiced	+	+
voiceless	+	–	voiceless	+	+

If indeed Tamil learners are responsible for the change from alveolar to dental due to their native phonologies, it would come as a surprise that the much more salient constraint against voiced stops in onset position should not have any effect. One would rather expect /#t/ for Tamil learners, but not /#d/, an absolute onset which is illicit in Tamil. On a more general level, the expansion of the possible onsets from two in Malay (t,d) to three in SLM (t,d,d) is difficult to trace to Tamil influence, where we find only one allophone in this position (t).

### 1.3.5 Animacy in the pronominal paradigm

Ansaldo (2008) argues that animacy distinctions in the pronominal paradigm are a possible result of Sinhala influence. Two steps must be performed in order to ascertain this: the relevant forms must be established and analysed correctly, and the structures on which this distinction is modelled in Sinhala must be identified. Unfortunately, Ansaldo's account fails on both points.

Ansaldo (2008:30) states that “The DAT marker appears in two alternations [...]: the form *day* is used for first and second pronouns singular exclusively, while for all the other pronouns *naŋ* must be used. [...] This suggests a distinction that may be captured along the lines of the Animacy Hierarchy where first and second person pronouns are assigned highest animacy and therefore differential marking”. Ansaldo's paradigm is repeated in Table 4, with a column for nominative added.

We see that the dative forms of first and second singular persons have *-dang* instead of *-nang*. The third person has *-nang*, so that one can analyse this as a differential treatment based on the Animacy Hierarchy. This becomes problematic in the plural though. There we find that *-dang* is never used. This is surprising under an account based on animacy. Under that account, we would expect *\*kitangdang* and *\*lorangdang* instead of *kitangnang* and *lorangnang*.

**Table 4** Nominal paradigm according to Ansaldo (2008) with nominative column added

	Nominative	Accusative
1s	go	godang
2s	lu	ludang
3s	dia	dianang
1p	kitang	kitangnang
2p	lorang	lorangnang
3p	derang	derangnang

One can modify the analysis by including number as an additional parameter and rank plural lower on the animacy hierarchy than singular, but this seems like an ad hoc solution. There might be a possible argument for this, but I am not aware of any cognitive work suggesting such an additional split. Still, the different allomorphs of the dative marker have to be explained. Since the semantic explanation is not fully convincing, insights can be sought from another domain of linguistics, e.g. phonology. It turns out that a phonological account is actually able to neatly explain the difference: monosyllabic pronouns like *go* and *lu* take *-dang*, polysyllabic pronouns (i.e. all others) take *-nang*.<sup>9</sup> This phonological account finds support from additional data in Nordhoff (2009:225), where we find the forms *dedang*, ‘3s.IMPOLITE.DATIVE’ and *diyanang*, ‘3s.POLITE.DATIVE’. Both forms are third person, but one is monosyllabic (*de-*) while the other one is disyllabic (*diya*). As predicted by the phonological analysis, the monosyllabic pronoun takes *-dang* and the disyllabic one, *-nang*. One could try to save the animacy account and argue that impolite forms are higher on the Animacy Hierarchy than polite forms but this clearly stretches this particular hierarchy beyond its capacities, besides being unintuitive in the first place.

Even if we accept the animacy account for the sake of discussion, we have to explain how this distinction was brought about by Sinhala influence. Ansaldo (2008:31) argues that “unlike Tamil, the Sinhala system has a natural gender in which the main distinction is between Animate and Inanimate. Inanimate nouns can take four cases [...], while Animate nouns have two additional cases [...]. The animacy effect noted in the SLM pronouns could have developed under influence of a more pervasive animacy feature found in Sinhala.” Sinhala thus does not have the difference in case allomorphs as argued for SLM. Rather, it has animacy effects in the number of cases available to different lexemes. The transfer can thus not be a direct influence from Sinhala, but must take some more intricate path involving “transfer of semantic features without corresponding syntactic categories” (Ansaldo, 2008:31). If we accept this particular possibility, it is clear that Tamil features could also be at the origin of the SLM distinction. Tamil has animacy distinctions in verbal agreement and number marking,<sup>10</sup> which could have made use of the same mechanism as proposed for Sinhala by Ansaldo. To sum up, the animacy analysis of the pronominal paradigm is problematic to begin with, but even if we accept it, a clear case for Sinhala influence rather than Tamil influence cannot be made.

<sup>9</sup> *Dia* is disyllabic.

<sup>10</sup> Sinhala furthermore also has distinctions along the lines of animacy in plural formation (Nitz & Nordhoff 2010) and selection of existential verbs.

### 1.3.6 Semantics of the instrumental case marker

Another domain of Sinhala influence argued by Ansaldo (2008:32,35) is the conflation of instrumental and ablative, both marked by *-riŋ* in the Kirinda dialect. This is what he finds in Sinhala, but crucially not in Tamil, where these two cases are distinguished. Ansaldo does not give the actual case forms, but I take the Sinhala forms to be *-ing* ([-anim]), *-geng* ([+anim]) for ablative/instrumental and the Tamil forms *-aale* for the instrumental and *-(i)runtu* for the ablative.

According to Ansaldo, SLM has a conflation of ablative and instrumental like Sinhala, and unlike Tamil. For this, a clear analysis of the Sinhala and the Tamil systems based on common criteria is necessary. It happens that Sinhala studies and Tamil studies use different definitions of case. This leads to the disappearing of two cases from the Sinhala paradigms found in the literature, which would have been there if the Tamil standards had been applied. An equivalence of the cases of interest here is given in Table 5.

We see that while Tamil has an instrumental suffix (*-aale*), and SLM has an instrumental enclitic postposition (=dering, Ansaldo's *-riŋ*). Sinhala has a circumlocution with the free word *atiŋ*, 'with the hand', which is paradigmatic, but not phonologically integrated, leading to the exclusion from case paradigms in the literature.

The same is true of the form *iñdalaā*, which is the conjunctive participle of the animate existential *innavaa*. In Sinhala studies, this form is regarded as an independent word and not analyzed as forming part of the case paradigm. The Tamil form *iruntu*, originally also the conjunctive participle of the existential (*iru*), on the other hand, is analyzed as a case suffix and found in the ablative cell of paradigms by Tamil scholars. As for SLM, the conjunctive participle of the animate existential, *asàduuduk*, is normally not seen as a case marker, but

**Table 5** Cases in Tamil, Sinhala, and SLM. Parentheses mark morphemes indicating thematic role which are not phonologically integrated.

	Tamil	Sinhala	SLM
NOM	Ø	Ø	Ø
ACC	-ei	-va	=yang
DAT	-kku	=t̪a	=nang
INSTR	-aale	=geng =ing (atiŋ)	=dering
ABL	-runtu	=geng =ing (iñdalaā)	=dering (asàduuduk)

rather as a free word as in Sinhala. It is clear, however, that all three languages have a clear ablative marking based on the very same process, the grammaticalization of the conjunctive participle of the existential. This is phonologically more integrated into the host word in Tamil, in line with the overall structure of this language, than in Sinhala or SLM, but the semantics in what concerns thematic role expressed are the same.

The choice of the ablative marker is obligatory in Tamil, but optional in Sinhala and SLM, where the instrumental marker can be used instead. The parallels between Sinhala and SLM are thus closer here. Furthermore, the original distinction between the old Malay ablative preposition *dari* and the instrumental preposition *pake* has been lost in Sri Lanka Malay, where only *dering*, the reflex of *dari*, remained. This conflation cannot be due to Tamil influence since Tamil keeps its instrumental and its ablative distinct at all times. It must be due to Sinhala influence, where *-geng* or *-ing* can be used for both meanings. On the other hand, there is a difference in animacy for the Sinhala instrumental forms not found in SLM. This suggests less influence from Sinhala here. To sum up, the conflation of instrumental and ablative posited for Sinhala and SLM is more involved than previously assumed, but a case for Sinhala influence to the exclusion of Tamil can still be made.

#### 1.4 Summary of previous claims

Taking a look at what has been said above about the mainly nominal categories of number, case, and definiteness, a muddled picture emerges. Accusative seems to be closer to Tamil, while the indefiniteness marker and the conflation of ablative and instrumental point towards influence from Sinhala (6). The other features discussed in the literature had to be discarded because retention cannot be ruled out (number marking), the scenario suggested would trigger other changes than the ones observed (dental stops) or the analysis of SLM data was problematic (animacy in pronouns). It is clear that more empirical data is needed. This will be addressed in the next section.

## 2 Broadening the empirical basis

Given that Sinhala does seem to have had some influence on the formation of Sri Lanka Malay, in indefiniteness marking and, to a lesser extent, in the instrumental, it is interesting to see whether other domains of Sinhala influence can be found. As noted by Smith (2003), it is important to assure that the surveyed areas are actually treated differently in Sinhala and Tamil so that

**Table 6** Summary of adstrate features proposed in the literature.

	Tamil	Sinhala	inherited?
optional number	+	–	+
accusative	+	(–)	–
indefiniteness	–	+	–
dental onsets	–	–	–
animacy (dubious)	pronouns	nouns existentials	–
conflation of ABL and INSTR	–	+	–

the differential input can be evaluated as to its outcome. Furthermore, the feature must not be a plain retention of an old Malay feature. The following features comply with these requirements.

- number of stop series
- gemination
- distribution of stops and nasals
- existentials
- non-past form used in past contexts
- instrumental and accusative subjects
- zero adclausal nominalization

In these domains, Tamil and Sinhala show different grammatical features. Additionally, the SLM facts are sufficiently distant from most varieties of Malay as to include language contact as a possible scenario for the SLM feature.<sup>11</sup>

### 2.1 *Number of stop series*

A phonological feature where Sinhala and Tamil differ is the number of phonemic stop series. Tamil has only one series (p, t, (r),<sup>12</sup> t, c, k). Depending on the phonological environment, these can surface either voiceless, voiced, or fricativized. Sinhala on the other hand is quite different; it has three series: voiceless (p, t, t, c, k), voiced (b, d, d, j, g), and prenasalized (mb, nd, nd, nj,<sup>13</sup> ng).

<sup>11</sup> There are areas where Tamil and Sinhala differ, but where language contact need not be assumed. One such feature are retroflex ʎ and ŋ, which are phonemic in Tamil but not in Sinhala or SLM. This does however not constitute a change towards Sinhala but rather a retention of the ancient feature: these phonemes are virtually unheard of in the Malay world.

<sup>12</sup> This sound is not phonemic in most dialects.

<sup>13</sup> The phonemic status of nj is unclear. The Sinhala alphabet contains a letter for this sound, but there seem to be no words using this letter. It is also unclear whether nj is a prenasalized stop

Most if not all Malay varieties have 2 series of stops: they contrast a series of voiceless stops (p, t, c, k) with a series of voiced stops (b, d, j, g).<sup>14</sup> In contrast to the two series commonly found in varieties of Malay, SLM has three series: voiceless (p, t, c, k), voiced (b, d, j, g), and prenasalized (mb, nd, nj, ng, Nordhoff, 2009:96–99). This is thus exactly like Sinhala (discounting the absence of nd). Given that the two series of Malay are situated in between the one series of Tamil and the three series of Sinhala, this criterion allows us to clearly state that SLM now looks more like Sinhala than before, and that it is now, at least in this phonological domain, farther removed from Tamil than most other varieties of Malay. Having prenasalized stops which are not allophones of the nasals (as in Guarani for instance) or of the voiced stops (as in several African languages) is typologically not very common. Downing (2005) remarks, based on Maddieson & Ladefoged (1993), that such a distinction is uncommon, citing only Fula as a case comparable to Sinhala<sup>15</sup> in having this threefold distinction of N, C, and C̣. This typologically marked status suggests that language contact is a more likely scenario for the development of these sounds than internal developments.

## 2.2 Gemination

Another phonological feature where Sinhala and Tamil differ is gemination. In Sinhala, geminate stops can be voiceless or voiced. In Tamil, geminate stops are always realized as voiceless. Most if not all varieties of Malay do not have phonemic gemination (See Table 7). In SLM however, gemination is now distinctive. This is shown by the following near-minimal pairs:

- (19) a. **baapi appi maatha atthas thiikam ikkang**  
bring fire eye top stab fish
- b. **uubar subbath suuda puddas thiiga siggar**  
chase because thus spicy three unwell



If gemination was due to Tamil influence, the second series would be surprising. We would expect devoicing of the geminates there, but what we find are voiced geminates. It is more likely that allophonic gemination of both voiced and voiceless consonants present in certain varieties of Malays spoken by some

in the same virtue as the others. Matzel (1983) sees it as heterosyllabic N.C sequence rather than a tautosyllabic sequence, NC̣ as the other stops in the series.

<sup>14</sup> I am not aware of any counterexample to this claim, but the dialectal variation in Malay is so great that positing complete absence is difficult.

<sup>15</sup> We have to add Maldivian to this list (Fritz 2002), which is historically related to Sinhala.

immigrants was reinforced by Sinhala to yield the phonemic distinction we find today.

### 2.3 Distribution of stops and nasals

The third and final phonological criterion are the phonotactic properties of stops and nasals. In Tamil, a consonant after a nasal is always voiced. In Sinhala, it can be voiceless or voiced. Additionally, voiced consonants can syllabify with either the preceding or the following syllable, yielding a difference in meaning. This leads to a threefold contrast in NC sequences (Table 8).

All Malay varieties allow a voiceless stop after a nasal. All varieties also allow a voiced stop after a nasal. Where they differ is how this sequence of nasal and voiced stop is syllabified. In some varieties the syllabification is N.Ç,<sup>16</sup> while in others (of e.g. Sumatra, David Gil p.c.), it is .NÇ.<sup>17</sup> In no Malay variety is there a threefold contrast of the type we find in Sinhala, where N.Ç, N.Ç, and .NÇ contrast (Uri Tadmor, p.c.). That is, in no variety outside of Sri Lanka, where we find exactly the same contrast as in Sinhala.

**Table 7** Gemination

schema	VÇÇV	VÇÇV	VNNV
model	appa	abba	amma
Malay	N/a	N/a	N/a
Tamil	<i>Appaa</i> ‘father’	N/a	<i>Ammaa</i> ‘mother’
Sinhala	<i>Baappaa</i> ‘uncle’	<i>Ibba</i> ‘tortoise’	<i>Ammaa</i> ‘mother’
SLM	<i>Kappal</i> ‘ship’	<i>Habbar</i> ‘news’	<i>Samma</i> ‘all’

**Table 8** Sequences of nasal+stop. *Sambal* is in either the middle column or the right column, depending on the dialect. The other cell is then empty.

schema	VNÇV	VNÇV	V <sup>N</sup> ÇV
Malay	<i>rumpu</i> ‘grass’	<= <i>Sambal</i> => ‘condiment’	
Tamil	N/a	<i>Ambadu</i> ‘fifty’	N/a
Sinhala	<i>Gampaha</i> (place name)	<i>Kamba</i> ‘cloth book cover’	<i>Kaṁba</i> ‘ropes’
SLM	<i>kampak</i> ‘axe’	<i>Sambal</i> ‘condiment’	<i>Gaṁbar</i> ‘picture’

<sup>16</sup> I use the full stop to mark the syllable boundary, as customary in works on syllabification.

<sup>17</sup> The data from Johore Malay (Onn 1980) are also amenable to such an analysis, although Onn does not do so.



## 2.4 Existentials

Malay varieties generally have only one existential marker, *ada*. This marker is also found in SLM, where it surfaces as *aada*. Form and semantics of this marker are clearly inherited. Compare Indonesian *rumah ada*, ‘There are houses’ with SLM *ruuma aada*, ‘id.’

In distinction to Indonesian, Sinhala and Tamil both have two existential markers, but their semantics are different. In Sinhala, the distinguishing parameter is clearly animacy (Karunatilake, 2004:12). Existence of animate referents is expressed by *innavaa*, while existence of inanimate referents is expressed by *tiyenaavaa*. Swapping the existentials leads to ungrammaticality, as the following example shows.

- (20) a. mee iskolee    sinhala panti    tunak tiyenaavaa/\*innavaa    SINHALA  
           this school.LOC Sinhala classes three exist.INANIM/exist.ANIM  
           ‘There are three Sinhala classes in this school.’

b. langkaava rilaw    huṅgak    \*tiyenaavaa/innavaa    SINHALA  
    Sri    monkeys many    exist.INANIM/exist.ANIM  
    ‘There are many (macaque) monkeys in Sri Lanka.’ (Gair 2003:790)

Such a clear semantic distinction is not found in Tamil. There, *iru* is the default marker. It is occasionally substituted by *undu*, but the conditioning semantic parameters are less clear than in Sinhala and in no case identical to what we find in the latter language.<sup>18</sup>

Varieties of Malay typically use *ada* as the existential. SLM however has also recruited *duuduk* for this purpose. *Duuduk* can only be used with animate referents. Besides the use as an existential, it also retains its old meaning of ‘to sit’ and the extended meaning of ‘to reside’. These two meanings are also found in other varieties of Malay. What is not found is the fully grammaticalized existential as in (21).

- (21) Kiccil vavvaal pada daalang=ka arà-duuduk    SLM  
       small bat    PL    inside=LOC NON.PAST-exist.ANIM  
       ‘There are small bats inside.’

The above example shows that the historical meaning of ‘to sit’ is not the one intended here since bats are physically unable to perform the action of sitting. *Duuduk* in (21) is rather a pure existential. The semantics of this lexeme have thus changed on Lankan soil. It is easy to come up with a hypothesis why this

<sup>18</sup> Tamil does mark animacy in verbal agreement suffixes, but here we are concerned with the existentials.

change would have taken place: the difference between animate and inanimate existential found in Sinhala served as role model for the distinction between *aada* and *duuduk* we now find in SLM. Tamil on the other hand does not provide grounds for a similar hypothesis since the distinction between animate and inanimate existentials is not found there. What is more, the Sinhala animate existential *innavaa* itself cannibalized the historic meaning of the lexeme, which was ‘to sit’. The change from sitting to animate existential found in Sinhala has thus been retraced in SLM. It is true that this development might have taken place without Sinhala influence because the relevant ingredients were already present in the Malay varieties spoken by the immigrants. After all, grammaticalization paths like sit → reside → existential are not unheard of elsewhere (Heine & Kuteva, 2002:278ff). But the seeds planted in the Malay varieties fell on a fertile soil in Sinhala, where the same grammaticalization path was already trodden farther. Sinhala thus reinforced the tendencies already present, while a similar argument could not be made for Tamil.

## 2.5 Non-nominative subjects

Turning to the domain of syntax, we find that Tamil and Sinhala differ in the kind and amount of non-nominative subjects. In Tamil “stative predicates expressing the notion of mental, emotional, and physical experience require the case marking pattern of DAT-ACC, stative predicates expressing the notions of possession, necessity, etc. require the case marking pattern DAT-NOM.” (Lehmann, 1989:180). Tamil does not use dative subjects for predicates other than mentioned above. An example is given in (22).

- (22) Kumaar-ukku inda uur-ai-t teri-y-um TAMIL  
 Kumar-DAT this place-ACC know-3SN.FUT  
 ‘Kumar knows this place’ (Lehmann, 1989:184)

Sinhala (23) uses dative subjects as well, but they are more prevalent there. Sinhala has three conjugation classes in *-anavaa*, *-inavaa* and *-enavaa*, and all of the latter can take dative subjects (Gair, 1971). Additionally, Sinhala has accusative (24) and instrumental (25) subjects, which are not found in Tamil:


- (23) ma=ʃa kolaṁba bas.eka hæmadaama varadinavaa SINHALA  
 1s=DAT Colombo bus every.day miss  
 ‘I miss the Colombo bus every day’ (Gair, 2003:791)

- (24) **miniha-va** liss-una SINHALA  
 man-ACC slip-PAST.INVOL  
 ‘The fellow slipped’ (Gair, 1976:51)

- (25) ehee **polisiy-enj** innavaa SINHALA  
 there police-INSTR exist.ANIM  
 ‘There are police there.’ (Gair, 1991:13)

No other variety of Malay has non-nominative subjects, which is due to the absence of case in those varieties to begin with, entailing that morphological marking for semantic roles is not available. SLM on the other hand has a good number of dative subjects, as well as the accusative and instrumental subjects found in Sinhala.

- (26) **Derang=nang** byaasa svaara=hatthu su-dinngar. SLM  
 3PL=DAT habit noise=INDEF PAST-hear  
 ‘They heard a familiar sound.’ (K070000wrt04)

- (27) *Titanic* **kappal=yang** su-thingalam SLM  
 Tita  ip=ACC PAST-drown  
 ‘The Titanic sank.’ (Nordhoff, 2009:483)

- (28) **Police=dering** su-dhaathang SLM  
 police=INSTR PAST-come  
 ‘The police came.’ (Nordhoff, 2009:483)

The conditioning factors for the use of accusative and instrumental seem to be the same in Sinhala and SLM. This pattern of case marking is a clear instance of Sinhala influence.

## 2.6 Semantics of the non-past form

Sinhala, Tamil and SLM all have grammaticalized tense. Malay varieties on the other hand usually have aspect as a more important parameter, and the markers are morphologically less integrated than in the Lankan languages. In this section, I will focus on the present tense form in Sri Lanka Malay, encoded by the prefix *arà-*. This marker derives from Malay *ada*, which, among other things, encodes progressive aspect in the relevant varieties (Adelaar, 1991). In Sri Lanka Malay, it is used to refer to the present and to the future. So, *karang arà-nyaanyi* means ‘I am singing now’ and *beesona arà-nyaanyi* means ‘I will sing tomorrow’. The time reference is clearly indicated by the adverbs *karang*, ‘now’ and *beesona*, ‘tomorrow’, but *arà-* is compatible with both. The same flexibility is true of the Sinhala present tense form in *-navaa* and the Tamil present tense form in *(kki)r*, the latter with some morphological complications. This does not seem spectacular up to here. The interesting point is the possibility to use the present tense form in subordinates with past reference as in (29).

- (29) Blaakang=jo incayang anà-kuthumung [moonyeth pada ...  
 after=EMPH 3S.POLITE PAST-see monkey PL ...  
 pohong atthas=ka arà-maayeng] SLM  
 tree top=LOC SIMULT-play  
 ‘Then only he saw that the monkeys were playing on the top of the trees.’  
 (K070000wrt01)

The reference time is past tense, since the action already happened. This is why the verb *kuthumung*, ‘to see’ is marked for past tense with the prefix *anà-*. The complement of this verb is an embedded clause. This clause inherits the reference time from the matrix clause, as it were. In the embedded clause, then, the only thing which is indicated is that the event of the monkeys playing was simultaneous to the man looking at them (rather than anterior or posterior). This simultaneity of the event expressed in the embedded clause with the event expressed in the matrix clause is expressed by *arà-*.

This is possible in the same contexts where English would use the *-ing*-form, as in *I saw you dancing, I heard you crying, I felt you coming* etc. The extension of a present tense form to a participle like form in the past is surprising. It suddenly makes more sense when we consider that the Sinhala form in *-navaa* actually started out as a non-finite form (Geiger, 1938:163) and only later acquired the present tense reading it has today, supplanting the older finite present tense forms. Vestigial participial uses of the *-navaa*-form are still found in Sinhala (30,31).<sup>19</sup>

- (30) Siri horaa par-ee duw-anavaa dækka SINHALA  
 Siri thief road-LOC run-NONPAST see\PAST  
 ‘Siri saw the thief running on the road.’ (Gair, 2003:806)
- (31) amma=[a maalini sindu kiy-anaava æhuna SINHALA  
 mother=DAT Malini song say-NONPAST hear\PAST  
 ‘Mother heard Malini singing songs.’ (Gair, 2003:806)

The semantics of the *-navaa*-form and the *arà*-form are thus closely parallel. Both can be used for present and future. Additionally, they can be used in

<sup>19</sup> A more explicit diachrony of the *navaa* forms is provided by an anonymous reviewer: “Sinhala *-navaa* forms are probably derived from verbal adjective (relativizing) forms plus a nominal (agentive) form (*v*)*aa*. Geiger derives them from an older genitive form *-dāa*, by consonantal loss and coalescence, which became an oblique affix and then the forms were unemployed as nominative. [...] The same affix is the usual m.sg. pres. animate affix on nouns and some pronouns. Thus the *-navaal-vaa* (past and present respectively) were, and to a degree, are, nominal, which accounts for the possibility of case affixes being directly applied. It also helps account for these forms being used in subordinate clauses *sans* further marking, even though they are the general present/past “finite” forms in Colloquial Sinhala.”



subordinate clauses in the past. While the *navaa*-form has changed from an erstwhile participle to a present/future form, SLM *arà*- is partly on the opposite direction and acquiring participle uses. This extension of the present tense form in the domain of subordinates finds clear role models in Sinhala, but could not possibly be explained by Tamil influence.

One could of course argue that the use of *arà*- in progressive contexts as in (29) is a retention of the old Malay use of *ada* to denote progressive (Adelaar, 1991). But this account falls short of explaining why the use of SLM *arà*- with past reference is only possible in *subordinate* contexts, and not in main clauses. ‘The monkeys were playing on top of the trees’ in (29) must be marked with a past-tense prefix unless it is found in a complement clause. This restriction to subordinate clauses is not found in the relevant Malay varieties, where *ada* can be found in main clauses with past tense reference. An example of this is given in (32).

- (32) Nene Luhu itu ada mau punya tunanga ... AMBON MALAY  
 Nene Luhu DIST PROG FUT POSS fiancé ...  
 ‘Nene Luhu was going to be engaged ...’ (Pauw, 2008:415)

## 2.7 Zero adclausal nominalization

Sinhala and Tamil differ in the way they encode argument clauses. In Tamil, all clauses which serve as an argument of a verb are necessarily nominalized, usually with *adu*. In the following example, the nominalizer is found between the stem and the accusative case marker. Leaving out the nominalizer would lead to ungrammaticality.

- (33) Kumaar [oru tiruan va-nt-at-ai-p] paar-tt-aan    
 Kumar a thief come-PAST-NMLZR-ACC see-PAST-3SM  
 ‘Kumar saw that a thief had come’ (Lehmann, 1989:302)

In Sinhala, the presence of an overt nominalizer is not obligatory (but possible Garusinghe, 1962).

- (34) mama Gunapaala ya-navaa-Ø=[a kəmatii SINHALA  
 1s Gunapala go-NONPAST(-NMLZR)=DAT like  
 ‘I like for Gunapala to go there’ (Gair, 2003:796)

In this example, the case marker attaches directly to the finite verb form, leading to the curious situation that the same lexeme receives verbal marking for non-past and nominal marking for dative, without any kind of mediating derivation (Gair, 1991). The same is also what we find in SLM. There, finite verbs like *asə* *dhaathang* can immediately be marked for case without further measures being taken.

- (35) Suda buthul suuka nyaari siini su-dhaathang=nang.  
 thus correct like today here past-come= DAT  
 ‘So I very much liked that you came here today.’ (G051222nar01)

Malay varieties in general do not have case marking, so that this feature cannot directly be traced to them. However, they allow the use of prepositions with argument clauses without overt nominalization. Example (36) shows the preposition *untuk*, ‘for’ governing an NP, while in (40), the very same preposition governs a clause, which shows no signs of nominalization (see Gil (forthcoming: (11, 12)) for a similar argumentation). What is more, the non-nominal status of the root *libat*, ‘involve’ is highlighted by the prefix *me-* and the applicative *-kan*, which are two of the most salient morphological signs for the verbal nature of the predicate discussed in the literature. Furthermore, *melibatkan* takes the verbal negator *tidak* rather than the nominal negator *bukan*. Note that while *for* is used in both English translations, in (37) the verb is derived with *-ing* in order to allow the clause to be treated as an NP in this context.

- (36) Yang seribu kan untuk [ongkos becak] COLLOQUIAL INDONESIAN  
 REL 1.000 PRT for fare pedicab  
 ‘The 1,000 (rupees) is for pedicab fare’ (Englebretson, 2003:142)
- (37) kecenderungan untuk [tidak me-libat-kan emosi] COLLOQUIAL INDONESIAN  
 tendency for NEG AT-involve-APP emotion  
 ‘a tendency for not involving their emotions’ (Englebretson, 2003:142)

Zero adclausal nominalization is thus again a feature that sets Sinhala apart from Tamil, but its presence in SLM might not be entirely due to Sinhala influence since seeds were already available in the Malay immigrants’ languages. It is possible to combine a preposition with a main clause in the same way that SLM combines a main clause with a postposition. However, the typologically very unusual combination of bound verbal morphology with bound nominal morphology, which we find in both Sinhala and SLM, suggests that these two languages have become to resemble each other more closely. In Malay varieties which permit argument clauses without nominalization, we do not find bound morphology so that this case is slightly different.

## 2.8 Summary

In this short survey of grammatical areas where Sinhala and Tamil differ, we have found two recurrent patterns: There are a couple of features, like accusative and instrumental subjects, the extension of the present tense form

**Table 9** Summary of features shared between SLM, Sinhala, and Tamil.

	Sinhala	Tamil	inherited?
number	–	(+)	+
accusative	(–)	+	–
indefiniteness	+	–	–
retroflex stops	+	–	(+)
# stop series	+	–	–
gemination	+	–	–
distribution of stops	+	–	–
existentials	+	–	(+)
relative and accusative subjects	+	–	–
present/future/past participle	+	–	–
zero adclausal nominalization	+	–	(+)
loan words	–	+	–

to subordinate clauses in the past and the distribution of consonants which are strong indicators of Sinhala influence. A number of other features, like prenasalized stops or adclausal nominalization function like in Sinhala, but an origin in Malay varieties cannot be excluded. In these cases, the grammatical patterns of some varieties of Malay were probably reinforced, spread and consolidated through contact with Sinhala, where the same structures were found. This process of ‘substrate reinforcement’ has also been found at work in other contact settings (Siegel, 1998, 1999, 2000, 2008). As for influence from Tamil, which is only peripherally relevant for this paper, we can note accusative marking as shown by Smith. A further candidate might be the use of the irrealis marker *anthi*, which is close to the Tamil future form, but not to the impoverished Sinhala future. However, the conflation of future, irrealis and habitual, as found in SLM and Tamil is typologically quite common (Givón, 1994), so that we cannot base a lot of conclusions on that. The lack of grammatical influence from Tamil does not invalidate the lexical influence from Tamil, which is by orders of magnitude greater than lexical influence from Sinhala. Table 9 summarizes the findings.

### 3 Discussion

I have shown above that we find clear or likely Sinhala influence in phonology, morphology, syntax, and semantics. The question which arises is at what

period of time these parts of SLM grammar changed towards Sinhala. Related questions concern the socio-historical setting at the relevant period and the processes of language change we find. As far as the common history of Sinhala and SLM are concerned, two periods have to be distinguished: the colonial period, when Sinhala was a majority language, but the official languages were European, and the post-colonial period, starting in the middle of the 20th century, when nationalist language policies favoured Sinhala in public life and increased the impact of this language on all other languages of the island (Nissan & Stirrat, 1990, Saldin, 2003). We can distinguish Early Sinhala Influence (EaSI) during colonial times from contemporary Late Sinhala Influence (LaSI). What is there to say about the relative importance of the two? There is no denying that LaSI has had an enormous impact on the linguistic ecology of the island. Some, mainly syntactic, features can without problems be explained by Late Sinhala Influence. For some others, especially for those which rely on substrate reinforcement, Early Sinhala Influence is more likely. The reason for this is that at that earlier point in time, the linguistic background of the Malays on Sri Lanka was still diverse. The language would have consolidated and stabilized well before independence, ~~the language would have consolidated and stabilized~~, so that the two-pronged combination of Malay+Sinhala features would no longer be possible. This will be discussed in more detail below.

Turning to the socio-historical processes which led to the change we observe, we can exclude intermarriage for the Sinhala case. While there were some cases of intermarriage between Sinhalese and Malays even in the early days, the numbers are simply too small. An alternative explanation at hand is metatypy. In the following section, I will sketch the Sri Lankan linguistic ecology and examine parallels with other settings where we find metatypy.

#### 4 Sociohistorical setting

Sinhala is the main language of the Southern regions of Sri Lanka, while (“Jaffna”)<sup>20</sup> Tamil is the language of wider communication in the North. Moorish Tamil is an important language in the East and for trading. As a trade language, it is overrepresented in urban areas. The following two graphics

<sup>20</sup> This designation is used in Sri Lanka to distinguish Hindu-Christian varieties from both Moorish Tamil and Indian Tamil. The terms “Sri Lankan Tamil” or “Ceylon Tamil” are not used and would be ambiguous. “Jaffna Tamil” is used regardless of whether the speaker actually comes from Jaffna.



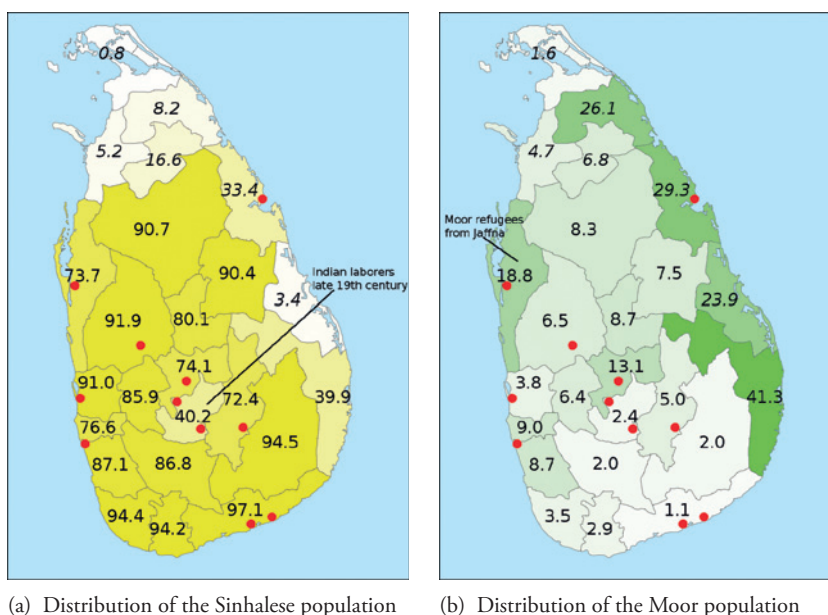
show the distribution of languages in the island for the 2001 census.<sup>21</sup> Figure 1 shows that the settlements with a Malay minority are exclusively found in the Sinhalese area. None are found in the “Jaffna Tamil” area, nor are there any in the Moor-dominated East.<sup>22</sup> In the relevant areas, the historical Sinhalese population ranges between 70% and 97%. Even if we allow for a somewhat lower proportion in the urban centers (and a higher proportion of Moors), it is clear that Sinhala as a language of wider communication cannot be dispensed with. It is hard to imagine that the Malays would not have had some form of competence in a language spoken by three quarters of the people on the street. The kind of influence discussed in the beginning of this paper is further evidence for an early ac[quaintance] with the Sinhalese language.

In the linguistic ecology of 19th century Sri Lanka, Sinhala (and different forms of Tamil) played the role of language of wider communication. Malay (and other minority languages like Sri Lanka Portuguese (Smith, 1979) or Malayalam (Senaratne, 2005)) were in-group languages. Settings of this kind have been investigated in other areas of the world and show recurring patterns with regard to the types of language change we find.

One of the earliest cases described in such a setting is Cappadocian Greek (Dawkins, 1916). The Greeks having settled down in Cappadocia in the 5th century found themselves surrounded by Turks after the Seljuk invasion in the 11th century, who spoke a language which differed widely from theirs in phonology, morphology and syntax. Being enormously outnumbered, the Greeks had to become fluent in this language of wider communication. This fluency in two languages did not fail to have an impact on their home tongue: Cappadocian Greek ~~has~~ changed its phonology (final devoicing, uvular stops), morphology (agglutinative structure, passive formation, pluperfect formation), syntax (left-branching word order), and semantics (accusative only for definite referents, Turkish idioms), all on the basis of a Turkish model.

<sup>21</sup> In order to apply this to the 19th century, two recent population movements have to be factored in: the massive influx of Indian labourers to the central tea plantations in the late 19th century, who brought yet another dialect of Tamil to the island. This led to the dilution of the former Sinhalese dominance in this area (probably around 75% in the 19th century). The second factor is the influx of Moor refugees expelled from Jaffna to the Negombo area in the West by the LTTE, which now suggests a higher Moor proportion in this area than what had been there in the 19th century. The numbers of 18.8% for the Negombo area and 1.6% for the Jaffna area are thus not representative of the historical facts but represent a skewed picture with regard to the 19th century.

<sup>22</sup> Romola Rassool informs me (p.c. 2010) that she found a Malay-speaking community in Trincomalee.



**Figure 1:** Demographics of Sri Lanka. Dots mark main towns with Malay population. The Malay population centers are all in the Sinhala speaking area. Recent shifts in population leading to a skewed picture with regard to the early 19th century are indicated. (Based on census data. Graphics modified from [http://commons.wikimedia.org/wiki/File:Sri\\_Lanka\\_Moor.svg](http://commons.wikimedia.org/wiki/File:Sri_Lanka_Moor.svg), [http://commons.wikimedia.org/wiki/File:Sri\\_Lanka\\_Sinhalese.svg](http://commons.wikimedia.org/wiki/File:Sri_Lanka_Sinhalese.svg))

Both files under CC-BY-SA license (<http://creativecommons.org/licenses/by-sa/3.0/deed.en>) Original creator <http://commons.wikimedia.org/wiki/User:Sadalmelik> ~~(a) Distribution of the Sinhalese population (b) Distribution of the Moor population.~~

This led Dawkins (1916:198) to the formulation that “the body has remained Greek, but the soul has become Turkish.” Crucially, “[t]here is no indication that Turkish speakers have shifted in any significant number to Greek” (Thomason & Kaufman, 1988:251f). This is identical to Sri Lanka, where there is no reason to assume that any significant number of Sinhalese acquired Malay.<sup>23</sup>

<sup>23</sup> The issue is more complicated for the Moors, but this group was not material in bringing about Sinhala influence.

The linguistic changes we find in Cappadocian Greek are similar in kind and extent to what we find in Sri Lanka Malay. There is not a complete adaptation of the larger language's grammar, but an impressive restructuring of all linguistic subsystems to align more closely with the language of wider communication. The social setting is comparable as well: The Greeks formed a close-knit community differentiated by the surrounding Turks by language, religion, and culture. However, as time progressed, acculturation took place and other parts of culture also became more Turkish, similar to what happened in language. This is also what we find in Sri Lanka, where dress and food of the Malays comply with the Sri Lankan model rather than with the Indonesian one. A difference is that religion is being maintained in Sri Lanka, while some Islamization of Cappadocian Greeks took place (Thomason & Kaufman, 1988:215).

The Cappadocian Greek setting is similar to Sri Lanka Malay in another respect: both communities are conscious that they are in the diaspora, and that somewhere out there, there is the metropolis where the homeland of the language lies. The Cappadocian Greek setting is well described and has received some discussion in the literature on language contact since Thomason & Kaufman (1988), but the precise processes leading to the particular structure are not explicated in detail.

This is different in the next language contact situation I want to discuss. On Karkar Island off the north coast of New Guinea, the language of the Takia people has changed its originally Oceanic grammar to comply with the surrounding Papuan languages, most notably Waskia. This has been described and analyzed extensively by Malcolm Ross under the rubric of metatypy (Ross, 1996, 1997, 2001, 2007). Takia adapted word order, semantics of adpositions, and lexical structure of compounds from Waskia. Ross argues that this can only happen when an internally tightly-knit group uses one or more secondary lects/languages to communicate with other groups they are in contact with. Their primary lect/language, however, is the in-group lect/language. Ross (2003:189) offers the following psychological explanation for this, based on principles of economy:

[L]exical calquing a[n]d metatypy are driven by a natural tendency to relieve the bilingual speakers' mental burden by expressing meanings in matching ways in both the primary and the secondary lect. The secondary lect wins out because it is the language of the larger community and is slower to change. Grammatical features of the primary lect are allowed to change because it is usually a lects' lexicon (and sometimes also its intonation) that is emblematic of its speakers' identity, not its syntax, so syntax may change without loss of emblematicity. Change begins at the level of the clause because the clause is, roughly, the unit which expresses the speech act, and speakers achieve their first reduction in burden by

achieving speech-act equivalence between the two lects. Each set of changes at the next lower syntactic rank represents a further reduction in burden.

Ross coined the terms ‘exoteric’ for the language of wider communication and ‘esoteric’ for the in-group language. Under metatypy, the esoteric language takes over the grammatical structures of the exoteric language due to cognitive pressure but resists language shift due to the emblematic function of the esoteric language for the group identity of its speakers.

It seems fair to say that the Malay language in Sri Lanka was an esoteric language. It was confined to Malay speaking homes and annex domain-like barracks of the Malay regiment. It was not used as a medium of inter-ethnic communication. This role was reserved for Sinhala (and Tamil, to which I will return below). However, their origin in the Indonesian archipelago was an important part of Malay identity. This can be seen from the fact that even today most Malays trace their ancestry to exiled Javanese princes, a link which is more often than not quite tenuous. The geographical origin of family names is also preserved and a topic of discussion. The in-group language thus served as a marker of identity, which could not be given up (cf. Rassool, 2010). However, the grammatical structure was slowly but surely invaded by Lankan structures, which encroached on the language in the same way Turkish encroached on Greek, and Takia, on Waskia.

This type of language change is also attested in South Asia: Nadkarni (1975) describes how the in-group language Karnataka Sarasawat Konkani changed its structure to comply with the structure of the language of wider communication, Kannada. Nadkarni analyzes this as a reduction of cognitive pressure: using the same constructions in both languages lessens the cognitive load speakers have to support.

As far as Malay varieties are concerned, we find a comparable case by the Nonthaburi Malays in Thailand (Tadmor 1992, 1995, 2004). The Nonthaburi Malays are descendants of war captives from the 18th century or earlier. They live in a Thai-speaking environment, but maintain the Malay language and Islamic faith. The parallels with Sri Lanka are obvious: the time scale is two to three centuries; the reason for the displacement is warfare; the speakers are isolated and remote from the homeland; they are in a country where their religion is a minority (Tadmor, 2004:527); the differences in size between the in-group languages and the language(s) of wider communication are several orders of magnitude. When we compare the linguistic outcomes, the number of parallels grows even further:

- Nonthaburi augmented the two Malay stop series by an additional series of aspirated stops under the influence of Thai; SLM added a series of prenasalized stops. Both features are typologically marked.

- Nonthaburi developed thriving verb serialization, adopting the semantic and discourse patterns of Thai; Sri Lanka Malay adopted conjunctive particles and clause chains formed with them.
- Nonthaburi changed the *kena*-passive to align with a Thai model; SLM split the *kena*-passive in *kàrà-*, influenced by Sinhala semantics (Nordhoff, 2009:407), and *kinna*, without Sinhala semantics (Nordhoff, 2009:187ff).

Tadmor (1995:322–323) summarizes the linguistic facts as:

Thai interference in Malay is very strong in grammar and semantics, strong in the phonology, and weak in the lexicon. Malay interference in Thai is nonexistent in the grammar and semantics, very weak (if at all) in the phonology, and moderate in the lexicon. In more general terms, Thai interference in Nonthaburi Malay is mostly structural, while Malay interference in Nonthaburi-Malay Thai is mostly lexical. (Tadmor, 1995:322–323)

In the context of this paper, we can replace Thai with Sinhala (or Tamil) and Nonthaburi with SLM and the exact same paragraph summarizes the present state of research.

According to Tadmor, the purely linguistic facts point to speakers having shifted from Thai to Malay. However, it is absolutely clear from socio-historical facts that that is not what happened. The direction of shift is rather the reverse one: Malays shift to Thai. The same is true for Sinhalese, who never shifted to Malay.<sup>24</sup> Rather, it is the Malays who become dominant in Sinhala. Tadmor goes on to argue that it is the language which is primary in every day use (which is not necessarily the maintained language, and not necessarily native) which exerts structural influence.

Structural interference is usually automatic, and associated with **primary** ☐ **secondary** language interference, while lexical interference may be controlled, and may therefore be associated with interference in either direction. Unlike what has been claimed in previous studies, it makes no difference if the direction of the interference is from the native (or original, or maintained) language to the second language (or new language, or language shifted to), or vice versa. What matters is which of the two languages, at the time the interference took place, was the primary language of the speaker of the community, and which was the secondary language. (Tadmor, 1995:325)

This distinction between primary/wide/exoteric and secondary/narrow/esoteric language can then explain why we find structural patterns from the adstrates in Nonthaburi Malay and Sri Lanka Malay without the slightest sociohistorical evidence of language shift from Thai to Malay, or from Sinhala to Malay.

<sup>24</sup> As noted above, the case is slightly more complicated for the Moors.

It appears that the interplay between languages of wider communication (exoteric) and in-group languages (esoteric), which is found in many other parts of the globe, can account for the Sinhalese influence in Sri Lanka Malay. Given the repeated exposure to certain morphosyntactic constructions and certain phonetic data, and the active production of those by the Malays, these forms became entrenched and eventually made it into the native grammars, where they managed to supplant the traditional forms, very similar to what happened in Cappadocian Greek, Takia, Konkani, or Nonthaburi.

The evidence presented above is difficult to reconcile with the view that Sinhala was not an important contact language. One could of course argue that this is all due to language change after 1956, when nationalist language policies favouring the use of Sinhala were enacted. There are two points which argue against this: The older generation born in the 1930s show the discussed features of Sinhala influence. These structures were already present when they acquired SLM, which was before 1956.

The second point is of phonological nature: the development of phonemic prenasalized stops and phonemic gemination crucially relies on the presence of subphonemic variation in some Malay dialects spoken by the immigrants. Some dialects syllabified clusters of nasal+consonant as .NC, others as N.C. Widely differing grammars are common among immigrant communities, but after some generations, the differences level out (Trudgill, 1986, Mesthrie, 1993). Given that outside of Sri Lanka, no Malay dialect has both N.C and .NC, we expect that Sri Lanka Malay would have chosen one of the two possibilities when the language stabilized,<sup>25</sup> all other things being equal. But this is not what happened, Sri Lanka Malay has a phonemic distinction. This distinction was clearly facilitated by Sinhala, which also makes this distinction. We thus need dialectal variation in Malay varieties as well as Sinhala influence at the same time. Since dialectal variation in SLM was levelled out by 1800, we have to posit that Sinhala had exerted influence before that date.

The argumentation for geminates is analogous: Voiced geminates are found in Sinhala, but not in Tamil. They occur subphonemically in Malay varieties, but no variety has a phonemic contrast. The most logical explanation for the phonemicization of SLM geminates is adstrate influence, forcing the interdialectal variation to phonemicize. This adstrate influence must have been present before the levelling was completed, i.e. before 1800, and it must have come from Sinhala, since Tamil does not have voiced geminates, unlike Sinhala and SLM.

<sup>25</sup> The precise date of the stabilization of the language is uncertain, but no later than 1800 is a date commonly assumed.

In this paper, I have surveyed a number of areas where Sinhala contact is sure, or at least very likely. The domains surveyed in this paper suggest that there is actually more influence from Sinhala than from Tamil, but this is due to the topic of this paper being Sinhala influence. Argumentations based on the number of features are of course a difficult enterprise, since it is too easy to stack the deck to come up with any configuration of numbers desirable to defend the thesis. I will therefore not use the data presented in this paper to make any quantitative claims. My claim is rather of a qualitative nature: there has been Sinhala influence in Sri Lanka Malay, and the kind of influence we find suggests that this influence is not a product of the twentieth century alone. This does not conflict with the view that Tamil also had an influence. This is most clearly seen in the lexicon, where Tamil loanwords for basic vocabulary terms like kin or household items are common, but Sinhala loans are all but absent.

How does this relate to the different theories of genesis proposed? Smith et al. (2004, 2006) propose that Sri Lanka Malay is the result of imperfect acquisition of Trade Malay by Moor wives. I have argued in Nordhoff (2009) that the proportion of Moor wives was around 20%, i.e. 10% of the total population. In my view, this is not sufficient to trigger the kinds of changes we find against the 90% of the population whose native language was not Tamil. Regardless of one's assessment of the marriage patterns, it is clear that local wives are not responsible for the Sinhala influence we find: there were close to no Sinhala wives. We must assume that Sri Lanka Malays had a good command of Sinhala through contact with the local population even in the early days, enough command to allow Sinhala phonological features to penetrate their language. Next to their knowledge of Sinhala, they were of course also fluent in Tamil, as attested by the loanwords. We must therefore assume a trilingual setting (cf. Ansaldo, 2008).

## 6 Conclusion

In the above section, I have sketched a scenario how Sinhala structures could have entered the Malay language in Sri Lanka. As for Tamil structures, the scenario most often assumed, namely shift from Moorish Tamil to Malay (Smith et al., 2004, Smith & Paauw, 2006, Slomanson, 2006), is different in nature, but led to remarkably similar results. This has led researchers like Bakker (2006), Ansaldo (2008) or Nordhoff (2009) to suppose that the processes of Tamil influence and Sinhala influence were indeed not that different. If one assumes that Tamil was a language of wider communication with social



functions similar to those of Sinhala, the metatypic explanation outlined above becomes available as well (Bakker, 2006, Ansaldo, 2008). More and deeper research into the demographics and social fabric of Ceylonese society of the 19th century will allow to ascertain whether the domains of Tamil use and the domains of the use of Sinhala suggest positing different processes for the respective languages or not. What I hope to have accomplished in this paper is a more modest goal: showing that Sinhala influence was present in Sri Lanka Malay from the early days onwards and outlining the social processes explaining this influence. This will serve as a base for future investigations of the matter.

## References

- Adelaar, Sander. 1991. "Some notes on the origin of Sri Lankan Malay". In Steinhauer 1991, pp. 23–37.
- Ansaldo, Umberto. 2008. "Sri Lanka Malay revisited: Genesis and classification". In Dwyer, A., D. Harrison and David Rood (eds.) "A world of many voices: Lessons from documented endangered languages", Studies in Language, Amsterdam, Philadelphia: John Benjamins. pp. 13–42.
- Ansaldo, Umberto. 2009. *Contact languages, Ecology and Evolution in Asia*. Cambridge: Cambridge University Press.
- Asher, Ronald. 1985. *Tamil*. Beckenham: Croom Helm.
- Bakker, Peter. 2006. "The Sri Lanka *Sprachbund*: The newcomers Portuguese and Malay". In Yaron, Matras, April McMahon and Nigel Vincent (eds.) "Linguistic Areas – Convergence in Historical and Typological Perspective", Houndmills, New York: Palgrave Macmillan. pp. 135–159.
- Chung, Siaw-Fong. 2005. "Kena as a third type of Malay passive". *Oceanic Linguistics*, 44(1):194–214.
- Dawkins, RM. 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.
- Deumert, Ana and Stefanie Durlmann (eds.) 2006. *Structure and Variation in Language Contact*. Amsterdam: John Benjamins.
- Downing, Laura. 2005. "On the ambiguous segmental status of nasals in homorganic NC sequences". In van Oostendorp, M. and Jeroen van de Weijer (eds.) "The Internal Organization of Phonological Segments", Berlin: Mouton de Gruyter. pp. 183–216.
- Durie, Mark and Malcolm Ross (eds.) 1996. *The Comparative Method Reviewed: Regularity and Irregularity in Language Change*. Oxford: OUP.
- Emerson Tennent, James. 1859. London.
- Englebretson, Robert. 2003. *Searching for Structure: The problem of complementation in colloquial Indonesian conversation*. Amsterdam: John Benjamins.
- Fritz, Sonja. 2002. *The Dhivehi Language. A Descriptive and Historical Grammar of Maldivian and Its Dialects*. Würzburg: Ergon Verlag.
- Gair, James W. 1971. "Action Involvement Categories in Colloquial Sinhala". In Zamorra, M., J. M. Mahar and H. Orenstein (eds.) "Themes in Culture: Essays in Honor of



- Morris E. Opler", Manila: Kayumanggi Publishers. pp. 238–256. Reprinted in Gair 1998, pp. 25–46.
- Gair, James W. 1976. "Is Sinhala a Subject Language?" In Verma 1976, pp. 39–63.
- Gair, James W. 1991. "Subjects, Case and INFL in Sinhala". In Manindra Verma and Karuvannur Mohanan (eds.) "Experiencer Subjects in South Asian Languages", Palo Alto: Center for the Study of Language and Information. pp. 13–41. Reprinted in Gair 1998, pp. 65–86.
- 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. 1998. *Sinhala and Other South East Asian Languages*. New York, Oxford: OUP.
- Gair, James W. 2003. "Sinhala". In Cardona, George and Dhanesh Jain (eds.) "The Indo-Aryan Languages", London: Routledge. pp. 766–818.
- Gair, James W. forthcoming. Sri Lankan Languages in the South-South Asia Linguistic Area: Sinhala and SL Malay. In Nordhoff, S. and Paauw, S. The Genesis of Sri Lanka Malay. Leiden: Brill.
- Gair, James W. and John C. Paolillo 1997. *Sinhala*. München: LINCOM.
- Garusinghe, D. 1962. *Sinhala, the Spoken Idiom*. München: Max Hueber.
- Geiger, Wilhelm. 1938. *A Grammar of the Sinhalese Language*. Colombo: Royal Asiatic Society, Ceylon Branch.
- Gil, David. forthcoming. "Riau Indonesian: A Language without Nouns and Verbs". In van Lier, Eva and Jan Rijkhoff (ed.) "Flexible word classes: a typological study of underspecified parts-of-speech", Oxford: OUP.
- Givón, Tom. 1994. "Irrealis and the subjunctive". *Studies in Language*, 18(2):265–338.
- Heine, Bernd and Tania Kuteva 2002. *World lexicon of grammaticalization*. Cambridge: CUP.
- Hopper, Paul and Sandra A. Thompson 1980. "Transitivity in Grammar and Discourse". *Language*, 56:251–99.
- Karunatilake, W. S. 2004. *An Introduction to Spoken Sinhala* Colombo: Gunasena, 3rd edn.
- Lassen, C. 1847. *Indische Alterthumskunde*. Bonn: Kitzler/Williams and Norgate.
- Lehmann, Thomas. 1989. *A Grammar of Modern Tamil*. Pondicherry: Pondicherry Institute of Linguistics and Culture.
- Maddieson, Ian and Peter Ladefoged 1993. "Phonetics of partially nasal consonants". In Huffman, Marie and Rena Krakow (eds.) "Nasals, Nasalization and the Velum", San Diego: Academic Press. pp. 251–301.
- Matzel, Klaus. 1983. *Einführung in die singhalesische Sprache*. Wiesbaden: Harrassowitz, 2nd edn.
- Mesthrie, R. 1993. "Koineization in the Bhojpuri-Hindi diaspora – with special reference to South Africa". *International Journal of the Sociology of Language*, 99(1):25–44.
- Michaelis, Susanne. (ed.) 2008. *Roots of creole structures: Weighing the contribution of substrates and superstrates*. Amsterdam: John Benjamins.
- Nadkarni, Mangesh V. 1975. "Bilingualism and syntactic change in Konkani." *Language*, 51:672–83.
- Næss, Å. 2007. *Prototypical Transitivity*. Philadelphia: Benjamins.
- Nissan, Elizabeth and RL Stirrat 1990. "The generation of communal identities". In Spencer 1990, pp. 19–44.
- Nitz, Eike. and Sebastian Nordhoff 2010. "Subtractive Plural Morphology in Sinhala". In Wohlgemuth, Jan and Michael Cysouw (eds.) "Rara & Rarissima. Collecting and interpreting unusual characteristics of human languages", Berlin: Mouton de Gruyter.
- Nordhoff, Sebastian. 2009. *A Grammar of Upcountry Sri Lanka Malay*. Ph.D. thesis, University of Amsterdam.
- Onn, Farid M. 1980. *Aspects of Malay Phonology and Morphology: A Generative Approach*. Ph.D. thesis, Universiti Kebangsaan Malaysia.

- Paauw, Scott. H. 2004. *A Historical Analysis of the Lexical Sources of Sri Lanka Malay*. Master's thesis, York University.
- Paauw, S. H. 2008. *The Malay contact varieties of Eastern Indonesia*. Ph.D. thesis, State University of New York at Buffalo.
- Rask, Rasmus. 1821. *Singalesisk Skrifflære* Kolombo.
- Rassool, R. 2010. "Studying contemporary usage of and attitudes to the SLM language: a preliminary discussion". Paper presented at the workshop 'Past, Present, and Future of Sri Lanka Malay'.
- Ross, Malcolm. 1996. "Contact-induced change and the comparative method: cases from Papua New Guinea". In Durie Ross 1996, pp. 180–217.
- Ross, Malcolm. 2007. "Calquing and metatypy". *Journal of Language Contact*, 1:117–143.
- Ross, Malcolm. 1997. "Social networks and kinds of speech-community event". In Blench R. M. Spriggs (eds.) "Archaeology and language 1: Theoretical and methodological orientations", London: Routledge. pp. 209–261.
- Ross, Malcolm. 2001. "Contact-induced change in Oceanic languages in North-West Melanesia". In Aikhenvald, Alexandra Y. RMW. Dixon (eds.) "Areal diffusion and genetic inheritance", Oxford: Oxford University Press. pp. 134–166.
- Ross, Malcolm. 2003. "Diagnosing prehistoric language contact". In Hickey, Raymond (ed.). "Motivations for language change", Cambridge: Cambridge University Press. pp. 174–198.
- Saldin, BDK. 2003. *Portrait of a Sri Lankan Malay*. Kurunegala: Nihon.
- Senaratne, Jagath. 2005. "The Malayalees – their origins, growth and decline". In Vijayalakshmi, E. (ed.) "Cultural minorities of Sri Lanka – Their growth, achievements & relevance today", Colombo: International Centre for Ethnic Studies. pp. 80–107.
- Siegel, Jeff. 1998. "Substrate reinforcement and dialectal differences in Melanesian Pidgin". *Journal of Sociolinguistics*, 2/3:347–373.
- Siegel, Jeff. 1999. "Transfer constraints and substrate influence in Melanesian Pidgin". *Journal of Pidgin and Creole Languages*, 14:1–44.
- Siegel, Jeff. 2000. "Substrate Influence in Hawai'i Creole English". *Language in Society*, 29:197–236.
- Siegel, Jeff. 2008. "In praise of the cafeteria principle: Language mixing in Hawai'i Creole". In Michaelis 2008, pp. 59–82.
- Slomanson, Peter. 2006. "Sri Lankan Malay morphosyntax: Lankan or Malay?" In Deumert & Durrleman 2006, pp. 135–158.
- Smith, Ian. 1979. "Convergence in Sri Lanka: A Creole Example". *Lingua*, 48:193–222.
- Smith, Ian. 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 and Scott Paauw 2006. "Sri Lanka Malay: Creole or convert". In Deumert & Durrleman 2006, pp. 159–182.
- Smith, Ian, Scott Paauw & BA Hussainmiya 2004. "Sri Lanka Malay: the state of the art." *Yearbook of South Asian Languages*, 2004:197–215.
- Sneddon, James Neil. 2006. *Colloquial Jakarta Indonesian*. Canberra: Australian National University.
- Spencer, Jonathan (ed.). 1990. *Sri Lanka – History and Roots of the Conflict*. London: Routledge.
- Steinhauer, Hein (ed.). 1991. *Papers in Austronesian linguistics, Pacific Linguistics*, vol. 1. Canberra: The Australian National University.
- Stoel, Ruben. 2005. *Focus in Manado Malay: Grammar, particles, and Intonation*. Leiden: CNWS Publications.

- Tadmor, Uri. 1992. "The Malay Dialects of Central Thailand: a Preliminary Survey". In "The Third International Symposium on Language and Linguistics, Bangkok, Thailand, pp. . Chulalongkorn University. cite.", Bangkok: Chulalongkorn University, pp. 1367–1378.
- Tadmor, Uri. 1995. *Language contact and systemic restructuring: The Malay of Nonthaburi, Central Thailand*. Ph.D. thesis, University of Hawai'i.
- Tadmor, Uri. 2004. "Dialect endangerment: The case of Nonthaburi Malay". *Bijdragen tot de Taal-, Land- en Volkenkunde*, 160(4):511–531.
- Thomason, Sarah and Terrence Kaufman 1988. *Language Contact, Creolization and Genetic Linguistics*. Berkeley: University of California Press.
- Trudgill, Peter. 1986. *Dialects in contact*. Oxford: Blackwell.
- van Minde, Don. 1997. *Malayu Ambong: phonology, morphology, syntax*. Leiden: Research School CNWS, School of Asian, African, and Amerindian Studies.
- Verma, Manindra (ed.). 1976. *The Subject in South Asian Languages*. Madison: University of Wisconsin South Asia Program.
- Wahab, Kartini A. & Hiroki Nomoto 2010. "Kena passives and unvoiced voice alternations". Paper presented at ISMIL XIV at the University of Minnesota.