

## **The Transmission of Substrate Features: The case of Malacca Portuguese**

**Robert Laub**

602735@soas.ac.uk

SOAS University of London, UK

### ***Abstract***

Theories of the formation of creole syntax have been proposed to explain whether substrates and superstrates influence the resultant creole structures, and if so, what the mechanisms are by which they influence them. Using the framework of feature transfer, this study investigates features in Malacca Portuguese (MP), a Portuguese-based creole spoken in Malaysia. Citing grammatical characteristics such as aspect particles, I investigate similarities between features in MP and Malay, and how these similarities could have transferred from Malay to a Portuguese lexical item to create a grammatical words in MP. . I also discuss short/long variation in personal pronouns, as well as other features such as TAM markers and auxiliaries, and investigate how they have come about from a partial feature transfer from Malay. I propose that in this particular case a substrate has affected creole syntax, and provide evidence of how this has occurred. While a much more substantial exploration of this phenomenon is essential, I put forth a hypothesis about how the short/long distinction in pronouns functions in MP.

Keywords: creole, feature transfer, feature pool, substrate, Malacca Portuguese

## **1. Introduction**

Malacca Portuguese (MP) is a Portuguese-based creole spoken in Malacca, Malaysia<sup>1</sup>. It belongs to a Southeast Asian subgroup of creoles, including Makista from Macau, China (Pinharanda-Nunes, 2008), and Batavia and Tugu Creole Portuguese, formerly spoken in Jakarta, Indonesia (Maurer, 2011). All of these are related to Portuguese-based creoles spoken in South Asia (Cardoso, 2009; Clements, 1996). Research on MP provides an opportunity to investigate creole structures, especially those of “fort-type” creoles (Holm, 2000, p. 41), which emerged out of fortified European ports where the economy was heavily based on commerce and trade.

To better understand the elements that led to MP, I begin this paper with an outline of the historical background of Malacca in Section 2. This city has a long history of overseas trade. Although a majority Malay-speaking city, it contains communities that evolved out of the history of the region, including the Portuguese period from which MP emerged. Section 3 outlines the theoretical framework and the points of view within creolistics that have informed this study. There is an argument in the literature over whether creoles, no matter where they are spoken and which substrates they have, are a typologically distinct category of human languages (see Bakker et al., 2011; McWhorter, 2001, 2005, 2018; Parkvall, 2008), or whether the particular languages spoken by those who first created the creoles played a role in the development of the resulting grammar (see DeGraff, 2003; Pinharanda-Nunes, 2012; Siegel, 2008). As MP is spoken in Malaysia, Malay has been involved in both the creation of the grammar and further influence on it throughout its history. Section 4 looks at the morphosyntactic ways in which Malay and MP are similar. Section 5 details the methodology as well as the data and findings relevant to the present study. I conducted fieldwork in Malacca in 2017 and looked at the ways short and long pronouns work in MP. The findings from this fieldwork will form part of the main evidence discussed in this chapter, along with analysis of previously recorded data (Baxter, 1988; Pillai, 2011).

Following this section, I will explain how the data relates to the previous sections, especially with regard to theories of substrate influence in the formation of creoles. We will see that although Malay structure does not exert its influence in all facets of MP grammar, it does manifest itself in important ways.

---

<sup>1</sup> Malacca is where MP originated and where most of its speakers currently live. During the colonial period the speakers spread across the Malay Peninsula, and there exist smaller MP-speaking communities elsewhere in Malaysia, and in Singapore.

## **2. Historical Background**

Malacca has a long history of being a hub for multiculturalism and trade. In the first millennium CE, Malacca was a center for trade that attracted sailors from various parts of Asia, including China, India, and Arabia. This resulted in an environment conducive to language contact, and led to the creation of multicultural communities whose descendants exist today. The environment of Malacca was one of intercultural commerce preceding the European colonial period. MP came out of this European period, but it was far from the first language to emerge out of cross-cultural contact in the city (McPherson, 1993).

MP can trace its roots to the Portuguese colonization of Malacca from 1514 to 1641. Portugal overthrew the Sultan of Malacca and assumed power, using the city as a base for its colonial exploits in the region. It was during this time that men from Portugal and other colonies in India mixed with local women, out of which a distinct Eurasian community emerged. The Portuguese empire at the time was mainly based on trade and not settlement, and was “heavily localized, decentralized, and integrated, out of necessity, with local political structures” (Ansaldo, 2009, p. 73). Baxter (1996) presents more information on the sociolinguistic settings of Southeast Asia. Malacca Eurasians’ subsequent culture developed out of a fusion of the cultures of the areas colonized by the Portuguese (the Indian subcontinent and Malay peninsula) with the Portuguese language and culture, including the Catholic faith. While the linguistic influence of Portuguese is no doubt prominent in MP, the language that had arrived in Malacca was an already a pidginized, mixed variety, developed in India and Africa, and far removed from the Continental standard.

Malacca was a heavily multicultural city due to its position as an inter-Asian trading center. Malay dominance throughout Southeast Asia led to the Malay language being a lingua franca in the region, especially in its pidginized version, Bazaar Malay (Adelaar, 1996). Structurally, Bazaar Malay is more isolating than literary Malay, and would have also been used in Malacca during the time the Portuguese had control (Ansaldo, 2009). In addition, Malacca lay near the eastern edge of the Portuguese trading empire, which stretched around the coast of Africa and Asia. Portugal proper, having a small population and unreliable supply of sailors and merchants, was required to engage people from the colonies further west from Malacca, so the language spoken on the European ships that arrived there was likely to have been a pidginized variety of Portuguese (whose creole descendants continue to be spoken in Sri Lanka and Southern India today). Because of this, MP did not emerge out of a strictly Portuguese-Malay blend, but rather from the input of

non-standard versions of both these two languages, including Bazaar Malay (see also Baxter, 1996; Cardoso, Baxter & Nunes, 2012; Clements, 2000), along with influences from contact with other languages in the city (Baxter, 2012).

Malacca's strategic location was alluring to other European powers, and the Dutch took control from 1641 to 1795, and again from 1818 to 1825, after a period of British hegemony from 1795 to 1818. The British displaced the Dutch in 1825 and continued to rule Malacca until 1957, at which time Malaysia became independent. The British periods saw additional migration of Indians and Chinese to British Malaya, many of whom settled in Malacca (Parthesius, 2010). The communities descended from these migrants formed their own distinct identities, which are separate from the Peranakan Eurasians, despite their having a shared Chinese heritage (Ansaldo, Lim, & Mufwene, 2007).

Continuing into the 21<sup>st</sup> century, the influences of migration and colonialism have left their marks in many ways, especially linguistically. Malacca is a multilingual city, and its residents are likely to regularly use more than one language in their daily lives (Kärchner-Ober, 2013). While MP was created from the lexifier pidginized Portuguese and the substrate Malay, the continuing social trends in Malacca's history show that other languages would have come into contact with MP as well. What is notable is that the presence of Portuguese itself diminished after the Dutch takeover, although it was not completely absent, especially through the vehicle of the Catholic church (see Baxter, 2018). The presence of Malay continues to this day (Baxter, 1988; Pillai, Soh, Kajita, 2014).

### **3. Theoretical Framework**

I turn to the theoretical framework that has informed this study, especially the argument that the concept of what makes something a creole is a sociological, rather than structural one, owing to their creation within communities that have emerged out of colonialism. This is especially advocated by DeGraff (2003, 2005, 2009), who argues that creoles are essentially a sociolinguistic grouping, and not a typologically distinct class of languages. DeGraff further states that the notion of creoles as typologically separate is a colonialist viewpoint, whereby the creoles are languages that have been created in the colonies by the Europeans' subjects. This way of thinking, he argues, derives from opinions that creoles are inferior versions of their European lexifiers.

This argument is paralleled in Mufwene (2000), who states that creolization is a social, not structural, process. In opposition to McWhorter (2001), Mufwene points out that creoles are not created in vacuums. The language of the colonizers was not formally taught but acquired through a desire to communicate. The situations in which different creoles emerged led to how they are structured, according to Mufwene (2000). The emergence of MP is, therefore, a very different situation from the emergence of Haitian Creole, which came out of a diverse group of West African slaves forced to live and work together under their French colonial masters.

These accounts by Mufwene (2000) and De Graff (2003) explain why creoles can be grouped separately from other languages, and other studies have investigated how languages that come into contact in such situations can affect the resultant creole structure. Superstratists (Chaudenson, 1979; 1992; Hall, 1966; Valdman 1978) argue that the lexifier (or superstrate) is the most important element in the creation of creole syntax, where non-standard varieties of the lexifier are acquired in successive waves and eventually evolve into a creole (DeGraff, 1999, p. 7). Substratists, meanwhile, view the (non-colonizers) substrate as being the most important element (Koopman, 1986; Lefebvre & Lumsden, 1989). They argue that the lexifier provides the bulk of the vocabulary and phonology, but the overall syntactic structure of the creole resembles more closely that of the substrate. Note that there is also a universalist approach (e.g. Bickerton, 1988), which argues that creole syntax reflects cross-linguistic similarities between all languages (part of ‘universal grammar’ or the ‘bioprogram’). The ways in which superstratism and substratism differ in their approaches to creole origins are small compared to how these two views differ from the perspectives of universalists, in that they acknowledge the input that lexifier and substrate varieties can have in the resultant structure.

From these viewpoints emerged theories explaining the mechanisms by which a creole acquires features, one of which is the Feature Pool (FP) hypothesis. FP states that a language that emerges out of a contact situation can build its structure through the inheritance of various features from the contact varieties (Mufwene, 2002). Many of these features will be replicated as identical to their input, while some will have a resemblance but also some differences. This approach is analogous to a gene pool, where genes are inherited from parents to offspring, with and without mutation.

Siegel (2007, 2008) proposed a constraint of the FP, namely that a feature must have a place for it to transfer to. For example, the language spoken by the Peranakan people of Malaysia,

known as Baba Malay, is a Malay variety that has undergone heavy Chinese influence as a result of the community's Chinese heritage. An example of a Baba Malay possessive construction is the following, compared to standard Malay and Hokkien Chinese taken from Ansaldo et al. (2007, p. 215):

1. Baba      Malay  
   dia      punya bilik  
   3SG      POSS room  
   “His/her room”
2. Malay  
   bilik      dia/nya  
   room      3SG  
   “His/her room”
3. Hokkien  
   i          e          pang-keng  
   3SG      POSS room  
   “His/her room”

Here we see how the Hokkien possessive construction has influenced the structure of Baba Malay. While the Malay possessive is expressed with [possessee+possessor] word order with no particle or overt morphology expressing possession, Hokkien has [possessor+possessee] order, and uses a particle *e*. The Baba Malay linking particle arises from the standard Malay *punya* verb meaning “to own”. We can think of *punya* as the destination feature, and the semantics and function of the Hokkien particle *e* as the source feature. In this case, the transfer from the substrate to the lexifier results in *punya* functioning as a possessive particle in the creole, used syntactically in a similar way to its correspondent in Hokkien (Ansaldo et al., 2007).

Relabeling is another process by which a feature is transferred from a substrate to a creole (Lefebvre 2008). While arguments exist that it is a strictly lexical process (Lumsden, 1999; Muysken, 1988), Lefebvre (1998, 2008) proposes that it can also be functional. Lefebvre (2008, p. 199) describes a process for relabeling in three representations: (a) the lexical entry as it is in the lexifier; (b) the assignment of a second phonological representation to the semantic and syntactic features of the item; and (c) abandonment of the phonology from the substrate, leaving the

phonological shape from the superstrate lexifier but the underlying syntax and semantics of the substrate.

#### 4. Malay Influence on MP

This section outlines some of the features that Malay has in common with MP, and presents features that are more well-documented than those discussed in Section 5 (the short and long pronoun distinction). It illustrates some of the ways that features have been transferred from Malay to MP.

I begin by considering how tense, aspect, and mood (TAM) are encoded as particles in MP. TAM particles are relevant to the FP hypothesis in that they have their origins in Portuguese adverbs and auxiliary verbs that are semantically related to tense and aspect, but do not in themselves convey tense. The MP TAM particles have similarities with pre-verbal aspect markers in Malay, pointing to one of the more obvious substrate-influenced features in MP. Standard Portuguese expresses tense using verbal morphology, in particular suffixes that express tense (as well as person and number), as in examples like the following:

4. Portuguese: (Ganho & McGovern, 2003, p. 178)

Comi.

Eat.1SG.PST

“I ate/I have eaten”

TAM in Malay, in contrast, is expressed through particles which are found in a pre-verbal position, while the verb itself undergoes no inflection. The particles only encode TAM, and not person or number, as in:

5. Malay (Nomoto, 2006, p. 98):

Dia            sudah membaca    buku    itu

3SG        PFV    read            book    that

“She has already read the book”

No tense, aspect or mood is expressed in the word *membaca*, and it is the pre-verbal particle *sudah* which indicates that the reading has completed. MP marks aspect in nearly the same way as Malay, although because of its lexifier, the phonological shapes of the particles have their origins in Portuguese, as in:

6. Malacca Portuguese (Baxter, 1988, p. 212):

Eli        ja        tomah faka    kotrah kandri  
3SG       PFV    take    knife    cut        meat  
“He cut the meat with a knife”

The MP perfective particle *ja* originates from Portuguese *já* “already”. Semantically, the two elements share similar meanings in that they indicate a sense of completedness, but Portuguese *já* is an adverb, while MP *ja* is an aspect particle. Notice that Portuguese *já* appears pre-verbally in a sentence like the following:

7. Portuguese (Ganho & McGovern, 2003, p. 178)

Já            comi  
Already    eat.1SG.PST  
“I already ate/I have already eaten”

However, *já* is not restricted to pre-verbal position in Portuguese, but may be found in other positions in the clause, such as post-verbally (unlike MP *ja*), with a meaning similar to Malay *sudah*. Nonetheless, the pre-verbal position is the least marked, and other positions carry other semantic implications (Martins, 2007). In this sense then, *já* aligns with the same position that a TAM marker is found in Malay. In terms of the framework in Siegel (2008), this provides the destination to which the syntactic and functional features of Malay *sudah* can transfer.

The structural similarities between MP and Malay are striking, but this alone is not sufficient evidence that Malay is the immediate origin of the similar features, especially considering that there are similar structures found in Portuguese creole varieties spoken in India, where the Portuguese presence has a longer history. While Malay is not necessarily the origin of these features, the centuries-long contact MP speakers had with Malay could have reinforced their continued use.

## 5. Methods, Methodology and Data Sources

The data for the present study comes from two main sources. The first comprise recordings made by Stefanie Pillai in the Portuguese Settlement of Malacca in 2011 (Pillai, 2011). These comprise interviews, monologues, and conversations. As far as I am aware, these data were not collected with the focus of the present paper in mind, meaning that the relevant examples are found outside of elicitation, although with the limitation that the number of tokens of them is not large.



The data second set is from my fieldwork in the Portuguese Settlement of Malacca in 2017, mainly from two separate recordings. The first one consists of a monologue by a single speaker. It was largely unstructured, but prompted by asking the consultant to talk about what a tourist should expect when they visit Malacca. The second recording involved four participants, all males living in the Portuguese settlement, with three aged in their sixties, and one in his thirties. This session consisted of elicitation exercises directed by me.

The first data set with one participant included grammaticality judgments, translation, and an unscripted monologue on the topic of what a tourist should do and expect to see during a visit to Malacca. The purpose of this activity was to generate instances of a non-specific, indefinite pronoun (corresponding to English general “you”). The second data set includes input from four participants, with at least two other MP speakers speaking at any given time. There was a marked difference in MP fluency between the three older men and the younger one, with the younger participant sometimes struggling to find the right word and to finish a sentence on his own. The second recording session was built around a guided activity using pictures and questions taken from the University of Potsdam’s *Questionnaire for Information Structure* (Skopeteas et al, 2006). The aim was to elicit examples of focus and contrast and to investigate where the participants would use different types of MP pronouns in various contexts, in particular the long and short forms of personal pronouns (see Section 6).

For the initial investigation into pronoun length variation, I searched for instances of the three pronouns in question (second person singular, third person singular, first person plural) in Pillai’s corpus, which is more naturalistic and less forced, as far as the pronoun variation is concerned. However, there is a limitation in that the actual number of pronoun tokens is too low for conclusive results when analyzed quantitatively. I decided to tally the tokens of short and long pronouns in Pillai’s (2011) data to see if any patterns emerged. Initial observations suggest that subjects take the short form, especially when they come before a negative particle, while in object position a long form is preferred. This will be discussed in depth in the results section.

## **6. Short and Long Pronoun Variation**

Personal pronouns, along with some other lexical items in MP, have at least two variants, one shorter than the other, although according to the existing descriptions it is uncertain what triggers this variation. There is evidence of the same speaker using both forms, and a short and long version

of the same lexical item can occur within the same utterance. In the following sections, I give an account of the MP pronoun inventory and which pronouns may take a short form. This is followed by a brief discussion of short/long variations that occur with other lexical items.

### 6.1. Malacca Portuguese Personal Pronouns

The personal pronoun inventory for MP is shown in Table 1.

Table 1: MP Pronoun Inventory

	Singular	Plural
1st person	yo	nus
2nd person	bos	bolutu
3rd person	eli	ilotu

While these originate from the subject form of the Portuguese pronouns (compare MP 1SG *yo* to Portuguese 1SG.SUBJ *eu*), they may be used in any syntactic function and do not change when they are in object position:

8. Malacca Portuguese (Stoltz, 1987 in Hancock, 1973, p. 26)

Yo        ja        kabah kumih  
 1SG       PFV    PAST eat  
 “I have just finished eating”

9. Malacca Portuguese (Pillai, 2011)

Ja        largah ku    yo    na    boboi  
 PFV       leave ACC 1SG in    cradle  
 “They left me in the cradle”

10. Malacca Portuguese (Laub, fieldwork October 2017)

Bo        ke        dah    ku    yo    prezanti    epal  
 2SG       want give to    1SG present    apple  
 “You want to give me an apple as a present”

These examples illustrate that in MP the personal pronouns do not change form whether they are the subject (as in Sentence 8) or the object (direct object in Sentence 9, indirect object in Sentence 10).<sup>2</sup>

Of the six personal pronouns in MP, four exhibit alternative forms which vary in length. These are: Second Person Singular with variants *bos~bo*; Third Person Singular with variants *eli~el~e*; First Person Plural with variants *nus~nu*; and Third Person Plural with variants *olotu~olotru~oló~ol~otu*. This variation in form does not appear to be related to grammatical function, as both the short form and the long form can be found in both subject and object positions in the data (relevant pronouns are in bold):

11. Malacca Portuguese (Pillai, 2011)

<b>Nus</b>	tokah	bai	buskah	kumih	is	sorti	lah
<b>1PL</b>	touch	go	find	eat	this	type	PRT

“We have to go find a living like this”

12. Malacca Portuguese (Pillai, 2011)

Singku	sentu	anu	ki	<b>nu</b>	ja	fikah	na	isti	tera
Five hundred	year	that	<b>1PL</b>	PFV	stay	in	this	land	

“Five hundred years that we have stayed in this country”

13. Malacca Portuguese (Laub, fieldwork 2017):

Yo	dah	lus	ku	<b>bos</b>
1SG	give	light	ACC	<b>2SG</b>

“I give a light to you”

14. Malacca Portuguese (Laub, fieldwork 2017):

Logu	nganah	ku	<b>bo</b>
FUT	deceive	ACC	<b>2SG</b>

“They will deceive you”

The above examples illustrate, respectively, the long form subject, short form subject, long form object, and short form object. Grammatical function is not expressed by the shape of the pronoun in MP, which is an SVO language and expresses functions using either word order or the object marker *ku*. The examples above unambiguously show both short and long pronoun forms in subject

<sup>2</sup> Note that for pronouns and NPs with human referents, the marker *ku* is used, regardless of whether it is a direct or indirect object (Baxter, 1988).

and object positions, so we can determine that this choice is not strictly related to grammatical function.

Another potential issue is individual variation. An initial hypothesis might be that the demographic backgrounds of speakers may affect pronoun use, however, it is not difficult to find examples of MP speakers using both forms, even within the same utterance:

15. Malacca Portuguese (Pillai, 2011)

**nu** nang ubih keng falah **nus** papiah Kristang

**1PL** NEG listen who say **1PL** speak Kristang

“We don’t listen to those who say we speak Kristang.”

Example (15) is from a monologue in which both short and long forms are used to express subjects by the narrator. Note, however, they are found in different phonetic environments: the short form occurs before a word that begins with a nasal consonant, while the long form occurs before a voiceless stop. Further explorations of the corpus suggest that phonological environment is not a relevant factor, however.

## 6.2. Other Long/Short Distinctions in MP

Aside from pronouns, there are instances of long and short variants of other items. For instance, the pre-verbal future marker, derived historically from Portuguese *logo* “later”, is expressed variably as long form *logu* or short form *lo*. Both can be found in pre-verbal position:

16. Malacca Portuguese (Pillai, 2011)

Mas tantu kaza logu kuzeh papa

Most many house FUT cook porridge

“Most of the homes will cook porridge”

17. Malacca Portuguese (Pillai, 2011)

Kristang lo mureh

Kristang FUT die

“Kristang will die”

As with the short and long forms of personal pronouns, position does not necessarily determine which form will be used. Both the above examples have the future marker in a pre-verbal position, following the subject. Note that, MP allows some pre-verbal elements to occur by themselves as a

sentence fragment, usually as an answer to a question. In the case of *logu*, only the long form may be used in answer fragments.

18. Malacca Portuguese (Laub, fieldwork 2017)

Bos lo/logu bebeh?

2SH FUT drink

“Would you like a drink?”

Logu/\*Lo

FUT

“Yes, I would”

In the question, both the short and long forms of *logu* are grammatical, but in the answer only the long form is accepted. This could imply that when not a fragment, the long form expresses prominence or focus. There is insufficient support for such an analysis in the present data, however, and it requires further investigation.

## 7. Findings

We begin by exploring the pronoun tokens found in Pillai (2011) to investigate if there is a relationship between the type of pronoun used and whether it functions as a subject or object. The reasoning behind this is that Portuguese, the main lexifier of MP, and English, a major contact language among residents of the Portuguese Settlement, have variants of personal pronouns dependent on grammatical function (though in English not for second person pronouns which are invariant). Table 2 sets out the token counts in Pillai’s corpus. Notice that apart from *bo* never occurring as object, the long and short forms are found in both functions.

Table 2: Pronoun forms by function

	Subject	Object
<i>bos</i>	4	6
<i>bo</i>	30	0
<i>nus</i>	53	6
<i>nu</i>	51	3

As MP is an SVO language, we might hypothesize that the long form comes at the end of an utterance or before a pause, independent of its grammatical role. I therefore tallied the pronouns based on where in the utterance they are found (see Table 3).

Table 3: Pronoun variants by position in utterance

	At the end of utterance	Before a pause	Beginning or middle of utterance
<i>bos</i>	4	1	5
<i>bo</i>	1	0	29
<i>nus</i>	2	11	46
<i>nu</i>	0	3	51

Since there are no examples of *bo* in the corpus functioning as an object I carried out translation elicitation to check if this was always the case. Consider the following elicited example where the same referent is picked out as both subject and object in two different clauses:

19. Malacca Portuguese (Laub, fieldwork 2017)

Kantu bo fazeh di tona yo lo dali ku bos  
 If 2SG do again 1SG FUT beat ACC 2SG  
 “If you do that again I will beat you.”

Here, as expected, the short form occurs as subject, and the long form occurs as object. Another example is the following:

20. Malacca Portuguese (Laub, fieldwork 2017)

Bo ke dah ku yo prezanti, yo lo ku bos prezanti ku<sup>3</sup> kuelu bedri  
 2SG want give to 1SG present 1SG FUT to 2SG present to rabbit green  
 “You want to give me a present, I will [give] you a present for the green rabbit”

While this example occurred in an elicitation recording session, the particular sentence was part of a discussion between two MP-speaking participants about what they needed to do in the exercise, so was not directly elicited. The relevant factor appears to be an element of contrast: the object in the second clause differs from that in the first clause, and the choice of pronoun seems to

<sup>3</sup> Note that *ku* here does not precede an NP which refers to a human, however the elicitation exercise involved imagining a birthday party with animals, so the example could be seen as a type of anthropomorphism.

emphasize this difference. Further exploration showed that the long form of variable pronouns is used when contrast or focus is expressed, as in the following unelicited exchange:

21. Malacca Portuguese (Laub, fieldwork 2017)

Agora ngkoza eli teng bo nteng eli lo pidi  
 Now something 3SG have 2SG NEG.have 3SG FUT ask  
 “Now he has something which you don’t have, so he will ask [you]”

22. Malacca Portuguese (Laub, fieldwork 2017)

Bos pidi?  
 2SG ask  
 “Do *you* ask?” (*i.e.*, he does not)

Example 21 has two referents, expressed with third-person singular and second-person singular pronouns. Both instances of the third-person pronoun use the long form, while the second-person takes the short one, even though all the pronouns are functioning as subjects of their clauses. Here it is the long form third-person pronoun which is in focus. The second speaker responds in example 22 with a long second person form to emphasize and contrast its subject referent.

Definiteness also appears to be a relevant factor. In a recording in the Pillai (2011) corpus, a woman is describing how to make curry paste. She uses the second-person singular and first-person plural as indefinite or general pronouns, as she is setting out a recipe that anyone might use. Throughout the recording, the short forms are exclusively used.

23. Malacca Portuguese (Pillai, 2011)

Achar pesi bo kereh jinjibri, kunyit, alu  
 “Fish pickle 2SG want ginger turmeric garlic  
 “For fish pickle, you need ginger, turmeric and garlic.”  
 Bo blend fazeh ungua  
 2SG blend make one  
 “You blend until everything is combined.”

24. Malacca Portuguese (Pillai 2011)

Padi achar \_pesi nu kereh subezu azeti  
 For fish \_pickle 1PL want extra oil  
 “For fish pickle we need extra oil.”

Within the context of the discussion, it is clear that the speaker is making impersonal reference using *bo* or *nu*, however, all instances are functioning as subjects. To explore whether non-specific pronouns would have the short form in object position in elicitation, I asked a participant to describe what a (generic) tourist in Malacca should do. She spoke for five minutes and every instance of the second-person singular pronoun was in the short form, even in object position, as in:

25. Malacca Portuguese (Laub, fieldwork 2017)

Logu      nganah      ku      bo,      rekah      ku      bo  
 FUT      deceive      ACC      2SG      deceive      ACC      2SG  
 “They will deceive you”

The short form is used after the accusative marker *ku*, and at the end of the sentence. The tendency to use a long form in these positions is overpowered by its function as a non-specific pronoun.

## 8. Discussion

As demonstrated in the previous section, there is variation between short forms and long forms for some personal pronouns in MP, as well as other elements, such as the future particle. In this section we explore how this variation ties into wider theories of creole formation, and what implications it might have for the Feature Pool theory.

Malay also has variation between short and long form pronouns. As in MP, this does not encode a difference in grammatical function, but rather morphosyntactic status. Sneddon (1996, p. 170) gives the inventory of Malay pronouns set out in Table 4.

Table 4: Malay/Indonesian pronouns

	long form	prefixed short form	suffixed short form
first person	aku	ku-	-ku
second person	engkau, kamu	kau-	-mu
third person	dia, ia	-	-nya

The short form pronouns are clitics, and always attach before or after the verb. In MP, clauses have a generally rigid structure, in that after the subject come the TAM markers, then an (optional) auxiliary, and finally the verb (Baxter, 1988). Nothing may come between a TAM marker and a



verb. The standard word order of SVO holds whether a full noun phrase, long form pronoun, or short form pronoun is used. In contrast, Malay subject clitic pronouns can occur post-verbally. We can also conclude that a short form is not a proclitic, at least not in all cases, as it may occur at the end of an utterance, not adjacent to the verb in a sentence (see Example 21).

Despite the differences, the tendency of MP speakers to vary the forms of personal pronouns could be a development arising out of the contact with Malay. As proposed in the Feature Pool (Mufwene, 2002; Siegel, 2007), features are transferred over but not always completely. In the case of the pronouns, the lexical and functional features of Malay pronouns may have transferred to the destination of those from Portuguese, carrying with them the ability to have variant lengths based on information structure.

Kroeger (2014), in a paper on the functions of the Malay clitic pronoun *-nya* (3SG), cites Givón (1983: 18) regarding the encoding of topics: “The more disruptive, surprising, discontinuous or hard to process a topic is, the more *coding material* must be assigned to it.” Givón’s “scale of phonological size” from most inaccessible to most accessible is: Full NP > stressed/independent pronoun > unstressed/bound pronoun > zero anaphora. Within MP, it may be the case that long-form pronouns are more towards the inaccessible side of the spectrum, while short-forms are more accessible.

Kroeger (2014, p. 17) cites examples showing constraints to usage of clitic pronouns in Malay:

26. Malay (Arka & Manning 1998, in Kroeger 2014, p. 17)

\*Diri=nya      di-serahkan              ke      polisi      oleh      Amir

Self=3SG      PASS-surrender              to      police      by      Amir

(for: “Himself was surrendered to the police by Amir.”)

27. Malay (Arka & Manning 1998, in Kroeger 2014, p. 17)

Diri=nya      selalu      di-utamakan=nya

Self=3SG      always      PASS=prioritize=3SG

“Himself is always prioritized by him.” (*i.e.* “He always gives priority to himself.”)

Kroeger (2014, p. 17) argues that “the contrast illustrated [in the previous examples] is due to discourse or pragmatic factors, rather than the syntactic status of the arguments.” We can see in example 26 that the clitic pronoun is ungrammatical because it precedes the full NP *Amir*, which

should be the antecedent, whereas in example 27 both are the topic and thus can be expressed as clitic pronouns.

The short forms of the MP TAM marker *lo* also has a parallel in Malay. In Malay, the pre-verbal negator *tidak* also occurs as *tak*, and the perfective marker *sudah* can be shortened to *dah* (Mintz, 1994). Therefore, this feature of short and long pre-verbal markers could have influenced MP, where there are short and long forms of some pre-verbal elements such as the future marker *logu~lo*. As with the personal pronouns, the ability to have short forms of these elements may have its origins in Malay, but the eventual functions of this distinction have become an MP element on its own.

## **9. Conclusion**

The findings here, that suggest short pronouns in MP are less prominent, while long forms are more likely to be the focus, and that this contrast arises as a transfer from Malay, are meant to be a preliminary exploration of the interaction between language contact, information structure, and creole formation and evolution. As argued by Ansaldo (2009), Fon Sing (2018) and others, the local linguistic environment can play a role in the formation of creole syntax, even when certain grammatical structures may have also formed through internal development

The investigation of short and long pronouns would benefit from a fuller quantitative analysis, which requires a corpus of a much larger size. In addition, there can be more elicitation exercises conducted with MP speakers to elaborate on how information structure plays a role in the choice of pronouns.

## **Acknowledgements**

I would like to thank the anonymous reviewers for their detailed and insightful comments, and the Editors of this issue for their help in improving my manuscript. I would also like to express my gratitude to Professor Stefanie Pillai from Universiti Malaya for letting me use her data at the Endangered Languages Archive at SOAS (<https://elar.soas.ac.uk/Collection/MPI130545>). This research would not have been able to even begin without her help.

## References

- Adelaar, K. A., & Prentice, D. J. (1996). Malay, its history, role and spread. In S. A. Wurm, P. Mühlhäusler & D. T. Tryon (Eds.), *Atlas of languages of intercultural communication in the Pacific, Asia and the Americas* (pp. 673-693). Berlin: Mouton de Gruyter.
- Ansaldo, U. (2009). *Contact languages: Ecology and evolution in Asia*. Cambridge: Cambridge University Press.
- Ansaldo, U., Lim, L., & Mufwene, S. (2007). The sociolinguistic history of the Peranakans. In U. Ansaldo, S. Matthews & Lisa Lim (Eds.), *Deconstructing creole* (pp. 203-226). Amsterdam: John Benjamins.
- Bakker, P., A. Daval-Markussen, M. Parkvall & I. Plag (2011). Creoles are typologically distinct from non-creoles. *Journal of Pidgin and Creole Languages*, 26(1), pp. 5-42.
- Baxter, A. N. (1988). *A grammar of Kristang (Malacca Creole Portuguese)*. Canberra: Australian National University.
- Baxter, A. N. (1996). Portuguese and Creole Portuguese in the Pacific and Western Pacific rim. In S. A. Wurm, P. Mühlhäusler & D. T. Tryon (Eds.), *Atlas of languages of intercultural communication in the Pacific, Asia and the Americas* (pp. 299-338). Berlin: Mouton de Gruyter.
- Baxter, A. N. (2012). Vestiges of etymological gender in Malacca Creole Portuguese. *Journal of Pidgin and Creole Languages*, 25(1), 120-154.
- Baxter, A. N., & Bastos, A. (2012). A closer look at the post-nominal genitive in Asian Creole Portuguese. In H. Cardoso, A. N. Baxter & M. Pinharanda-Nunes (Eds.), *Ibero-Asian creoles: Comparative perspectives* (pp. 47-79). Amsterdam: John Benjamins.
- Bickerton, D. (1988). Creole languages and the bioprogram. In Frederick J. Newmeyer (Ed.), *Linguistics: The Cambridge survey 2* (pp. 268-284). Cambridge: Cambridge University Press
- Cardoso, H (2009). *The Indo-Portuguese language of Diu*. [Unpublished PhD thesis]. University of Amsterdam.
- Cardoso, H., A. N. Baxter, & M. Pinharanda-Nunes. (Eds.). (2012). *Ibero-Asian Creoles: Comparative perspectives*. Amsterdam: John Benjamins.
- Chaudenson, R. (1979). *Les créoles français [French creoles]*. Paris: Fernand Nathan.

- Chaudenson, R. (1992). *Des îles, des hommes, des langues [Islands, men, languages]*. Paris: L'Harmattan.
- Clements, J. C. (1996). *The genesis of a language. The formation and development of Korlai Portuguese*. Amsterdam: John Benjamins.
- Clements, J. C. (2000). Evidência para a existência de um pidgin português asiático. In E. d'Andrade [Evidence for the existence of an Asian Portuguese pidgin], M. A. Mota & D. Pereira (Eds.), *Crioulos de Base Portuguesa. Actas do Workshop sobre Crioulos de Base Lexical Portuguesa [Portuguese-based Creoles. Workshop Proceedings on Portuguese Lexical Based Creole]* (pp. 185-200). Braga: Associação Portuguesa de Linguística.
- DeGraff, M. (1999). Creolization, language change, and language acquisition: A prolegomenon. In M. DeGraff (Ed.), *Language creation and language change: Creolization, diachrony and development* (pp. 1-46). Cambridge, MA: The MIT Press.
- DeGraff, M. (2003). Against Creole exceptionalism. *Language*, 79(2), 391-410.
- DeGraff, M. (2005). Linguists' most dangerous myth: The fallacy of creole exceptionalism. *Language in Society*, 34(4), 533-591.
- DeGraff, M. (2009). Creole exceptionalism and the (mis)education of the creole speaker. In Jo Anne Kleifgen & George Bond (Eds.), *The languages of Africa and the diaspora: Educating for language awareness* (pp. 124-144). Bristol: Multilingual Matters.
- Fon Sing, G. (2018). Creoles are not typologically distinct from non-Creoles. *Language Ecology* 1(1), 44-74.
- Ganho, A. S. & T. McGovern. (2003). *Using Portuguese: A guide to contemporary usage*. Cambridge: Cambridge University Press.
- Hall, R. (1966). *Pidgin and creole languages*. Ithaca, NY: Cornell University Press.
- Hancock, I. (1973). Malacca Creole Portuguese: A brief transformational account. *Te Reo*, 16, 23-44.
- Holm, J. (2000). *An introduction to pidgins and creoles*. Cambridge: Cambridge University Press.
- Kärchner-Ober, R. (2013). Linguistic quandary in multilingual Malaysia: Socio-political issues, language policy, educational changes. In D. Singleton, L. Aronin, J. A. Fishman, and M. Ó Laoire (Eds.). *Current multilingualism: A new linguistic dispensation* (pp. 297-315). Berlin: DeGruyter.

- Koopman, H. (1986). The genesis of Haitian: Implications of a comparison of some features of the syntax of Haitian, French, and West African Languages. In P. Muysken & N. Smith (Eds.), *Substrata versus universals in creole genesis* (pp. 231-258). Amsterdam: John Benjamins.
- Kroeger, P. (2014). Passive agents in Malay: The binding properties and discourse functions of agentive =*nya*. In S. Chung and H. Nomoto. (Eds.). *Current Trends in Malay Linguistics*.
- Lefebvre, C. (1998). *Creole genesis and the acquisition of grammar. The case of Haitian Creole*. Cambridge: Cambridge University Press.
- Lefebvre, C. (2008). On the principled nature of the respective contribution of substrate and superstrate languages to a creole's lexicon. In S. Michaelis (Ed.), *Roots of creole structures: Weighing the contribution of superstrates and substrates* (pp. 197-224). Amsterdam: John Benjamins.
- Lefebvre, C. & J. Lumsden. (1989). Les langues créoles et la théorie linguistique [*Creole languages and linguistic theory*]. *Revue Canadienne de Linguistique* 34, 319-377.
- Lumsden, J. (1999). Language acquisition and creolization. In M DeGraff (Ed.), *Language creation and language change: Creolization, diachrony and development* (pp. 129-158). Cambridge, MA: The MIT Press.
- Martins, A. M. (2007). Double realization of verbal copies in European Portuguese emphatic affirmation. In N. Corver and J. Nunes (Eds.), *The copy theory of movement* (pp. 77-118). Amsterdam: John Benjamins.
- Maurer, P. (2011). *The former Portuguese creole of Batavia and Tugu (Indonesia)*. London: Battlebridge.
- McPherson, K. (1993). *The Indian Ocean: A history of people and the sea*. Oxford: Oxford University Press.
- McWhorter, J. (2001). The world's simplest grammars are creole grammars. *Linguistic Typology*, 5(2/3), 125-166.
- McWhorter, J. (2005). *Defining Creole*. New York: Oxford University Press.
- McWhorter, J. (2018). *The Creole debate*. Cambridge: Cambridge University Press.
- Mintz, M. (1994). *A student's grammar of Malay & Indonesian*. Singapore: EPB Publishers.
- Mufwene, S. (2000). Creolization is a social, not a structural process. In I. Neuman-Holzschuh & E. Schneider (Eds.), *Degrees of restructuring in creole languages* (pp. 65-84). Amsterdam: John Benjamins.

- Mufwene, S. (2002). Competition and selection in language evolution. *Selection* 3, 45-56.
- Muysken, P. (1988). Parameters for serial verbs. In K. Demuth & V. Manfredi (Eds.), *Niger-Congo syntax and semantics 1* (pp. 65-75). Boston, MA: Boston University.
- Nomoto, H. (2006). Voice in colloquial Malay relatives. In Y. Tsuruga, N. Miyake & Y. Kawaguchi (Eds.), *Gengojouhougaku Kenkyuuhoukoku*, 12, (pp. 97-116). Tokyo: Tokyo University of Foreign Studies.
- Ostler, N. (2005). *Empires of the World: A language history of the World*. New York: Harper Collins.
- Parkvall, M. (2008). The simplicity of creoles in a cross-linguistic perspective. In M. Mestamo, K. Sinnemäki & Fred Karlsson (Eds.), *Language complexity: typology, contact, change* (pp. 265-285). Amsterdam: John Benjamins.
- Parthesius, R. (2010). *Dutch ships in tropical waters: The development of the Dutch East India Company (VOC) shipping network in Asia 1959-1660*. Amsterdam: Amsterdam University Press.
- Pillai, S. (2011). Malaccan Portuguese Creole: A Portuguese-based creole. Endangered Languages Archive at SOAS University of London. <<https://elar.soas.ac.uk/Collection/MPI130545>>
- Pillai, S., Soh, W. Y., Kajita, A. S. (2014). Family language policy and heritage language maintenance of Malacca Portuguese Creole. *Language & Communication*, 37, 75-85.
- Pinharanda-Nunes, M. (2008). Os demonstrativos em maquista: uma análise morfossintáctica contrastiva [Demonstratives in Makista: a contrastive morphosyntactic analysis]. In *Papia, Revista Brasileira de Estudos Crioulos e Similares*, 18, 8-22.
- Pinharanda-Nunes, M. (2012). Traces of superstrate verb inflection in Makista and other Asian-Portuguese creoles. In H. Cardoso, A. N. Baxter, & M. Pinharanda-Nunes (Eds.), *Ibero-Asian Creoles: Comparative perspectives* (pp. 291-326). Amsterdam: John Benjamins.
- Siegel, J. (2007). Sociohistorical contexts: Transmission and transfer. In Umberto Ansaldo, Stephen Matthews, & Lisa Lim (Eds.), *Deconstructing creole* (pp. 167-202). Amsterdam: John Benjamins.
- Siegel, J. (2008). *The emergence of pidgin and creole languages*. Oxford: Oxford University Press.
- Skopetas, S., Fiedler, I., Hellmuth, S., Schwarz, A., Stoel, R., Fanselow, G., Féry, C., & Krifka, M. (2006). *Questionnaire on information structure (QUIS): Reference manual*,

*Interdisciplinary Studies on Information Structure (ISIS) 4.* Universitätsverlag Potsdam, Germany.