

Variability or its loss in creole endangerment

The case of Baba Malay

Nala H. Lee

National University of Singapore

A language's endangerment is said to be typified by variation accompanied by some degree of unpredictability. This paper examines if this characterization is appropriate of Baba Malay, an endangered creole, as it is spoken in Singapore. Recent work suggests that the language is much less variable than one expects of a creole. A comparison between historical Baba Malay material and newer Baba Malay material, focusing on the language's pronominal system, and aspectual and tense system, demonstrates that this is indeed the case. Such variation or an increasing lack of it is not unpredictable, stemming not only from natural language change, but also from changes in the language's socio-environment. Increasing lack of proficiency in the component languages and a desire among speakers to distinguish themselves from speakers of the lexifier variety may explain some loss in Baba Malay's variability.

Keywords: creole, language endangerment, variation, lexifier, substrate, lexicon, syntax

1. Introduction

It has been observed that endangered languages can be typified by variation linked to some degree of unpredictability, in the sense that the variation does not have to be conditioned by social factors (see Campbell & Muntzel, 1989; Palosaari & Campbell, 2012). Such variability is also linked to the notion of reduction or shrinkage, often said to be due to imperfect learning (Dorian, 1977; Dressler, 1972; Dressler & Wodak-Leodolter, 1977; Palosaari & Campbell, 2012). For example, excessive glottalization was reported in the speech of a speaker of Xinca, and in Ocuilteco, the rule that voices stops after nasals was sometimes not applied in the speech of imperfect speakers (Campbell & Muntzel, 1989, p.189).

There is however awareness, that while “imperfect learning may be sufficient to explain many of these cases”, other factors such as the absence of a relevant contrast in the dominant language (in the common cases of language shift in bilingual situations) may also contribute to contrast losses, such as in the case of the merger of uvular and velar consonants in the endangered Mam of Tuxtla Chico due to the absence of such a contrast in Spanish (Palosaari & Campbell, 2012, p.111).

There is more recent sociolinguistic work on endangered languages that goes beyond the notions of unpredictability and imperfect learning, and this shows that variation in the traditional sociolinguistic sense abounds in these languages – variation in these cases can be explained by social factors and is used in the additive sense to imply additional ways of meaning-making, not less. Work in this tradition includes that of Nagy (2009), who explains how contextual and social dimensions feature in the development of a sociogrammar of Faetar, as well as that of Meyerhoff (2015), who shows that the Nkep vowel system can be described using attention to speech, and who explains Bislama borrowing in the language using the notion of apparent time. In more recent work, Kasstan (2019) focuses on a previously obligatory palatalization-rule in Francoprovençal, demonstrating that among the Lyonnais speakers, palatalization was used to represent their ideal Francoprovençal in careful speech, although it was no longer common in spontaneous speech. The current article situates itself somewhere in between both camps. It observes the loss of variability in Baba Malay, an endangered creole, but attempts to rationalize these losses within the social context of its speakers and in natural mechanisms of language change, rather than through the explanatory device of unpredictability.

Recent work on Baba Malay, spoken in the Malay peninsula, mainly in the city-state of Singapore (1.3521° N, 103.8198° E) and Malacca (2.1896° N, 102.2501° E), a state within Malaysia, has been met with surprise – the language appears to be much less variable than expected from a creole in particular, considering that creoles usually exhibit considerable amounts of variation, attributable to their inherent continuums (Bickerton, 1973; DeCamp, 1971; Le Page, 1960; Rickford, 1987). Modern Baba Malay as it is currently spoken, has been characterized as a rather singular linguistic entity with an ascribed list of lexical and grammatical qualities in the literature (Ansaldi & Matthews, 1999; G.A.M.-E. Lee, 1999; N.H. Lee, 2014; Pakir, 1986). In contrast, work on early Baba Malay, based on written sources from the late 1800s and early 1900s, paints a more dynamic picture. While Thurgood’s (1998) aim was to present early Baba Malay as a language that Hokkien speakers were switching to, it is clear from her data and analysis that there was a wider range of variation in the early variety of the language. For example, while the classifier phrase-head noun word order was predominant in her data, the existence of counter data was also noted. Similarly, N.H. Lee

(forthcoming) explains how the linguistic ideologies of early Baba Malay speakers and competing pressures in their group identities led to a wide span of variation along its continuum, adhering to and detracting from its substrate and lexifier languages, as well as attesting to the growing dominance of English at that time. For example, while some authors only used a progressive aspect marker derived from Hokkien, others embraced both Hokkien and Malay-derived progressive aspect markers. The current paper is an extension of that work, illustrating how Baba Malay has lost some of this early variability.

For this article, comparisons are carried out between early Baba Malay material (written sources from the late 1800s and early 1900s) and modern Baba Malay material (collected through primary fieldwork for a language documentation project that took place between 2012 and 2019), focusing predominantly on two subsystems of the language – the pronominal system as well as the aspectual and tense system. Both sets of data were published in and collected in Singapore. These two specific language subsystems are chosen to illustrate the fact that the loss of variability can permeate various areas, which span from the more lexical (the pronominal system) to the more grammatical (the aspectual and tense system). Explanations for the diachronic patterns observed are found not only in mechanisms of natural change, but also stem from changes that have occurred in the language's sociological environment.

The rest of the paper is organized as follows: Section 2 introduces the sociological context of the language. Section 3 introduces the data that is used. Analysis of the data and a discussion of the findings ensue in Sections 4 and 5 respectively.

2. Sociological background of Baba Malay then and now

Baba Malay dates back to the 15th century (Ansaldò, Lim, & Mufwene, 2007; C.-B. Tan, 1988). During the Ming dynasty, particularly between 1405 and 1433, the Chinese were known to have led trading exchanges in the South China Sea, Java Sea and the Malacca Strait (Wade, 2015; Wang, 1964; Widodo, 2002). In effect, records show that communities of Chinese male traders were established in Malacca at that time (Fei, 1436; Ma, 1416), concurrent with the period during which women rarely travelled out of China – women were expected to stay behind to take care of their households, observe filial piety, and carry out ancestor worship (J.-H. Lim, 1967). The Qing dynasty that followed then saw a ban on emigration, which was enforced especially strictly on women (C.-B. Tan, 1979). As such, the Chinese men who had established themselves in Malacca, and who predominantly spoke Hokkien, a variety of Southern Min, married indigenous women (B.-K. Lim, 1917; Purcell, 1980; C.-B. Tan, 1979). While it has been suggested

that the women may have been of other origins, such as Batak, Balinese, and Javanese (Purcell, 1980), in addition to being of Malay origin (C.-B. Tan, 1979), a recent genetic study shows that the mixed community that ensued comprises a specific admixture of Chinese and Malay ancestries, with Malay ancestry traceable primarily through the maternal lineage (Wu et al., 2021). The community that resulted through these early mixed marriages is referred to as the *Peranakan*,¹ the *Baba* ‘male’ and *Nyonya* ‘female’, or the Straits Chinese. The Peranakan community, culture and language thus came to be. The language, referred to either as *Peranakan* or *Baba Malay*, has a predominantly Hokkien substrate (Ansaldò & Matthews, 1999) and a Malay-based lexifier (N.H. Lee, 2022). While the community originally emerged in Malacca, a number of Peranakans moved to Singapore after it was established as an entrepôt, given their interest in trade and commerce (Skinner, 1996). Siah (1848) observes that the greatest number of married men in Singapore by the mid-1800s were actually Peranakans who had moved from Malacca.

The sociohistory of the Peranakans as a creolized community is unique for being non-traumatic. There was neither enslavement, nor oppression (Ansaldò et al., 2007). The Peranakans were instead an affluent group of traders and merchants, perceived to be “the best educated and wealthiest and most intelligent section of the Chinese community” (Nathan, 1922, p. 77), due in large part to the fact that they were among the first in Singapore to receive an English-medium education, following the establishment of English-medium schools in Malacca and Singapore in the 1800s (Ansaldò et al., 2007). With their command of English, the Peranakans were valued middlemen who mediated between the British administrators and the other newcomers to the region (L. Lim, 2016). Where allegiances were concerned, some Peranakans had aligned themselves with the British so much that they viewed themselves as British subjects rather than Chinese subjects, referring to themselves as “King’s Chinese” (Song, 1967, p. 38), or “Queen’s Chinese” (C.-B. Tan, 1979, p. 68).

In reality, the actual sociolinguistic situation and identity alignments of the Peranakans were much more complex. Before the founding of English-medium schools such as the Anglo-Chinese College in Malacca in 1814, and the Singapore Free School in 1834, well-off Peranakans were known to send their sons to China for education (C.-B. Tan, 1979, p. 70). This inadvertently reflects early ties that some Peranakans had with China. After the establishment of English schools

1. It has been conjectured that the term *Peranakan* comprises the middle voice prefix *ber-*, which combines with *anak* ‘child’, giving rise to *beranak*, which means ‘to give birth’. The verb is then combined with the nominalizer *-an* to indicate ‘womb’. The word initial *b* may then have dissimilated in voicing with the following sequence (Lee, 2022).

however, there was no longer any need for the Peranakans to send their sons to China.

The turn of the 20th century was also a time when the Peranakans began to forge a more distinct identity from the newer immigrants from China, who they called *sinkhek* 'newcomer'. While the Peranakans perceived themselves as being superior to the *sinkhek* who worked as coolies in construction, at the port, or in the mines (Vaughan, 1879), they realised that the Peranakans no longer constituted the majority of the Chinese population locally, and that there was also a new class of Chinese capitalists among the newer immigrants who were directly challenging their elite status (C.-B. Tan, 1979, p.70).

It was also at this time that the Peranakans began laying claim to their status as "British subjects", "Malayans", and "sons of the soil" (C.-B. Tan, 1979, p.71), emphasising not only their alignment with the British, but also their localness, and hence special status in the region, especially in comparison to the *sinkhek*. It must be said however, that the situation was a lot more nuanced than one in which the Peranakans were trying to emphasize that they were not the "other". Crucially, there were also Peranakans at that time who saw the need to identify themselves as being culturally Chinese, and who felt the need to learn a Chinese language. In particular, Hokkien was becoming more central in daily interactions, with recent Chinese immigrants outnumbering the Peranakans, and with the former becoming increasingly prosperous. Other influential events that took place in the late 1800s and early 1900s include the nationalist movement of 1894 to 1895, and the reform movement of 1898 to 1912, these movements resulting from China's defeat in the First Sino-Japanese War. It became more important than ever for the Peranakans to pick an allegiance, to be Chinese or otherwise.

The complex ideological alignments of the Peranakans came across in very public ways. While some Peranakans embraced the calls for Chinese reform, and were in favour of studying Mandarin as a common language (B.-K. Lim, 1899), others went as far as discouraging the use of Baba Malay. A contributor to *The Straits Chinese Magazine* writes, "We Straits Chinese fathers, should therefore encourage our girls in every possible way to give up the Malay language and revert to the Chinese tongue" (Soh, 1907, p.143). In the opposing vein, there were also those who objected to Peranakans studying any form of Chinese. In a 1929 issue of the *Malacca Guardian* newspaper, the editor writes that "Malaya is not a part of China", and that "the language of this country is either English or Malay, and most of the Straits-born Chinese are content to learn well those two languages" (cited in C.-B. Tan, 1979, p.79).

What is clear is that the Peranakans were negotiating their identities not only between the two poles of Britishness and Chineseness, but also that of localness. The situation is made even more complex by the fact that some Peranakans did

not want to be associated with speaking pure Malay, even if they saw its usefulness in aligning away from the Chinese. These were the Peranakans who called the language “*Malayu hutan* – the language of the jungle” (Shellabear 1913, p. 146, italics in original). Crucially, the linguistic ideologies of early Baba Malay speakers and competing pressures in their group identity manifested in an overt Baba Malay continuum with variation adhering to and detracting from its substrate and lexifier languages, as well as a language that was developing under the influence of English (N.H. Lee, forthcoming).

In contrast to how Baba Malay clearly thrived in more vibrant linguistic circumstances of the late 1800s and early 1900s, the language is now critically endangered, with fewer than 1,000 speakers each in Singapore and in Malacca (N.H. Lee, 2022). These speakers are mostly above the age of 50. Monolingual speakers are also rare and would be in their 90s at the very least. Three main factors have been identified for the language’s endangerment (N.H. Lee, 2019). First, the dominance of British influence over the lives of the Peranakans (as seen for example, in the history of the Peranakan Association which was initially set up as the Straits Chinese British Association) meant that at some point, the Peranakans began to embrace English as their home language. The dominance of English continued into the 1960s and beyond. 1966 was the year during which the bilingual language education policy was implemented in Singapore (Pakir, 1994) – students are categorized into broad ethnic categories and assigned a ‘mother tongue’ that they have to learn in addition to English. Accordingly, Peranakans are classified as Chinese, and are required to learn Mandarin as a subject at school. Today, Baba Malay has lost even more ground in the home domain, with a recent census reporting that the Chinese (the label under which Peranakans are subsumed) literate population aged 15 years and over are mostly literate in English and Mandarin, at 58.0% and 62.3% respectively (Department of Statistics Singapore, 2020), although it is often anecdotally observed that the Peranakan community continues to value English above all other languages. The final consideration, which is equally as important, is the fact that the community is not expanding. While Chinese-Malay intermarriages still occur, these intermarriages no longer result in Peranakan ethnicity or culture. It was popular for Peranakans to marry among themselves previously (Pakir, 1986), but intra-community marriages have since become uncommon (C.-B. Tan, 1979, p.19). With the progression of time, and the decrease in community numbers, most Peranakans marry outside their own community and do not pass on Baba Malay as a home language. Loss is further fuelled by the reality that the language is no longer spoken by most people of child-bearing age, as speakers are mainly aged 50 and above. At this point, it is notable perhaps that most Peranakans in Singapore, which provides the locus of this study, are no longer profi-

cient in Malay and in Hokkien, even if there were Peranakans who were proficient in these languages in the past.

Baba Malay, as it is stands, is undergoing revitalization efforts. Among these efforts is the development of pedagogical materials, which includes a short guide to the language, *Speak Baba Malay: The Easy Way* (P. Chan, 2007), and a textbook (K. Chan, 2018), among others. A dictionary has also been produced of the language (Gwee, 2006). Utilizing resources such as these, the language is being taught in unofficial language classes in Singapore (N.H. Lee, 2019). Learners include heritage speakers, as well as members of the public with an interest in Peranakan culture, such as museum docents. These efforts are driven by the community, in hopes that the language's endangerment trends can one day be reversed.

3. Data

The sociological background of the language presents two very different scenarios: one of language vitality then, and one of language endangerment now. Two different sets of data are then used to represent each period so that the diachronic progression of the language can be observed.

The early Baba Malay data utilized for this paper comes from a corpus of texts dating from 1892 to 1931, 1899 to 1939 being a defining period for the publication of *chrita dulu kala*, literally “stories before time” or “stories from long ago” (Yoong & Zainab, 2002). The period of publication falls within the early period discussed in Section 2 and attests to the vitality of the language then, with Baba Malay being used in additional domains such as literature. *Chrita dulu kala* refers to serialized Chinese novels translated into Baba Malay. These volumes were enjoyed widely among Peranakans, most of them being “historical and chivalrous stories” (Yoong & Zainab, 2002, p. 11). Estimates provided for the number of *chrita dulu kala* titles published range from 60 to 88 (Salmon, 1977; B. Tan, 2007; Yoong & Zainab, 2002). All titles ever listed were published in Singapore, and none in Malaysia. Notably, while there were also *chrita dulu kala* published in Indonesia, these were written in Peranakan Indonesian, and not in Baba Malay. Alongside the *chrita dulu kala*, fan letters to the writers of *chrita dulu kala* also exist as an early written resource. While there might be some concern about the use of written resources, the corpus of fan letters and *chrita dulu kala* provides a rare insight on the early form of the language. This corpus was used by N.H. Lee (2021) to show how much variation there was in early Baba Malay, with writers detracting from and adhering to the substrate and lexifier languages to different degrees. To that end, and for the purposes of this article, the data utilized from the late 1800s and early 1900s includes:

- 20 fan letters included in 9 volumes of *Chrita Dahulu-Kala, Nama-nya Sam Kok, atau Tiga Negri Ber-prang: Siok, Gwi, sama Gor di jaman “Han Teow”*,² by Batu Gantong (1892–1896)
- *Chrita Dulu Kala dari Mula Pertama di Jaman Tong Teaw bernama Seeh Jin Quee Cheng Tang (pukol negri sebala timor)*³ – Book 1, by unknown author (1895)
- *Chrita Dulu-Kala “Ow Liat Kok Chee” atau Penghabisan Liat Kok di Zeman Chin Kok: Chin Si Ong Menjadi Raja Dewa Hye Tiow Seng Jin Berprang Besar sama Lam Koon Ong Soon Pin*⁴ – Book 1 by Wan Boon Seng (1931)
- *Chrita Dahulu-Kala bernama “Tai Lau Sam Boon Kuay” atau “Lee Kong” di Zeman Beng Teow*⁵ – Book 1, by unknown author (1939)

In total, the early Baba Malay corpus comprises 65,159 words. For the purpose of this paper, the multiple conventions used by the writers have been standardized to those of N.H. Lee (2022).

Where modern Baba Malay is concerned, conversations, narratives, and elicitation data were collected in Singapore for a language documentation project between 2011 and 2019. The data comprised speech from three main language consultants (two males and one female) aged 60 and above, as well as that of their interlocutors. These main language consultants served as contact points to other speakers from the community. They had learnt Baba Malay as their first language, and identified it as their dominant language, even though they were also bilingual in English. None of the speakers were monolinguals, given that monolingual speakers would have been in their late nineties at the time of recording, and could not be imposed upon. The language documentation project itself was a response to threats to the language’s vitality and is representative of Baba Malay at this juncture of endangerment. The corpus in this instance comprises 60,600 words.

There might be some concern regarding the reliability of contrasting two different forms of data (written and spoken), especially when “basilectal exaggeration” has been recognized to have occurred in early popular genres written in creoles (Corcoran & Mufwene, 1998). The exaggerated portrayals were the work

2. Story from long ago, called Three Kingdoms, or Three Fighting Kingdoms: Siok, Gwi and Gor in the time of the Han Dynasty.

3. Story from long ago, from the start of the Tang Dynasty, known as Seeh Jin Quee conquers the East (Beating the countries to the East).

4. Story from long ago, the end of the Liat Kok era, or the conquest of Liat Kok in the time of Chin Kok: Chin Si Ong becomes the King, Deity Hye Tiow engages in a big fight with Lam Koon Ong Soon Pin.

5. Story from long ago, known as “The Big Fight in Sam Boon Street” or “Lee Kong” in the time of the Ming Dynasty.

of non-native speakers who had become familiar with these languages, and these portrayals could essentially be viewed as early examples of linguistic “black-face” (Corcoran & Mufwene, 1998, p. 81). If there was any exemplification of such features in the written early Baba Malay dataset, analysis would have been affected. One might expect there to be fewer such features in the spoken modern-day dataset, thus pointing to a reduction in variation that has occurred diachronically. However, the authors of the serialized novels were native speakers of the language, and authentic representatives of the language that they wrote in.

Additionally, on the issue of reliability, there might be concern that the stories were of antiquarian Chinese interest, which might have increased the usage of Hokkien, but in reality, it has been observed that there was much variability in the amount of Hokkien or Malay that was used by each author. For example, authors such as Wan Boon Seng were known to have used more Hokkien, while others such as Chan Kim Boon were known to have been more influenced by Malay (C.-B. Tan, 1979). The letters in the dataset also varied in terms of how much or how little authors used various features from different component languages. It is hoped that when considered together, both genres provide an overall more balanced depiction of early Baba Malay. Without these written resources, it would not be possible to postulate what early Baba Malay might have been like, and also the types of changes that have taken place.

4. Variability or the lack of it

Two linguistic dimensions are chosen here for analysis. These include the pronominal system of Baba Malay as well as its aspectual and tense system, for the reason that these represent the lexical and syntactic dimensions respectively.

4.1 Variability in Baba Malay’s pronominal system

The pronominal system of early Baba Malay can be represented by the information in Table 1.

Table 1. Usage patterns of the pronominal system of early Baba Malay

Pronoun	Function	Origin	Percentages (based on	
			Tokens	function)
<i>saya</i>	1sg (refined)	Malay	79	12.5%
<i>gua</i>	1sg (coarse)	Hokkien	547	86.8%

Table 1. (continued)

Pronoun	Function	Origin	Tokens	Percentages (based on function)
<i>aku</i>	1SG (coarse)	Malay	4	0.6%
<i>kita</i>	1PL.INCL	Malay	346	94.8%
<i>kami</i>	1PL.EXCL	Malay	7	1.9%
<i>kita-orang*</i>	1PL.INCL	Malay (lexical); Hokkien (structural)	11	3.0%
<i>gua-orang</i>	1PL.EXCL	Malay (lexical); Hokkien (structural)	1	0.3%
<i>lu</i>	2SG	Hokkien	500	99.6%
<i>anda</i>	2SG (refined)	Malay	2	0.4%
<i>lu</i>	2PL	Hokkien	15	62.5%
<i>lu-orang</i>	2PL	Malay (lexical); Hokkien (structural)	9	37.5%
<i>dia</i>	3SG	Malay	1503	100.0%
<i>dia</i>	3PL	Malay	32	11.6%
<i>dia-orang</i>	3PL	Malay (lexical); Hokkien (structural)	245	88.4%

* *Orang* indicates ‘people’.

In early Baba Malay, except for 3SG pronoun, *dia*, all other pronouns can be expressed in multiple ways, with direct inputs from Malay and from Hokkien, as well as through Malay-based calques of Hokkien forms. For example, while *saya* as a *halus* ‘refined’⁶ form for 1SG is derived directly from Malay, *gua* is derived from Hokkien as its *kasar* ‘coarse’ counterpart. To a lesser extent, the Malay coarse form, *aku* appears as well. Similarly, while *lu* is the main form for 2SG derived from Hokkien, there are at least 2 tokens of Malay-derived *anda* observed. These are refined forms, although in Baba Malay, the interlocutor’s salutation and name is mainly used in refined speech, rather than a second person pronoun.⁷

6. *Halus* ‘refined’ and *kasar* ‘coarse’ registers are utilized depending on whether a speaker wants to come across as being particularly refined or otherwise, rather than social settings that are formal or informal. For the categorization of *halus* and *kasar* personal pronouns, see Yusoff (2007). For more information on the *kasar* and *halus* registers in Baba Malay, see Lee (2020).

7. Also see Table 3, where K. Chan (2018) has indicated *lu* to be a familiar 2SG form.

The Malay-based calques of Hokkien pronominal forms involve *orang* ‘people’ as a marker of plurality. The 2PL and 3PL pronouns in Hokkien are *lin-lâng* ‘2SG-people’, and *i-lâng* ‘3SG-people’ respectively. Likewise, in early Baba Malay, 2PL and 3PL can be derived by conjoining forms for 2SG and 3SG with ‘people’, forming *lu-orang* and *dia-orang* respectively. Also observed are instances of unaffixed *lu* and *dia* used for 2PL and 3PL, even though *lu* and *dia* are more commonly used as forms for 2SG and 3SG.

The first-person plural forms are the most numerous in types in early Baba Malay’s pronominal system, with two inclusive-exclusive sub-systems. *Kita* for 1PL.INCL and *kami* for 1PL.EXCL originate in Malay, while *kita-orang* for 1PL.INCL and *gua-orang* for 1PL.EXCL parallel *lán-lâng* 1PL.INCL and *goá-lâng* 1PL.EXCL in Hokkien. *Kita*, usually used to express 1PL.INCL on its own can be found with *orang* ‘people’, which resembles the *lán-lâng* form found in Hokkien, given that *lán* expresses the first person plural inclusive with or without *lâng* ‘people’ attached (see Pakir 1986). Note that what is evident from Table 1 is that where variation occurs for a particular pronoun, not all forms are equally used, and what is also apparent is that the usage patterns of early Baba Malay pronouns paint a picture of considerable variability as compared to their usage patterns in modern Baba Malay (see Table 2).

Table 2. Usage patterns of the pronominal system of modern Baba Malay

Pronoun	Function	Lexical origin	Percentages (based on	
			Tokens	function)
<i>saya</i>	1SG (refined)	Malay	19	2.6%
<i>gua</i>	1SG (coarse)	Hokkien	720	97.4%
<i>kita</i>	1PL	Malay	145	100.0%
<i>lu</i>	2SG	Hokkien	282	100.0%
<i>lu</i>	2PL	Hokkien	10	45.5%
<i>lu-orang</i>	2PL	Malay (lexical); Hokkien (structural)	12	54.5%
<i>dia</i>	3SG	Malay	1322	100.0%
<i>dia</i>	3PL	Malay	18	8.5%
<i>dia-orang</i>	3PL	Malay (lexical); Hokkien (structural)	193	91.5%

A comparison of Tables 1 and 2 shows that the pronominal system of modern Baba Malay has contracted in the number of variants available for some pronouns, including the 1SG, the 1PL, and the 2SG pronouns. In the case of 1SG and 2SG, the Malay variants that were already among the least commonly found options in early Baba Malay (*aku* at 0.6% of the 1SG tokens, and *anda* at 0.4% of the 2SG tokens) are the ones that are completely lost. The 1PL subsystems have also undergone losses. Only *kita*, which previously was the most frequently observed variant (at 94.8% of the 1PL tokens) is left. Whereas it previously was used as an inclusive form in relation to its exclusive counterpart form, *kami*, it has evolved into a general 1PL form in modern Baba Malay, with the complete loss of *kami*.

Also lost from modern Baba Malay is the other pair of 1PL inclusive and exclusive forms, *kita-orang* and *gua-orang* respectively. Such losses in pronominal inclusive-exclusive distinctions are known changes that have been observed in other languages, whether it be in other non-standard varieties of Malay/Indonesian (Donohue & Smith, 1998), or in other contact languages, such as in Walpiri (Bavin & Shopen, 1991) and in Tok Pisin (Paliwala, 2020), although explanations for these sorts of changes differ. While Bavin & Shopen (1991) recognize the absence of these categories in English and the changes in young people's Walpiri due to language endangerment, they mostly attribute these changes to internal mechanisms that promote greater semantic transparency and fewer oppositions. Paliwala (2020) on the other hand, states that the influence of the English pronominal system could be responsible for the loss in these distinctions.

Note interestingly, that while these calqued forms are lost, *lu-orang* for 2PL and *dia-orang* for 3PL, are still used in modern Baba Malay. In fact, use of these forms has proportionately increased – *lu-orang* accounts for 54.5% of the total number of 2PL tokens in the modern Baba Malay dataset, as compared to 37.5%, and *dia-orang* accounts for 91.5% of 3PL tokens in the same dataset, as compared to 88.4% previously.

Of significance perhaps is the observation that the pronominal system of modern Baba Malay depicted in Table 2 is reflected in pedagogical material developed for Baba Malay language classes aimed at language revitalization, but not the additional variants that existed in early Baba Malay. The exact same pronouns appear in *Speak Baba Malay: The Easy Way* (P. Chan, 2007), and in *Mari Chakap Baba: A comprehensive guide to the Baba Nyonya language* (K. Chan, 2018). Specifically, the information, as it appears in the section on pronouns for *Mari Chakap Baba: A comprehensive guide to the Baba Nyonya language*, is presented in Table 3.

Table 3. Personal pronouns presented in *Mari Chakap Baba: A comprehensive guide to the Baba Nyonya language* (K. Chan, 2018)

Pronoun	Description	Pronoun	Description
<i>gua</i>	I (familiar form)	<i>dia</i>	he/she (singular)
<i>saya</i>	I (polite form)	<i>dia-orang, diorang, jorang</i> *	they (plural)
<i>lu</i>	you (singular familiar form)	<i>kita</i>	we
<i>lu-orang</i>	you (plural)		

* *Jorang* has been cited for 2PL by language consultants, and in playscripts written in Baba Malay, but does not appear in the language documentation dataset collected for the modern Baba Malay corpus.

A comparison of Table 2 with Table 3 shows that the pronominal forms overlap almost entirely with some exceptions. The forms that are prescribed for 2PL and 3PL in Table 3 are the calqued forms *lu-orang* and *dia-orang* ‘you-people’ and ‘they-people’ literally, but not the bare forms *lu* and *dia* which appear in the modern Baba Malay corpus data.

4.2 Variability in Baba Malay’s aspectual and tense system

The loss of variability can also be observed of Baba Malay’s aspectual and tense system. Aspect and tense markers are optionally used across both early and modern Baba Malay. Table 4 lists the various aspect and tense markers that can be found in early Baba Malay and their patterns of use in the corpus.

Table 4. Usage patterns of the aspectual and tense system of early Baba Malay

Aspect or tense marker	Function	Origin	Tokens	Percentages (based on function)
<i>sudah/sua</i>	perfective	Malay	1191	93.9%
<i>ada</i>	perfective	Malay (lexical) Hokkien (structural)	74	5.8%
<i>telah</i>	perfective	Malay	4	0.3%
<i>pernah</i>	experiential perfect	Malay	5	100.0%
<i>baru</i>	recent perfect	Malay	111	100.0%
<i>ada</i>	progressive	Malay (lexical) Hokkien (structural)	83	20.3%
<i>men-</i>	progressive	Malay	300	73.5%

Table 4. (continued)

Aspect or tense marker	Function	Origin	Tokens	Percentages (based on function)
<i>tengah</i>	progressive	Malay	25	6.1%
<i>ada</i>	habitual	Malay (lexical) Hokkien (structural)	44	100.0%
<i>nanti</i>	future	Malay	100	97.1%
<i>akan</i>	future	Malay	3	2.9%
<i>belum</i>	future (not yet)	Malay	12	100%

A considerable amount of variability is seen, with multiple ways of expressing the perfective, the progressive, and the future specifically. The perfective aspect can be expressed by using forms that are directly derived from Malay or with a structure transferred from Hokkien. While Malay-derived *sudah* and its contracted form *sua* is mostly used, instances of *ada*, whose form is derived from Malay, but whose structure is derived from Hokkien (N.H. Lee, 2022), and to a lesser extent, instances of Malay-derived *telah* are also observed. The progressive patterns similarly. While the most common form in early Baba Malay data is Malay-derived *men-*, *ada* is also used here, with its structure derived from Hokkien, but form derived from Malay. To a lesser extent, Malay-derived *tengah* is also found. There are then three ways to indicate the future. *Nanti* means ‘later’ or ‘to wait’ in Malay and can be used in the manner of an adverb to indicate that an event will occur in the future. *Akan* on the other hand indicates ‘will’ directly. The function of *belum* is different as it indicates that an event is expected to occur in the future but has not occurred at the time of speech.

The aspectual and tense system of modern Baba Malay is illustrated in Table 5.

Table 5. Usage patterns of the aspectual and tense system of modern Baba Malay

Aspect or tense marker	Function	Origin	Tokens	Percentages (based on function)
<i>sudah/sua</i>	perfective	Malay	664	98.7%
<i>ada</i>	perfective	Malay (lexical) Hokkien (structural)	9	1.3%
<i>pernah</i>	experiential perfect	Malay	5	100.0%

Table 5. (continued)

Aspect or tense marker	Function	Origin	Tokens	Percentages (based on function)
<i>baru</i>	recent perfect	Malay	25	100.0%
<i>ada</i>	progressive	Malay (lexical) Hokkien (structural)	23	100.0%
<i>ada</i>	habitual	Malay (lexical) Hokkien (structural)	20	100.0%
<i>nanti</i>	future	Malay	239	100.0%
<i>belom</i>	future (not yet)	Malay	13	100.0%

Similarly to the pronouns, a comparison of the aspectual and tense system of early Baba Malay with that of modern Baba Malay shows that the system is contracting overall. In addition to there being fewer tokens of aspect and tense markers used (998 tokens in the modern Baba Malay data set as compared to 1952 tokens in the late Baba Malay data set), there is also a clear reduction in the number of variants. *Men-* as a prefix indicating the progressive is no longer used for this purpose, leaving only *ada* to express the progressive. Malay-derived forms *telah* for perfective and *akan* for future ‘will’, that were minimally used in the early Baba Malay corpus at 0.3% of all perfective variants and 2.9% of future variants respectively, do not show up in the modern Baba Malay corpus as well, leaving *sudah/sua* and *ada* to express the perfective and *nanti* to express the future. While *men-* was most commonly used for the progressive at 73.5% of the overall progressive tokens in the early Baba Malay corpus, its function has been assumed by *ada*, which derives its form from Malay but its structure from Hokkien. The perfective and progressive uses of *ada*, which literally means ‘to possess’, and their congruence with the *Hokkien* possessive *u* are demonstrated by Example (1) and (2), and (3) and (4) respectively. The congruence between *u* and *ada* has also been previously noted by Ansaldo & Matthews (1999). Examples whose sources are not cited are collected from a consultant for each language and have been checked for well-formedness by at least three other consultants of that same language.

- (1) Baba Malay
Gua ada tutup.

1SG PFV close

‘I closed (the door).’

(N.H. Lee, 2014, p. 354)

(2) Hokkien

Goà ǔ khi hê-chiā-cām

1SG PFV go train-station

'I went to the train station.'

(Bodman, 1955, p.18)

(3) Baba Malay

Lu ada bikin apa?

2SG PROG make what

'You are making what?'

(N.H. Lee, 2014, p.371)

(4) Hokkien

Lì ǔ sīu bèq khi bou?

2SG PROG think want go NEG

'You are thinking of going, or not?'

(Bodman, 1958, p.2)

Notably, variants of other forms that still exist in the modern corpus, which are directly derived from Malay (perfective *sudah/sua* and future *nanti*) are forms that correspond structurally and functionally with similar Hokkien counterparts. While *tengah* for progressive shares the same distribution with progressive *ada*, it literally means 'middle', indicating that an event is in the midst of unfolding – there is no semantic equivalent that expresses the progressive in Hokkien.

Examples (5), (6) and (7) demonstrate the congruence across Baba Malay and Malay *sudah*, and Hokkien *í-keng*. Both literally indicate 'already' and both occur preverbally.

(5) Baba Malay

Gua sudah tau.

Mary already know

'Mary knew.'

(Lee, 2022, p.172)

(6) Malay

Saya sudah tahu.

1SG already know

'I knew.'

(7) Hokkien

Goà í-keng zai

1SG already know

'I know.'

While *telah* shares the same distribution as Hokkien *í-keng*, it is functionally different semantically, connoting formality in Malay, and does not appear in the corpus of modern Baba Malay at all, having made up only 0.3% of the perfective variants in early Baba Malay.

Nanti in Malay is used as a general adverbializer meaning ‘later’ or ‘to wait’. A similar counterpart in Hokkien is *tán*, which indicates ‘to wait’. Examples (8), (9) and (10) demonstrate a similar structural congruence across Baba Malay and Malay *nanti*, and Hokkien *tán*.

(8) Baba Malay

Nanti lu tanya.

later/wait 2SG ask

‘Later you will ask.’

(9) Malay

Nanti saya bertanya.

Later/wait 1SG ask

‘Later I will ask.’

(10) Hokkien

Tán goà mēng

Wait I come

‘Later I will ask.’

There is no equivalent counterpart of *akan* ‘will’ in Hokkien, which comprised only 2.9% of the future variants in early Baba Malay.

As opposed to the content on pronominals, the aspectual and tense system of Baba Malay is less expanded as a whole in revitalization material. In the guide-book, *Speak Baba Malay: The Easy Way*, verbs are described as “not (necessarily) chang[ing] forms to indicate tense”, as in *Gua kreja* ‘1SG work’ to mean ‘I work or am working; or I work or was working’ (P. Chan, 2007, p.21). This corresponds with the observation that overall, fewer aspect and tense markers appear in the modern Baba Malay data; verb markers being used more optionally than in the early Baba Malay dataset.

5. Discussion

The loss in variability that has become apparent in modern Baba Malay is not random. The variants that are lost are of three sorts. First, the Malay-derived variants that were used negligibly in early Baba Malay are lost. These include the 1SG pronoun *aku* and 2SG pronoun *anda*, which accounted for less than 1% of the total number of tokens for each function. Also, in the case that there is competition between the various forms for the same function, those that were derived solely from Malay, but do not correspond functionally to anything in Hokkien, are also lost. These include *men-* as a progressive prefix, *tengah* as another progressive

marker, *telah* as a perfective marker, and *akan* as a future marker. Finally, even where form and function overlap in both substrate and lexifier language, semantically more opaque forms that involve more distinctions have been lost, such as in the case of the first person plural inclusive and exclusive forms.

Explanations for the loss of some of the early variants in Baba Malay may be found in conventional language change, wherein words can fade altogether from a language (Campbell, 2020), without that language necessarily being in any peril. In the case of Baba Malay, early variants that were in very low usage could have fallen out of use due to natural change, wherein more salient and dominant forms become the only forms. However, this does not explain the loss of forms that were more frequently used, such as *men-* as a progressive prefix.

Crucially, losses do not happen inexplicably, and a primary reason for the loss of vocabulary can be found in historical changes in society (Campbell, 2020). Explanations for the patterns of variant loss in Baba Malay can be found in developments that have occurred in the social narrative of the Peranakans. First, Peranakans were once much more multilingual and proficient in Malay and/or Hokkien, as evidenced by the wide creole continuum that existed in the turn of the 20th century (N.H. Lee, forthcoming). The bilingual education policy, introduced in 1966 (see Section 2) meant that Peranakan students post-1966 were and are mandated to learn Mandarin as a second language, as opposed to Malay as a possible second language of choice. Census data of 2020 also shows that only 11.8% of the Chinese population (under which the Peranakans are subsumed) frequently speak a Chinese dialect at home, which would include Hokkien (Department of Statistics Singapore, 2020). Likewise, in the same census data, only 2.0% of the Chinese population responded that they speak another language that is not English, Mandarin, or a Chinese dialect – the chance of them speaking Malay at home is extremely low. Overall, these changes in the linguistic circumstances of the Peranakans have repercussions on Baba Malay, and one of the ways in which these changes manifest is in its loss in variability. Malay-derived forms that were already used negligibly in early Baba Malay are completely dropped in modern Baba Malay, with speakers plausibly no longer as aware of these forms that exist in the lexifier language, Malay. The loss in first person plural inclusive and exclusive distinctions appears to be part of a more general language change motivated by semantic transparency and fewer oppositions. Such a change is most probably further triggered by decreasing proficiency in both of Baba Malay's component languages in which these distinctions are observed, as well as continued English dominance among the Peranakans. This is similar to Paliwala's (2020) observations of the loss of inclusive-exclusive pronominal distinctions in Tok Pisin due to the influence of English, a language in which these distinctions are not observed.

The loss of less commonly used Malay-derived variants can perhaps also be viewed in the same perspective as the loss of variants derived solely from Malay that do not correspond functionally with Hokkien counterparts. It is possible that Malay-derived forms that have equivalents in Hokkien, such as perfective *sudah/sua* (which finds an equivalent in *í-keng*), are reinforced and are less likely to be lost in Baba Malay because of that equivalence. However, it may also be that Malay forms that are outrightly Malay, with no correspondences in Hokkien, are dispreferred when there are other options.

The reasons for the dispreference of these variants might stem from the observation that when given a choice, Baba Malay speakers prefer to avoid sounding too Malay. In a previous matched-guise study on matching coarse and refined pairs in Baba Malay (coarse forms here being words that end with [-al], [-ar], and [-as], and refined forms being their counterparts that end with [-ε]), younger community members did not identify the coarse forms as being spoken by a Baba Malay speaker. These coarse forms also occur in Malay, but with neutral (not coarse or refined) meanings. The results suggested that younger speakers of Baba Malay desired to keep the language, culture and community distinct from that of others, particularly the Malays, who also use the same forms.





In the face of language endangerment, it is more important than ever for these speakers to emphasize their unique identities as Baba Malay speakers (N. H. Lee, 2020). While the current study does not focus on the effects of age, the comparison between early and modern Baba Malay data seems to reflect a similar trend. The early Baba Malay corpus of the late 1800s and early 1900s reasonably represents the language as it was used in its heyday, when the language thrived to such an extent that literature was produced in it. The modern corpus on the other hand, was produced as part of a language documentation project, which was in turn developed in response to the language's endangerment, and it thus represents a period in which the language is under threat. Speakers captured in the modern dataset may attempt to avoid sounding too Malay, for reasons of keeping their identity distinct, perhaps as a means of ensuring that the language does not simply fade, much less into the mainstream Malay variety that exists in the same socio-environment. Further sociolinguistic work that goes beyond the previous language documentation work and probes the issue in a more targeted manner can be useful.






Finally, it is interesting to consider the effect that language revitalization may have on variation. To that effect, it has been shown that “new speakers” can themselves be “agents of sociolinguistic change in variationist terms” (Kasstan, 2017, p.7). For instance, in a study on new speakers of Scottish Gaelic, speakers who did not acquire the language via intergenerational transmission preferred using weakly rhotic or non-rhotic variants where native speakers would have other-





wise produced palatalised rhotics in the word final position – this phenomenon was not driven by low proficiency, but rather that the new speakers viewed the native-speaker target as inauthentic for themselves (Nance, McLeod, O'Rourke, & Dunmore, 2016). In contrast, interspeaker variationist work on New York Yiddish showed that the Yiddishists who are overtly committed to speaking “correct” Yiddish produced significantly more normative number agreement than the Hasidic Jews, who do not share this overt goal (Bleaman, 2022).

In the case of Baba Malay, it may be interesting to consider the pedagogical material that is available. It is not the case that the loss of variability was brought upon by language revitalization efforts, given that the speakers who contributed to corpus of modern Baba Malay are proficient in the language as it is currently spoken, but conceivably, such material can further reinforce the lack of variability. Standardization becomes part of the process of language revitalization, as rules for language usage are prescribed (see for example England, 1996 on Maya language maintenance in Guatemala). The question of which form of the language should be taught in is an inevitable one, with different approaches entailing consequences of different sorts (Whaley, 2011). In the case of Baba Malay, observations of the available pedagogical material show that these are even more restrictive than what is currently reflected in the corpus of modern Baba Malay usage. For example, while bare forms *lu* ‘2SG’ and *dia* ‘3SG’ have also been used by speakers to refer to 2PL and 3PL respectively, these do not appear as prescribed norms, and the lack of variability in Baba Malay may be further consolidated. The actual consequences of language revitalization efforts on variability await to be seen, with more community members likely to be learning the language outside of traditional means of language transmission.

References

-  Ansaldo, Umberto, Lim, Lisa, & Mufwene, Salikoko S. (2007). The sociolinguistic history of the Peranakans: What it tells us about “creolization.” In Umberto Ansaldo, Stephen Matthews, & Lisa Lim (Eds.), *Deconstructing Creole. Typological Studies in Language*, 73 (pp. 203–226). Amsterdam: John Benjamins.
- Ansaldo, Umberto, & Matthews, Stephens (1999). The Minnan substrate and creolization in Baba Malay. *Journal of Chinese Linguistics*, 27(1), 38–68.
-  Bavin, Edith L., & Shopen, Tim (1991). Walpiri in the 1980s: An overview of research in language variation and child language. In Suzanne Romaine (Ed.), *Language in Australia* (pp. 104–117). Cambridge: Cambridge University Press.
-  Bickerton, Derek (1973). The nature of a creole continuum. *Language*, 49, 640–669.
-  Bleaman, Isaac L. (2022). Minority language maintenance and the production-prescription interface: Number agreement in New York Yiddish. *Journal of Sociolinguistics*, 26(2), 221–245.

- Bodman, Nicholas C. (1955). *Spoken Amoy Hokkien, Volume 1*. Kuala Lumpur: The Government of the Federation of Malaya.
- Bodman, Nicholas C. (1958). *Spoken Amoy Hokkien, Volume 2*. Kuala Lumpur: The Government of the Federation of Malaya.
- Campbell, Lyle (2020). *Historical Linguistics: An introduction* (Fourth edition). Edinburgh: Edinburgh University Press.
-  Campbell, Lyle, & Muntzel, Martha C. (1989). The structural consequences of language death. In Nancy Dorian (Ed.), *Investigating obsolescence: Studies in language contraction and death* (pp. 181–196). Cambridge: Cambridge University Press.
- Chan, Kenneth (2018). *Mari Chakap Baba: A comprehensive guide to the Baba Nyonya language*. Singapore: Gunong Sayang Association.
- Chan, Philip (2007). *Speak Baba Malay: The easy way*. Singapore: Baba Nyonya Sayang.
- Corcoran, Chris, & Mufwene, Salikoko (1998). Sam Matthew's Kittinian. What is it evidence of? In Philip Baker & Adrienne Bruyn (Eds.), *St Kitts and the Atlantic Creoles: The texts of Samuel Augustus Mathews in perspective* (pp. 75–102). London: University of Westminster Press.
- DeCamp, David (1971). Toward a generative analysis of a post-creole speech continuum. In Dell Hymes (Ed.), *Pidginization and Creolization of Languages* (pp. 349–370). Cambridge: Cambridge University Press.
- Department of Statistics Singapore (2020). *Singapore Census of Population 2020, Statistical Release 1: Demographic Characteristics, Education, Language and Religion*. Retrieved from http://www.singstat.gov.sg/publications/reference/cop2020/cop2020-sr1/census20_stat_release1
-  Donohue, Mark, & Smith, John C. (1998). What's Happened to Us? Some Developments in the Malay Pronoun System. *Oceanic Linguistics*, 37(1), 65–84.
- Dorian, Nancy C. (1977). The problem of the semi-speaker in language death. *International Journal of the Sociology of Language*, 12, 23–32.
- Dressler, Wolfgang U. (1972). On the phonology of language death. *Chicago Linguistic Society Papers*, 8, 448–457.
- Dressler, Wolfgang U., & Wodak-Leodolter, Ruth (1977). Language preservation and language death in Brittany. *International Journal of the Sociology of Language*, 12, 33–44.
-  England, Nora C. (1996). The role of language standardization in revitalization. In Edward F. Fischer & R. McKenna Brown (Eds.), *Maya Cultural Activism in Guatemala* (pp. 178–194). Texas: University of Texas Press.
- Fei, Hsin (1436). *Hsing-ch'a sheng-lan: The overall survey of the Star Raft. (Republished in South China and Maritime Asia. 4. Translated by John Vivian Gottlieb Mills. Edited and annotated by Roderich Ptak.)*. Wiesbaden: Harrassowitz Verlag.
- Gwee, William T. H. (2006). *A Baba Malay Dictionary: The First Comprehensive Compendium of Straits Chinese Terms and Expressions*. Singapore: Tuttle Publishing.
-  Kasstan, Jonathan (2017). New speakers: Challenges and opportunities for variationist sociolinguistics. *Language and Linguistic Compass*, 11(8), e12249.
-  Kasstan, Jonathan (2019). Emergent sociolinguistic variation in severe language endangerment. *Language in Society*, 48(5), 685–720.
- Le Page, Robert (1960). *Jamaican Creole*. London: Macmillan.

- Lee, Gwyneth A. M.-E. (1999). A descriptive grammar of spoken Peranakan (Honour's thesis). National University of Singapore, Singapore.
- Lee, Nala H. (2014). A grammar of Baba Malay with sociophonetic considerations (PhD dissertation). University of Hawai'i, Manoa.
- Lee, Nala H. (2019). Peranakans in Singapore: Responses to language endangerment and documentation. *Language Documentation & Conservation. Special Issue on Documentation and Conservation of Contact Languages in Southeast Asia and East Asia: Current Issues and Ongoing Initiatives*, 19, 123–140.
-  Lee, Nala H. (2020). Utilizing the matched-guise as a method of examining perceptual change in an endangered creole. *Applied Linguistics*, 42(2), 207–229.
-  Lee, Nala H. (2022). *A grammar of modern Baba Malay*. Berlin: Mouton de Gruyter.
- Lee, Nala H. (forthcoming). The early Baba Malay continuum. *Journal of Pidgin and Creole Languages*.
- Lim, Boon-Keng (1899). Straits Chinese Reform IV: Religion. *The Straits Chinese Magazine*, 3(12), 163–166.
- Lim, Boon-Keng (1917). The Chinese in Malaya. In W. Feldwick (Ed.), *Present day impressions of the Far East and prominent and progressive Chinese at home and abroad; the history, people, commerce, industries and resources of China, Hong Kong, Indo-China, Malaya and Netherlands India* (pp. 875–882). London: Globe Encyclopedia Co.
- Lim, Joo-Hock (1967). Chinese female immigration into the Straits Settlements 1860–1901. *Journal of the South Seas Society*, 22, 58–110.
- Lim, Lisa (2016). Multilingual mediators: The role of the Peranakans in the contact dynamics of Singapore. In Wei, Li (Ed.), *Multilingualism in the Chinese diaspora world-wide* (pp. 216–236). New York: Routledge.
- Ma, Huan (1416). *Ying-yai sheng-lan: The overall survey of the ocean's shores. Republished by the Hakluty Society, London in 1970, and reprinted by the White Lotus Press in 1997*. Translated by J. V. G. Mills.
- Meyerhoff, Miriam (2015, October 22). *Keeping a foot in the door: Variation and language documentation*. Presented at the New Ways of Analyzing Variation 44.
-  Nagy, Naomi (2009). The challenges of less commonly studied languages: Writing a sociogrammar of Faetar. In James N. Stanford & Dennis R. Preston (Eds.), *Variation in indigenous minority languages* (pp. 397–417). Philadelphia: John Benjamins.
-  Nance, Claire, McLeod, Wilson, O'Rourke, Bernadette, & Dunmore, Stuart (2016). Identity, accent aim, and motivation in second language users: New Scottish Gaelic speakers' use of phonetic variation. *Journal of Sociolinguistics*, 20(2), 164–191.
- Nathan, J. (1922). *The Census of British Malaysia*. London: Waterloo & Sons.
- Pakir, Anne (1986). A linguistic investigation of Baba Malay (PhD dissertation). University of Hawai'i, Manoa.
- Pakir, Anne (1994). Educational linguistics: Looking to the East. In James Alatis (Ed.), *Georgetown University Round Table on Languages and Linguistics* (pp. 370–383). Washington DC: Georgetown University Press.
- Paliwala, Adam A. H. B. (2020). Language contact and Tok Pisin. In Anthony P. Grant (Ed.), *The Oxford Handbook of Language Contact* (pp. 606–626). Oxford: Oxford University Press.

- Palosaari, Naomi, & Campbell, Lyle (2012). Structural aspects of language endangerment. In Peter K. Austin & Julia Sallabank (Eds.), *The Cambridge Handbook of Endangered Languages* (pp. 100–119).
- Purcell, Victor (1980). *The Chinese in Southeast Asia*. Issued under the auspices of the Royal Institute of International Affairs [by] Oxford University Press. Retrieved from <http://books.google.com.sg/books?id=yZmTnQEACAAJ>
- Rickford, John R. (1987). *Dimensions of a Creole Continuum*. Stanford: Stanford University Press.
-  Salmon, Claudine (1977). La littérature en malais romanisé des Chinois de Malaisie, première enquête. *Archipel*, 14(1), 79–109.
- Shellabear, William G. (1913). Baba Malay: An introduction to the language of the Straits-born Chinese. Attached as appendix in John R. Clammer (1980). In *Straits Chinese Society* (pp. 153–165). Singapore: Singapore University Press.
- Siah, U-Chin (1848). General sketch of the numbers, tribes and avocations of the Chinese in Singapore. *Journal of the Indian Archipelago*, 2, 283–289.
- Skinner, G. W. (1996). Creolized Chinese Societies in Southeast Asia. In Anthony Reid & Kristine Aililunas-Rodgers (Eds.), *Sojourners and settlers: Histories of Southeast Asia and the Chinese* (pp. 51–93). Hawai'i: University of Hawai'i Press.
- Soh, Poh-Thong (1907). Concerning Our Girls. *The Straits Chinese Magazine*, 11(4), 139–143.
- Song, Ong-Siang (1967). *One hundred years' history of the Chinese in Singapore. First published in 1923*. Singapore: University of Malaya Press.
- Tan, Bonny (2007). *A Baba Malay Bibliography: A select annotated listing of sources on the Peranakan Chinese in Singapore and Malaysia*. Singapore: National Library Board.
- Tan, Chee-Beng (1979). *Baba and Nyonya: A study of the ethnic identity of the Chinese Peranakan in Malacca*. Cornell University, New York.
- Tan, Chee-Beng (1988). *The Baba of Melaka: Culture and identity of a Chinese Peranakan Community in Malaysia*. Petaling Jaya: Pelanduk Publications.
- Thurgood, Elzbieta (1998). *A description of nineteenth century Baba Malay: A Malay variety influenced by language shift*. University of Hawai'i, Hawai'i.
- Vaughan, J. D. (1879). *Manners and customs of the Chinese in the Straits Settlements*. Singapore: The Mission Press.
-  Wade, Geoff (2015). Chinese Engagement with the Indian Ocean during the Song, Yuan, and Ming Dynasties (Tenth to Sixteenth Centuries). In Michael Pearson (Ed.), *Trade, Circulation, and Flow in the Indian Ocean World. Palgrave Series in Indian Ocean World Studies* (pp. 55–81). New York: Palgrave Macmillan.
- Wang, Gungwu (1964). The opening of relations between China and Malacca, 1403–5. In John Bastin & R. Roolvink (Eds.), *Malayan and Indonesian Studies: Essays presented to Sir Richard Windstedt on his 85th birthday* (pp. 87–104). Oxford: Clarendon Press.
-  Whaley, Lindsay J. (2011). Some ways to endanger an endangered language project. *Language and Education*, 25(4), 339–348.
- Widodo, Johannes (2002). A celebration of diversity: Zheng He and the origins of the pre-colonial coastal urban pattern in Southeast Asia. *Journal of Southeast Asian Architecture*, 6, 11–22.



- Wu, Degang, Li, Peter Y., Bangfen, Pan, Tiang, Zenia, Dou, Jinzhuang, Williantarra, Ivanna, ... Chaolong, Wang (2021). Genetic Admixture in the Culturally Unique Peranakan Chinese Population in Southeast Asia. *Molecular Biology and Evolution*, 38(10), 4463–4474.
- Yoong, S. K., & Zainab, A. N. (2002). Chinese literary works translated into Baba Malay: A bibliometric study. *Malaysian Journal of Library & Information Science*, 7(2), 1–23.
- Yusoff, Radiah (2007). Translating kinship terms to Malay. *Language and Communication Translation Journal*, 11(3).

Abstract (Baba Malay)

Kalo bahasa mo ilang, tu bahasa mesti ada variasi. Sini, kita tanya kalo Bahasa Melayu Baba kat Singapore pun ada variasi. Kerja baru tunjok ni bahasa kreol kurang variasi daripada kreol biasa. Bandingkan dulu lama mia Bahasa Melayu Baba sama sekarang mia Bahasa Melayu Baba, ni artikel tunjok sekarang mia bahasa kurang variasi daripada dulu lama mia bahasa, pasair bahasa tukair untuk sebab natural, pasair bahasa tukair untuk sebab sosial.

Address for correspondence

Nala H. Lee
 Department of English, Linguistics and Theatre Studies
 National University of Singapore
 Block AS5, 7 Arts Link
 117570
 Singapore
 nala.lee@nus.edu.sg
<https://orcid.org/0000-0002-3807-5193>



Publication history

Date received: 23 December 2021

Date accepted: 1 August 2022