

QUANTIFYING MANGLISH CREOLE CONTINUUM AND ITS CORRELATION WITH-FORMALITY FACTOR

Cheng Siew May, Kuek Chian Shiun

Department of Languages and Linguistics, Faculty of Arts and Social Science,
Universiti Tunku Abdul Rahman, Kampar, Perak, Malaysia

E-mail: chengsm@utar.edu.my

ABSTRACT

Under the backdrop of multi-cultural, multi-ethnicity and multi-lingual Malaysia, the Malaysian English or more commonly known as Manglish has bloomed into a broad creole spectrum. Bearing the features of distorted language and creative communication at the same time, the Manglish creole continuum can be categorised into acrolect, mesolect and basilect Manglish. To map this continuum, a sampling study was taken on Malaysians of various age, ethnicity and educational background. An assessment model based on mathematical and computing theory was then developed to quantify the level of mixture with local languages and culture. A number of formulation models were also proposed to translate corresponding environment affecting Manglish creole into numeric scales. Pearson Correlation Coefficient between various scales were computed and compared with the proposed models. Finally, hypotheses were derived on the existence of Malaysian creole continuum and predictions made of its future development.

Keywords: acrolect, mesolect, basilect, Manglish, creole continuum, communication

INTRODUCTION

Quantifying the unquantifiable is the holy grail of many artsy fields of studies while the tools to achieve such goal are parallel to the electronic analogue-to-digital converters that translate our fluid, fuzzy world into discrete, measurable data for our computers to process. Relationship between vaguely connected areas can then be calculated with mathematical models such as the correlation coefficient while formulas and computer coding can be derived to predict the behaviours. Finally, a consistent and replicable intelligent response can be automated. In this study, the goal is being executed in the context of Manglish creole continuum. It is therefore crucial to first understand the background of the continuum and the fundamentals of the components that construct our tools and algorithms.

In this article, Manglish or Malaysian English (an English-based creole spoken in Malaysia) will be used interchangeably with Manglish creole. Manglish

creole continuum refers to the acrolect, mesolect and basilect Manglish.

BACKGROUND

Multi-Cultural, Multi-Ethnicity Malaysia

Generally, ethnic groups in Malaysia can be classified into four major groups namely Bumiputera (inclusive of Malay and Indigenous), Chinese, Indians and others. The recent statistics revealed that the population of Malaysia in 2019 was 32.6 million of which Bumiputera, the main ethnic constituted 69.3 per cent of the total population followed by Chinese and Indians at 22.8 and 6.9 per cent respectively (Department of Statistics Malaysia, 2020). Malaysia is a democratic country that allows its citizens to freely practise ethnical cultures and traditions. Pursuing education under various language-bases and the use of multiple languages in mass-media, formal and informal settings are part of the national heritage and identity of the country.

Education Policy and Multi-Lingual Malaysia

When Malaysians declared independence in 1957, *Bahasa Malaysia* became the national language in the country. The National Education Policy, the Education Act 1961, was formed thus divided the schools in Malaysia into two types: (1) Malay medium schools as national schools and (2) Vernacular and English medium schools as national type schools. The national English schools were converted to national Malay schools in stages while national Chinese and Tamil schools were retained. In national Malay schools, *Bahasa Malaysia* gradually took over English as the medium of instruction and English became the second language in the country (Foo & Richards, 2004). English became solely a compulsory school subject that students need to take but not to require pass in examinations (Asmah in Gill, 2002).

Meanwhile, Malaysians were also influenced by other languages or dialects such as Hokkien, Cantonese, Malay and Tamil (Platt & Weber, 1980). Such phenomenon worsened the usage of English in Malaysians. Following that, the Malaysian government established the Education Act 1996 and Private Higher Education Institutions Act 1996 to allow English to be used as the medium of instruction for technical areas (Gill, 2002).

Acrolect, Mesolect, Basilect and the Mode-Switching Capability

According to Platt & Weber (1980), the continuum of varieties exists in post-creole situations, as in this case, the educated, uneducated as well as the substandard varieties in Malaysian English (Baskaran, 1987, 1994). Crystal (2003) describes them as mixture of sociolects instead of a single dialect. In Manglish creole continuum, acrolect is defined as the highest and most prestigious variety of a continuum, basilect being the lowest and least prestigious variety while mesolect is in-between both extremes (Stewart, 1969, as cited in R. Shuy, 1969).

Acrolectal Malaysian English is formally used and it is internationally intelligible (Gill, 2002). Baskaran (1994) rendered it somehow possible for Malaysians to use acrolectal Malaysian English that is very close to the British, American and Australian Englishes for educational and official purposes except for the accent. As for mesolectal 'Malaysia dialect English', according

to Gill (2002), the dialectal Malaysian English is used informally and it is nationally intelligible. The basilect (the 'low' social dialect), on the other hand, is used informally and colloquially as a 'patois', and shades into a pidgin used mostly, for instance, by village pedlars when talking to tourists and other potential customers (Baskaran, 1994).

According to Baskaran (1987), Malaysians feel free to use the variants not because they do not know any other but because they prefer to use their own. Kachru in Gill (2002) would agree that speakers of a language are indeed able to switch from a variety to another, depending on the participants as well as the context. Gill (2002) analysed that majority of the Malaysians in the 1960s and 1970s who were educated through English-medium education were able to switch between acrolect, mesolect and basilect in different situations for varied functions and purposes. However, in the 1980s and 1990s, the number of the internationally intelligible acrolects reduced while the nationally intelligible mesolects and basilects rose. Many speakers did not have the ability to switch from the mesolect or basilect to acrolect, suggesting that being highly educated or just being educated does not guarantee them speaking the acrolectal variety (ibid, p. 58) in such 'unidirectional downward switch' (Vethamani, 1996). Moreover, English is becoming more like a foreign language rather than a second language for the local people (Platt, Weber, Ho, & Crystal, in Preshous, 2001).

Hamming Distance

Hamming distance is a mathematical-computing theory in detecting, calculating and correcting error in coding (Hamming, 1950). It measures the minimum number of correction steps required when an error is detected. In this article, Hamming distance is defined as the minimum number of words to be added or removed to elevate a sentence to acrolect status.

Algorithmic State Machine

Algorithmic state machine (ASM) is a type of flow-chart that defines the sequential operations of a digital system (Christopher, 1973). ASM is used here to map the prestige of a Manglish sentence according to its Hamming distance.

Pearson Correlation Coefficient

Pearson Correlation Coefficient or Pearson Correlation is a statistical method used to measure the linear correlation between two sets of variables (Pearson, 1896). Its equation produces the value between +1 to -1, whereby 1 denotes a complete positive linear correlation, 0 denotes no linear correlation and -1 denotes absolute negative linear correlation. The Pearson Correlation is used here to gauge the relationship of Manglish Creole with various environment settings and attitudes. Pearson Correlation can be described using the following formulation:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \quad (1)$$

where n , x and y denote population size and two allegedly correlated variables respectively.

Objectives of this Study

Studies have been done on Malaysian English including studies on people's attitudes towards Manglish, (e.g., Abdul Razak, 2015; Zhia, 2015) and impact of language on learners' identities (Lee, 2003). To our best knowledge, this is the first study that seeks to quantify the Manglish Creole continuum among Malaysian speakers using mathematical and computing approach

METHODS

Participants

Participants in this study were Malaysians from all walks of life (i.e. the general public) who used English as one of their medium of communication. The participants ranged from secondary and tertiary students; working professionals from multiple industries to those workers in general servicing line. The population size was 173 participants (50% males and 50% females) whose age group ranged from 18 to 72 years old with the average of 29.14 years old. There were 36% participants from the age group of 18-22 years old, 25% from 23-30 years old, 27% from 31-40 years old, 9% from 41-50 years old and 3% from above 50 years old.

Ethnicity was broken down into 40% for Malays, 45% for Chinese, 10% for Indians, 2% for Indigenous and 3% for others. This might not seem to be parallel with the Malaysian demographics in general. However, the

sampling group displayed good distribution in mother-tongue and their most spoken language. In terms of mother-tongue, 43% were from Malay-medium, 37% from various Chinese dialects (inclusive of Mandarin, Cantonese, Hokkien etc.), 7% from Tamil, 8% from English and 5% from other languages background. As for the most spoken language category, 22% were English-speaking group, 40% Malay-speaking, 31% Chinese-speaking, 6% Tamil and 1% for other speaking groups.

In the background education system category, 66% of the participants had undergone Malay-medium schools, 29% had undergone Chinese-medium schools, 2% were from Tamil-medium, 2% were from English-based and 1% was from other languages based schools.

Materials and Data Analysis

The main instrument used for this study was self-completed questionnaires being distributed to general public in soft or hard-copy forms. Such survey design was suitably done in correspondence with Creswell's (2003) claim that "a survey design provides a quantitative or numeric description of trends, attitudes, or opinions of a population by studying a sample of that population". This study was also done alongside other Malaysian English studies ranging from a corpus study on Malaysian English newspaper, an examination on sociolinguistics status of Malaysian English to efforts in understanding Manglish jargon used in social media (Rusli et al., 2018; Tan, 2009; Vollmann & Soon, 2019). As such, the questionnaire in this study was broken into four parts, starting with preliminary information, definitions and guidelines in conducting the survey and the second part concerning demographic information of participants. Henceforth, the main focus was on the remaining sections, with section three looking into the severity, frequency and occurrence of Manglish.

In the study of Manglish creole continuum, the ratio among acrolect, mesolect and basilect Manglish used in formal, semi-formal and informal occasions was gathered and compiled into a matrix. Participants were inquired to sort a list of scenarios according to their levels of formality. A normalised scale from 0% to 100% was then being formulated by factoring the collective perception towards each scenario with a weight-factor. 100 was assigned for formal occasion, 50

for semi-formal and 0 for informal votes. The Formality Factor (FF) was thus computed using the following formula:

$$FF = \frac{[(F \cdot 100) + (SF \cdot 50) + (I \cdot 0)]}{(T \cdot 100)} = [F + 0.5SF]/T \quad (2)$$

where F, SF and I denote formal, semi-formal and informal votes respectively. T denotes total samples. All variables were in the unit of percentage (%).

Five scales of Formality Factor were being defined as follows: 100-80% being formal, 79.99-60% being weak-formal, 59.99-40% as semi-formal, 39.99-20% as weak semi-formal and 19.99-0% in the case of informal occasion. The scales could be shown more clearly in the following formulation:

$$FF = \begin{cases} 80\% \leq x \leq 100\% & , \text{ formal} \\ 60\% \leq x < 80\% & , \text{ weak - formal} \\ 40\% \leq x < 60\% & , \text{ semi - formal} \\ 20\% \leq x < 40\% & , \text{ weak semi - formal} \\ 0\% \leq x < 20\% & , \text{ informal} \end{cases} \quad (3)$$

Pearson Correlation Coefficient is also used to compute the strength of correlation between formality setting and the usage of Manglish creole.

Next, the analysis of Manglish creole continuum would begin with the compilation of Sentences-Under-Test (SUT) by picking a list of common Manglish sentences randomly used on the social media. An analysis was conducted to group the sentences with similar traits and degree of mutation based on the defined Malaysian culture. Next, each group was being given a title to describe how severe it has swayed from Standard English. Since Manglish creole - like any other language creole in the world - is a continuum, five categories have been established to provide a better mapping for the sentences-group mentioned earlier. The categories were acrolect, acro-mesolect, mesolect, meso-basilect and basilect, with acrolect having compatible standing with Standard English while basilect being the worst-case corruption possible for an English sentence to be remained recognisable as an English sentence (Platt, 1975). A "corruption scale", which shall be referred to as the Manglish Creole Factor (MCF), was being assigned

for each category, with 0-20% for acrolect, 20-40% for acro-mesolect, 40-60% for mesolect, 60-80% for meso-basilect and 80-100% for basilect. The classification could be more accurately described in the following mathematical equation:

$$MCF = \begin{cases} 80\% \leq x \leq 100\% & , \text{ basilect} \\ 60\% \leq x < 80\% & , \text{ weak - basilect} \\ 40\% \leq x < 60\% & , \text{ mesolect} \\ 20\% \leq x < 40\% & , \text{ acro - mesolect} \\ 0\% \leq x < 20\% & , \text{ acrolect} \end{cases} \quad (4)$$

The system to define and classify Manglish creole continuum established in this article shall thus be referred to as the *Ernest-SiewMay Model for Manglish Creole Assessment* in the rest of the report. Further description of this system shall be discussed in the next section.

Having established a model to define Manglish, the following phase was to test the validity of the model. Two groups of panel have been established, each with 7 members comprising university lecturers in linguistics field, veteran secondary school teachers with more than 30 years of teaching and 20 years of assessment experience in major nationwide English examination besides one working professional from the field of Engineering (to enhance analytical capability). Both panel groups, deemed as the "experts", have compatible credentials in the command of English and were inquired to scale each sentence within SUT. However, the first group has been briefed with *Ernest-SiewMay Model for Manglish Creole Assessment* prior to analysing SUT while the second group was left with individual intuition to scale the sentences within SUT. The grading process was three-fold, starting with each panel individually assigning a category for each Manglish sentence. Then, the verdicts for all panels were accumulated. Second phase was executed by removing one vote from each extreme end (similar to the grading system of Olympics diving event). Finally, the number of votes were converted to percentage and factored with a weight-constant to compute Manglish Creole Factor (MCF). In mathematical form, it would be described as the following:

$$MCF_{panels} = \frac{[(A \cdot W_A) + (AM \cdot W_{AM}) + (M \cdot W_M) + (MB \cdot W_{MB}) + (B \cdot W_B)]}{(T \cdot 100)} \quad (5)$$

where A, AM, M, MB and B denote the votes for acrolect, acro-mesolect, mesolect, meso-basilect and basilect in percentage. W_A , W_{AM} , W_M , W_{MB} and W_B denote the weight-constant for each corresponding category. T denotes total votes.

In the selection of weight constant, one option is to use the median of each scale, with $W_A = 10$, $W_{AM} = 30$, $W_M = 50$, $W_{MB} = 70$ and $W_B = 90$. However, this technique is inappropriate as there is a finite upper and lower limit to the scale, resulting in distortion in computation as the population curve approaches the edges. The problems are illustrated in Figure 1(a)–1(d).

As a matter of fact, votes are likely to be manifested in a normal distribution as shown in Figure 1(a). However, boundary votes are likely to experience "stack-effect" when the curves swing toward the edges as there are no further categories to accommodate them. This causes the stacked votes to be factored with disproportionate weight-constant, causing a diminishing contribution in the computation of overall behaviour. A biased-weight was therefore employed in the MCF computation

to compensate the stack-effect, resulting in the weight-constant of $W_A = 0$, $W_{AM} = 25$, $W_M = 50$, $W_{MB} = 75$ and $W_B = 100$. The MCF formula was therefore revised as:

$$\begin{aligned} MCF_{panels} &= [(A \cdot 0) + (AM \cdot 25) + (M \cdot 50) \\ &\quad + (MB \cdot 75) + (B \cdot 100)] / (T \cdot 100) \\ &= [0.25AM + 0.5M + 0.75MB \\ &\quad + B] / T \end{aligned} \quad (6)$$

To further illustrate the advantage of Biased Weight over Median Weight, consider a certain SUT receiving 7 votes as Basilect (100% vote). Removing one vote from each extreme end (phase 2) and computing using Biased Weight would result in $MCF = (5)/5 = 100\%$. Likewise, computing using Median Weight would result in $MCF = (0.9 \times 5)/5 = 90\%$, therefore giving the implication that it is "not completely Basilect".

The MCF ratings for SUT given by both panels were compared with the expected outcome of *Ernest-SiewMay Model for Manglish Creole Assessment*.

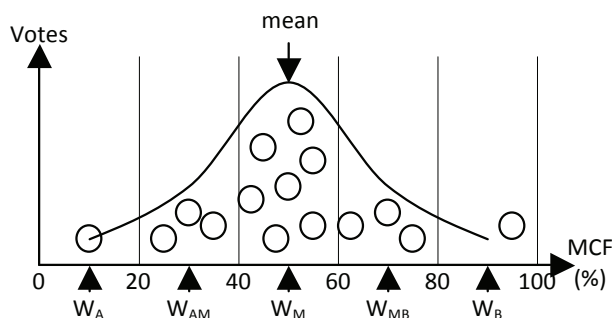


Figure 1(a) Normal Distribution with Median Weight

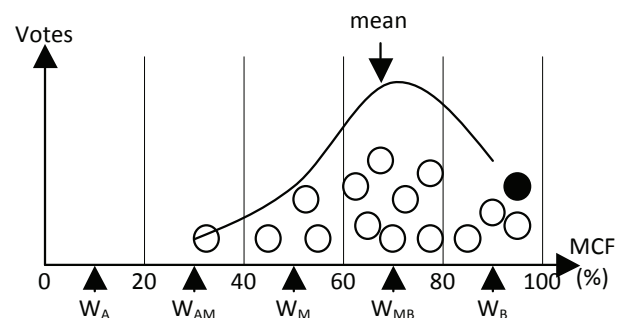


Figure 1(b) Partially Distorted Normal Distribution with Median Weight

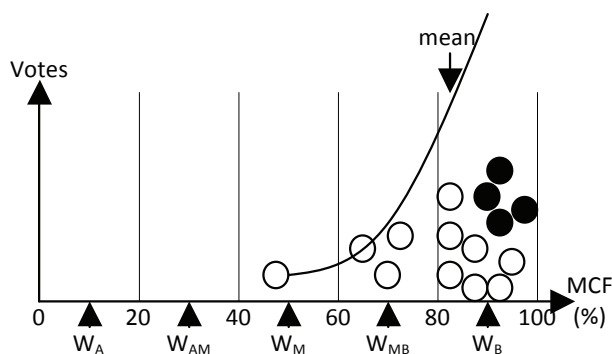


Figure 1(c) Distorted Normal Distribution with Median Weight

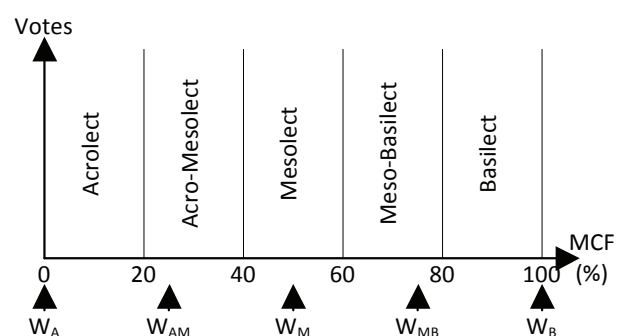


Figure 1(d) Biased Weight

To assess the correlations, Perception Error Factor (PEF) was introduced whereby the MCF for each test-sentence was being compared with expert panel's verdict. If the MCF was within range, PEF would be awarded 0% for the particular test sentence. However, if MCF was out of range, the gradient between closest boundary and MCF was taken as PEF. For example, if sentence A was defined as mesolectal, MCF was to be in the range of 40-60%. However, if MCF was being computed as 32%, PEF would be taken as 40%-32% = 8%. If MCF was found to be 75% then PEF = 75%-60% = 15%. Finally, PEF for each category was being averaged to determine overall panel's capability and consistency in differentiating Manglish creole continuum, with lower PEF showing better performance. PEF could be described in the following formula:

$$PEF = |MCF_{px} - MCF_{model\ boundary}| \quad (7)$$

where MCF_{px} denotes the MCF computed from public or panels while $MCF_{model\ boundary}$ denotes the closest boundary defined by *Ernest-SiewMay Model for Manglish Creole Assessment* to MCF_{px} . Pearson Correlation Coefficient is also computed and compared to PEF.

To conclude the Manglish creole continuum analysis, the sampling result of Malaysian general public was summed and the corresponding MCF and PEF were generated. This in turn determined the validity of the claims made by general public in their use of various Manglish creole on various occasions. In view of the much larger population size, Manglish creole continuum was simplified to only three categories, being the acrolect, mesolect and basilect. Consequently, the MCF computation was also being simplified to:

$$MCF_{panels} = [(A \cdot 0) + (M \cdot 50) + (B \cdot 100)] / (T \cdot 100) = [0.5M + B] / T \quad (8)$$

RESULTS

Manglish Creole Occurrence and Frequency

In general, the use of all Manglish variants was found in every genre of occasion with acrolect Manglish having major occurrences in formal occasions, mesolect Manglish in semi-formal occasions and basilect Manglish in informal situations. Yet, the bias towards a particular Manglish creole was most pronounced in formal settings with acrolect Manglish taking 69% of the overall distribution, a 60% difference from the

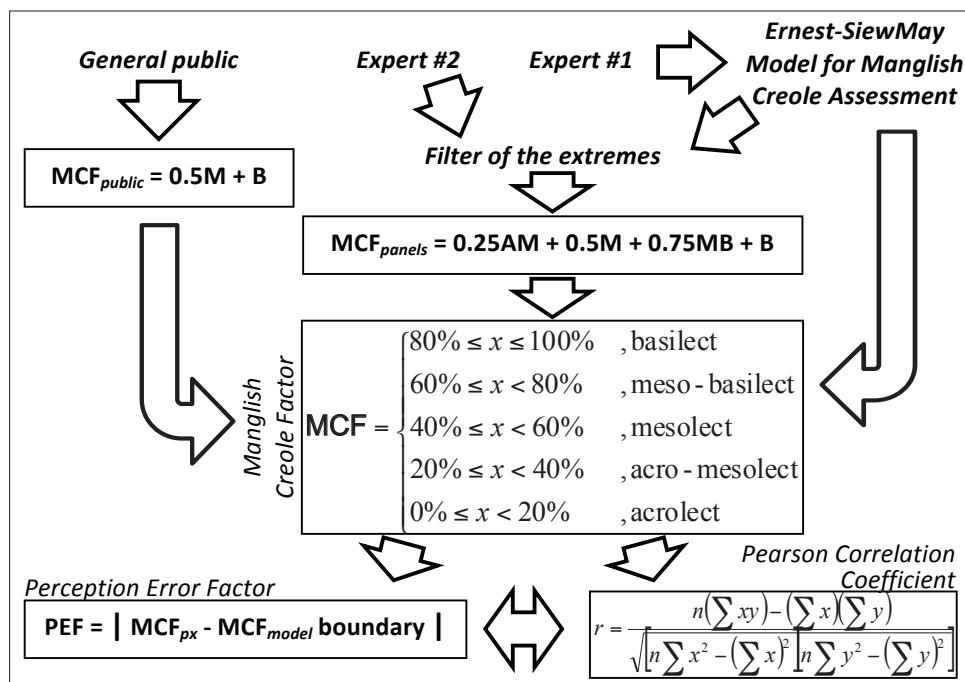


Figure 2 Manglish Creole Continuum Analysis

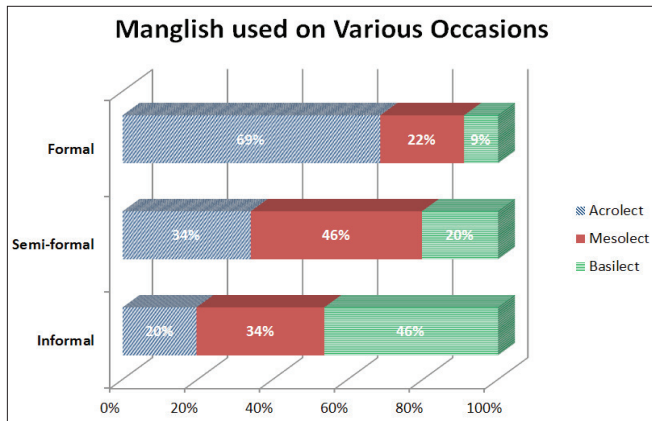


Figure 3 Manglish Used on Various Occasions

basilect sector. The continuum in semi-formal scenario was likewise more evenly distributed with the gradient of 26% between the biggest sector (mesolect, 46%) and smallest sector (basilect, 20%). As for informal occasions, a difference of 26% was observed in the acrolect-basilect comparison. Malaysians in the sample population also exhibited reasonable agility in exercising various Manglish creole, with 49% difference in the use of acrolect Manglish and 37% variation in the use of basilect Manglish between formal-informal settings.

The data was converted into scatter plot (Figure 4) by assigning a relative weight of 1, 2, 3 to informal, semi-formal and formal settings as well as basilect, mesolect and acrolect Manglish respectively.

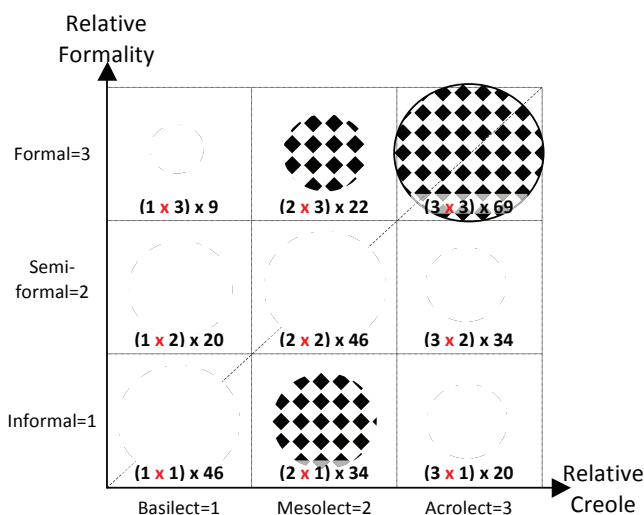


Figure 4 Scatter Plot of Formality Settings versus Manglish Creole

Each element of the Pearson Correlation Coefficient (PCC) can be computed as follows:

$$n(\sum xy) = 300 \left[\begin{matrix} (1 \times 3 \times 9) + (2 \times 3 \times 22) + (3 \times 3 \times 69) + \\ (1 \times 2 \times 20) + (2 \times 2 \times 46) + (3 \times 2 \times 34) + \\ (1 \times 1 \times 46) + (2 \times 1 \times 34) + (3 \times 1 \times 20) \end{matrix} \right] = 414600$$

$$\sum x = [(9 \times 20 \times 46) \times 1] + [(22 \times 46 \times 34) \times 2] + [(69 \times 34 \times 20) \times 3] = 648$$

$$\sum y = [(46 \times 34 \times 20) \times 1] + [(20 \times 46 \times 34) \times 2] + [(9 \times 22 \times 69) \times 3] = 600$$

$$n\sum x^2 = 300 \left[\begin{matrix} (1^2 \times 9) + (2^2 \times 22) + (3^2 \times 69) + \\ (1^2 \times 20) + (2^2 \times 46) + (3^2 \times 34) + \\ (1^2 \times 46) + (2^2 \times 34) + (3^2 \times 20) \end{matrix} \right] = 477000$$

$$n\sum y^2 = 300 \left[\begin{matrix} (3^2 \times 9) + (2^2 \times 22) + (1^2 \times 69) + \\ (2^2 \times 20) + (1^2 \times 46) + (3^2 \times 34) + \\ (1^2 \times 46) + (3^2 \times 34) + (2^2 \times 20) \end{matrix} \right] = 420000$$

$$\therefore r = \frac{n(\sum xy) = (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} = 0.4408$$

It seems that the hypothesis of using more prestigious English relative to the formality of environment is only moderately correlated (PCC = 0.4408). Whether this is due to capacity limit as argued by Gill (2002, p. 58) or personal choice as proposed by Baskaran (1987, p. 49) and Kachru in Gill (2002, p. 55) is left for future exploration.

Formality Factor of Various Occasions

In the aspect of formality settings, most work-related functions were defined as formal, with greater gradient in authority or internationalisation settings displaying a higher Formality Factor. "Teacher-students interaction in class" was at the boundary between formal and weak-formal, which may be due to the presence of a sizable tertiary student population (36%). A significant drop in formality was also observed in identical setting but different community – between 15-22% drop in Formality Factor when an international community was replaced with locals. On the other hand, a small but obvious difference was found between colleagues and customers' circles. Whether the 8% drop an intentional or subconscious inclination to bridge the gap with customers would open the doors for future research.

The rest of the results displayed unsurprising figures, with TV advertisements, dramas and programmes as

MALAYSIANS' PERCEPTION ON FORMALITY SETTING					
FORMALITY FACTOR	OCCASIONS				
98%	Principal speaking on stage				
95%	TV News				
95%	Meeting with international colleagues				
94%	Meeting with international customers				
81%	Teacher-students interaction in class				
80%	Meeting among Malaysian colleagues				
72%	Meeting among Malaysian customers				
58%	TV Advertisements				
47%	TV Dramas / Programmes				
43%	Teacher-students interaction after class				
5%	Hang-out (shopping, coffee-shop ...etc)				
5%	Communication with family members				
FF Scale	Informal	Weak SF	Semi-Formal	Weak Formal	Formal
	0-20%	21-40%	41-60%	61-80%	81-100%

Figure 5 Malaysians' Perception on Formality Setting

well as teacher-students interaction after class being placed under semi-formal setting. On the other hand, friendship and family interactions were at the lowest end of the list with 5% Formality Factor, almost hitting the ultimate end of the scale. The only note-worthy behaviour was the teacher-students interaction which still held a certain degree of formality, in or out of class.

Ernest-SiewMay Model for Manglish Creole Assessment

Upon analysing various Manglish sentences from social-media, the *Ernest-SiewMay Model for Manglish Creole Assessment* has been established based on the following principles:

1. Manglish is an independent entity in a sentence. While it contributes to the accumulative quality of a sentence, it does not correlate directly with other elements such as grammatical or terminology error induced by non-Malaysian culture. The overall quality of a sentence can be loosely described by the following mathematical model:

$$\text{Total Error of a Sentence, TES} = (G + T + M...) \quad (9)$$

where G, M, T denote the intensity of error contributed by grammar, terminology and Manglish.

To further illustrate the concept, consider the sentence "Improvident *obtuse* are his unbecoming". The correct manner to express this sentence should be "Improvident *and obtuseness* are his

unbecoming", which suggests a grammatical error. The situation can be illustrated in graphical form as shown in Figure 6(a). As for the case of "Please *gostan* your car", the correct sentence should be "Please *reverse* your car", as illustrated in Figure 6(b) since the word "*gostan*" is an invalid English word used only by locals. On the other hand, a compound error is demonstrated by the sentence "You want go where *eating*?". The correct sentence ought to be "Where would you like to go to eat?", where the phrase "you want go where" is parallel with Malay "*kamu mahu pergi mana*" or Chinese "你要去哪里" (Nǐ yào qù nǐ lǐ), which exhibits the influence of popular languages used in Malaysia. "...*eating*" is a genuine English grammatical error. This sentence therefore comprises multi-dimensional errors, as illustrated in Figure 6(c).

As a wrap-up, the first sentence would be considered acrolect Manglish since it does not comprise any Malaysian culture induced error. Sentences two and three would be non-acrolect Manglish with the severity to be defined in the next section. Likewise, the total error of an English sentence can be concluded as a vector, having not only the magnitude of contributing elements but also the direction of error with Manglish having its own independent axis.

2. The main pillars to define magnitude of Manglish error are sentence structure, terminology and correction effort. As the main purpose for the existence of spoken sentences is communication, distorted sentence structure can lead to misinterpretation while undefined terminology might result in meaningless sentence. However, undefined terminology error can easily be fixed with description in preliminary or subsequent sentences. The undefined terminology can then be broken into two categories: terms that carry some unique local meanings that will otherwise be lost when compromised or terms that can be easily replaced with a Standard English word without affecting the overall flavour. Replaceable terms are deemed unnecessary and therefore rated lower in Manglish creole continuum, meanwhile irreplaceable unique terms are just one step away from Standard English, therefore are highly regarded in Manglish creole continuum.

Table 1 Ernest-SiewMay Model for Manglish Creole Assessment

Manglish-variant	The sentences in this category generally...	Sample Manglish sentences
Acrolect	– have no sentence structure error with respect to English (except for “Improvident obtuse is his unbecoming”, which should have been “Improvident and obtuseness...”, which is a non-local induced error). – comprise words found in English dictionary (except “Manglish”, which is assumed to be defined by some prequel texts and “SMS” (should have been “text you”), which is an internationally known abbreviation rather than local influence). – can be understood easily on the international level.	Where shall we eat? I will SMS you. Do you serve coffee? Gosh! He is really late. Manglish is the lingua franca among Malaysians. Contrived gesticulation is a mettlesome strategy. He is the de facto president of the party. Improvident obtuse are his unbecoming.
Acro-Mesolect	– have no sentence structure error with respect to English (or slight concern as in the case of “correct, correct, correct...” and “listen, listen, listen”). – can be understood on the international level except description required for the keywords. – comprise words not found in English dictionary but are unique to local culture. The keyword loses meaning when being translated or cannot be translated easily without a lengthy description (i.e.: “sekolah agama” can be translated to “religious school” but is generally understood as “Islamic school” in local culture).	The government will set-up a <i>sekolah agama</i> here. The restaurant is serving <i>nasi lemak</i>. We want <i>kopi-O kau</i>. The <i>Malaysia Boleh</i> spirit is empowering her. Correct, correct, correct, correct, correct... Listen, listen, listen, listen, listen... That boy is a <i>banana</i>.
Mesolect	– have minor sentence structure error with respect to English. – can be understood on the international level. – comprise unnecessarily added words, mostly particles (due to local culture), which can be easily ignored without altering the overall meaning.	Please forgive me <i>la</i>. I’m sure he can solve the problem <i>one</i>. Hard work will lead to good result <i>ma</i>. You’re still required to work on Saturday <i>meh?</i> This <i>one</i>, we have it here. The students always like to do <i>like</i> that.
Meso-Basilect	– have no sentence structure error with respect to English. – can be understood on the international level except description required for the keywords. – comprise words not found in English dictionary but can be found in dictionary of certain local language (therefore, a “Manglish” rather than “some-other-lish”). Alternatively, the keyword produced an unclear meaning (as in the case of “You looked so blur”) that would otherwise be meaningful when translated into another language used locally. – contain keywords that can be replaced easily into English without altering the overall meaning (therefore, unnecessary use in the sentence).	You looked so <i>blur</i>. He is very <i>action</i>. This system is so <i>canggih!</i> <i>Wah lau!</i> She is so early. <i>Ai-yo!</i> He lost the gold medal. Would you like to <i>makan?</i> Please <i>gostan</i> you car. I would like to <i>tapau</i> for lunch. The children were so <i>syok</i>. I’m so <i>paiseh</i> to look for him.
Basilect	– have major sentence structure error or distorted sentence structure with respect to English. However, it is aligned with the sentence structure of another language used locally (therefore, a “Manglish” rather than “some-other-lish”). – are possible to be misinterpreted by people unfamiliar with the local culture.	Why you’re so like that? Like this also can? Why take so long to eat? Sure got one <i>la</i>. Where you want to <i>makan?</i> Where got? How can? Never mind, still got chance.

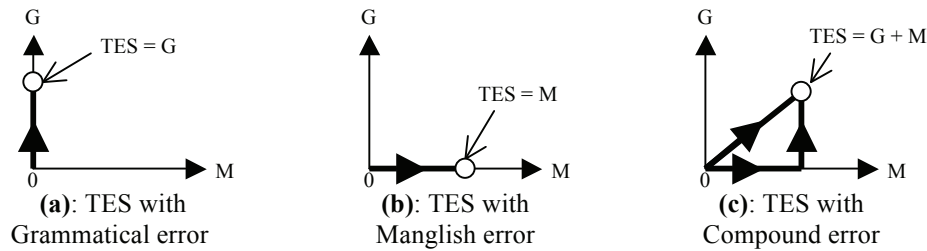


Figure 6 Total Error in a Sentence (TES)

In regard to correction effort, Hamming distance methodology shall be employed. In the sentences "Please forgive me *la*" and "You looked so *blur*", the former only requires the action of removing "*la*", therefore causing Hamming distance of one. The latter likewise requires the removal of "*blur*" followed by addition of the word "*confused*", resulting in Hamming distance of two. Sentences with mutated structure such as "*Why you're so like that?*" requires much longer Hamming distance for correction, therefore being placed last in the hierarchy, the basilect.

In a nutshell, sentences with no influence from Malaysian culture are deemed "Acrolect Manglish". These include sentences with grammatical errors that are not related to Malaysian culture. "Acro-mesolect Manglish" would likewise be sentences with no error caused by Malaysian culture and include local words not found in English dictionary.

However, the words must play an important role in describing some Malaysian phenomena, incidents or tastes that otherwise require lengthy description with alternative text. Besides, the term must be defined in prequel or subsequent discussions. The rest of the classifications would be based on the Hamming distance to elevate a sentence to the level of acrolect or acro-mesolect Manglish. Hamming distance of one classifies the "Mesolect Manglish", Hamming distance of two for "Meso-Basilect Manglish" while greater Hamming distance draws the territory of "Basilect Manglish". Figure 7 summarised the cognitive flow with an Algorithmic State Machine (ASM) chart.

Quantitative Analysis on Manglish Creole Continuum

Table 2 displays the list of SUT offered to panels and sample group. "Expert #1" denotes the panel group that has been briefed with *Ernest-SiewMay Model* for

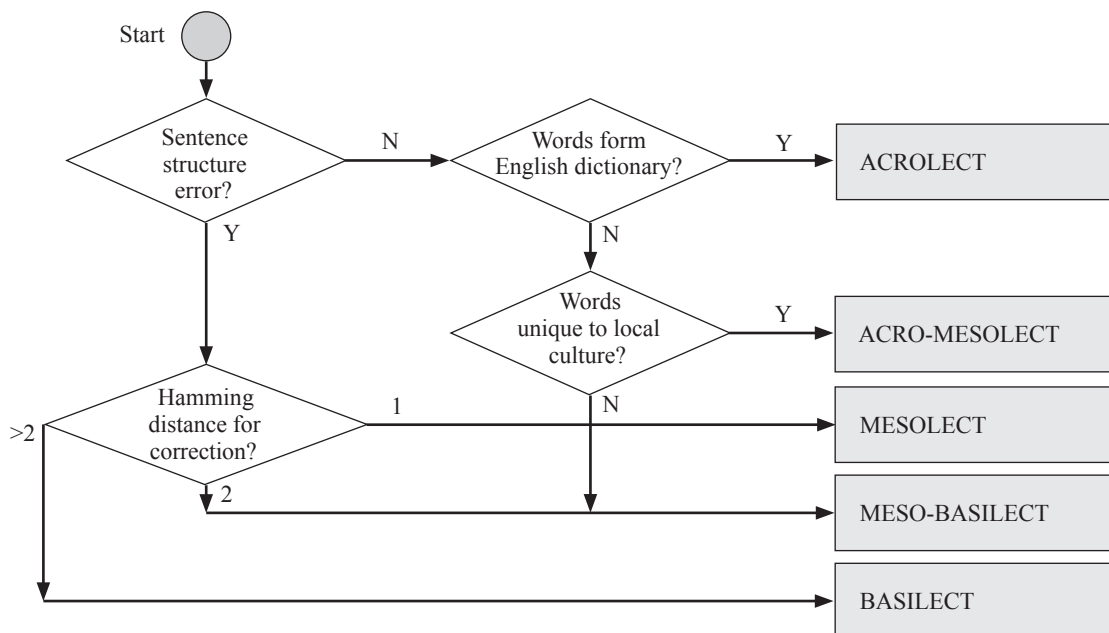


Figure 7 Ernest-SiewMay Model for Manglish Creole Assessment using ASM Chart

Manglish Creole Assessment prior to assessing SUT while "Expert #2" denotes the panels unexposed to the model. The result displayed a 4% difference between the average expert #1 MCF rating (54.0%) and average median MCF defined by guideline (50.0%). The error in classifying Manglish Creole as measured by PEF was low at 2.0%. MCF rating for experts #2 likewise showed somewhat similar outcome as the general public with 68.8% and 64.0% MCF score respectively. This shows that experts were 4.4% stricter than general public in assessing Manglish. Classification error was also similar with PEF of 11.3% for expert #2 and 10.5% for general public.

For expert #1, largest discrepancy was found in the case of the terms "*sekolah agama*" and "*kopi-O kau*" which demonstrated a split opinion on the importance of the terms, which would result in heavy consequence of them being graded as acro-mesolectal or meso-basilectal. For expert #2, largest discrepancy was found in two distinct areas. First was a heavy punishment imposed on the usage of "*SMS*", a term introduced by technological influence rather than Malaysian culture. Second was the usage of redundant notations such as "*one*", "*ma*" and "*like*".

In both cases, the penalty causes demotion by two grades in the scale of Manglish creole continuum, which may be too harsh for a Hamming distance = 1 error or error unrelated to Malaysia at all. For general public likewise, discrepancy was found in more than half of the cases. However, an unusual distain was recorded towards the work "*kopi-O kau*" and in lesser sense for "*SMS*" and "*Wah lau*".

In terms of the practicality of having a model to govern the assessment of Manglish creole, its absence created a gap of 45% (50-95% range) and 36% (49-85% range) from unguided experts and public against 20% (45-65% range) for guided experts in the case of having redundant particles such as "*one*", "*ma*", "*la*", "*meh*" and "*like*" in sentences. Another example was imposing unnecessary penalty on non-English term such as "*SMS*", up to 55% and 48% from unguided groups against 20% by guided group. Nevertheless, it was not sufficient to cover the ground of acceptable local terms, giving a range of 25% (60-35%), 30% (60-30%) and 48% (77-29%) by guided experts, unguided experts and general public respectively in the case of "*nasi lemak*", "*sekolah agama*" and "*kopi-O kau*". As a conclusive remark, it is highly recommended to have

Table 2 Assessment Results of Manglish Creole Based on Various Techniques

Sentences-Under-Test (SUT)	GUIDELINE		EXPERT #1		EXPERT #2		PUBLIC	
	MCF range	median MCF	MCF	PEF	MCF	PEF	MCF	PEF
17. Do you serve coffee?	0-20	10%	0%	0%	10%	0%	13%	0%
1. Where shall we eat?	0-20	10%	0%	0%	15%	0%	20%	0%
18. Gosh! He is really late.	0-20	10%	15%	0%	20%	0%	34%	14%
10. I will <i>SMS</i> you.	0-20	10%	20%	0%	55%	35%	48%	28%
14. The restaurant is serving <i>nasi lemak</i> .	20-40	30%	35%	0%	30%	0%	29%	0%
9. The government will set-up a <i>sekolah agama</i> here.	20-40	30%	50%	10%	55%	15%	60%	20%
20. We want <i>kopi-O kau</i> .	20-40	30%	60%	20%	60%	20%	77%	37%
5. I'm sure he can solve the problem <i>one</i> .	40-60	50%	45%	0%	95%	35%	82%	22%
7. Hard work will lead to good result <i>ma</i> .	40-60	50%	45%	0%	95%	35%	69%	9%
4. Please forgive me <i>la</i> .	40-60	50%	50%	0%	75%	15%	76%	16%
8. You're still required to work on Saturday <i>meh</i> ?	40-60	50%	50%	0%	75%	15%	77%	17%
19. The students always like to do <i>like</i> that.	40-60	50%	50%	0%	90%	30%	65%	5%
16. <i>Wah lau</i> ! She is so early.	40-60	50%	60%	0%	85%	25%	85%	25%
15. This <i>one</i> , we have it here.	40-60	50%	65%	5%	50%	0%	49%	0%
11. This system is so <i>canggih</i> !	60-80	70%	75%	0%	75%	0%	84%	4%
6. Why take so long to eat?	80-100	90%	75%	5%	90%	0%	68%	12%
13. Where you want to <i>makan</i> ?	80-100	90%	85%	0%	100%	0%	84%	0%
2. Why you're so like that?	80-100	90%	100%	0%	100%	0%	81%	0%
3. Like this also can?	80-100	90%	100%	0%	100%	0%	91%	0%
12. Sure got one <i>la</i> .	80-100	90%	100%	0%	100%	0%	88%	0%
AVG		50.0%	54.0%	2.0%	68.8%	11.3%	64.0%	10.5%

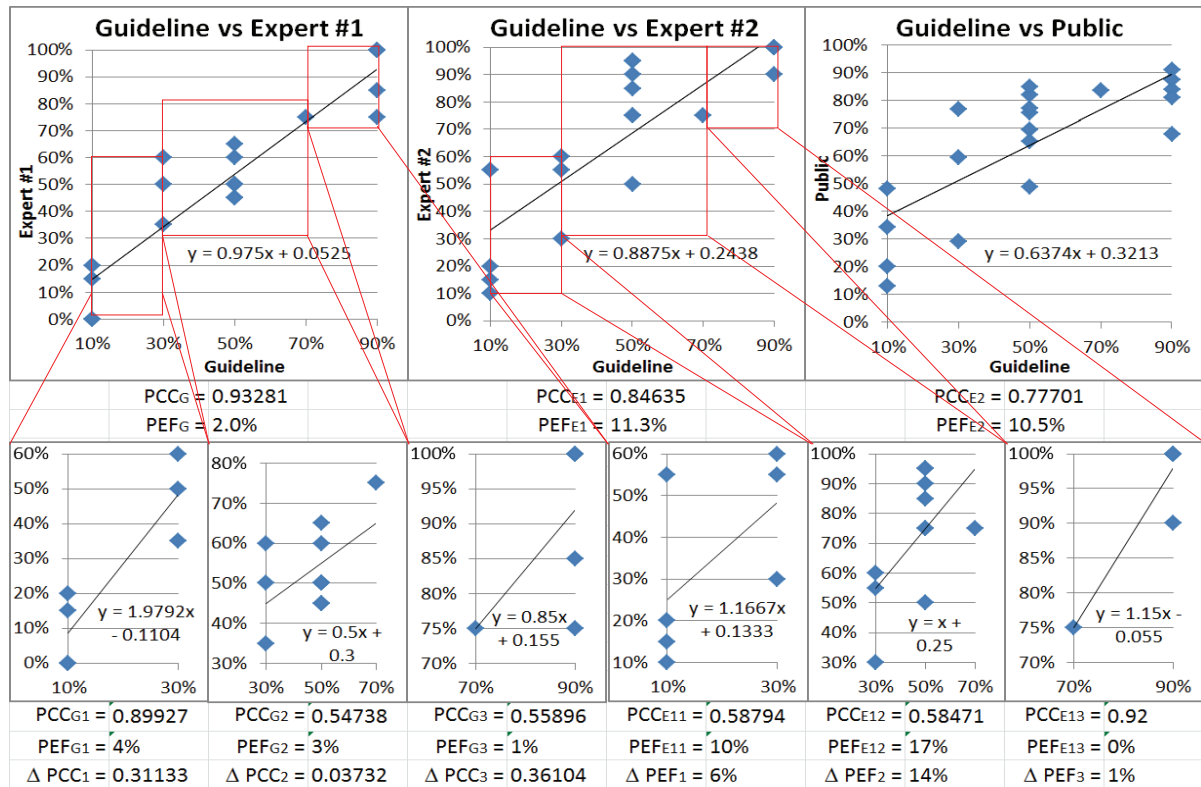


Figure 8 Scatter Plot for Guidelines versus Expert #1, Expert #2 and Public along with corresponding Pearson Correlation Coefficients

a guiding model to promote consistent assessment of Manglish creole coupled with a list of acceptable local terms.

Scatter plot for the outcome of guideline versus Expert#1, Expert#2 and Public were plotted and Pearson Correlation Coefficient (PCC) for each case were computed. It showed that PCC for Expert#1 was the most closely correlated with guideline. However, the PCC for Expert#2 and Public were not far behind. Compared to Perception Error Factor (PEF), PCC did not reflect a compatible difference between Expert#1 and Expert#2. Going deeper into analysis, most discrepancies were detected around the assessments of mesolectal sentences ($DPEF_2 = 14\%$). However, PCC did not reflect such difference ($DPCC_2 = 0.037$). As a matter of fact, the difference was much lower than the acrolectal and basilectal sentences ($DPCC_1 = 0.311$ and $DPCC_3 = 0.361$). In contrast, PEF only displayed fractional differences for such sentences ($DPEF_1 = 6\%$ and $DPEF_3 = 1\%$). It is therefore concluded that PEF is a much better model to detect discrepancies in assessment results.

CONCLUSION

This study dealt with the severity, frequency and occurrence of Manglish creole continuum among Malaysians. To achieve this, a holistic survey was done on Malaysian general public from various age, ethnicity, educational background, mother tongues and most frequently spoken language. The frequency and occurrence of Manglish was compiled by comparing the usage of acrolect, mesolect and basilect Manglish against formality of occasions. On the other hand, the investigation on severity was done through an analysis on the Manglish creole continuum whereby a model to govern its assessment was introduced. Mathematical models were also proposed to compute the magnitude of formality as well as the intensity of deterioration in Manglish creole. Parallel computation were done using Pearson Correlation Coefficient and compared against proposed models found in favour in selective cases. Generally, the results revealed the tendency of standard of language to gradually degrade as formality of events decrease. However, the correlation was only mediocre, suggesting either a limited capability

among the participants to switch between Manglish creole or to accept lower Manglish creole in daily usage. As a whole, each Manglish variant is predicted to continue playing its respective roles in this multi-cultural and multi-racial country. Most remarkably, Malaysians own a local variety which truly signifies their identities.

REFERENCES

- Abdul Razak, R.A. (2015). *Attitudes towards non-standard Malaysian English*. University of Malaya, Malaysia.
- Baskaran, L. (1987). *Aspects of Malaysian English syntax*. PhD. University of London.
- Baskaran, L. (1987). Indigenization of English (Malaysian English—Its Development & Features. In Proceedings of the Modern Language Association Conference on “The Language Situation in Malaysia?” Malaysia: Kuala Lumpur.
- Baskaran, L. (1994). The Malaysian English mosaic. *English Today*, 10(1): pp. 27-31.
- Christopher, R. C. (1973). *Designing Logic Systems Using State Machines*. McGraw-Hill
- Crystal, D. (2003). *English as a global language*. Cambridge: Cambridge University Press.
- Department of Statistics Malaysia. (2020). Current Population Estimates, Malaysia, 2018-2019. Retrieved from https://www.dosm.gov.my/v1/index.php?r=column/cthemByCat&cat=155&bul_id=aWJZRkJ4UEdKcUZpT2tVT090Snpydz09&menu_id=L0pheU43NWJwRWVSZklWdzQ4TlhUUT09
- Foo, B. & Richards, C. (2004). English in Malaysia. *RELC Journal*, 35.
- Gill, S. K. (2002). *English language challenges for Malaysia: International communication*. Selangor: Universiti Putra Malaysia Press.
- Hamming, R. W. (1950). Error detecting and error correcting codes. *Bell System Technical Journal*, 29(2): pp. 147-160.
- Lee, S. K. (2003). Multiple identities in a multicultural world: A Malaysian perspective. *Journal of Language, Identity and Education*, 2(3), pp. 137-158.
- M. F. Rusli, M. A. Aziz, S. R. S. Aris, N. A. Jasri, & R Maskat. (2018). Understanding Malaysian English (Manglish) jargon in social media. *Journal of Fundamental and Applied Sciences*, 10(2S), pp. 116-125. <http://dx.doi.org/10.4314/jfas.v10i2s.10>
- Pearson, K. (1986). Mathematical Contributions to the Theory of Evolution: Regression, Heredity and Panmixia. *Philosophical Transactions of the Royal Society of London*, 187, pp. 253-318.
- Platt, J. T. (1975). The Singapore English speech continuum and its basilect ‘Singlish’ as a ‘Creoloid’. *Anthropological Linguistics*, 17(7): 363-374.
- Platt, J. T. & Weber, H. (1980). *English in Singapore & Malaysia*. Kuala Lumpur: Oxford University Press.
- Preshous, A. (2001). ‘Where you going ah?’. *English Today*, 17(1), pp. 46-53.
- Stewart, W. (1969). Urban Negro speech: Sociolinguistic factors affecting English teaching. In R. Shuy, A. Davis & R. Hogan (Eds.). *Social dialects and language learning* (pp. 10-18), National Council of Teachers of English.
- Tan, S. I. (2009). Lexical borrowing from Chinese languages in Malaysian English. *World Englishes*, 28(4), pp. 451-484.
- Vethamani, M. E. (1996). *Character presentation and interaction: Styles of minority discourse in the Malaysian English novel*, PhD. Thesis, University of Nottingham.
- Vollmann, R. & Soon, T. W. (2019). The sociolinguistics status of Malaysian English. *Grazer Linguistische Studien 91 (Frühling 2019)*, S. pp. 133-150. doi:10.25364/04.46:2019.91.5
- Zhia, E. L. (2015). *Colloquial Malaysian English (CMaE): A problem or a cool phenomenon?* Spain: Universitat Jaume.