#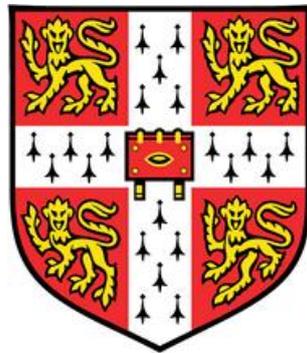 Investigating the role of an indigenised variety of English in the acquisitional and sociolinguistic contexts of the Malaysian ecology

Samantha Li Wen Sie

Newnham College

Theoretical and Applied Linguistics Section

Faculty of Modern and Medieval Languages and Linguistics

University of Cambridge

This thesis is submitted for the degree of

Doctor of Philosophy

January 2023

DECLARATION

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the preface and specified in the text. I further state that no substantial part of my thesis has already been submitted or is being concurrently submitted for any such degree or any other qualification at the University of Cambridge or any other University of similar institution except as declared in the preface and specified in the text. It does not exceed the prescribed word limit of 80,000 words, including footnotes, references, and appendices but excluding bibliographies, for the Faculty of Modern and Medieval Languages and Linguistics Degree Committee.

<div align="right">

Samantha Li Wen Sie

January 2023

</div>

Investigating the role of an indigenised variety of English in the acquisitional and sociolinguistic contexts of the Malaysian ecology

Samantha Li Wen Sie

ABSTRACT

The realm of New Englishes offers enriching avenues to explore the interplay between language acquisition and sociolinguistic influences in linguistically diverse ecologies. Yet research into this interdisciplinary arena remains lacking. Accordingly, this thesis addresses this paradigm gap by focusing on the Malaysian ecology. One of the three empirical studies conducted as part of this project is i) a CASE STUDY which examines the morphosyntactic properties of an indigenised variety of English viz., Colloquial Malaysian English (CME). The data generated from naturalistic conversations came from two pairs of adult Malaysians with different L1 backgrounds (i.e., Malay and Chinese). While many of the non-standard features supplied could be explained by substrate influence, there were also features resembling general second language (L2) behaviours and creative innovation. The MAIN STUDY adopts a concurrent embedded design, which comprises ii) an ACQUISITIONAL STUDY and iii) a SOCIOLINGUISTIC STUDY. The ACQUISITIONAL STUDY investigates the roles of the first language (L1) and CME in the ultimate acquisition of finiteness in Standard English (StE). The adult participants recruited for this study were 145 Malaysians and 30 British (control). Malaysians who acquired English as (one of) their L1(s) (L1-MalE(+)) were predicted to have less difficulty than their L1-Malay and L1-Chinese peers and perform more similarly to the British English (BritE) monolinguals. This is because, despite the prevalence of CME in the local environment, L1-MalE(+) learners would merely have to reset the optional features of finiteness in CME to obligatory, as required in StE. Meanwhile, L1-Malay and L1-Chinese learners would be faced with an additional learnability burden of acquiring finiteness as a new functional feature, given its absence in their L1s. Findings from a grammaticality judgement task and narrative task revealed that although the Malaysian cohort behaved statistically differently from the L1-BritE control, the L1-MalE(+) groups outperformed the L1-Chinese and L1-Malay groups across the board. That said, the L1-Malay group fared considerably better than its L1-Chinese counterpart and was about on par with the L1-MalE(+) peers. These findings indicated clear L1 effects modulated by typological proximity. Meanwhile, the SOCIOLINGUISTIC STUDY explores Malaysians' attitudinal behaviours towards CME and StE. The same participants from the acquisitional study undertook a sociolinguistic survey administered for this study. Findings revealed that the

participants were non-discriminatory towards CME and StE, and that they were aware of when to use these varieties across different social settings. Altogether, this thesis demonstrates the facilitative role of CME in the acquisition of StE, and concurrently vindicates the functional importance of CME and StE as legitimate varieties in the Malaysian milieu.

ACKNOWLEDGEMENTS

This thesis would not have been possible without the generous help and support of the following people, to whom I give my sincerest thanks.

I am incredibly indebted to my supervisor, Prof Ianthi Tsimpli, for taking me under her wing, and for allowing me to explore and find my voice in this project. Her input has been immensely invaluable. All the meetings we have had never failed to leave me in awe of her academic prowess and incredible insights. Also, I can never thank her enough for her patience and determination to see me through the end of this thesis. Thank you for believing in me.

I would like to thank my advisor, Dr Theodora Alexopoulou, for welcoming me to her amazing EF lab group. I owe her my deepest gratitude to her and her team for giving me the opportunity to showcase my work-in-progress and for providing invaluable feedback on it. I am also grateful to Prof Henriëtte Hendriks and Prof Margaret Deuchar for offering their guidance on transcription work and the use of TalkBank-related facilities.

I would also like to thank the Faculty of Modern and Medieval Languages and Linguistics and Newnham College for supporting me financially for the fieldtrip and conferences, as well as the University of Cambridge for providing financial assistance due to the impact of Covid.

This research would not have been carried out all the more successfully without the generous support of my host supervisor, Prof Dr Stefanie Pillai, who made it possible for me to do my fieldwork at the University of Malaya; as well as Madam Rema Menon and her wonderful team of English Language teachers, who relentlessly promoted my study to their students. My stay in Kuala Lumpur would not have been smoother without the help of Uncle Aaron Tham, Aunty Salina Wong, and their families, who not only went through great lengths to help find an accommodation for me in KL, but also extended their warm hospitality to me whilst I was there. And, very importantly, this research would not have been possible without the participants who willingly took part in all my studies. Having the interview sessions with them made the research experience extremely pleasant. To these main players who contributed to the success of this fieldwork, I owe you a great debt of gratitude.

I would also like to thank to my colleagues and friends who selflessly extended their helping hand whenever I got stuck at work. Special thanks are due to Dr Margreet Vogelzang, Dr Jasper Sim, Dr Kate Derkach, Dr Itamar Shatz, Andromachi Tsoukala, Dr Tsen Vei Lim, Tze Vun Voo, and Dr Anne Thomas for being patient in helping me understand all things statistics; Dr Mandy Wigdorowitz and James Algie for being meticulous in reviewing my

research instruments; Dr Carlos Balhana, for teaching me how to do annotation work for the corpus study; colleagues at Prof Tsimpli's research group and Dr Alexopoulou's EF group for giving me good feedback on my work; and Cikgu Wong Fui Lin, Jennica Chin, Wan Ji Pang, and Cynthia James, for providing me clarifications about the education system in Malaysia.

Furthermore, I am deeply grateful for my friends who are always there for me. I would especially like to thank Dr Tsen Vei Lim, Joanna Worthington, Dr Laura Achtelstetter, Nazia Jassim, Karim Hyde-Smith, Dr Maria Zacharopoulou, Vivian Lam, Jessica Chin, Sam Welch, Adriel Wong, and Julia Heine, not only for all the good times we share, but also for staying close to me especially in some of my lowest moments. I would like to express my gratitude to my family in Christ at the Christian Graduate Society and at Ridley Hall, who have so faithfully demonstrated the love of God by loving others sacrificially. Special thanks go to Prof John and Dr Diane Lister, Teacher Gan Poh Choo, Dr Lizelke Klindt, Mary Brown, Emily Chan, Bea Simpson, and Dr Gamal Crichton for always checking in and keeping me in your prayers. I would also like to thank Aaron Raj for all his help and good company.

I could not have undertaken this thesis without the unconditional love and support of my family. To Sarah, who cheered me on all the way to the finish line; Samuel, whose pep talks kept me going; Mom, who always looks out for me; and, above all, Dad, who is an avid supporter of the work I do – you have no idea how blessed I am to have you all in my life. Thank you for always being there for me.

Finally, I am eternally thankful to Yeshua, who has granted me His wisdom, strength, and joy to complete this thesis. Above all, thank You for bringing me home.

Contents

11

List of tables and figures

List of abbreviations and acronyms

3rd.SING.-*s*	Third-person singular -*s*
3PL	Third-person plural
3S	Third-person singular
3SG	Third-person singular feature
ACC	Accusative case
Adj	Adjective
AdjP	Adjectival phrase
AspP	Aspect Phrase
BLP	Bilingual Language Profile
BritE	British English
CEFR	Common European Framework of Reference for Languages
CHAT	Codes for Human Analysis Transcripts
CILD-Q	Contextual and Individual Linguistic Diversity Questionnaire
CL	Classifier
CLAN	Computerized Language Analysis programme
CLIL	Content and Language Integrated Learning
CME	Colloquial Malaysian English
COP	Copula
CP	Complementizer Phrase
CSE	Colloquial Singapore English
DUR	Durative
EPP	External Projection Principle
EXP	Experiential
FA	Full Access
FEM	Feminine gender feature
FFFH	Failed Functional Features Hypothesis
FTFA	Full Transfer Full Access Hypothesis
GCSE	General Certificate of Secondary Education
GJT	Grammaticality judgement task
H$_n$	Hypothesis
IELTS	International English Language Testing System

IH	Interpretability Hypothesis
INFL	Inflection
IP	Inflectional Phrase
KBSR	Kurikulum Bersepadu Sekolah Rendah (Malay)
	In English: Integrated Curriculum for Primary Schools
KSSR	Kurikulum Standard Sekolah Rendah (Malay)
	In English: Standards-based Primary School Curriculum
L1	First language
L2	Second language
LHQ 2.0	Language History Questionnaire
MalE	Malaysian English
MBMMBI	Memartabatkan Bahasa Malaysia dan Memperkukuh Bahasa Inggeris (Malay)
	In English: Upholding the Malay Language and Strengthening the English Language
MLU	Mean length of utterance
MSIH	Missing Surface Inflection Hypothesis
MUET	Malaysian University English Test
N	Noun
NEG	Negation
NegP	Negative phrase
NEU	Neuter gender feature
NOM	Nominative case
NP	Noun phrase
NT	Narrative task
OI	Optional infinitive
past	Past tense feature
pres	Present tense feature
prog	Progressive feature
PA	Partial Access
POSS	Possessive
PP	Prepositional phrase

PPSMI	Pengajaran dan Pembelajaran Sains dan Matematiks dalam Bahasa Inggeris (Malay)
	In English: Teaching and Learning of Science and Mathematics in English
PERF	Perfective
PROG	Progressive
PTH	Prosodic Transfer Hypothesis
PTR	Pronounce Tense Rule
RC	Relative clause
RD	Representational Deficit
RP	Relative pronoun
RQ	Research question
SAI	Subject-auxiliary inversion
SD	Standard deviation
SE	Standard error
SES	Socioeconomic status
SL	Substrate language
Spec	Specifier
SPM	Sijil Pelajaran Malaysia (Malay)
	In English: Malaysian Certificate for Education
SingE	Singapore English
SLA	Second Language Acquisition
SLI	Specific language impairment
SMG	Standard Modern Greek
StE	Standard English
t	Trace
T	Tense
TD	Typically developing
TEGI	Test of Early Grammatical Impairment
TOEFL	Test of English as a Foreign Language
TP	Tense phrase
TTR	Type-token ratio
u	Uninterpretable
*u*CASE	Uninterpretable case feature

*u*INFL	Uninterpretable inflectional feature
*u*N	Uninterpretable noun (EPP) feature
UG	Universal Grammar
UK	United Kingdom
*v*P	'Little' *v*(erb) phrase
V	Verb
VP	Verb phrase
WE	World Englishes

1 Introduction

1.1. Second Language Acquisition and World Englishes

Owing to colonisation and globalisation, the world has witnessed the spread of English, through which the language has attained the status of the world's lingua franca (Crystal, 2012). As of 2022, English is the most widely spoken language in the world. This includes non-native (n = 1.08 billion) and native (n = 373 million) speakers, with the former group forming a substantial majority of the English-speaking population (Eberhard et al., 2022). Zeroing in on former Anglophone colonies (e.g., Jamaica, Singapore, Uganda), English enjoys an elevated status as an official language in many of these linguistically and culturally diverse territories (Crystal, 2012:151). Indeed, English is not commonly acquired as a first language (L1) as it finds itself rivalling other local or heritage languages integral to ethnic identities. However, given its significance as a functional language, it observes a growing number of people acquiring it as an, or one of the, L1(s) (Buschfeld, 2020:1). Furthermore, as political independence fosters unity and identity rewritings, national and socio-cultural sentiments inculcated by the local communities spur the restructuring of the ex-colonial language, engendering new indigenised varieties of English (Schneider, 2003, 2007, 2008). For the multilingual ecologies it offers, the realm of New Englishes promises enriching avenues to explore language acquisition and sociolinguistic variation. Yet, acquisitional research into such settings is hitherto lacking (De costa & Crowther, 2018:21; Gass, 2018; Mesthrie & Bhatt, 2008:159; Sridhar & Sridhar, 1986). This is where this thesis comes in, addressing what Sridhar & Sridhar (1986) refer to as the paradigm gap between the fields of (*Second*) *Language Acquisition* and *World Englishes*.

As the disciplinary approach taken by the current research is predominantly an acquisitional one, the field of Second Language Acquisition (SLA) and some key concepts merit a brief introduction. Studies in this field investigate how individuals acquire a second (L2) or more languages in addition to their L1.[1] Under the generative approach (Chomsky, 1981, 1986, 1993, 1995), language acquisition is guided by a genetically endowed component of the language faculty known as Universal Grammar (UG). It has been claimed to supply a genetic blueprint of grammatical categories and features necessary to form internal grammars. It also lays out universal rules (i.e., *principles*) and enables options (i.e., *parameters*) that

---

[1] In this thesis, L2 refers more generally to a subsequent language (being) acquired.

constrain the diversity of languages. In L1 acquisition, linguistic input that is available in a child's language environment serves a vital role in setting the parameters relevant to the language(s) being acquired. That said, while convergence on the adult native speaker (L1) grammar is always guaranteed, it is usually not the case in L2 acquisition (Selinker, 1972; Slabakova, 2016; White, 2003). This is because, by the time L2 learners start acquiring the target language, their L1s have been established and, correspondingly, influence the development of the L2 grammar. This postulation, which will be central to the current research, is known as *L1 transfer* or *influence*.[2] If the parametric value (e.g., feature, feature strength) of an L2 property is instantiated in the L1 grammar, then the transfer of properties from the L1 to the L2, a.k.a. positive transfer, facilitates acquisition. Contrariwise, if the L2 property is parametrically different or not instantiated in the L1, then negative transfer from the L1 instigates L2 errors; in this case, parameter-(re)setting is required to restructure the learner grammar (Bardovi-Harlig & Sprouse, 2018; Slabakova, 2016:45; White, 2003). However, the question of whether parameter-(re)setting is still possible in L2 acquisition has spurred much contention in SLA research. In this regard, the issue of accessibility to UG is brought to the fore for further scrutiny. Accounts arguing for or against UG accessibility in L2 acquisition, which is reviewed later in **§1.2.**, will be tested against findings from the acquisitional study of this thesis.

While the field of generative SLA primarily focuses on linguistic factors and cognitive mechanisms propelling the acquisition of an additional language, the World Englishes (WE) enterprise tends to take on a sociolinguistic approach to investigate the spread and effects of English on society. Various proposals have been put forth to categorise the different English varieties attested around the world (Kachru, 1992b:3; Görlach, 1990; McArthur, 1987:11; Schneider, 2003, 2007; Szmrecsanyi & Kortmann, 2009a, 2011), but the Three Concentric Model postulated by Kachru (1992b) will be used to refer to the target varieties (i.e., Malaysian English, and British English) considered in this thesis. The three circles depicted in Figure 1.1 correspond to the different sociolinguistic statuses of English across various countries. The Inner Circle encompasses Anglophone countries where English is the L1 of the majority population (e.g., UK, USA, Australia). The Outer Circle subsumes former Anglophone (i.e.,

---

[2] There are other factors such as input (Paradis et al., 2017; Unsworth, 2014) and age effects (DeKeyser et al., 2010; Johnson & Newport, 1989) that play important roles in mediating L2 development (Tsimpli, 2014; Unsworth et al., 2014). However, as it is beyond the scope of the current research to include them in our investigation, they shall not be expounded.

British/American) countries and territories where English enjoys the status of an (additional) official language (e.g., Malaysia, India, South Africa). The Expanding Circle refers to other countries where English is not used at a societal level but mostly confined in a classroom setting (e.g., China, Indonesia, Japan). To wit, these circles correspond to the respective distinctions of English as a native (ENL), second (ESL), and foreign (EFL) language (Kachru, 1992b:3).

Figure 1.1: The Kachruvian Concentric Circles of World Englishes

The "Expanding Circle"

China	1,088,200,000
Egypt	50,273,000
Indonesia	175,904,000
Israel	4,512,000
Japan	122,620,000
Korea	42,593,000
Nepal	18,004,000
Saudi Arabia	12,972,000
Taiwan	19,813,000
USSR	285,796,000
Zimbabwe	8,878,000

The "Outer Circle"

Bangladesh	107,756,000
Ghana	13,754,000
India	810,806,000
Kenya	22,919,000
Malaysia	16,965,000
Nigeria	112,258,000
Pakistan	109,434,000
Philippines	58,723,000
Singapore	2,641,000
Sri Lanka	16,606,000
Tanzania	23,996,000
Zambia	7,384,000

The "Inner Circle"

USA	245,800,000
UK	57,006,000
Canada	25,890,000
Australia	16,470,000
New Zealand	3,366,000

(Kachru, 1992b:3)

Delving into the Outer Circle domain, the linguistically diverse yet intricate ecologies of former Anglophone colonies offer enriching avenues to investigate language acquisition because it is in many of these contexts that we find indigenised varieties of English (IVEs) being acquired alongside Standard English (StE) at a societal level. Regarding the evolution of postcolonial Englishes, Schneider (2003, 2007, 2014) raises the importance of historical (i.e., from colonisation to independence) and socio-cultural (including identity (re)constructions) events – on top of acquisitional considerations (e.g., transfer, L2 acquisition, multilingualism) – in their contribution to the indigenisation of English. As mentioned earlier, although English

may be an official (L2) language in many of the former colonies, it is not the L1 of the local communities. That said, when accounting for non-standard features in such settings, not only is it justified to say that they reflect L2 behaviours, but it must also be pointed out that some of them have undergone structural innovation to serve as markers of group identity. Accordingly, this thesis deals mainly with the Outer Circle domain, more specifically with Malaysia, for which the acquisitional-cum-sociolinguistic ecology will be extensively investigated.

Having reviewed the fields of SLA and WE, we now address the paradigm gap between these disciplines identified by Sridhar & Sridhar (1986). Coming from the WE domain, the authors criticise the SLA approach by accentuating the lack of attention being paid to IVE contexts. They also point out that the field mainly focuses on L2 acquisition in monolingual realities and base successful attainment on native-speaker models, which are not representative of many multilingual societies such as those in IVE settings. While their remarks are not unsound, they have come under fire by SLA researchers. Gass (2018:122) contends that the goal of the discipline is essentially to understand the mechanisms underlying language learning in the context of L2 acquisition, rather than to get learners "to acquire native-like competence in the target language" as stated by the authors (Sridhar & Sridhar, 1986:5). Additionally, Larsen-Freeman (2018) defends that L2 learning is contingent not upon native-speaker competence but rather on the learner/user's agency in using the L2 in a meaningful and intentional way (ibid., p.83; see also Wee, 2018). Indeed, in spite of Sridhar & Sridhar's (1986) criticisms of SLA, the authors do acknowledge the shortfalls of WE as far as research into acquisition is concerned (see also Mesthrie & Bhatt, 2008:159). They concede that much of the work being done in WE is "descriptive and atheoretical, rather than based on rigorous and systematic empirical research" (ibid., p.12). Taken together, the lack of collaborative research between the disciplines of WE and SLA is precisely due to the "different goals and foci" which these disciplines independently strive towards (Gass, 2018:123; see also De Costa & Crowther, 2018:23; Larsen-Freeman, 2018:81; Ortega, 2018:65). Notwithstanding, this does not diminish Sridhar & Sridhar's (1986) criticism of the neglect of IVE realities in SLA research. To this end, this thesis comes in with the acquisitional (i.e., generative SLA) approach to examine the acquisition of English in the IVE context of Malaysia.

There are a few key aspects pertaining to language acquisition with which both disciplines of SLA and WE concur. One of them concerns the role of crosslinguistic influence/transfer. In WE research, transfer is often referred to as *substrate transfer*, namely linguistic influence from other local or heritage languages (viz., substrate languages). This acquisitional phenomenon, extensively examined in SLA studies (Hermas, 2019; Hopp, 2010;

Schwartz & Sprouse, 1994, 1996), is also underscored in the literature in accounting for the indigenisation of English (Bao, 2005; Sand, 2004; Sharma, 2009). Non-target features reflecting overgeneralisation processes (e.g., overmarking of tense in DO-constructions "…*did not tried*…"; subject-verb inversion in indirect questions "I wonder why *is she* angry") have also been attested in WE studies (SLA: Taylor, 1975; Pozzan & Quirk, 2014; WE: de Klerk, 2003; Hilbert, 2011). Concerning the role of UG (highly contested in SLA research), it is unclear as to the extent to which WE researchers would subscribe to this notion. What is clear, however, is that there are proponents in the WE enterprise arguing for vernacular universals (Chambers, 2004; Kortmann & Szmrecsanyi, 2004, 2011; Szmrecsanyi & Kortmann, 2009b), which are linguistic features shared by various non-standard varieties (e.g., lack of inversion in direct questions "ø You get the point?"; see Kortmann & Szmrecsanyi, 2004). Whether or not such non-standard features can directly be explained by substrate/crosslinguistic influence (Kortmann & Szmrecsanyi, 2004), what has been agreed upon is that they generally reflect linguistic behaviours shared with those in non-WE contexts, implicating the presence of universal mechanisms underlying (L2) acquisition (Buschfeld & Kautzsch, 2016:111; Kortmann & Szmrecsanyi, 2011:286; Mesthrie & Bhatt, 2008; Schneider, 2007:89).

However, the epistemological approaches taken by SLA and WE on other (socio)linguistic aspects remain unequivocally distinct. Corroborating Sridhar & Sridhar (1986), the types of linguistic input considered in each of these fields are different. The L2 input referred to in SLA tends to conform to a monolithic target language. On the other hand, the target input (i.e., English, including the standard and colloquial varieties) in the IVE milieu is highly variable, with the indigenised variety comprising the "primary input" (ibid., p.6). Furthermore, given the multilingual ecology of IVE settings, (L2-)English remains susceptible to restructuring as it continues to receive influence from the contact languages. Therefore, investigation into the acquisition of English in IVE settings will need to consider the different input varieties of English as well as other substrate languages as potential sources of influence on the target language. Another aspect not taken into account in SLA research but honoured in the WE enterprise regards bilingual creativity and the innovative features it generates (Bamgbose, 1998; Bhatt, 2001; Bolton, 2010; Kachru, 1985). In SLA, features deviant from the target variant are deemed to be learner errors. In WE, non-standard features are seen as potential avenues for linguistic innovation. When such features are being shared and used extensively amongst the local communities, not only does it entail that they are constrained by "cognitive-economy considerations", but it also reflects "social-functional requirements" that propagate acceptance and normalisation of these features (Bhatt, 2001:536). Ergo, in

recapitulating Kachru (1985:24), Bolton (2010:458) succinctly states that there is a need for "a greater awareness of the sociolinguistic context of bilingual communities, in order to identify local norms of usage, and to differentiate between 'errors' and creative innovations." A final aspect pointing to the departure between SLA and WE concerns language attitude. Attitudinal factors may not be as commonly probed in SLA research, but they are important in WE studies. In WE settings, different varieties of English are attached different types of social prestige. Non-standard varieties have been observed to be associated with traits of solidarity (covert prestige; e.g., friendliness, group identity), whereas standard varieties tend to be associated with traits of status (overt prestige; e.g., educatedness, professionalism) (Ng & Diskin-Holdaway, 2021; Tan & Tan, 2008). From an acquisitional standpoint, not much interdisciplinary research has been done to examine the extent to which attitudes towards standard and non-standard varieties might influence acquisition in IVE realities. However, it might cast some light on the motivations underlying learners of English and, potentially, envisage the sociolinguistic evolution of the target language and/or language shift that might take place in the local milieu.

As the primary aim of the current thesis is to investigate the acquisition of StE in Malaysia, a concurrent embedded design is adopted, with heavier weightage being placed on the acquisitional study. Simultaneously, given the IVE reality of Malaysia, a sociolinguistic study is also incorporated to elicit attitudinal behaviours towards the standard and colloquial varieties of English being spoken in the Malaysian context.

At this stage, it is worth clarifying the terms *non-standard*, *non-target*, *ungrammatical*, and *learner errors* when describing linguistic forms. If the language variety under inquiry is a colloquial variety (e.g., Colloquial Malaysian English), then the term *non-standard* shall be employed to refer to linguistic forms not prescribed in the standard variety. If the target language variety is the standard variety (e.g., StE), then *non-target/ungrammatical* features or *learner errors* shall be used interchangeably.

Finally, it is imperative to mention that this thesis does not endorse any of these varieties as being more important than the other, nor does it promote linguistic purism or imperialism in any way. What this thesis endeavours to show is that the standard and colloquial varieties serve different yet important sociolinguistic functions in society, and it is at the heart of this research to vindicate both of them as legitimate varieties in their own right.

1.2. SLA theories

The SLA theories to be considered in this thesis conform to the generative approach. Concerning the issue of UG accessibility in L2 acquisition, the two main camps to be reviewed are the Full Access (FA) and Partial Access (PA) views. According to the former view, UG operations (i.e., principles and parameters) governing L1 acquisition are also fully available for L2 acquisition; this implicates successful attainment at the steady-state interlanguage (Goad & White, 2006; Prévost & White, 2000; Schwartz & Sprouse, 1994, 1996). Regarding the latter view, UG still constrains the development of L2 grammars but will not be available in the acquisition of certain linguistic (e.g., morphosyntactic) domains, thereby implicating permanent representational deficits in the interlanguage (Hawkins & Chan, 1997; Tsimpli & Dimitrakopoulou, 2007). Accordingly, the theories subsumed under each of these views shall be introduced in the remainder of this section.

Under the FA view, the Full Transfer Full Access Hypothesis (FTFA) postulated by Schwartz & Sprouse (1994, 1996) proposes that the initial state of the L2 grammar anticipates heavy L1 transfer. However, exposure to the target input, which cannot be supported by the L1 grammar, will trigger parameter-(re)setting under the guidance of UG. In this instance, the interlanguage restructures away from the L1 grammar and develops towards the steady state. Convergence on the target-language grammar is implicated but not guaranteed. This is in part due to the different L1s constituting different starting points in L2 development, and in part due to the lack of input cues needed for further restructuring in the language-learning environment (Schwartz & Sprouse, 1996:42). In Yuan's (1998) study, adult learners of different L1 backgrounds (i.e., English, Japanese) and proficiency levels (i.e., intermediate, advanced) were tested on long-distance reflexivity in L2-Mandarin. Results from a multiple-choice comprehension test revealed that learners of L1-Japanese did not perform divergently from the native Chinese controls because the Japanese reflexive *zibun* was similar to the (Mandarin) Chinese reflexive *ziji* in that they both allowed for long-distance binding. In this case, positive transfer was indicated. Between the English-speaking groups of intermediate and advanced proficiencies, those who were advanced in L2-Chinese performed better than those at the intermediate level in accepting long-distance binding of the L2-Chinese reflexive, which is not instantiated in L1-English; the difference in linguistic performance between these learner groups therefore alludes to full access to UG.

Another hypothesis corroborating the FA view is the Missing Surface Inflection Hypothesis (MSIH). Formulated by Prévost & White (2000), it postulates that L2 learners have underlying morphosyntactic representations in the L2 grammar but have issues supplying target-like morphological exponents, which the authors ascribe to "processing reasons" or "communication pressures" being some of the plausible factors (ibid., p.129). Supporting MSIH is a study by Geçkin & Haznedar (2008), which examined longitudinal data elicited from the oral production of three Turkish-speaking children acquiring L2-English. The authors observed that these children consistently produced copula BE in appropriate contexts and hardly produced agreement errors despite omitting tense inflections frequently. Taken together, the authors suggested that the morphological variability observed in the children's suppliance of tense marking did not indicate syntactic deficits per se, since there were other pieces of evidence suggesting successful acquisition of those features. Therefore, the authors maintained that morphological variability in the L2 lay in a mapping issue between fully established abstract features and their morphological exponents.

Goad and White's (2006) Prosodic Transfer Hypothesis (PTH) is more specific than MSIH in that it zeroes in on the phonological inventory of the L1. That is, if the prosodic representation of an L2 property is not instantiated in the L1, then learners will not be able to realise the morphophonological exponents of the L2 property in question. To wit, the linguistic deficit lies not in the morphosyntactic domain but rather in the phonological domain. What PTH (and/or MSIH) further entails is that there will be differences in linguistic performance depending on the modality of the linguistic task(s) administered (Slabakova, 2016:193). Correspondingly, Goad & White employed a combined sentence completion and elicited production task to examine L1-Mandarin learners' metalinguistic knowledge and oral production of tense inflections in L2-English. The rationale underlying the authors' study was that the prosodic structure of Mandarin, which disallows consonant clusters in coda positions, would constrain the morphophonological exponence of tense inflections in L2-English. Findings revealed that the learners were accurate in choosing appropriate answers and producing the completed stimulus sentences. Interestingly, the learners' target-like production of the English infection surpassed Goad & White's prediction, but this also suggests successful acquisition of the English tense feature in the target L2.

Turning to theories supporting the PA view, the Failed Functional Features Hypothesis (FFFH) by Hawkins & Chan (1997) posits that access to UG is subject to maturational constraints determined by the critical period for L2 acquisition. If the acquisition of L2 functional features takes place beyond the critical (i.e., post-pubescent) period, UG becomes

unavailable for parametric options to be (re)set. This entails permanent syntactic deficits in the interlanguage, which means that the end-state interlanguage of L2 learners will diverge from that of their native counterparts. To demonstrate their point, Hawkins & Chan looked into the acquisition of English relative clause constructions by Cantonese and French-speaking learners of different proficiency levels (i.e., elementary, intermediate, advanced). The motivation underlying the authors' investigation was based on the parametric differences between Chinese and English in relation to the [±wh] feature in the Complementizer Phrase (CP). As the [wh] feature is absent in Chinese, the authors predicted that acquisition of this feature would pose learnability issues for the L1-Cantonese speakers. L1-French learners, on the other hand, would not have much of an issue acquiring it, because it would have already been instantiated in their L1. Results from a grammaticality judgement test showed that the L1-French learners outperformed their L1-Cantonese counterparts and were about on par with the native-English controls in their judgement across different un/grammatical conditions tested on English relative clauses. Meanwhile, the performances of the L1-Cantonese learners varied between the proficiency groups, leading Hawkins & Chan to conclude that L1-Cantonese learners failed to acquire the L2-English [wh] feature.

Similar to FFFH, Tsimpli & Dimitrakopoulou's (2007) Interpretability Hypothesis offers a more specific postulation. In their proposal, although syntactic (i.e., uninterpretable) features of the L2 will not be acquirable past the critical period, learning of those features could be remedied by relying on their semantic (i.e., interpretable) counterparts. While Tsimpli & Dimitrakopoulou further predict L1 effects throughout L2 acquisition, they also expect development to take place in the interlanguage grammar. In the authors' investigation of *wh*-subject/object extraction in L2-English interrogatives, L1-Greek learners of intermediate and advanced proficiencies were recruited to see if they would transfer the resumptive strategy in L1-Greek to L2-English, which would render ungrammaticality in the target language. Findings from the study revealed that while a significant number of L1-Greek learners from the intermediate group favoured resumptive pronouns regardless of the site of extraction in *wh*-interrogatives, learners from the advanced group rejected the resumptive strategy, but only when it involved resumptive object pronouns. Accordingly, Tsimpli & Dimitrakopoulou suggested that partial success in the linguistic judgement of the advanced learners was not due to their syntactic knowledge of *wh*-movement in L2-English, but rather due to the compensatory roles of the interpretable features [±animacy] and [d-linking] which the learners resorted to in their attempt to learn the grammatical structure of the English *wh*-interrogatives.

Taken together, while the FA and PA views expect the role of L1 influence in the development of the interlanguage, they predict different linguistic outcomes informed by different acquisitional trajectories in relation to UG availability. Theories advocating the FA view (i.e., FTFA, MSIH, PTH) maintain that the mechanisms of UG driving L1 acquisition are also fully available throughout L2 development. Contrariwise, theories defending the PA position (i.e., FFFH, IH) contend that acquisition of uninterpretable features is unattainable post-critical-period due to inaccessibility to UG in the morphosyntactic domain. These hypotheses will be evaluated when we arrive at the acquisitional study of this thesis.

1.3. Aims and objectives

As mentioned in **§1.1.**, the current thesis comprises two aims.

The primary aim is to investigate the acquisition of StE in the Malaysian context. Given its linguistically diverse ecology as an IVE reality, the contributory roles of the L1 and the indigenised variety known as Colloquial Malaysian English (CME) in the acquisition of StE will be investigated with consideration to the SLA theories introduced in **§1.2.**. The morphosyntactic phenomenon to be examined is the English finiteness alongside its associated features. To elicit this phenomenon, a grammaticality judgement task and a narrative task are administered. The different modalities employed by these tasks provides an opportunity to test any task-based effects, which will be relevant to evaluate MSIH and PTH reviewed in **§1.2.**.

The secondary aim is to tap into the attitudinal behaviours of Malaysians towards StE and CME. To execute this, a sociolinguistic survey is administered, which includes questions eliciting speakers' awareness, attitude, and preferences across different sociolinguistic dimensions.

Altogether, the acquisitional and sociolinguistic studies in this thesis will endeavour to capture and represent the multilingual and multicultural dynamics of IVE realities, with Malaysia being the site of study.

1.4. Thesis structure

The remainder of this thesis proceeds as follows.

Chapter 2 introduces the sociolinguistic landscape of Malaysia. The subvarieties of Malaysian English identified in this linguistically diverse ecology are described in relation to

sociohistorical and political events as well as acquisitional mechanisms, all of which are crucial in informing the indigenisation of English. It is also in this chapter where CME is introduced.

In relation to the CME literature, **Chapter 3** reports a small case study which aims to document, analyse, and showcase naturalistic data on CME. By understanding the morphosyntactic properties and linguistic mechanisms involved in steering the development of CME, findings from this case study help set the scene for the acquisitional study, which constitutes the bulk of this thesis.

The morphosyntactic phenomenon examined in the main study is finiteness, viz., tense and agreement. Accordingly, **Chapter 4** reviews the theoretical and acquisitional studies investigating this phenomenon. It also looks into and compares the languages of interest, namely StE being the target language, Malay, Chinese, and CME, in terms of whether and/or how finiteness is grammatically represented. This lays the groundwork for the acquisitional study by anticipating the learnability issues faced by speakers of these languages.

Chapter 5 establishes the methodological design of the main study, covering its research questions alongside their predictions, instruments administered, procedure, and participant profiles.

As the main study employs the linguistic tasks of grammaticality judgement and narrative-telling to elicit the knowledge and use of finiteness in StE, empirical findings from these tasks are reported in **Chapters 6** and **7**, respectively.

Meanwhile, **Chapter 8** reports attitudinal responses towards the English subvarieties spoken in Malaysia which are elicited from the sociolinguistic survey.

In **Chapter 9**, the key findings reported in Chapters 6 through 8 are discussed with respect to the research questions laid out for the main study. Limitations and directions for future research are also presented herein.

Finally, **Chapter 10** summarises the key findings of this thesis and concludes with a final remark on the paradigm gap between (second) language acquisition and World Englishes.

2 Malaysia

2.1. Introduction

To better understand the prominence of English in Malaysia, the country's socio-political and linguistic landscapes propagating the rooting of English deserve their mention. This chapter traces the trajectory of English through the sociohistorical events (**§2.2.**) and policy making pertaining to language and education (**§2.3.**) which have developed over the decades. The chapter then unravels the subvarieties of English that have arisen not only from these external events but also through extensive language contact and other acquisitional mechanisms (**§2.4.**), which will be demonstrated by looking into the morphosyntactic structure of the indigenised subvariety of interest known as Colloquial Malaysian English (**§2.5.**).

2.2. Sociohistorical background

In light of its multi-ethnic composition, Malaysia takes pride in its linguistically multifarious backdrop, with an estimation of 132 languages and dialects attested in the country (Eberhard et al., 2021). Out of the 32.66 million population, 91.73% (i.e., 29.96 million) are Malaysian citizens. The three main ethno-linguistic groups are the *Bumiputera* 'sons of the soil' (69.8%), Chinese (22.4%), and Indians (6.8%) (Department of Statistics Malaysia, 2021).

Bumiputera is a nomenclature given to the Malay majority and indigenous minorities. The languages spoken by this group find their roots in the Austronesian and Austroasiatic families. The Austronesian language family consists of the Malay dialects spoken across Malaysia and the aboriginal languages mostly spoken in East Malaysia (e.g., Iban, Kelabit, Penan). The Austroasiatic family comprises *Aslian* 'aboriginal' languages mostly spoken in West Malaysia (e.g., Kensiu, Semai, Temiar) (Haji Omar, 2004). Of all these languages, Malay is the dominant societal language due to the ethnic Malay majority. From a historical perspective, it had been the language of governance, diplomacy, and trade in the Malay Archipelago since as early as the 7th century A.D. during the Srivijaya Empire (Haji Omar, 1985:45, 2015:8-9; Hashim & Leitner, 2015). It maintained its status, albeit restricted, as a regional lingua franca in commerce during the Dutch occupation in Malacca (17th – 19th century) and British colonisation in the whole of the Malayan Peninsula (19th – 20th century) (Hashim,

2020:374). By the end of the colonisation era, Malay regained its sovereignty and became the sole national and official language of Malaysia.

The Chinese communities are generally associated with the Sinitic languages such as Cantonese, Hokkien, Hakka, and Teochew, which are the more populous dialects spoken locally (Chong, 2004:52). Mandarin is also widespread across Malaysia, since it is the language of education in Chinese-medium schools. Meanwhile, the majority Indian population traces its heritage back to the Southern regions of India, where Tamil (Dravidian) is spoken by 80% of the local communities (Haji Omar, 2015). It also serves as the medium of instruction in Tamil-medium national primary schools. Other Indian languages include Malayalam, Punjabi, Telugu, and Urdu (Thilagawathi & Ramasamy, 2004). In addition, there were earlier albeit smaller groups of Chinese and Indians who came to settle down as early as the 15th century during the Malaccan empire. These communities have assimilated into the Malayan culture, from which Malay-based creoles emerged. For instance, Baba Malay, spoken by the Baba Chinese communities, has influence from the Hokkien dialect, whereas Chitty Malay, spoken by the descendants of traders from Panai in Tamil Nadu, has traces of the Tamil language. Due to language loss, however, these acculturated communities are no longer identified with their original ethno-linguistic families (Haji Omar, 2015:3). The influx of Chinese and Indians in the 19th century was mainly attributed to the opening of tin mines and rubber estates. These immigrations revamped the demography of the Malayan and Bornean lands during the British rule (Hashim & Leitner, 2015:47; Hashim, 2020:374; Schneider, 2007:145), which is why Malaysia today hosts Malay, Chinese, and Indian as the three major ethnic groups.

Moreover, there are languages that are not ethnically tied to any particular speech communities. Arabic, for instance, is affiliated with the official religion of Malaysia, viz. Islam. Arabic loanwords have been adopted into the Malay and English vocabularies in Malaysia and are used in the media, formal and informal speeches, legal systems (especially with regards to Syariah law), and Islamic financial services (Hashim & Leitner, 2016; Hashim et al. 2017). It is also the language of instruction in Islamic religious schools normally attended by Malay students (Haji Omar, 1985:40). However, the religious schools represent only a small percentage (e.g., 2.67%) as compared to non-religious government(-aided) ones (e.g., 89.23%) (Department of Statistics, 2020), whose languages of instruction are conducted in the national (i.e., Malay) or the official vernacular (i.e., Mandarin or Tamil) languages. As the learning and use of Arabic is restricted to a minority of Malaysians, it does not leave as huge of a sociolinguistic impact as English, which takes precedence after Malay due to its importance as the language of globalisation. According to Crystal (2003:108-109), it is estimated that 31.5%

of the Malaysian population in 2001 would speak English as the L2 whereas 1.8% would have acquired it as their L1. However, this estimation is not accurate as it is purely based on the politico-historical reality instead of a sociolinguistic one. Also, there has not been any census that records the languages used in Malaysia since 1990 (Kärchner-Ober, 2013:298; Tan, 2003:176). It is therefore difficult to report the current linguistic landscape of the country. Notwithstanding, given the widespread accessibility of English in the local mass media (Schneider, 2007:149) and national education system, it would not be surprising if the percentages of L1 and L2 English speakers are larger than what was predicted by Crystal (2003).

Taken together, Malaysia hosts a diversity of languages, many of which are not genetically related to one another. Some languages bear more significance than others due to the functional roles they serve in administration and/or education.

Regarding the role of English in Malaysia, the next section expands on the various language policies that have been implemented to retain the relevance of English as an instrumental language for the country.

2.3. Language policies and education reforms

During the British colonisation era, English was not only the language of administration but also one of the languages of education alongside Malay, Mandarin Chinese, and Tamil. The latter three vernacular schools were mostly attended by students who spoke those respective languages as their first or dominant language. Meanwhile, English-medium schools hosted students from various ethnolinguistic backgrounds (Haji Omar, 1982:84; Hirschman, 1972:488). The vernacular schools, however, were not as established and did not share similar prestige as the English-medium schools. For instance, the Malay and Tamil schools could only offer primary education despite being funded by the colonial government. Although Chinese schools did provide secondary education (Hirschman, 1972:488), they were not granted any support from the administration and therefore could not set up avenues for students to advance into tertiary education (Haji Omar, 1996:514, 2015:10-11). On the other hand, English-medium schools were more equipped in preparing students for higher education overseas and for better career options such as administrative positions (Haji Omar, 1996, 2015). However, due to high tuition fees, only students from affluent families could enjoy the British education system and the socioeconomic benefits that ensued; students attending vernacular schools were suppressed of such privileges (Haji Omar, 1996). There was also a lack of uniformity in the curricula

amongst schools of different mediums of instruction (Haji Omar, 1985:40; Pillai & Ong, 2018:147). Altogether, these inequalities called for the need for an education reform as part of nation-building in preparation for Independence from the British.

In 1951, the British colonial administration in Malaya issued the *Barnes Report* in favour of the Malay education. Malay and English were proposed to be the languages of education in primary schools, and English was to be used in secondary education (Federation of Malaya, 1951:4). This suggestion was contested by the Chinese communities who wished to protect and preserve their cultural identity (Kärchner-Ober, 2013; Samuel & Khan, 2013, Samuel et al., 2014:208). Maintenance of the Chinese-medium education was called for via the *Fenn-Wu Report* (Federation of Malaya, 1951). Consequently, the *Education Ordinance* was implemented in 1952 to accommodate the needs of the plurilingual communities to maintain their vernacular languages in education, as well as to establish a national school system. Its proposal included making Malay or English the medium of instruction and additionally offering Mandarin and Tamil as language subjects in national schools (Federation of Malaya, 1952; in Hashim, 2020:375; Kärchner-Ober, 2013:308). However, the proposal was not well received by the Chinese and Indian communities as Chinese and Tamil-medium schools were still not considered in the plan (Hashim, 2020:275; Tan, 2009:8).

In the efforts to appease the Chinese and Indians (Hashim, 2020:375) and to preserve the cultural heritage of the plurilingual communities (Samuel et al., 2017:3), the 1956 *Razak Report*, also known as the *Report of the Education Committee*, recommended the establishment of national and national-type schools under a common curriculum (Ministry of Education, 1956). That is, national schools were to be conducted in Malay, whereas national-type schools would comprise English, Mandarin, and Tamil as the other languages of instruction. The proposal also aimed to gradually introduce Malay into the curriculum in order to make it the main language of education. Upon the implementation of *Article 152* of the Federal Constitution (1957, amended in 2009), Malay was granted status as the national language in Malaya. On the other hand, the official status of English would be revoked 10 years after the Malayan Independence on 31 August 1957. This enactment later applied to Sabah and Sarawak, which joined Malaya to form the federation of Malaysia on 16 September 1963. By 1973 and 1985, English ceased to be the official language in Sabah and Sarawak, respectively. However, it is still permitted to be used for legislative purposes (National Language Act, 1967; see also David, 2003; Haji Omar, 2012:158; Hashim, 2020:376-379). The shift in official languages left an impact on the country's education system as Malay became the sole medium of instruction across national schools. It was also made a compulsory language subject in

national-type primary schools (*Education Act*, 1961). The official status of Malay was consolidated through the *National Language Act* (1967). This motivated the transition from English to Malay as the language of education in 1970, thereby phasing out all English-medium schools by 1983 (Haji Omar, 2015:14; Hashim, 2020:379). Although English has lost its role as the main educational language, it remains a compulsory language subject in the national education system (*Education Act*, 1996, amended in 2016).

The lack of emphasis given to English has led to a proficiency decline over the years in Malaysia (Haji Omar, 2015:14; Kärchner-Ober, 2013). Acknowledging the importance of English as the language of globalisation, the *Teaching and Learning of Science and Mathematics in English* policy (in Malay: *Pengajaran dan Pembelajaran Sains dan Matematiks dalam Bahasa Inggeris*) (hereafter *PPSMI*) tabled in 2003 elevated the educational role of English in Science and Mathematics across all national primary and secondary schools. PPSMI was driven by the *Vision 2020* policy proposed in 1991, whose aspiration was to transform Malaysia into a fully developed country by 2020. As science and technology are regarded as pivotal fields for national development, it was hoped that the reintroduction of English in Science and Mathematics would not only help students to improve their command of English but also familiarise them with technical terminologies for better information access to these fields. However, the implementation of PPSMI had obverse effects on the performances of both students and teachers. Since English was acquired as an additional language for most of the Malaysian population, many students were still struggling to overcome the language barrier and therefore could not perform well in Science and Mathematics. Teachers had not been given sufficient training to teach these subjects using English, thus many had to resort to codeswitching between English and the vernacular school language. This in turn posed added repercussions to the students' learning (Haji Omar, 2015:21; Hashim & Leitner, 2015:50; Tan, 2011). Due to the shortcomings of PPSMI, English was once again rescinded of its role as the medium of instruction for Science and Mathematics (Ministry of Education Malaysia, 2019; cf. Gill, 2012), and the phasing-out process began in 2012 under the new *Upholding the Malay Language and Strengthening the English Language* policy (in Malay: *Memartabatkan Bahasa Malaysia dan Memperkukuh Penguasaan Bahasa Inggeris*) (hereafter *MBMMBI*). By 2021, Science and Mathematics were once again taught using Malay or the vernacular languages in primary schools, and only Malay was used in secondary schools (Ministry of Education Malaysia, 2014).

The reduced status of English means that there are now reduced hours of exposure to English in the classroom. This raises concerns as the need to improve on English nationwide

remains. According to the *Malaysian Education Blueprint 2013-2025* (hereafter *Blueprint*), only 28% of Form 5 secondary school students managed to pass the English Language subject in the 2010 SPM exam (in Malay: *Sijil Pelajaran Malaysia*; in English: the Malaysian Certificate of Education) (Ministry of Education Malaysia, 2013, exhibit 4-11). Moreover, the *Blueprint* attributes low student performance in English to low proficiency amongst many English Language teachers who failed to achieve the minimum requirement for the Cambridge Placement Test (ibid.). Following this, the *English Language Education Reform Roadmap 2015-2025* (hereafter *Roadmap*) was formulated in 2015 to enumerate phase-by-phase plans to be carried out in the hope to remedy the proficiency issues faced by both students and teachers. Some of the plans include introducing the Common European Framework of Reference for Languages (CEFR) as a standard scale to measure English proficiency, making the English subject a compulsory pass for students by the end of their secondary education, setting minimum proficiency requirements for English Language teachers, and increasing remedial intervention programmes and training courses for students and teachers, respectively (Ministry of Education Malaysia, 2015; Mohd. Don & Abdullah, 2019).

Since Independence, constant changes in the language education policy have been made in the attempt to enhance the quality of education in Malaysia. However, given the phasing in and out of English as the medium of instruction and insufficient exposure to English in the classroom, Malaysia is unfortunately witnessing a deterioration in the quality of English (Adi Badiozaman, 2019; Hashim, 2014; see also Gooch, 2009). Notwithstanding, one cannot deny that the Ministry has always recognised the importance of English in the workforce, since English is needed to conduct the business of government and to compete successfully on the international arena (Government of Malaysia, 1976:384-387; Ministry of Education, 2015). Hence, while efforts to improve English proficiency are still under way, English is here to stay.

Thus far, we have seen how the socio-historical and political factors have contributed to the learning of English in Malaysia. The next section takes us into sociolinguistic territory and reveals how English became an indigenised variety in the local context.

2.4. English in Malaysia

The English language may have its historical connections with the British, but over the last two centuries it has undergone linguistic decolonisation and indigenisation in Malaysia. In linguistic research, this localised variety of English has been given the generic term *Malaysian*

English (MalE). MalE has earned its status as an institutionalised variety (Baskaran, 1987; Siegel, 2003:180) not only because it bears an instrumental purpose nationwide but also because it consists of subvarieties that are used in different social settings (Kachru, 1992a:55). Furthermore, there are other intertwining factors such as crosslinguistic influence, proficiency, and bilingual creativity (Kachru, 1985; Bhatt, 2001) that contribute to the complexification of MalE. The goal of this subsection, therefore, is to unravel the intricacies of MalE by introducing the subvarieties that have emerged from the sociological-cum-acquisitional dimensions.

According to Baskaran (1987, 1994, 2008a; see also Govindan & Pillai, 2009:74; Platt & Weber, 1980; Bickerton, 1973, 1975), MalE comprises subvarieties that fall along a three-tiered lectal continuum. The *acrolect*, which is perceived as the 'high' or prestigious sociolect, is the nativized standard model of English. De facto, it is employed in official undertakings and formal education. It may tolerate slight phonological variation and allow the use of local loanwords, but its morphosyntactic structure strictly adheres to Standard English (StE). On the opposite end of the spectrum, the *basilect* is viewed as a patois or 'broken' English. This is because its linguistic properties reflect those of untutored L2-English and it involves a heavy load of lexical transfer from other languages (Baskaran, 1987:44, 2008a:281). Therefore, due to its substantial divergence from the acrolectal variety, the use of the basilect may generally impede intelligibility (Baskaran, 1994:29). Meanwhile, the *mesolect* falls between the acrolectal and basilectal varieties. It diverges quite substantially from the standard variety but is not as stigmatised as the basilect. Known commonly as *Manglish*, or within the academic circle as *Colloquial Malaysian English* (CME) (Govindan & Pillai, 2009; Pillai & Ong, 2018; see also Hashim, 2020; Lowenberg, 1993; Pillai, 2013; Rajadurai, 2007; Thirusanku & Md. Yunus, 2012, 2013), this is the subvariety that is most familiar to many Malaysians.[3] It is used in casual settings and is acceptable in some (semi-)formal situations, locally (Baskaran, 1987, 1994, 2008a; Gill, 1999; Morais, 1998; Nair-Venugopal, 2003). Given that StE is taught as a compulsory language subject in school, many linguistic properties of CME resemble those of tutored L2-English (Baskaran, 1987:44). To wit, the association between the sociolects of MalE and proficiency levels of StE cannot be clearer. Speakers who are more proficient in StE can

---

[3] There has been resistance within the academic sphere to use the word *Manglish*. This is due to the pejorative connotation it seems to carry (Pillai et al., 2010:159; Pillai & Ong, 2018:149). To appease scholars who are affected by it, the nomenclature *Colloquial Malaysian English* shall be adopted throughout this thesis. Notwithstanding, in the efforts to destigmatise the putatively pejorative term *Manglish*, attitudinal data from the main study (see **Chapter 8**) will be used to challenge such misconceptions about the term.

style-shift between the acrolectal variety and the colloquial ones, depending on the social context (Baskaran, 2008a:281; Pillai, 2013; Pillai & Ong, 2018). Those who are less proficient in StE are restricted from using the full range of the lectal cline, regardless of the social context they are in.

Moreover, proficiency and social context are not the only factors engendering the subvarieties of MalE. Crosslinguistic influence from the substrate languages (mainly Malay, Chinese, and, to a certain extent, Tamil) provide an added dimension to MalE (Morais, 1998; Nair-Venugopal, 2000; Ng & Diskin-Holdaway, 2021; Pillai & Ong, 2018). At the same time, language mixing is common when conversing in English (Baskaran, 1987, 1994; Thirusanku & Md. Yunus, 2012, 2013). Such extensive contact with the substrate languages further facilitates the linguistic restructuring of English, giving rise to various ethnolects of MalE.

Taken together, MalE is not a homogeneous linguistic product per se but a miscellany of subvarieties engendered by social factors, proficiency, and language contact.

2.5. The morphosyntax of Colloquial Malaysian English

As one of the lectal varieties of interest in this thesis is CME, this section describes its morphosyntactic properties as attested in the MalE literature. To reiterate, CME is a developing variety of English shaped by internal (e.g., L2 behaviours, proficiency, bilingual creativity) and external conditions (e.g., amount of exposure to English, societal attitude) modulated by the Malaysian ecology. Since English is acquired as an L2 by the majority Malaysian population (Crystal, 2003:108-109), it follows that the L1 or dominant language will inevitably have an influence on the linguistic restructuring of English at individual and societal levels. Moreover, CME is claimed to be spoken by Malaysians who have received classroom instructions on StE (Baskaran, 1987:44). This means that its linguistic system also possesses features resembling L2 learner behaviours that cannot be accounted for by L1 effects such as overgeneralisations (Hashim, 2020; Lim, 1976; Lowenberg, 1986). Given the L2-like profile of CME, one may argue that this localised variety is grammatically more impoverished than StE. However, this view is not entirely accurate because, as we shall soon discover, there are structural innovations that show morphosyntactic complexity governed by "internal change and linguistic creativity" (Schneider, 2007:102; see also Kachru, 1985; Bhatt, 2001).

2.5.1. Nominal domain

2.5.1.1. Omission of articles

One of the nominal features attested in CME is the omission of articles (Baskaran, 2008b:611-612; Hashim, 2020:386-387; Percillier, 2016b:90; Pillai, 2013:575; Pillai & Greig, 2020; Schneider, 2003:58-59):

(1) a. Ming: Maybe we need to put ϕ soul inside the AI [artificial intelligence] to solve this problem.
b. Anna: It's just like ϕ walking distance from my husband's house lah.[4]

This feature may be attributed to crosslinguistic influence from the substrate languages of Malay (Baskaran, 2008b:611-612; Hashim, 2020:386-387) and Chinese (Robertson, 2000; Wong & Quek, 2007), as they do not have an equivalent article system to that of English.

2.5.1.2. Number agreement

2.5.1.2.1. Omission of plurality

The marking of English plurality is optional in CME (Percillier, 2016b:68-71; Schneider, 2003:57):

(2)   Ming: Because there's a lot of male gamer-ϕ mah.

Entertaining the plausibility of substrate influence, plurality in Malay can be expressed using numeral classifiers (3a), noun reduplication (3b), or numeral adverbs (3c) (Nomoto & Soh, 2019). Chinese also employs a classifier system (3d) and allows for the use of adverbials (3e) but additionally has a plural suffix (们 *men*) reserved only for animate NPs (3f) (Sun, 2006).

---

[4] Unless cited, the examples demonstrating the non-standard features of CME were taken from the corpus of this thesis's small case study (see **Chapter 3**).

(3) a. Malay: dua   buah   buku
                    two   CL   book
                    "two books"

     b. Malay: buku-buku
                    book-book
                    "books"

     c. Malay: semua   buku
                    all   book
                    "all books"

     d. Chinese: 两   位   同学
                    liǎng   wèi   tóng xué
                    two   CL   student
                    "two students"

     e. Chinese: 全部   同学
                    qúan bù   tóng xué
                    all   student
                    "all students"

     f. Chinese: 同学   们
                    tóng xué   men
                    student   PL
                    "students"

Furthermore, indicating plurality in Malay and Chinese is not obligatory. For this reason, the lack of plural marking observed in CME may be another instance attributed to substrate transfer.

2.5.1.2.2. Pronominal concord

The lack of pronominal concord – attested in CME, (Baskaran, 2008b:612; Hashim, 2020:387; Newbrook, 1997:248; Pillai & Greig, 2020) – is a non-standard phenomenon where a pronoun and its corresponding antecedent do not agree in number:

(4) *These books* are very informative. *It* can be obtained at Dillon's. (Baskaran, 1987:348)

41

According to Baskaran (1987:348-350), the lack of pronominal concord may be attributed to partial transfer from Malay, where the third-person invariant pronoun *ia* is used to refer to non-human or inanimate antecedents irrespective of their number (i.e., singular/plural) distinctions:

(5) [*Surat-surat itu*]$_i$ baru   sampai – mungkin   *ia*$_i$ dari   ayah   saya.

  [*Letters   those*]$_i$ just   arrive – maybe   *it*$_i$ from   father   POSS.1S

  "Those letters have just arrived – they must be from my father."

(Adapted from Baskaran, 1987:349)

2.5.1.3. Pronominal omission

The omission of pronominal arguments is argued to be transferred from the pro-drop languages of Chinese (Huang, 1984; see also Sato & Kim, 2012 on CSE) and Malay (Mustaffa, 2020; Nomoto & Soh, 2019:489-499). In CME, pronominal omission happens not only at subject positions (6a-b) but also at object positions (6c-d) (Baskaran, 2008b:618; Newbrook, 1997:246; Hashim, 2020:387; Percillier, 2016b:90; Pillai, 2013:574; Pillai & Greig, 2020; Schneider, 2003:58):

(6) *Subject ellipsis*

  a. Ming: Black Panther$_i$, ø watch already t_i ah?

  b. Amir: Fukuoka, um ø never been, but many friends from Fukuoka, yeah.

  Object ellipsis

  c. Ming: I also don't think I'll watch ø alone.

  d.   Amir: I believe you go ø every day.

2.5.1.4. Omission of possessive 's

There is a tendency for the English genitive/possessive marker *'s* to be dropped in CME (Platt et al., 1983, 1984; Schneider, 2003:57; cf. Pillai & Greig, 2020):

(7) Ming: Joy-ø boyfriend go to Taiwan with his own friends.

The Chinese genitive structure follows that of StE in that the particle 的 *de* links the pre-modifying NP-possessor to the NP-possessee that is head noun. While *de* is usually required to mark genitive phrases (Shi & Li, 2002), there are certain conditions where it can be optional, such as when it links a pronominal possessor (e.g., *wo* "my") to its head noun denoting kinship (e.g., *baba* "father") in (8a) (Zhang, 2005:167) or when the head noun (e.g., *erzi* "son") is preceded by a demonstrative modifier (e.g., *ne* "that") in (8b) (Shi & Li, 2002:8):

(8) a. 我（的） 爸爸   的   书
       wǒ (*de*) bàba de   shū
       my POSS father POSS book
       "my father's book"

                                                          (Adapted from Zhang, 2005:167)

     b. 张三   (的)   那   儿子   常常   抽烟
       zhāng sān (*de*) nà   érzi   chángcháng chōuyān.
       Zhangsan POSS that   son   often   smoke
       "Zhangsan's son often smokes."

                                                          (Adapted from Shi & Li, 2002:8)

Meanwhile, Malay has two types of genitive constructions. In standard constructions such as (9a), the NP-possessor post-modifies the head noun without a genitive marker (Tija, 2004:54). The other option, as seen in (9b), follows the StE and Chinese patterns in that the pre-modifying possessor and its head noun on the right are linked by the Malay *punya*; that said, this structure is used in colloquial Malay (Nomoto & Soh, 2019:489; Tija, 2004:54).

(9) a. ayah   saya
       father   1S
     b. saya   *punya* ayah
       1S   POSS father
       "my father"

                                                          (Adapted from Tija, 2004:53)

Given the conditional omission of the genitive markers in Chinese and Malay, it is tempting to assume that the optionality of the English *'s*-genitive observed in CME is a result of partial

transfer from these substrate languages. As this non-standard feature is not widely attested in the CME literature, more research needs to be done to (dis)confirm this claim.

2.5.1.5. Omission of relativizers

In StE, relative clauses (RCs) – i.e., clausal modifiers of NPs – are headed by a relativizer (e.g., *that*, *which*) which can be dropped if an object-NP is relativized. Subject extraction without a relativizer would render the sentence ungrammatical. However, this non-standard construction is attested in CME (Percillier, 2016b:90-91; Platt et al., 1984:8-9; cf. Pillai & Greig, 2020):

(10) Xing: Yalah, *people*ᵢ [RC [RP ∅] *t*ᵢ dry and not dry] is very different eh.

In Malay and Chinese, complementizers (i.e., Malay *yang*, Chinese 的 *de*) are employed to head RCs. Their difference, however, lies in the positions of the RC and relative markers around the head noun (Alsagoff & Ho, 1998:130-131). Malay is similar to English in that the RC post-modifies the nominal head; they are also similar in that the relativizer *yang* occupies the left-periphery of the RC (11a). Chinese differs from both languages in that the RC pre-modifies the head noun and that the relativizer *de* proceeds the RC (11b). Notwithstanding, unlike English, the Malay *yang* (Awang, 1981; Wong, 2008:115) and Chinese *de* (Chen, 2008; Wang, 2016) are obligatory markers of relativization. Therefore, it is unlikely that the optional feature of the English relativizer in CME is an effect of substrate transfer.

(11) a. Budak   ituᵢ   [RC [RP yang] *t*ᵢ mencubit   ibu   saya] sangat jahat.
          Child   that   REL   pinch   mother my   very   naughty

     b. [RC *t*ᵢ 捏   我的   妈妈   [的]] *那个   孩子*ᵢ 很   坏蛋。
         [RC *t*ᵢ niē   wǒde   māma   [RP de]]*nàgè   háizi* hěn   huàidàn
             Pinch   my   mother REL   that   child   very   naughty
         "*The child*ᵢ [RC [RP *who*] *t*ᵢ pinched my mother] is very naughty."

                                                          (Adapted from Alsagoff & Ho, 1998:129)

2.5.2. Verbal domain

2.5.2.1. Omission of tense inflections

While tense marking is obligatory in StE, it is often dropped in CME (Baskaran, 2008:614; Hashim, 2020:389-390; Percillier, 2016b; Pillai, 2013:575; Pillai & Greig, 2020; Pillai & Ong, 2018:150). The examples in (12a-b) present omissions of the third person singular -*s* (3rd.SING.-*s*), whereas (12c-d) and (12e-f) respectively show regular and irregular verbs being unmarked:

(12) *Omission of 3rd.SING.-*s

    a. Ming: Every year she just send-*ø* message two time.

    b. Amir: Yeah, even though it's on your responsibility, if anything happen-*ø*, you still need to be involve.

    Omission of regular past tense

    c. Xing: No lah, Chan just finish-*ø* studying Dentist.

    d. Anna: They held this event but they ask-*ø* for my help lah, because they never plan such a big event like this, kan?

    Omission of irregular past tense

    e. Ming: You know or not, we went to the train right, she *don't* even want to sit with us oh.

    f. Anna: When I delivered Siti, I *begin* to develop that craving for sweet foods every day.

Unlike English, Malay and Chinese do not have inflectional and agreement systems for tense (Pillai, 2013:575-576; Platt & Weber, 1980; Wong, 2012). These contact languages rely on adverbial phrases (e.g. Malay: *semalam* "yesterday", Chinese: 刚才 *gāng cái* "just now") to convey temporal information (Hashim, 2020:389; Pillai, 2013:575).

2.5.2.2. Omission of copula BE

Omission of the English copula BE is another CME feature (Newbrook, 1997:246; Percillier, 2016b:93-95; Pillai, 2013:576-577; Pillai & Greig, 2020):

(13)   Ming: I actually pity Ruby but I ∅ also happy for her.

This feature is likely to be a by-product of substrate transfer, since copulas can be, conditionally, optional in Malay (Mustaffa, 2018; Wong, 2012:8) and Chinese (Li & Thompson, 1981; Wong, 2012:8-9; Zhan & Sun, 2013:762-763).

2.5.2.3. Omission of auxiliary verbs

In CME, the English auxiliary verbs (i.e., progressive/passive BE, perfect HAVE, auxiliary DO) can be dropped (Hashim, 2020; Pillai & Greig, 2020; Schneider, 2003:58), not only in declarative sentences (14a-b) but also in interrogative constructions (14c-d).

(14)   *Omission of auxiliary verbs in declaratives*
    a.   Ming: Maybe I ∅ still thinking want to get a Switch or not.
    b.   Amir: I ∅ never been to Mount Kinabalu.

    Omission of auxiliary verbs in interrogatives
    c.   Xing: Why ∅ you want to message me oh?
    d.   Amir: What kind of food ∅ you like, seafood?

An explanation for this is that there are no auxiliary counterparts in Malay and Chinese. In these contact languages, adverbials (e.g., Malay: *sedang* "currently") or grammatical particles (e.g., Chinese: 了 *le* perfective marker) are usually employed to convey manners of time.

2.5.2.4. Stative progressivity

The overextension of the progressive aspect to stative verbs, a.k.a. stative progressivity, has not been widely recognised as a feature in the CME literature. Baskaran (1987, 1994, 2008b) is amongst the few that attests to its presence in the discourse (cf. Pillai & Greig, 2020):

   (15)   That bottle *is containing* sulphuric acid. (Baskaran, 2008b:615)

According to Baskaran (1987:290-294), stative progressivity may have received influence from the Malay *meN-* prefix, which changes the stative readings of certain verbs (e.g., *mempunyai* "to own") to have a dynamic or progressive-like interpretation (see also Soh & Nomoto, 2009):

   (16)   Dia   (sedang)   *mem*punya*i*   tiga   buah   rumah.
          He/she   (currently)   possess   three   cl.   houses
          "He/she (currently) possesses three houses."

                                                         (Adapted from Baskaran, 1987:294)

2.5.2.5. Double-tense marking in DO-constructions

The overgeneralisation of past tense on lexical verbs in DO-constructions is not, to my knowledge, widely attested in the CME literature except by Pillai & Greig (2020), who acknowledge it as a feature that is "neither pervasive nor extremely rare":

   (17)   The boy *didn*'t even really *got* angry. (MM032, corpus of the main study)

2.5.2.6. Structural innovations

Features of structural innovation are features that contribute to the complexification of indigenised Englishes such as CME (and CSE). According to Bamgbose (1998), innovations are non-standard features that encompass features of L2 behaviours at the morphosyntactic level [e.g., I cannot be able to go (ibid., p.3); furnitures (ibid., p.4)], "culture induced" expressions at the lexical level [e.g., go-slow (=traffic jam) (ibid., p.3)], and features from other

47

linguistic domains. He further proposes that non-standard features are likely to be accepted as innovations if they are widely used and/or accepted by the speech community (ibid., p.3). While this study concurs with Bamgbose (1998) that innovation is necessitated by the extensive use of a non-standard feature, we assign a narrower definition to the term *structural innovation*. That is, structural innovation is a linguistic mechanism that not only allows for substrate influence and/or other L2 behaviours in the restructuring of a vulnerable feature at a morphosyntactic level, but also involves bilingual creativity that marks the sociocultural identity of the speech community using it (Kachru, 1985).

2.5.2.6.1. Already

A notable case of structural innovation is the adverbial *already* (Baskaran, 1987; Pillai, 2013:575; Pillai & Greig, 2020; Platt, Weber, & Ho, 1984:71). It has undergone grammaticalization to compensate for the lack of grammatical aspect in CME. Corresponding to the functions of the Chinese particle 了 *le*, *already* bears not only the standard meaning denoting the completion of an event (18a-b) but also inchoative (18c) and inceptive (18d-f) readings that are non-standard (Bao, 2005; Li & Siemund, 2021:522:526; Ziegeler, 2020; Ziegeler & Lee, 2019):

(18) *Perfective* already *(standard meaning)*
   a. Ming: I finish *already* mah.
   b. Anna: It's been like five months *already*.

   Inchoative already
   c. Xing: I mean, if you go UK come back, different *already*.

   Inceptive already
   d. Ming: What is the next level after chef?
       Xing: I don't know.
       Ming: Waiter, can meet people *already*.
          ≈ Waiter, can *start to* meet people.
   e. Ming: Maybe you got too many girlfriend, you got bored *already*.
          ≈ Maybe if you have too many girlfriends, you'll *start to* get bored.

48

f. Anna: And then, if the person stay here around like one week, uh he or she will get bored *already* because we don't know what else to eat.

2.5.2.6.2. Got

GOT is another prevalent feature of structural innovation in CME (Baskaran, 2008b:618; Hashim, 2020:388; Percillier, 2016b:71-72; Pillai, 2013:577; Pillai & Ong, 2018:151; Pillai & Greig, 2020; Schneider, 2003:58). In this context, the past tense form of the English verb *get* has received grammatical and semantic reformulations under the substrate influence of the Chinese 有 *yǒu* "to have" (Bao, 2005, 2014) and, to a certain extent, the Malay *ada* "to have" (Goddard, 2002; see also Hiramoto & Sato, 2012 and Lee et al., 2009 on Baba Malay in the CSE literature). Similar to *already*, GOT can be employed to mark perfective aspect (19a). Moreover, it carries stative/habitual (19b) and existential (19c) interpretations, resembling the usage patterns of Chinese *yǒu* (Bao, 2014). The Malay *ada* functions as an existential predicate as well (Goddard, 2002:114-117), but its semantic patterns are less varied than those of its Chinese counterpart.

(19) a. *Perfective* GOT

Xing: But I don't know if her sister *got* receive any cards or not lah.

   ≈ But I don't know if her sister *has* received any cards.

b. *Stative/Habitual* GOT

Xing: You still *got* play Pokémon meh?

   ≈ Do you still play Pokémon?

c. *Existential* GOT

Ming: What do you think ah, Eve and Adam, *got* any chance ah?

   ≈ What do you think about Eve and Adam, *is there* any chance [between them]?

2.5.3. Clausal and discoursal domains

2.5.3.1. Interrogative constructions

2.5.3.1.1. Subject-auxiliary (non-)inversion in interrogatives

The interrogative structures of CME do not conform to those of StE in terms of subject-auxiliary inversion (SAI). In StE, SAI is obligatory in direct questions but is not allowed in indirect ones. Contrariwise, in CME, inversion is optional in direct questions (20a-d) but overgeneralised in indirect ones (20e-f) (Baskaran, 2008b; Govindan & Pillai, 2009; Pillai & Greig, 2020):

(20) *Non-inversion in direct yes/no- (a-b) and wh- (c-d) questions*
   a. Ming: You *ø* going to watch Black Panther?
   b. Anna: Eh, *this is* you being nice or you being honest?
   c. Ming: What *ø* he say, get wreck?
   d. Anna: When *he's* coming back?

   Inversion in indirect yes/no- (a) and wh-(f) questions
   e. Ming: Actually I don't know *is it* they go Singapore work as admin ah, janitor ah, like that.
   f. Ming: I don't know *what's* the name of the station.

The non-standard non-/inversion patterns are rather common L2 behaviours observable also in other indigenised varieties such as the colloquial varieties of Indian English (Bhatt, 2004:1019-1020) and Singapore English (Gupta, 1994). The lack of SAI in direct questions in CME may be attributed to the absence of such syntactic equivalents in the interrogative structures of the local languages. Meanwhile, the overgeneralisation of SAI in indirect questions could be due to formulaic language learning habits, whereby certain combinations of words (i.e., [interrogative word] + [subject NP] + [verbal element]), if occurring frequently enough, would motivate inversion in embedded questions (Hilbert, 2011).

2.5.3.1.2. Wh-*in-situ constructions*

Another interrogative-related phenomenon is the optionality of *wh*-movement. It is argued to reflect substrate transfer since *wh*-in-situ is exhibited in Malay (Baskaran, 2008b:615-616; Sato, 2013) and Chinese (Bao, 2001; Gupta, 1994):

(21) a. Anna: Osaka is in *which part of Japan* ah?

    b. Xing: I wonder the one with the bag of biscuit one, those biscuit go to *where*.

2.5.3.1.3. Yes/no-question particles

In CME, it is common to find question tags and particles assigned at the end of direct interrogatives, especially given that in-situ question constructions are allowed in CME. One of the common ones is the question tag *or not* (Baskaran, 2008b:616; Hashim, 2020:388; Percillier, 2016b:112-113; Pillai, 2013:577; Pillai & Ong, 2018:151; Pillai & Greig, 2020):

(22) Ming: Sonia got boyfriend *or not*?

Or not marks direct yes/no-questions and is argued to have been influenced by the Malay in-situ interrogative structure (Baskaran, 2008b:616):

(23) Dia   makan *atau*   *tidak*?

     He   eat   or   not?

     "Did he eat?"

<div align="right">(Baskaran, 2008b:616)</div>

There are other yes/no-question particles transferred directly from Malay and Chinese. These include *ah* (24a), *kah* (24b), and *meh* (24c):

(24) a. Ming: You want to go toilet *ah*?

    b. Anna: It's not halal *kah*?

    c. Ming: Eh, you met his girlfriend before *meh*?

The particle *ah* is one of the most versatile and widely used particles in CME (and CSE) (Gupta, 1992; Ler, 2006:149-150; Smakman & Wagenaar, 2013:315-316). It is likely to have come from the Chinese dialect of Hokkien or Bazaar Malay (Lim, 2007:460; cf. Smakman & Wagenaar, 2013:316). When used to type a yes/no-question, it is marked with a low pitch (Lim, 2007:460). *Kah* can be traced back to the Malay language (Tay et al., 2016:483, 504). Unlike the multifunctional particle *ah*, *kah* can only be used as a question particle (Nomoto & Soh, 2019:503-504). *Meh* carries a narrower connotation than *ah* and *kah* in that it expresses surprise to a contradictory proposition (Gupta, 1992:44; Ler, 2006:157) or hostility to a remark (Tay et al., 2016:493). *Meh* is believed to have come from the Chinese dialect of Hokkien (Tay et al., 2016:493) and is pronounced with a high tone (Wong, 2004:780).

2.5.3.2. Discourse particles

The question particles above are merely some of the cornucopia of discourse particles attested in CME. For instance, in Tay and colleagues' (2016:504) study, at least 20 particles were documented among Malaysian Facebook users, with the polyfunctional particle *lah* being the most common one in their data (see also Ler, 2006:150). Particles usually appear in clause or sentence-final positions, although they can be attached to (topicalised) phrases as well:

  (25)   But now *oh*, a lot of people do that *mah*.

This study will not expound on discourse particles. However, it is worth noting that while the meanings particles carry can be distinguished intonationally, the pragmatic context in which they are used remains important in disambiguating the intended meaning (Tay et al., 2016:483-484). The linguistic origins of many CME particles can be traced back to Malay and Chinese (Gupta, 1992; Lim, 2007; Tay et al., 2016), yet the origins of some other particles, such as *lah* and *ah*, are blurred as they have been shared extensively between the Malay and Chinese substrates (Smakman & Wagenaar, 2013).

2.5.3.3. Language mixing

Given the linguistic diversity in Malaysia, language mixing is bound to happen regardless of what the main language in use may be. Being a product of language contact itself, CME is not

exempt from creating a space for language mixing to take place (Baskaran, 1994; Rajadurai, 2007; Thirusanku & Md. Yunus, 2012). A couple of instances of language mixing include *lexical borrowing* (26) and *codeswitching* (27). Lexical borrowing is when words or phrases from other languages are integrated into the lexical inventory of the target language (Myers-Scotton, 1992; Pfaff, 1979; Poplack, 2018). These borrowings may be related to institutional concepts (e.g., *bumiputera*, Malay: son of the soil), material-related terms [e.g., *sambal*, Malay: a type of chilli paste (26a)], or words associated with the local culture [e.g., *tahan*, Malay: "to endure" (26b)] (Baskaran, 2008b:619-620):

(26) a. Amir: I would like to have *sambal* if you have lah.

     b. Ming: […] she cannot *tahan* already mah.

Meanwhile, codeswitching is the dynamic alternation between two or more languages. It can take place within a clause/sentence (27a) (i.e., *intrasentential codeswitching*) or at a sentential boundary (27b) (i.e., *intersentential codeswitching*):

(27) a. Xing: But then the children very big already. I think now is 一岁 ah.

      [Mandarin Chinese: *yī suì*, lit. one year old]

     b. Amir: *Kalau* Japanese, they prefer well balance food. *Diaorang simpan sikit-sikit garam*.

      "*If* [it were the] Japanese, they prefer well-balanced food. *They add a bit of salt*."

2.5.4. Review summary

We have seen that the morphosyntax of CME has been influenced mainly by the main substrate languages of Malay and Chinese. Figure 2.1 recapitulates the CME features which have been derived from the linguistic mechanisms of substrate transfer and other L2 acquisitional behaviours not accounted for by transfer effects. Features that are plausibly a result of covert transfer include omissions, the lack of pronominal concord, and in-situ interrogative constructions. Cases of overt transfer include codeswitching, lexical borrowing, and the use of discourse particles, which can be subsumed under codeswitching or lexical borrowing depending on the degree of use. Features of CME resembling fossilizations of L2 behaviour include overgeneralisations of double-tense marking and inversion in embedded interrogatives.

More interestingly, although structural innovations, such as *already* and GOT, have indeed arisen from substrate transfer, they have additionally undergone relexification signalled by bilingual creativity. Despite being an L2 variety, CME allows much room for language mixing and creative innovation. This does not strictly entail morphosyntactic simplification per se but illuminates complexification that is uniquely tailored to CME.

Figure 2.1: Summary of CME features derived from the plausible linguistic mechanisms of substrate transfer, L2 behaviour, and bilingual creativity (marked with asterisks)

Linguistic mechanisms

Transfer

L2 behaviour

Covert

Overt

Nominal

Verbal

Clausal

Clausal/ Discoursal

Verbal

Clausal

Omissions of articles, arguments, possessive 's, relativizers; Pronominal concord

Omissions of tense, copula, and auxiliary verbs; Stative progressivity *Already*; *Got*

No SAI in direct questions; *Wh*-in-situ

Language mixing Lexical borrowing; Codesswitching; Discourse particles

Double-tense marking in DO-support

Overgeneralisation of SAI in indirect questions

2.6. Conclusion

This chapter has unveiled the socio-historical and linguistic tapestries of Malaysia leading to the emergence of an indigenised variety of English. It has also expounded on the morphosyntactic properties of CME by presenting a review on the literature, demonstrating that while many CME features do reflect transfer effects from the main substrate languages of Malay and Chinese and other L2 behaviours such as overgeneralisations, there are also features that have evolved beyond these mechanisms due to structural innovation propagated by bilingual creativity. This therefore counters the putative view that CME is merely a simplified version of StE and advocates the complexification of its linguistic structure.

The next chapter demonstrates some of the CME properties through a small case study, which provides not only supplementary evidence to the MalE literature but also a comparison for the main PhD study investigating the use of StE in the Malaysian context.

3   Colloquial Malaysian English: A small case study

3.1. Introduction

A small case study was carried out in February 2018 with the aim to explore and provide an empirically based description of CME using naturalistic data. Not only would this corroborate features already attested in the MalE literature, but it would also lay out a verifiable grounding for the main study.

A note to flag up is that the features described in **Chapter 2** were not all analysed in this case study. This is because the results to be presented were taken from the first-year PhD report (2018), which was exploratory in nature. There were also features not attested in the MalE literature which will be reported herein.

As literature on CME was already reviewed in the previous chapter, the next section (**§3.2.**) jumps to the methodology of the case study, covering its research questions and hypotheses, site of study, and participant profiles, etc. Results of the naturalistic data are presented in (**§3.3.**) and discussed in **§3.4.** The chapter wraps up in **§3.5.**.

3.2. Methodology

3.2.1. Research questions and hypotheses

RQ1:   **What are the stable features of CME shared between speakers of different substrate language (SL) groups?**
This exploratory question seeks to identify stable features of CME by gleaning non-standard features shared by speakers irrespective of SL group.

RQ2:   **Which features of CME are instances of (a) substrate influence or (b) general L2 behaviours?**
In relation to RQ2(a), it investigates if there are certain features that are predominantly or exclusively used by speakers of a certain SL group. With regard to RQ2(b), it examines if there are non-standard features that cannot be accounted for by substrate influence but by other L2 learning behaviours.

(a) <u>Substrate influence</u>

Null hypothesis:	There are no structural differences between CME and StE.
Alternative hypothesis:	If a non-standard feature can be explained by the presence or absence of the corresponding SL feature, then it is likely a result of substrate transfer.

(b) <u>L2 behaviour</u>

Null hypothesis:	There are no structural differences between CME and StE.
Alternative hypothesis:	If an English feature absent in the SL is overgenerated, then it likely reflects entrenched L2 learning behaviours.

RQ3: Which features of CME are a result of structural innovation?

This question aims to examine features that cannot be sufficiently explained by substrate influence and L2 behaviour but plausibly by bilingual creativity.

Null hypothesis:	A non-standard feature is not a case of structural innovation if it can be explained solely by substrate influence or general L2 behaviours.
Alternative hypothesis:	A non-standard feature is likely to be a case of structural innovation if it cannot be explained solely by substrate influence and general L2 behaviours.

3.2.2. Naturalistic conversations

As CME is usually spoken in everyday interactions, eliciting naturalistic conversations will be one of the most effective ways to document its linguistic properties. This type of observational data gives us a better understanding not only of the social context in which the variety is spoken, but also of how the variety itself is linguistically and pragmatically represented in its users (Hoffman, 2014:26; Shively, 2018:202).

A classic tool to collect naturalistic data is the *sociolinguistic interview*, during which the fieldworker engages in structured conversations with individuals who speak the vernacular of interest (Labov, 1972, 1984). However, the conversation being recorded might be adulterated due to the presence of the fieldworker (this is known as the *observer's paradox*) (Labov, 1972:113). To mitigate this effect, the sociolinguistic interviews conducted in this study were restructured by reducing the participatory role of the interviewer. By doing so, the

participants would be put at ease when conversing with each other (Schilling, 2013:109-110). This procedure is described further in **§3.2.5**; but in the meantime, the next subsections introduce the data collection site (**§3.2.3.**) and participant profiles (**§3.2.4.**).

3.2.3. Site of study

This study took place in my hometown, Tawau, which is a seaside city located southeast of Sabah (East Malaysia) and shares a border with Kalimantan, Indonesia. As of 2018 (Population and Housing Census of Malaysia. 2010), Tawau hosted a population of 506,700, 56.82% ($\frac{287,900}{506,700}$) of whom were Malaysian citizens. Regarding the local ethnic composition, 44.23% ($\frac{224,100}{506,700}$) were of Bumiputera status and 8.92% ($\frac{45,200}{506,700}$) were of Chinese ethnicity. The number of Malaysian Indians was scant and were included in the "Others" category ($\frac{18,600}{506,700}$, [3.67%]). Another note worth mentioning is that although the majority of Bumiputeras were Malay, there was also a considerable number of Bumiputeras from other indigenous tribes, such as Kadazan/Dusun, Bajau, and Murut. Non-citizens ($\frac{218,800}{506,700}$, [43.18%]) residing in Tawau mostly came from neighbouring countries, such as Indonesia and the Philippines. Given the geographical location and the ethnic makeup of Tawau, the local languages and dialects coming into contact with one another inevitably conjure a linguistic landscape different from other regions in Sabah and the rest of the Malaysian states (Pillai, 2013:574; Pillai & Ong, 2018:153). It is therefore worth noting that the linguistic data of this corpus may, to a certain extent, be different from the CME spoken elsewhere.

3.2.4. Participants

Four adult Malaysian participants took part in the case study. As they were friends of mine, they were contacted personally. Table 3.1 presents their profiles, which include basic demographic information, language history, language use, and English proficiency.

Table 3.1: Participant profiles in the CME case study

Participants	Ming	Xing	Anna	Amir
Age	25	25	29	30
Sex	Male	Male	Female	Male
Ethnicity	Chinese	Chinese	Malay (Bumiputera)	Sino-Dusun (Bumiputera)
Acquired languages				
L1	English	English	Malay	Malay
L2+ (age of onset, in years)	Chinese (1) Malay (2)	Chinese (4) Malay (5)	English (4)	English (5) Chinese (5) Japanese (19)
Medium of instruction				
Primary education	Mandarin	Mandarin	Malay	Malay
Secondary education	Malay	Malay	Malay	Malay
Tertiary education	English	English	English, Malay	English, Malay
Language use				
Family	English, Chinese	English, Chinese	Malay, English	Malay, English
Friends	Chinese, English	Chinese, English	Malay, English	Malay, English
At school (primary to tertiary education)	Chinese, English	Chinese, English	Malay, English	Malay, English
Strangers	Chinese, Malay, English	English	Malay, English	Malay, English
English proficiency	MUET: 4	SPM: A	MUET: 4	SPM: B

Ming and Xing are twin brothers of Chinese ethnicity. Anna and Amir are cousins of Bumiputera status.[5] Anna is ethnically Malay, whereas Amir identifies as Sino-Dusun (i.e.,

---

[5] In adherence to the UK's Data Protection Act, pseudonyms were given to the participants to anonymise their identities. On a different note, the participants' mean ages were calculated during the time of the study (2018).

partly Dusun and Chinese). All participants are at least bilingual and have acquired English and Malay. For Ming and Xing, English was acquired as their L1; Chinese and Malay were also acquired early in their childhood. Anna and Amir's L1 and L2 are Malay and English, respectively. Amir acquired Chinese at the same time as English fairly early in life and started learning Japanese on his own in his late adolescent years.

Although Ming and Xing's ethnic language, Chinese, was acquired later than L1-English, it was used as much as their L1. In fact, both English and Chinese were used with their family members. Outside their home, Chinese was used when interacting with friends, classmates, and educators of the same ethnicity. However, when interacting with people from different ethnic backgrounds, English was their preferred mode of communication. For Anna and Amir, Malay and English were both spoken at home and with friends regardless of their ethnicities; but beyond that, their linguistic experiences were relatively different to each other. Concerning Amir's language profile, Chinese is one of his heritage languages. Although it was acquired much earlier than Japanese, it was Japanese that was used more frequently with friends; Chinese, on the other hand, was rarely used. At school, Anna used English alongside Malay with her classmates and teachers more frequently during her secondary through tertiary education than during her primary education. For Amir, English was used more often at college and university than his earlier education.

Finally, concerning English proficiency, the participants were asked to provide their most recent formal assessment on the English Language. Only Ming and Anna undertook the Malaysian University English Test (MUET). [6] Both of them scored Band 4, which roughly corresponds to B2 of the CEFR (Malaysian Examination Council, 2022). Meanwhile, Xing and Amir reported the grades they obtained from the Malaysian Certificate of Education (hereafter *SPM*), which were grades A (CEFR B2) and B (CEFR B1), respectively (for grade-to-CEFR correspondences, refer *Roadmap*, Ministry of Education Malaysia, 2015:26).[7] Taken together, the proficiency levels of these participants fell within the CEFR "Independent User" category.

---

[6] The Malaysian University English Test (MUET) is a local English language proficiency test administered at (pre-)university levels. There are six bands altogether, with Band 6 indicating the highest proficiency category and Band 1 being the lowest.

[7] The Malaysian Certificate of Education (in Malay: *Sijil Pelajaran Malaysia* [SPM]) is equivalent to the UK's General Certificate of Secondary Education (GCSE). Its formal examination is taken during the fifth (final) form of secondary education. English is one of the compulsory subjects in SPM. The highest grade awarded is A1/A+, whereas the failing grade is G9/G.

3.2.5. Procedure

For the conversational task, the participants were paired according to their kinship. Each pair of participants was invited to interact with each other in what they understood as "Manglish". Two separate conversational sessions were run for each participant pair, and each session lasted around one hour. The participants were encouraged to explore any topic(s) of their preference.

The conversations took place in a public but quiet space and were recorded using a Sony ICD-UX560F sound recorder. Throughout each interview, I was present in the recording space. My role was to moderate the participants' conversation and resolve any technical issues that might arise during the live recording.

Before the recording, the participants were asked to read the participant information sheet (**Appendix B.1.1.**) and consent form (**Appendix B.2.1.**) before signing to agree to partake the study. At the end of the session, the participants were invited to fill out a short questionnaire (**Appendix A.1.**) for language profiling purposes.

3.2.6. Transcription

All speech data were transcribed and coded using the software tools and manuals provided by TalkBank (MacWhinney, 2000). To promote consistency and ease of readability, the transcriptions followed the format outlined by the facility's Codes for Human Analysis Transcripts (CHAT). Its integrated analytic tool, the Computerized Language Analysis programme (CLAN), was used to run searches and analyses (e.g., word frequency, mean length of utterances) on the transcripts.

The default language of the transcripts was English. In cases of *intrasentential codeswitching* (i.e., language switching taking place within an utterance), the @s marker was used to tag codeswitched items, including discourse particles (28a). Regarding *intersentential codeswitching* (i.e., utterances switching from one language to another), language pre-codes were assigned at the beginning of the speaker's utterance (28b).

(28) a. *ANA: what is taosa_pao@s:yue$n?
              "What is [red bean bao]?"
       b. *AMI: [- ind] kacang@s:ind$n hitam@s:ind$adj. [+ exc]
              "[Black bean]."

Code	Description
@s:yue$n	‣ Cantonese noun
@s:ind$adj	‣ Indonesian adjective
[-ind]	‣ Pre-code: Bahasa Indonesia
[+ exc]	‣ Post-code: exclude utterance

As the focus of the analysis was on English-based utterances, instances of intersentential and intrasentential codeswitching where the matrix language was not English were excluded. Also discarded from the analysis were imitations, self-repetitions, and utterances containing unintelligible words. Given the morphosyntactic complexity of CME, the transcripts hosted many linguistic features deviant from those of StE. Error codes prescribed by CHAT were initially assigned. However, due to much variation, they had to be removed to improve the readability of the transcripts. The next alternative taken was coding non-standard features on the printed transcripts manually.

For cross-checking purposes, each transcript was linked to its digitalised audio file. Written utterances were timestamped to their corresponding audio segments using the Walker Controller function in CHAT. The transcripts had been checked at least thrice on different occasions for any transcription errors and spelling errors before carrying out the data analysis.

3.2.7. Data analysis

The analysis of the current data involved two steps of quantification. In the first step, frequencies of the features of interest were quantified using CLAN's FREQ function. Specific words, alongside the utterances in which they occurred, were extracted using the KWAL function. As the error codes supplied by CHAT were not deployed in the transcripts, most of the features were quantified manually.

The second step involved running the statistical analysis. Participants were grouped according to their ethnic or substrate languages (SL), SL-Chinese and SL-Malay. This step compared SL proportions against the null hypothesis (i.e., no SL effects are found in the use of CME). The test statistic below was applied to the quantified observations tabulated in Microsoft Excel.

$$z = \frac{(P_x - P_y)}{\sqrt{P^*(1 - P^*)(\frac{1}{N_x} + \frac{1}{N_y})}}$$

where:

z	value on the z-distribution
P_x	proportion of sample x
P_y	proportion of sample y
P*	total proportions of samples $\{x+y\}$
N_x	total observations of sample x
N_y	total observations of sample y

Z-scores obtained from the formula above were then computed in R using its `pnorm()` function to generate p-values.

3.2.8. Ethical considerations

The case study obtained ethical approval from the Research Ethics Committee, University of Cambridge, and was conducted in line with the UK Data Protection Act.

Participants were recruited on a voluntary basis and were informed about the study before giving their signed consent to partake in the study. Paper copies of the participant information sheet and consent form were distributed to each participant for his or her own record. During the briefing, participants were made aware about the audio-recording activity that was about to take place and were reassured that all their personal data would be anonymised and protected. To minimise data sensitivity, they were strongly discouraged from discussing personal matters and using full names.

All files are stored in the author's passcode-locked laptop and hard drive. As some of the conversations contain sensitive information (e.g., gossips, personal matters), the transcripts will not be uploaded on any public repositories. Only aggregated data shall be made available through publications, such as this thesis.

3.3. Results

3.3.1. A descriptive overview of the corpus

This section reports participants' suppliance of non-standard morphosyntactic features in the nominal (**§3.3.2.**), verbal (**§3.3.3.**), and clausal/discoursal (**§3.3.4.**) domains. Before delving into these analyses, a general description of the participants' speech data is presented in terms of their mean length of utterance (MLU) and type/token ratio (TTR). In brief, MLU refers to the average number of free and bound morphemes per utterance (Brown, 1973), whereas TTR measures lexical or vocabulary variation by dividing the total number of word types by the total number of word tokens (Johnson, 1944; Hess et al., 1984).

There were altogether 876 English-based utterances coming from the SL-Chinese group and 649 utterances from the SL-Malay group. Despite its considerably higher number of utterances, the SL-Chinese group produced slightly less morphemes (MLU = 5.35) than its SL-Malay cohort (MLU = 6.06); the number of different lexical items used by the SL-Chinese participants (TTR = 0.176) was also smaller than their SL-Malay peers (TTR = 0.252) (Table 3.2). What these results show is that the speech data of the SL-Chinese group were less complex than those of their SL-Malay peers in terms of grammatical (MLU) and lexical (TTR) variation.

Table 3.2: Descriptions of the naturalistic speech data in terms of MLU and TTR

	SL-Chinese	**SL-Malay**
$\text{MLU} = \frac{Total\ number\ of\ morphemes}{Total\ number\ of\ utterances}$	$\frac{4711.5}{876} = \mathbf{5.354}$	$\frac{3933}{649} = \mathbf{6.06}$
$\text{TTR} = \frac{Total\ number\ of\ different\ item\ types}{Total\ nmber\ of\ tokens}$	$\frac{903}{5132} = \mathbf{0.176}$	$\frac{721}{2857} = \mathbf{0.252}$

3.3.2. Nominal domain

3.3.2.1. Omission of plural -*s*

One of the morphosyntactic properties of CME is the lack of plurality. The distribution of this feature quantified in this study only considers countable nouns that take the regular plural -*s*

marking. Excluded from the count were phonologically ambiguous cases where potentially pluralised nouns were followed by a word whose onset was a (post-)alveolar fricative (e.g., the girl**s s**ang). Of the total number of NPs where plural *-s* was expected, the omission rate of the SL-Chinese group (65.52%) was substantially higher than that of the SL-Malay group (31.11%) (Table 3.3). The between-group statistical comparison revealed a significant difference between the two groups ($z = 5.75, p < 0.001$).

Table 3.3: Omission of plural *-s*

	SL-Chinese	SL-Malay
Omission	95	45
	(65.52%)	(31.11%)
Total obligatory contexts	**145**	**135**

3.3.2.2. Omission of articles

The omission of articles, another property of CME, was also observed in the current data. As shown in Table 3.4, the rate of article ellipsis committed by the SL-Chinese group (24.8%) was higher than that of the SL-Malay group (18.1%), and the difference was statistically significant ($z = 1.762, p = 0.039$). Upon closer scrutiny, the proportions of indefinite and definite articles between the two groups were not significantly different from each other. More interestingly, the rates of omitted indefinite articles were considerably higher than those of omitted definite articles for both SL groups.

Table 3.4: Omission of indefinite and definite articles

Article ellipsis	**SL-Chinese**	**SL-Malay**
Indefinite articles	35/78	17/54
	(44.87%)	(31.48%)
Definite articles	27/172	23/167
	(15.70%)	(13.77%)
Total omissions in obligatory contexts	**62/250**	**40/221**
	(24.80%)	**(18.10%)**

3.3.2.3. Relative clause constructions

Turning to relative clause (RC) constructions, Table 3.5 shows the distribution of different RC types. The standard type included constructions which not only had an overt relative pronoun but also allowed for null (object) relativizers. The production of standard RCs was relatively low amongst the SL-Chinese (41.67%) and SL-Malay (35.71%) groups. Between the groups, the difference in suppliance was not significant ($z = 0.440$, $p = 0.33$). Concerning the non-standard omission of subject relativizers (attested in CME), the SL-Chinese group produced a significantly higher proportion of this feature (58.33%) than its SL-Malay counterpart (25%) ($z = 3.67$, $p < 0.001$). The *yang*-construction, found only in the SL-Malay data, is not an attested feature of CME but clearly reveals a case of intrasentential codeswitching from Malay.

Table 3.5: Types of RC constructions

	SL-Chinese	SL-Malay
Standard construction	10	10
e.g. "*something that/ø you've been craving for…*"	(41.67%)	(35.71%)
Absence of a subject relativizer	14	7
e.g. "*People ø dry and not dry is very different eh.*"	(58.33%)	(25%)
***Yang*-construction**	0	11
e.g. "*Which part of Japan yang consider as rural?*"	(0%)	(39.29%)
Total obligatory contexts	**24**	**28**

3.3.2.4. Possessive constructions

Moving on to possessive marking, four types of constructions were observed in this study. Note that the frequency count of this feature only considered possessive constructions with a full NP-possessor (e.g., *Adam*'s attention). Pronominal possessors (e.g., *his* girlfriend) were not included in the quantification.

As shown in Table 3.6, the standard use of the possessive ′s occurred less than 20% in the speech data of both SL groups. Omission of the possessive clitic (attested in CME) was committed only by the SL-Chinese cohort. Concerning uses of the Malay *punya* and the Chinese 的 *de* possessive markers, each of these markers was found only in the respective SL groups. The SL-Malay group supplied a substantially high proportion of the *punya* construction (84.62%), whereas the SL-Chinese group merely produced the *de* construction 25% of the time.

Table 3.6: Types of possessive constructions

	SL-Chinese	SL-Malay
Standard possessive ′s	2	4
e.g. "*Adam*'s *attention*"	(16.67%)	(15.38%)
Absence of possessive ′s	7	0
e.g. "*Joy-ø boyfriend*"	(58.33%)	(0%)
***Punya* possessive marker (Malay)**	0	22
e.g. "*Hida* punya *beef*"	(0%)	(84.62%)
的 **'*de*' possessive marker (Chinese)**	3	0
e.g. "*Ruby* 'de' *taste*"	(25.00%)	(0%)
Total obligatory contexts	**12**	**26**

3.3.3. Verbal domain

3.3.3.1. Omission of tense/aspect inflections

In CME, tense/aspect inflections are frequently dropped. Before looking into their distribution in the current data, a couple of notes ought to be brought to the fore. Firstly, regarding the quantification of past tense and 3rd.SING.-*s*, only regular verbs were considered. Secondly, inflections that shared similar phonological properties as the onset of the next word (e.g., consider**ed to**; he like**s S**onia) were excluded from the frequency count.

In Table 3.7, the omission rate of tense/aspect inflections was significantly higher in the SL-Chinese group (63.3%) than in the SL-Malay group (23.66%) ($z = 6.98$, $p < 0.001$). This applied to most of the features listed in Table 3.7, except for progressive *-ing* and passive *-en*, in which the omission rates produced by the SL-Chinese and SL-Malay groups were not significantly different to each other.

Table 3.7: Omission of tense/aspect inflections on lexical verbs

	SL-Chinese	SL-Malay
Past tense *-ed*	63/66	13/20
e.g. "*When we were small, we always play-ø mah.*"	(95.45%)	(65.00%)
3rd.SING.-*s*	45/55	8/22

e.g. "*She like-ø Anime also mah.*"	(81.82%)	(36.36%)
Progressive -*ing*	0/41	1/29
e.g. "*…instead of look-ø at the gadget…*"	(0%)	(3.45%)
Perfect -*en*	7/13	2/20
e.g. "*…they haven't* teach *immortality.*"	(53.85%)	(10.00%)
Passive -*en*	4/13	7/40
e.g. "*…scared they got harass-ø*"	(30.77%)	(17.5%)
Total omissions in obligatory contexts	**119/188**	**31/131**
	(63.30%)	**(23.66%)**

3.3.3.2. Omission of copula BE and auxiliary verbs

Turning to the omissions of copula BE and auxiliary verbs (attested in CME), Table 3.8 presents their distributions. Note that, for auxiliary DO, instances were only counted when it involved yes/no-questions.

Overall, the absence of copula BE and the auxiliary verbs of BE, HAVE, and DO was observed in both SL groups. That said, the SL-Chinese group dropped these suppletive morphemes significantly more than their SL-Malay peers (38.61% vs. 11.14%, $z = 8.35$, $p < 0.001$). This covered most of the features, except for perfect HAVE, where both SL groups did not behave significantly different from each other.

Table 3.8: Omissions of copula BE and auxiliary verbs

	SL-Chinese	SL-Malay
Copula BE	79/227	29/302
e.g. "*I ø lazy to go out…*"	(34.80%)	(9.60%)
Progressive/Passive BE	20/39	3/19
e.g. "*What ø you doing?*"	(51.28%)	(15.79%)
Perfect HAVE	3/12	6/20
e.g. "*Mak ø been there before.*"	(25.00%)	(30.00%)
Auxiliary DO	15/25	3/27
e.g. "*ø you know that there was a year Bella go to UK?*"	(60.00%)	(11.11%)
Total omissions in obligatory contexts	**117/303**	**41/368**
	(38.61%)	**(11.14%)**

3.3.3.3. GOT

The CME GOT construction was attested in the current data. As seen in Table 3.9, the occurrences of GOT were mostly found amongst the SL-Chinese participants ($n = 52$) but few amongst the SL-Malay peers ($n = 4$). Concerning the non-standard uses of GOT produced by the SL-Chinese group, the semantic extension of GOT, which denotes possession (similar to the lexical verb *have*), constituted the highest rate (30.77%). This was followed by the aspectual marker GOT (21.15%). The existential use of GOT was produced not only by the SL-Chinese group ($n = 8$) but also by the SL-Malay peers ($n = 3$). Similarly, the elided GOT "yes" responses were also attested in the speech of both SL-Chinese ($n = 10$) and SL-Malay ($n = 1$) groups.

Table 3.9: *GOT* constructions

	SL-Chinese	SL-Malay
Standard use	7	0
e.g. "*He* got.PAST *it.*"	(13.46%)	(0%)
Substitute of lexical verb 'have'	16	0
e.g. "*I think the girlfriend* got *problem one oh.*"	(30.77%)	(0%)
Existential use	8	3
e.g. "*My colleague,* got *one is working in Singapore one mah.*"	(15.39%)	(75.00%)
≈ There is *a colleague of mine who is working in Singapore*.		
Aspectual use	11	0
e.g. "*You* got *take your medicine?*"	(21.15%)	(0%)
≈ Have *you* taken *your medicine?*		
Elided *got* (≈ "yes")	10	1
e.g. A: "*Sonia* got (≈have) *boyfriend or not?*"	(19.23%)	(25%)
B: "Got (≈have)*!*" = Yes		
Total instances of constructions involving *got*	**52**	**4**

3.3.4. Clausal and discoursal domains

3.3.4.1. Non-/inversion in interrogatives

In CME, it is not uncommon to find the lack of subject-auxiliary inversion (SAI) in direct questions and overgeneralisation of SAI in indirect questions. In Table 3.10, the suppliance of standard direct *wh*-questions (84.96%) and indirect yes/no-questions (90.48%) was high amongst the participants. Regarding non-standard constructions, the production rate of non-inversion in direct questions was significantly higher in yes/no-questions (61.54%) than in *wh*-questions (15.04%) ($z = 6.49$, $p < 0.001$). Concerning inversion (overgeneralisation) in indirect questions, the occurrences of this feature were significantly higher in *wh*- (37.04%) than in yes/no-constructions (9.52%) ($z = 2.18$, $p = 0.01$).

Table 3.10: Distribution of (non-)SAI in direct and indirect questions

			Yes/no	**Wh**
Direct questions	SAI	Yes/no: "*Have you had your lunch today?*"	35/91	79/93
		Wh: "*How was it lah?*"	(38.46%)	(84.96%)
	*No SAI	#Yes/no: "*You didn't tell them fail already meh?*"	56/91	14/93
			(61.54%)	(15.04%)
		*Wh: "*When he is coming back?*"		
Indirect questions	No SAI	Yes/no: "*Not sure if you can eat or not lah.*"	19/21	17/27
		Wh: "*I don't even know how the game works.*"	(90.48%)	(62.96%)
	*SAI	*Yes/no: "*Actually I don't know* is it *they go Singapore work as admin(istrator) ah…*"	2/21	10/27
			(9.52%)	(37.04%)
		#Wh: "*Later I don't know* what's *her reaction, you know?*"		

Turning to between-group comparisons (Table 3.11), the SL-Malay group generally produced more standard interrogative constructions than the SL-Chinese group. Regarding the suppliance of direct questions, the SL-Chinese group produced higher rates of non-inversion than the SL-Malay counterpart. Meanwhile, concerning the production of indirect interrogatives, both groups conformed to the standard non-inversion structure for yes/no-

constructions. However, overgeneralisation of SAI was committed more often in indirect *wh*-interrogatives by both groups.

Table 3.11: Distribution of (non-)SAI between SL groups

			SL-Chinese	SL-Malay
Direct questions	Yes/no	SAI	9/47	26/44
			(19.15%)	(59.09%)
		#No SAI	38/47	18/44
			(80.85%)	(40.91%)
	Wh	SAI	28/40	51/53
			(70.00%)	(96.23%)
		*No SAI	12/40	2/53
			(30.00%)	(3.77%)
Indirect questions	Yes/no	No SAI	12/14	7/7
			(85.71%)	(100%)
		*SAI	2/14	0/7
			(14.29%)	(0%)
	Wh	No SAI	11/16	6/11
			(68.75%)	(54.55%)
		#SAI	5/16	5/11
			(31.25%)	(45.45%)

3.3.4.2. Question particles

Upon noticing the lack of SAI in direct questions, particularly in yes/no-interrogatives, the analysis was now taken further to investigate the use of question-related particles. The particles included *ah*, *meh*, *kah*, and the yes/no-invariant tag *or not*. In Table 3.12, the question type with the highest number of particles was indeed the yes/no-type. A between-group comparison revealed a significant difference between the yes/no- and *wh*-types ($z = 5.57$, $p < 0.001$).

Table 3.12: Distribution of particles across question types

		Yes/no	Wh
SAI	Yes/no: "*Can we talk about her boyfriend* ah, *later?*"	3	12
	Wh: "*What's the name* ah*?*"	(3.30%)	(12.90%)
No SAI	#Yes/no: "*You still got play Pokemon* meh*?*"	36	5
	*Wh: "*Why you want to message me* oh?"	(39.56%)	(5.38%)
Total tokens of direct questions		**91**	**93**

Turning to between-group comparisons (Table 3.13), a significantly higher proportion of question particles was produced by the SL-Chinese group (45.98%) as compared to their SL-Malay counterpart (16.49%) ($z = 4.34$, $p < 0.001$). Regarding the question type in which particles tended to occur, it was the non-inverted yes/no-questions that invited the highest usage of particles by both SL-Chinese and SL-Malay groups. That said, the suppliance of particles by the SL-Chinese group (55.32%) was significantly higher than that by the SL-Malay group (22.73%) ($z = 3.18$, $p < 0.001$).

Table 3.13: Between-group comparison in the use of particles in direct questions

		SL-Chinese	SL-Malay
Yes/no	SAI	3/47	0/44
		(6.38%)	(0%)
	#No SAI	26/47	10/44
		(55.32%)	(22.73%)
Wh	SAI	6/40	6/53
		(15.00%)	(11.32%)
	*No SAI	5/40	0/53
		(12.50%)	(0%)
Total tokens of particles		**40/87**	**16/97**
		(45.98%)	**(16.49%)**

3.3.4.3. Discourse particles

The use of discourse particles, which is one of the most prominent features of CME, shall be presented briefly in this report. Figure 3.1 illustrates a total of 18 particle types attested in this case study. The highest suppliance once again came from the SL-Chinese participants ($n = 910$), which was more than twice than that from the SL-Malay peers ($n = 366$). Of all particles, *lah* constituted the highest frequency, and its suppliance was more or less proportionate between the SL-Malay (58.88%) and SL-Chinese (41.12%) groups. Although the SL-Chinese group supplied lesser instances of *lah* than the SL-Malay group, it generated a substantially higher number of particles which were not found in the SL-Malay data.

Figure 3.1: Distribution of discourse particles in the CME discourse

	lah	ah	oh	mah	kan	one	eh	loh	what	kah	meh	leh	bah	woh	hoh	hah	wah	gua
SL-Malay	199	31	0	0	102	0	0	0	1	21	0	0	12	0	0	0	0	0
SL-Chinese	139	198	185	135	0	91	50	24	23	0	21	16	1	12	8	3	3	1
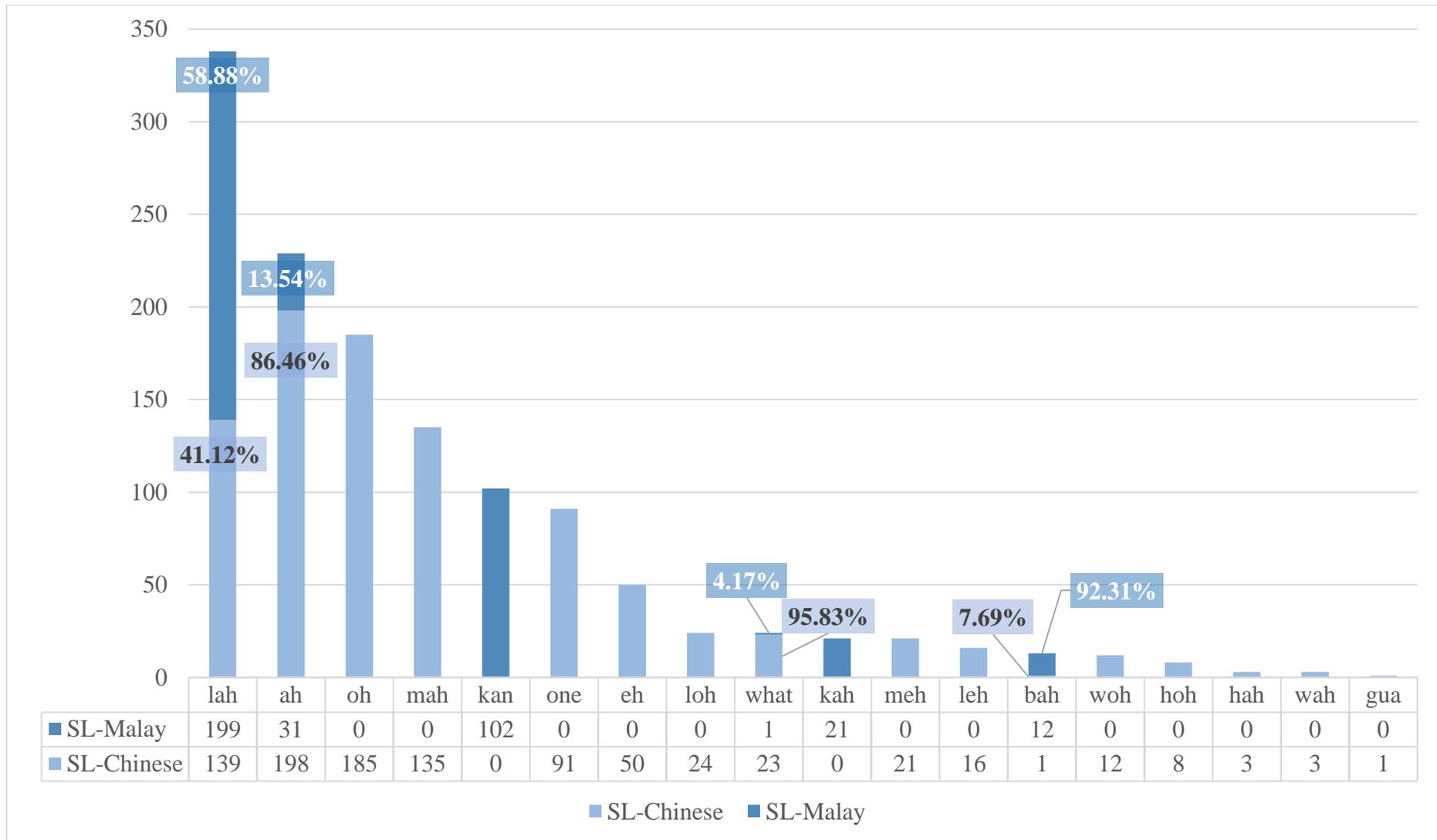
3.3.5. Summary of findings

The findings of the non-standard features are summarised in Table 3.14, with shaded regions indicating statistical significance. Not only were there many non-standard instances attested in the CME literature [e.g., morphosyntactic omissions, GOT, (no-)SAI, discourse particles], but there were also instances of overt transfer (e.g., *yang*-constructions, *de*/*punya*-possessives) not attested in CME. Between the SL-Chinese and SL-Malay cohorts, the former group produced substantially higher rates of grammatical omissions than the latter group. While the SL-Malay participants supplied considerably more lexical-based instances of overt transfer (e.g., *yang*-relativizer, *punya*-possessive) than the SL-Chinese counterparts, they did not generate as many discourse particles as the SL-Chinese. The CME GOT was attested in both SL groups, but its occurrence was markedly higher in the SL-Chinese group. Meanwhile, non-inversion in direct questions was more likely to occur in yes/no-questions, but this non-standard construction type was also marked with question-related particles. On the other hand, the overgeneralisation of SAI in embedded interrogatives was produced more often in *wh*-indirect questions than in yes/no-ones. Altogether, these observations shall be discussed in the next section.

Table 3.14: Summary of non-standard features attested in the CME discourse

	Features	SL-Chinese	SL-Malay
Nominal	Omission of plural *-s*	✓	✓
	Omission of articles	✓	✓
	Relative clauses		
	Standard constructions	✓	✓
	Absence of a relativizer	✓	✓
	SL-transferred yang-*constructions*	0	✓
	Possessives		
	Standard 's	✓	✓
	Absence of a possessive	✓	✓
	SL-transferred possessives	✓	✓
	De-*possessive*	✓	0
	Punya-*possessive*	0	✓
Verbal	Omission of tense/aspect inflections	✓	✓
	Omission of copula and auxiliary verbs	✓	✓
	Got constructions	✓	✓
Clausal / Discoursal	Subject-auxiliary inversion in interrogatives		
	No inversion in direct interrogatives	✓	✓
	SAI overgeneralisation in indirect interrogatives	✓	✓
	Question particles		
	With SAI	✓	✓
	No SAI	✓	✓
	Discourse particles	✓	✓

3.4. Discussion

To reiterate the research questions, the case study sought to inquire if there were features that were shared between speakers of different SL groups (RQ1), which features of CME were instances of substrate influence (RQ2a) or other L2 learning behaviours (RQ2b), and which features seemed to have developed beyond substrate transfer and L2 behaviours (RQ3).

RQ1: What are the stable features of CME shared between speakers of different SL groups?

With respect to RQ1, many of the attested CME features were present in the conversations of the SL-Chinese and SL-Malay cohorts. One of the features that stood out was the discourse particle *lah*. Of all the particles attested in the current corpus, *lah* was shared more or less proportionately by both SL groups. Its extensive use could be associated with overt transfer from Chinese and Malay (Lim, 2007; Smakman & Wagenaar, 2013), which instigated convergence into a stable feature of CME. However, concerning the optional uses of morphosyntactic omissions and non-standard (non-)inversion in interrogatives, one might ask if such optionality would disqualify these features as being stable in CME. It has been argued that the Minimalist Program does not give much account for optionality in favour of a comprehensive computational system governed by economy principles (see Papp, 2000:174). However, in light of language variation and change, more and more studies advocating the generative approach recognise optionality as a valid linguistic phenomenon (Adger & Smith, 2005; Biberauer & Richards, 2006). For instance, in Adger & Smith's (2005) investigation of the *was/were* alternation and optional *do* in negative declaratives in the Buckie dialect, the authors demonstrate that these variations can be explained in terms of retaining their respective semantic interpretation (i.e., interpretable features) but exhibiting more than one syntactic functions (i.e., uninterpretable features), thereby giving rise to different morphophonological expressions. Biberaurer & Richards (2006) claim that, in Afrikaans, the optional placement of auxiliary verbs in embedded clauses and expletives in impersonal passive constructions do not give rise to different semantic readings but satisfy the Extended Projection Principle all the same. Furthermore, in SLA research, (stable) optionality has been argued to persist even amongst L2 learners of advanced proficiency levels (Papp, 2000; Parodi & Tsimpli, 2005; Sorace, 2000, 2003, 2005). While this learner language phenomenon may be attributed to

representational (e.g., interface-related) and/or processing issues (Sorace, 2005:70-73), it does not occur in free variation but under certain constraints (Papp, 2000; Parodi & Tsimpli, 2005). In Papp's (2000) study, morphosyntactic optionality exhibited by advanced L2-Hungarian learners of L1-English in certain sentences (i.e., double focus and double *wh* constructions) – through an acceptability judgement task – is attributed to insufficient input of those constructions. In Parodi & Tsimpli's (2005) study, optionality in the (un)acceptability between empty categories and pronominal clitics in the null operator structures of L2-English and L2-Spanish/Greek is constrained by L1 effects, particularly with reference to their morphological richness and structural (dis)allowance for the option between clitics and empty categories in their respective null operator constructions. Returning to the current study, it is fair to claim that CME is indeed an L2 variety (Kortmann & Szmrecsanyi, 2004; Szmrecsanyi & Kortmann, 2009a), since English has always been an L2 for the majority Malaysian population. It therefore comes as no surprise if morphosyntactic optionality persists even amongst advanced learners and makes its way to become a fossilized, stable syntactic property of CME.

RQ2: Which features of CME are instances of (a) substrate influence or (b) general L2 behaviours?

In response to RQ2, many of the CME features reported in this case study have been attributed to substrate transfer and other L2 learning behaviours. Instances of overt transfer encompassed codeswitching that took place at intrasentential (e.g., *yang*-relative constructions; *punya-* and *de*-possessive constructions) and discoursal (e.g., discourse particles) levels. Instances of covert transfer included morphosyntactic omissions (i.e., of plural *-s*, possessive markers, relativizers, tense inflections, copula BE, and auxiliary verbs) and the lack of syntactic movements (e.g., non-inversion in direct interrogatives). Other L2-based instances which could not be explained by substrate transfer were those pertaining to features of overgeneralisation (e.g., SAI in indirect interrogatives).

The suppliance of non-standard features observed between the SL-Chinese and SL-Malay groups was disproportionate and therefore could not be overlooked. For example, why were omissions significantly higher in proportion in the SL-Chinese group than in the SL-Malay group, even though both languages do not have such counterparts in their morphosyntactic inventories? A plausible explanation for such discrepancy lies in the degree of morphological richness between Chinese and Malay, which might influence the processing and production of the L2 morphosyntax. Sagarra & Ellis (2013) entertained the plausible

effects of morphological richness by comparing between learners of L1-English (lacking in morphology) and L1-Romanian (rich in morphology) in their sensitivity towards the L2-Spanish (rich in morphology) temporal agreement. Findings from their eye-tracking study showed that the L1-English group relied more heavily on lexical cues (i.e., temporal adverbs) whereas the L1-Romanian group paid more attention to morphological cues (i.e., tense inflections), and that these respective preferences were attributed to learned attention during L1 acquisition. In Ellis & Sagarra's (2010) earlier study, the authors also examined learned attention effects between learners of L1-Chinese (lacking in morphology) and L1-English (richer in morphology, by comparison) in the acquisition of temporal reference in L2-Latin. Their results not only showed that the L1-Chinese participants were less able to perceive and produce verbal morphology than their L1-English peers but that they were also observed to depend more on temporal adverbs. Back to the current study, although the languages of Chinese and Malay do not have many grammatical counterparts of StE, that Malay is a morphologically richer (i.e., agglutinative) language than Chinese indicates that the L1-Malay speakers might have exhibited a more heightened sensitivity towards morphological cues and were thus able to produce more English verbal morphology than their Chinese-dominant peers.

Concerning the balanced suppliance of non-standard inversion in embedded (*wh-*)questions between the SL-Chinese and SL-Malay groups, this feature is also found in learner Englishes (e.g., L1-Spanish and L1-Chinese in Pozzan & Quirk, 2014) and, interestingly, many other World Englishes such as Philippine English (Munalim, 2019), Indian English, and Irish English (Hilbert, 2011). In an L2 oral elicitation study run by Pozzan & Quirk (2014), non-standard inversion tended to occur more frequently in embedded *wh-*questions than in yes/no-ones, similar to what we have found in this case study. Under the *wh-*condition, Pozzan & Quirk (2014) observed that L2-English learners of L1-Spanish (26%) and L1-Chinese (21.1%) supplied roughly proportionate rates of embedded inversion. This pattern was also similar to that in our study, although the SL-Malay (45.45%) and SL-Chinese (32.25%) groups produced relatively higher proportions. The authors claimed that such findings were not likely due to L1 effects, especially given the target-like performance by the L1-Spanish group. However, the authors argued that non-standard embedded inversion occurred more frequently when it involved *why-*items. In Hilbert's (2011) empirical investigation on Indian English, Singapore English, and Irish English, the author makes a claim against the notion of universalism or rule overgeneralisation concerning embedded inversion and, instead, argues for frequency effects. That is, embedded inversion usually occurs with BE or other auxiliary verbs. As the current case study did not go in-depth into the analysis of interrogative

constructions, it could be that the instances of embedded inversion supplied by the SL-Chinese and SL-Malay groups involved certain sequences of interrogative pronouns and verbs. However, unless a systematic comparison between both main and embedded interrogative constructions is conducted empirically across different English varieties, the notion of entrenched L2 overgeneralisation or even the universal status of embedded inversion cannot be swiftly repudiated.

RQ3: Which features of CME are a result of structural innovation?

Addressing RQ3, a couple of instances indicating structural innovation have been identified. To reiterate, structural innovation stipulated in this study refers to the linguistic mechanism through which features have undergone change as a result of more than just substrate transfer or any other L2 behaviours. That is, structural innovation involves linguistic creativity motivated by "psychological, sociological, and attitudinal reasons" (Kachru, 1985:20). To wit, features of structural innovation belong exclusively to the speech community of a language variety, whose sociocultural context has contributed to their emergence and/or development.

A clear example of structural innovation is GOT (Bao, 2005:257), which is widely attested in the CME and CSE literature. Its innovative uses have been found in the production of the SL-Chinese and SL-Malay groups, which warrants GOT as a stable feature of CME. At first blush, one may argue that the significantly higher uses of GOT by the SL-Chinese group than the SL-Malay group are more likely a result of transfer from the Chinese substrate. While that is indeed true, the restructuring of GOT might have also received influence from Malay (**§2.5.2.6.2.**; see also Goddard, 2002; Hiramoto & Sato, 2012; Lee et al., 2009). What makes GOT more interesting is that it has undergone relexification. Adapting Lefebvre's (1986, 1999) relexification hypothesis in pidgin and creole studies, Bao (2005) postulates that relexification in the WE context encompasses two processes: i) *System Transfer*, which involves transfer of a substrate language's entire grammatical subsystem; and ii) *Lexifier Filter*, which involves maintenance of the lexifier language's morphosyntactic expression or exponence (ibid., p. 258).[8] While this renders the relexified superstrate feature typologically closer to its substrate

---

[8] While Bao's (2005) contribution to Lefebvre's (1986, 1999) relexification hypothesis deserves its merit, it is not without a couple of scholastic concerns. For instance, Buschfeld (2020:267) calls the author out for not acknowledging Hulk & Müller's (2000) hypothesis, which postulates that when features of two languages exhibit

counterpart (ibid., p.258), the exponence still "conforms to the morphosyntax of the lexical-source language" (ibid., p.259). In other words, the relexified English GOT not only complies with the syntactic functions of the Chinese/Malay substrate but also adheres to the morphosyntactic structure of English. Additionally, whenever GOT is used by the speech communities in Malaysia and Singapore, it carries a unique set of semantic and grammatical configurations heavily influenced by the substrate counterparts (Bao, 2005, 2014; Lee et al., 2009). For these reasons, I argue that GOT is an epitome of bilingual creativity that reflects the complexification of CME (and CSE).

Another feature that might be on its way to undergoing structural innovation is the non-inverted yes/no-direct interrogative structure with substrate-transferred question particles. On the one hand, in-situ question formation is an attested feature in several World Englishes such as CME, CSE, Pakistani English, and Black and Indian South African Englishes (Mesthrie, 2004:1137; see also Kortmann & Szmrecsanyi, 2004). On the other hand, that the SL groups of the current study additionally employed question particles to in-situ yes/no-questions more than the *wh*-counterparts implies linguistic systematicity and bilingual creativity when such particles were used with the former type of question construction. Upon closer inspection, the markedly higher suppliance of question particles came from the SL-Chinese participants as compared to their SL-Malay peers. While this clearly demonstrates overt substrate transfer of question particles, mainly from Chinese, it would not be surprising if the interrogative structure in question becomes an entrenched feature of structural innovation, provided it is used extensively by the multilingual speech communities in Malaysia (see Bamgbose, 1998:3-4).

---

overlapping functions, crosslinguistic transfer can take place at the surface or underlying structures (see also Nicoladis, 2006:16).

Furthermore, Bao's (ibid.) concepts of System Transfer and Lexifier Filter bear some resemblance to Myer-Scotton's (1993, 2001, 2002) Matrix Language Frame (MLF) model of codeswitching, which posits that morphemes of an embedded language (corresponding to the substrate language) must adhere to the morphosyntactic structure of the matrix language (corresponding to the superstrate language). Despite recognising different disciplinary goals, Myers-Scotton (2002) critiques Lefebvre's (1986) relexification hypothesis as lacking explanation on how mental processes can give rise to relexification and offers how "the formation of a composite Matrix Language" can potentially enhance Lefebvre's theory. Whether or not Bao's (ibid.) postulation of the System Transfer and Lexifier Filter was inspired by Myer-Scotton's MLF model, it does supplement the relexification hypothesis.

3.5. Conclusion

This chapter has demonstrated that many morphosyntactic features of CME attested in the MalE literature have also been attested in this small case study. The disproportionate suppliance of non-standard features (including grammatical optionality, discourse particles, and codeswitching) by the SL groups of Chinese and Malay evinced substrate influence propagated by bilingual creativity. While it cannot be denied that this entails the dynamicity of CME, as represented differently in the linguistic repertoires of both SL-Chinese and SL-Malay groups, that the non-standard features were used systematically by the SL groups indicates the presence of inherent linguistic mechanisms constraining the production of those features.

Having demonstrated the naturalistic use of CME, we move on to the investigation of StE viz. the main study.

4 Finiteness

4.1. Introduction

For the main PhD study, the morphosyntactic phenomenon under investigation is *finiteness*. The next section establishes the theoretical conceptualisations revolving around finiteness, including the notions of tense and aspect (**§4.2.**). The following section (**§4.3.**) presents the typologies of StE, Malay, Chinese, and CME, focusing on their structural representations on temporality. And having reviewed L1 and L2 acquisition studies on finiteness (**§4.4.**), the final section (**§4.5.**) lays out the learnability tasks plausibly faced by Malaysians as they acquire finiteness marking in StE.

4.2. Theoretical conceptualisations

4.2.1. Finiteness

There has been a lack of consensus on the notion and syntactic representation of finiteness (Adger, 2007; Nikolaeva, 2010). But, crudely put, finiteness is a linguistic property associated with tense and agreement manifested on verbal elements (Nikolaeva, 2007:1). In (29), the English verb "likes" is said to be finite whereas the verb "play" is non-finite.

(29)   Sam *likes* to *play* badminton.

According to Radford (2004:337-338), finiteness depends on whether the verb – lexical or auxiliary – takes a subject NP that is nominative. If it does, then it is finite. Otherwise, it is non-finite. Finite verbs are also associated with tense marking alongside the phi(φ)-features of person and number valued by the subject NP (Huddleston, 1988:44; see also Hogg, 1992:541). To demonstrate, in the finite clause in (29) above, "Sam" takes up the subject NP position as it has been structurally assigned a nominative case (Chomsky, 2015:101). The finite verb "likes" is inflected with -*s*, which concomitantly encodes i) present tense and ii) the third-person and singular φ-features valued by "Sam". Regarding the *to*-infinitival clause, the verb "play" is not inflected for tense because it is not preceded by a nominative subject.

In Minimalist syntax, the structural composition of a clause has been argued to host three main projections, namely the Verb Phrase (VP), Tense Phrase (TP; a.k.a. Inflectional Phrase [IP]), and Complementizer Phrase (CP) (Adger, 2007:29; Haegeman, 1997:33-34):

Figure 4.1: The clausal structure
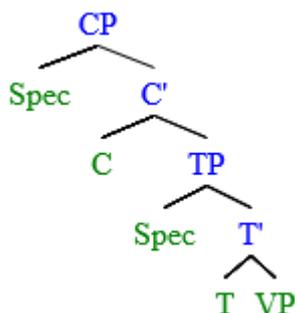
(Adapted from Haegeman, 1997:34)

As summarised by Adger (2007:29), the VP layer, headed by the main verb, is responsible for the selection of complements and the assignment of theta(θ)-roles. The TP layer hosts features pertaining to finiteness i.e., tense (e.g., [past], [present]), agreement [$u\varphi$:], nominative case (e.g., [$u$case:NOM], and the Extended Projection Principle (EPP) feature [uN*]). The CP layer deals with clause-related constructions such as illocutionary force (e.g., declarative, interrogative, imperative) and topicalization, etc. In the interest of this thesis, we will focus mainly on the TP layer (Parodi, 2006:666), and refer finiteness to the linguistic property associated with the morphosyntactic features of tense and agreement.

Having established our position on finiteness, the next two subsections introduce the concepts of tense and aspect. While tense is a feature of finiteness whereas aspect is not, it nonetheless merits a brief introduction on the notion of aspect due to its interrelation with tense.

4.2.2. Tense

Temporality can be expressed by locating an event or situation at a specific or relative point in time during the time of utterance (viz., *tense*) (Klein, 1995:22). It can also be described in terms of *how* time is represented in the event, such as whether the event is ongoing or completed (viz., *aspect*) (Hewson, 2012:511). We first look at tense in this subsection.

Tense is a deictic category that orders time into the three general segments: present, past, and future. Fabricius-Hansen (2006:567) succinctly defines these tenses in Table 4.1. That

is, present tense is used when the event being described takes place during the time of utterance (30a). Past tense refers to an event that already happened by the time it is being relayed to the interlocutor (30b). Future tense expresses an event that has yet to happen (30c).

Table 4.1: Schematic representations of present, past, and future tenses

Tense	Definition	Formalism[a]	Example	
Present	Event time overlaps utterance time	E o U	(30a)	The pope is visiting South America today.[b]
Past	Event time precedes utterance time	E < U	(30b)	The pope visited South America a couple of years ago.
Future	Event time follows utterance time	E > U	(30c)	The pope will visit South America next month.

[a] E, event; U, time of utterance; o, overlaps; <, precedes; >, follows

[b] The adverb "today" was added by present author to emphasise the temporal manner of the event.

(Adapted from Fabricius-Hansen, 2006:567)

There are several linguistic devices through which tense can be expressed. One of them is through inflectional morphemes employed by tensed languages such as English, Japanese, and Hindi. There are also tenseless languages, such as Chinese, Malay, and Thai, which do not have tense inflections in their morphosyntactic inventories. Such languages rely on aspectual markings (Smith, 2019) and other means such as contextual information to express time. A more robust way to convey temporal information is through the use of time adverbials (e.g., today, a couple of years ago, next month), which are not only universally available across languages but which also play a prominent role in learner languages (Klein, 1995:25).

Returning to Table 4.1, the English examples in (30a-c) are morphologically marked with different tenses. In (30a), the present tense is manifested on the progressive BE as "is". In (30b), the past tense -*ed* is inflected on the verb VISIT. In (30c), futurity is indicated via the use of the modal verb WILL. These sentences also distinguish the event in different temporal orientations: (30a) indicates that the event is ongoing via the use of the progressive BE…-*ing*; (30b) denotes completion of the event via past -*ed*; and (30c) signifies incompletion of the event, as indicated by the modal WILL. In other words, the sentences in (30a-c) are not only marked for tense but also for aspect, a concept which we expand in the next subsection.

4.2.3. Aspect

Aspect refers to the temporal orientation or manner of an event/situation or state. Unlike tense, it is non-deictic in that it "does not anchor the situation to the time axis" but "may affect temporal structure" (de Swart, 2012:753; see also Klein, 1995:22). It is generally represented through *grammatical aspect* or *lexical aspect* (Andersen, 1991:308; de Swart, 2012:753).

Grammatical aspect is encoded by means of functional morphology. It can be broadly classified into *perfective aspect* and *imperfective aspect* (Gvozdanović, 2012:784; see also Jakobson, 1971). Perfective aspect marks the completion of an event, which is conceived to be temporarily bounded (31a). Imperfective aspect denotes an event that is not temporarily bounded: it can be ongoing (31b), recurrent (31c), or habitual/generic (31d).

(31) a. Mary *built* a tent.
      b. Mary *is/was building* a tent.
      c. Mary *has/had built* a few tents.
      d. Mary *builds* tents.

In English, the perfective aspect is indicated by the simple past tense form, e.g., *built* in (31a). Also, English has at least three types of imperfective aspect. Ongoing events are grammatically encoded using the progressive BE…-*ing* (31b). Recurrent events are expressed with the perfect HAVE…-*en* (31c). Habituality can be conveyed using the simple tense (31d). Indeed, while there are languages that have different markers and distinctions of grammatical aspect (e.g., English, Chinese, Italian), there are also languages that lack it (e.g., Dutch, German, Norwegian). Notwithstanding, aspect can be conveyed through other means such as adverbials and lexical aspect expressed in all languages.

Lexical aspect (a.k.a. *aktionsart*) refers to the inherent semantic meaning pertaining to "properties of eventualities" (Filip, 2012:721) that can be encapsulated by the verb alongside its argument structure (Sasse, 2006:535) or by the whole proposition (Andersen, 1991:310-311). Following Vendler's (1967) seminal work on lexical aspect, verbs may be categorised into state, activity, accomplishment, and achievement. As shown in Table 4.2, stative verbs are distinct from the other dynamic verb categories (i.e., activity, accomplishment, and achievement) in that they denote a state of a situation or being, whereas dynamic verbs describe an action. Statives are durative/non-punctual, i.e., they continue over a period of time. They are also atelic, meaning that they have no endpoint (unless a change occurs). Activity verbs are

associated with habituality and are thus said to be atelic. Accomplishment verbs involve the goal of accomplishing something and are said to be telic. Achievement verbs differ from the other dynamic counterparts in that they are punctual (i.e., an action happens abruptly).

Table 4.2: Featural specifications of Vendler's (1967) lexical aspectual categories

	Lexical aspectual categories			
Features	**State** *e.g., believe, contain*	**Activity** *e.g., jog, sing, write*	**Accomplishment** *e.g., write a chapter*	**Achievement** *e.g., spot, bark*
punctual	-	-	-	+
telic	-	-	+	+
dynamic	-	+	+	+

Adapted from Bardovi-Harlig & Reynolds (1995:109)

The lexical aspect of verbs may change in relation to its propositional interpretation via the employment of different complements (e.g., *John ate* vs *John ate an apple*) or grammatical aspectual markers (e.g., *John ate* vs *John was eating*). In some languages, especially those from the Indo-European family, the marking of aspect is conflated with grammatical tense. There are also other languages, such as English and the Slavic languages, that tease apart tense and aspect grammatically. The use of aspectual markers may even indicate tense (e.g., Malay, Chinese) and vice versa (e.g., English; compare (31a) vs (31d)), illuminating the interplay between tense and aspect.

To sum up, different languages express tense and aspect differently. At the same time, tense and aspect can be intertwined since they both encode temporality.

Following this, the next section presents the temporal representations of English (StE and CME), Malay, and Chinese in relation to finiteness and, more specifically, tense, and aspect.

4.3. Finiteness and typology

4.3.1. Standard English

In English, finiteness is encoded through tense and agreement marking. It is manifested on the leftmost c-commanded verbal element in a finite clause.

While tense marking in English is relatively salient in terms of its morphophonological manifestations, agreement marking is not. To illustrate, the English tense system distinguishes between past and non-past, usually through past tense marking on regular (e.g., walk-*ed*) and irregular (e.g., *wrote*) verbs. However, concerning non-past, English lacks inflections to express it. A couple of exceptions are the 3rd.SING.-*s* suffix and the copula/auxiliary BE, which we shall get to shortly. Agreement marking in English is generally impoverished. It only has one inflectional exponent, namely -*s*, which can only be marked in non-past contexts where the subject NP carries the third person and singular φ-features. Another occasion where agreement is salient is when the BE forms are used. In non-past contexts, BE is manifested as *am*, *are*, or *is* when it is in agreement with first-person singular, second-person singular/plural, and third-person singular subjects, respectively. In past tense environments, the BE exponent *was* is used with first/third-person singular subjects whereas *were* is employed with first/second-person singular/plural subjects. Finally, it is worth noting that the tense and agreement features in English are conflated, as demonstrated with the 3rd.SING.-*s* inflection and BE.

Looking at the structural representation of English, the morphosyntactic features pertaining to finiteness – viz. *tense* (e.g., [past], [pres]), *agreement* [uφ:], nominative *case* [$u$case:NOM], and the Extended Projection Principle (EPP) feature [uN*] – are associated with the T node in the TP projection (see **§4.2.1.**). Other features associated with finiteness are: i) the closest NP to TP, which becomes the subject of the finite clause in adherence to the EPP; and ii) the nearest verbal element to TP, which receives tense and agreement valuation from T. Following the syntactic derivation proposed by the Minimalist programme (Adger, 2003; see also Chomsky, 1995, 2015), uninterpretable features, indicated as *u* in [uF], must be checked and deleted in order for the derivation to be successful to be pronounced at spell-out.

Consider the sentence: *Mary is baking cookies*. Before its spell-out, the morphological components required to form the sentence undergo syntactic derivation, as seen in Figure 4.2. Focusing on the finiteness-related derivation, the uninterpretable φ-feature on T is checked and valued as [uφ:3SG] by the interpretable φ-features of the closest NP i.e., *Mary*. Simultaneously,

T values the NP *Mary* as nominative, checking both uninterpretable case features in T and the valued NP. In turn, the valued agreement [~~u~~φ:3SG] and [pres] tense features in T both value and check the uninterpretable inflectional feature [uINFL:] on the nearest verbal element that is the progressive BE. As BE is an auxiliary, T values [uINFL:] as strong [~~uINFL~~:3SG,pres*]. BE is thus raised to T and is pronounced as 'is'. Finally, as the uninterpretable EPP feature [uN*] on T has yet to be checked, it motivates a T′ projection – to which it percolates – and moves *Mary* to the specifier position of TP to check the [uN*] feature.

So far, we have looked at the derivation of an aspectual (i.e., progressive) construction. Should sentences with simple tense be considered, the derivation is similar. The difference, however, lies in the valuation of feature strength on the verbal element receiving finiteness marking. If finiteness is valued on a non-lexical verbal element (e.g., modal, aspectual auxiliary, copula), it assigns a strong value and raises the verbal element from its respective projection to T, as seen with the progressive BE in Figure 4.2. If only the lexical verb is present in the numeration, the feature strength of finiteness values it as weak, and hence the lexical verb does not move from *v* to T. If the chain between *v* hosting the lexical verb and T bearing finiteness is intervened by a negator (i.e., not), or if the lexical verb being the only tense-bearing element in the sentence is elided, then DO-support must apply. This is because the Pronounce Tense Rule (PTR) in the English derivation fails to apply when *v* is no longer the head of T's sister. This also means that T cannot c-command a tense-bearing element. In this last resort, the dummy auxiliary DO will be invoked and bear the finiteness feature in T.

Figure 4.2: The syntactic derivation of *Mary is baking cookies*

TP
- NP
  - Mary
- T′ [*u*N*]
  - T
    [pres, *u*φ:3SG, *u*case:NOM, *u*N*]
    - T
    - be
  - ProgP
    - <be>
      [prog, *u*INFL:3SG,pres*]
    - vP
      - <Mary>
        [N, 3SG.FEM, θ:agent, *u*case:NOM]
      - v′ [*u*N, *u*θ]
        - v
          [*u*N, *u*θ, *u*V*, *u*INFL:prog, *u*case:ACC]
          - v
          - bake
        - VP
          - <bake>
            [V, *u*N, *u*θ]
          - cookies
            [N, 3PL.NEU, θ:theme, *u*case:ACC]
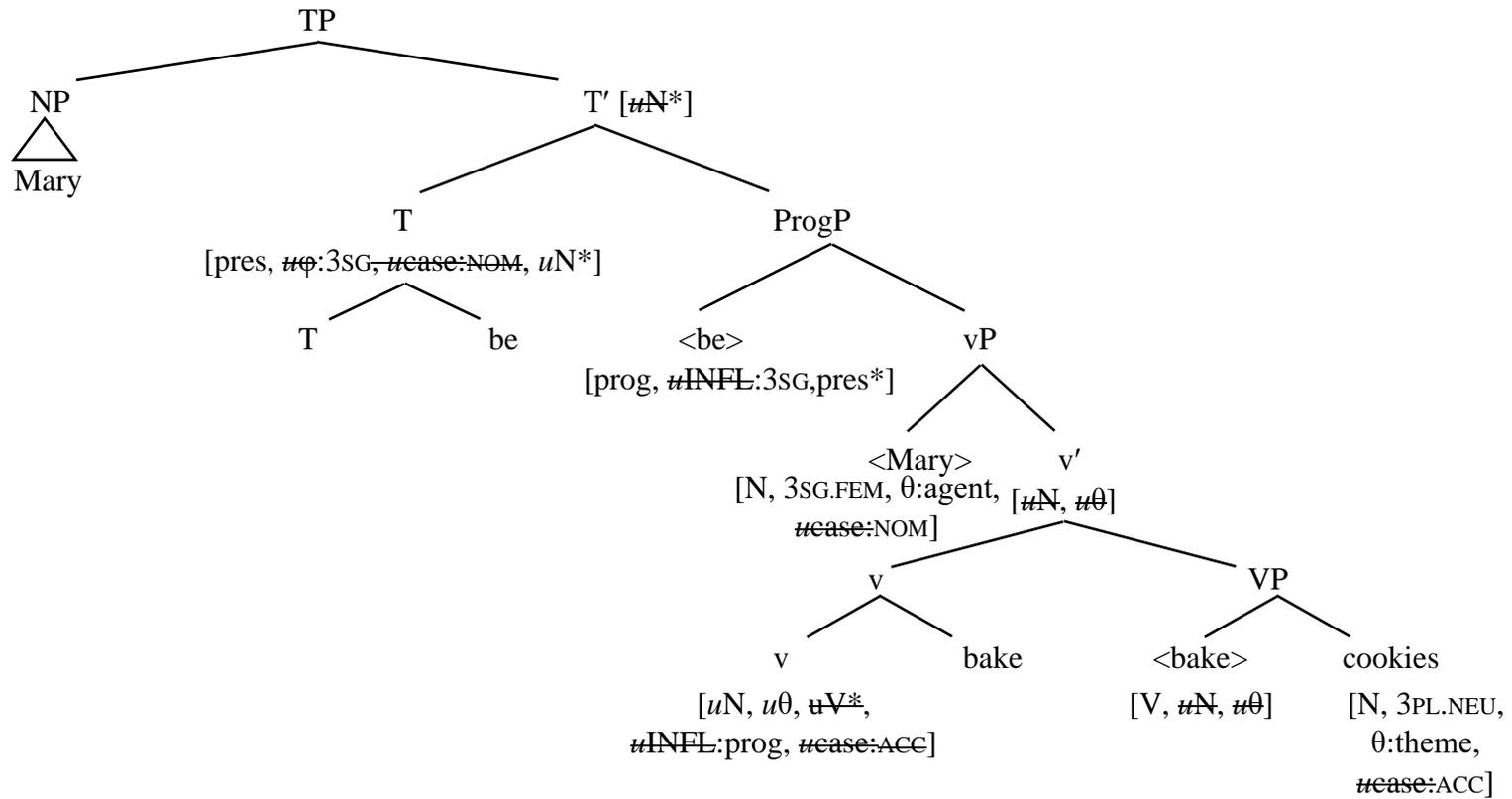
4.3.2. Malay

The Malay language (hereafter *Malay*) belongs to the Austronesian family and is spoken in the Malay Archipelago, especially in Malaysia, Indonesia, Brunei, and Singapore. In Malaysia, Malay consists of many regional dialects including Kedahan, Kelantanese, and Sarawak Malay. It also has a standard variety codified and regulated by the Institute of Language and Literature (in Malay: *Dewan Bahasa dan Pustaka*). While the colloquial varieties of Malay are widely used in casual settings, Standard Malay is employed in formal contexts such as in governance and education (Haji Omar, 2004:28-29). In this thesis, we mainly refer to the standard variety when demonstrating the morphosyntactic structure of Malay.

In terms of morphology, Malay is an agglutinative language rich with derivational and inflectional affixations. However, Malay does not inflect for tense and agreement (Wong, 2012:7-8; see also Goddard, 1996:431). Instead, time is encapsulated via the use of temporal adverbials (e.g., *semalam* 'yesterday' in (32)) or contextual information (Nomoto & Soh, 2019).

(32) Ali   tiba   di Tawau *semalam*.
       Ali   arrive in Tawau yesterday
       "Ali arrived in Tawau yesterday."

Malay also does not have grammatical aspect, but it does have several periphrastic aspectual markers functioning like auxiliaries (Karim et al., 2008:258-259). Some examples are *sudah* (perfective; see (33a)), *pernah* (experiential; see (33b)), *sedang* (progressive; see (33c)), and *akan* (futurity).

(33) a. Sarah   *sudah* cuci   pinggan.
        Sarah   PERF wash   plate
        "Sarah already washed the plate(s)."
    b. Siti (tidak) *pernah* makan durian.
        Siti (NEG) EXP   eat   durian
        Siti (NEG) EXP   eat   durian
        "Siti has (not) eaten durian before."
    c. Sanjit (bukan)   *sedang* buat   kajian.
        Sanjit (NEG)   PROG do   research

"Sanjit is/was (not) currently doing his research."

The Malay aspectual marking is not obligatory. However, when aspectual markers are employed, they must always precede the verb (Goddard, 1996:431). If a negator such as *tidak* (33b) or *bukan* (33c) is present, they can only appear after it.

Furthermore, the Malay aspectual markers may allude to the time of the event. In isolated sentences like (33a), the perfective *sudah* indicates that the activity of washing plates is already complete, indicating that it was done in the past. However, in (33c), the progressive *sedang* does not suffice in signalling whether the ongoing activity of doing research is located in the present or past context. In the lack of grammatical tense and aspect, contextual information – including the use of temporal adverbs, lexical aspect, and the organisation of the narrative – in the discourse is important to disambiguate the time reference.

Since Malay does not inflect for finiteness, the question arises as to whether or not a tenseless language like Malay exhibits a TP projection. Assuming that TP is a universal functional category (Ritter & Wiltschko, 2014), it follows that Malay has TP (see Mustaffa, 2018). However, in the absence of finiteness marking, tense and agreement are underspecified in T. Since Malay has aspectual markers, they head and project an Aspect Phrase (AspP) below the TP (Mustaffa, 2018; see also Grangé (2013) on Bahasa Indonesia). Furthermore, as aspectual markers can only appear after the negator, they do not raise from their respective positions over the Negative Phrase (NegP) projection to TP.

At this point, we ought to bring up the Malay copular constructions as well, since we will be dealing with the English copula BE (one of the finiteness features of investigation) in the experimental study.

First of all, the Malay copula is unlike its English counterpart in that it does not bear finiteness. No affixes or clitics can be attached to them (Mustaffa, 2018:28). Secondly, it has two morphological exponents: *ialah* and *adalah*. While *ialah* generally links a subject NP to an equative nominal predicate (34a), *adalah* takes predicates that are adjectival (34b) or prepositional (34c) (Karim et al., 2008:494-496; cf. Uzawa, 2009). Furthermore, the copulas in (34a-c) can be dropped without rendering the sentences ungrammatical:

(34) a. Ahmad   *ialah*/ø seorang   guru.
          Ahmad   COP   CL   teacher

"Ali is a teacher."

b. Kulit   Izzati   *adalah*/∅   cerah.

Skin   Izzati   COP   fair

"Izzati's skin is fair."

c. Cincin   emas   itu   *adalah*/∅   untuk   suami   saya.

Ring   gold   that   COP   for   husband   1.SG

"That gold ring is for my husband."

However, there are certain syntactic environments in which the copula cannot be morphologically manifested. In negated sentences, *adalah* can appear before a negator (35a), but *ialah* cannot (35b) (Karim et al., 2008:496; Ramli & Ab. Hamid, 2020:40):

(35) a. Berita   itu   *adalah*/∅   tidak   benar.

News   that   COP   NEG   true

"That news is not true."

b. Nama   saya   **ialah*/∅   bukan   Sofia.

Name   1.SG   COP   NEG   Sofia

"My name is not Sofia."

Additionally, where an aspectual marker (e.g., perfective *sudah* in (36a)) or a modal verb (e.g., deontic modal *boleh* 'can' (36b)) is present, the copula must be morphologically null:

(36) a. Lantai   sudah   **adalah*/∅   bersih.

Floor   PERF   COP   clean

"The floor is now clean." (Mustaffa, 2018:31-32)

b. Lantai   ini   boleh **adalah*/∅   bersih lagi.

Floor   this   can   COP   clean   more

"This floor can be cleaner." (Mustaffa, 2018:33)

Taken together, the omission of *ialah*/*adalah* does not apply across the board. As demonstrated in (35b) and (36), there are certain conditions where the morphological expression of either copula would give rise to ungrammaticality. This suggests that the Malay copula is obligatory in the syntactic derivation. It is only at spell-out that in certain contexts *ialah*/*adalah* can be phonologically dropped, but in other contexts they must (not at all) be pronounced.

94

4.3.3. Chinese

The Chinese language, which belongs to the Sino-Tibetan family, is an isolating language in that words typically are monomorphemic, and that the language is not contingent on affixation to encode grammatical information. Also, it is impoverished in terms of inflections (Li & Thompson, 2009:715; Packard, 2006:355). Dialects of Chinese include Mandarin, Hokkien, Hakka, and Teochew. In this thesis, the variety we will be focusing on is Mandarin (hereafter *Chinese*).

Like Malay, Chinese does not mark for tense and agreement and is thus said to be a tenseless language. Temporal adverbials (e.g., *míngnián* "next year" in (37)) and/or discoursal information are used to determine the event time (Chen & Shirai, 2010; Liu, 2015).

(37) 张三　　　　*明年*　　　毕业。

zhāngsān　　*míngnián*　　bìyè

Zhangsan　　next year　　graduate

"Zhangsan will graduate next year."

There are, however, grammatical markers of aspect. The four prominent ones are the perfective 了 *-le* (38a), the experiential 过 *-guò* (38b), the durative 着 *-zhe* (38c), and the progressive (正)在 (*zhèng*)*zài* (38d):

(38) a. 我 看了　　　　那部　　　电影。

wǒ kàn-*le*　　nàbù　　　diànyǐng

I　watch-PERF　that-CL　movie

"I watched that movie."

b. 李四　　　(没)　吃过　　　榴莲。

lǐsì　　　(méi)　chī-*guò*　líulián

Lisi　　　(NEG)　eat-EXP　durian

"Lisi has (not) eaten durian before."

c. 黑板　　　　　上　写着　　　数学　问题。

hēibǎn　　　　shàng　xiě-*zhe*　shùxué wèntí

blackboard	on	write-DUR	math problems

"Math problems are written on the blackboard."

d. 妈妈　　*在*　缝　衣服。

m*ā*ma	*zài*	féng	yīfú
Mom	PROG	sew	clothes

"Mom is sewing some clothes."

-Le (38a) signifies the completion of an event. *-Guò* (38b) marks the experiential status of an activity. The imperfective markers *-zhe* (38c) and *zài* both mark durativity, but *-zhe* is associated with stative events whereas *zài* denotes an ongoing action (Kwan-Terry, 1979; Li & Thompson,1981; Smith, 1997). Unlike English and Malay, most of the Chinese aspectual markers in (38a-c) follow the verb, with the exception of *zài* (38d), which precedes it. Finally, although Chinese does not have grammatical tense, its aspectual markers may help to locate the event time, though reliance on contextual information remains crucial in determining temporal location.

Similar to what has been proposed for the Malay syntactic representation, we maintain that Chinese has a TP projection [Ernst, 2009; Sybesma, 2007; see also Huang (1998) and Li (1990) on finite/non-finite distinctions, Sybesma, 2004 on Cantonese; cf. Lin, 2006b; Wu, 2009], but that TP is not specified for grammatical finiteness (Hawkins & Liszka, 2003:25-26). Chinese also has an AspP which is projected below NegP, since Chinese aspectual markers do not occur before the Chinese negators (example in (38b); Lin, 2006a).

Turning to copular constructions, the Chinese copula is 是 *shì*. To a certain extent, it bears some resemblance to the English copula BE in that it is predicative. *Shì* also behaves like BE in that it takes a nominal predicate. But, unlike BE, it only allows for the nominal predicate to be non-referential in relation to the subject NP that is the referent (Li & Thompson, 1981):

(39) 张三　　　(不)　*是*　老师。

zhāngsān	(bú)	*shì*	lǎoshī
Zhangsan	NEG	COP	teacher

Zhangsan is not a teacher.

Also, *shì* is used as a marker of affirmation to a proposition, which roughly translates to the {it *is/was* (true) that XP} construction in English:

(40) 他　　　　是　　　　没　　　　钱。

tā　　　*shì*　　　méi　　　qián

3SG　　COP　　NEG　　money

It is true that s/he does not have any money.

(Adapted from Li & Thompson, 1981:151)

Moreover, similar to English, *shì* may be used as an existential verb in place of the default 有 *yǒu*. However, the pragmatic reading of the existential construction changes as *shì* adds an emphatic effect to it:

(41) 外面　　　　　是　　　　一只　　　　　狗。

wàimiàn　　　*shì*　　　yizhī　　　　gǒu

outside　　　　COP　　one.CL　　dog

What's outside is a dog.

(Adapted from Li & Thompson, 1981:514)

If the predicate is adjectival or locative, *shì* cannot be employed to link it to the subject NP. What link the adjectival and locative predicates, instead, are degree adverbs (e.g., 很 *hěn* "very" (42a)) and the coverb 在 *zài* ≈"at" (42b), respectively.

(42) a.　他　　　　很/*是　　　　狡猾

tā　　　　hén/*shì　　jiǎohuá

3SG　　　very/*COP　cunning

S/He is cunning.

b.　小明　　　在/*是　　　英国。

xiǎomíng　zài/*shì　　yīngguó

Xiaoming　in/*COP　　the UK

Xiaoming *is* in the UK.

Other characteristics of the Chinese *shì* that are distinct from the English copula BE include the following: i) *shì* is not a tense-bearing copula; ii) it cannot occur with aspectual markers and most modal verbs (Li & Thompson, 1981:148-149); and iii) in negated copular constructions, it can only appear after the negator 不 *bù* (39), meaning that it does not raise to T in the Chinese syntactic structure.

Taken together, the Chinese copula *shì* is more restrictive than the English BE in that *shì* can only be phonologically manifested in certain contexts whereas BE must be overtly pronounced across all copular constructions.

4.3.4. Colloquial Malaysian English

In the previous chapters, we have discussed how CME has undergone restructuring partially as a result of cross-generational (L2) acquisition processes including crosslinguistic influence (mainly from the substrate languages of Malay and Chinese), learner errors of overgeneralisation, and bilingual creativity.

Having looked at how temporality is represented in StE, Malay, and Chinese, and how the latter two languages lack grammatical finiteness, it is unequivocal that CME has diverged from StE since it has, for the most part, received protracted influence from the Malay and Chinese substrates. The optional uses of tense and agreement markings, copula BE, and aspectual auxiliaries bearing finiteness in CME are likely to stem from either the absence or apparent optionality of the StE counterparts in Malay and/or Chinese:

Obligatory finiteness features in StE	**StE counterparts in Malay and Chinese**
• Tense and agreement morphemes	Absent
• Copula BE	Restrictive manifestations
• Auxiliary verbs (e.g., BE, HAVE, DO)	Absent

The next question concerning the syntactic makeup is: How is the syntactic structure of CME represented in relation to finiteness? We surmise that the CME structure is the same as that of its StE counterpart in that it has a TP projection bearing tense and agreement features. However, the parametric difference between the two English varieties is that while finiteness marking is obligatory in StE, it is optional in CME.

Having described the typologies of StE, Malay, Chinese, and CME, their summaries will be presented in **§4.5** where we formalise our predictions of the learnability tasks faced by Malaysian learners as they acquire obligatory finiteness in StE. In the meantime, we turn to the next section which delves into studies investigating the acquisition of finiteness amongst English monolinguals and bilinguals of different L1 backgrounds.

4.4. Finiteness and language acquisition

4.4.1. Acquisition of finiteness by L1-English children

It has been attested in many L1 acquisition studies that children tend to omit verbal inflections in contexts where finiteness marking is expected (Bar-Shalom & Snyder, 1998; Harris & Wexler, 1996) ; Paradis & Crago, 2001 ; Pratt & Grinstead, 2007 ; Rice et al., 1997 ; Schütze & Wexler, 2000 ; Wexler et al., 1998; Wijnen et al., 2001). Wexler (1994, 2011; see also Poeppel & Wexler, 1993; Rizzi, 1993) terms this phase as the optional infinitive (OI) stage, which usually occurs in the (L1) acquisition of tensed languages such as English, French, German, Dutch, Swedish, Danish, and Norwegian. In this subsection, we explore the characteristics of the OI stage in L1-English and theories accounting for the monolingual acquisition of finiteness. While it has been claimed that commission errors of finiteness (e.g., *he *are* running; *he *not is* running) are hardly found in child speech (Paradis et al., 2008:694; Rice et al., 1995; Rice & Wexler, 1996), they will nonetheless be discussed as they provide further insight about the representation of finiteness.

Amongst typically developing monolingual children, the alternation between finite verbs (e.g., she *goes*) and non-finite forms (e.g., *she *go*) – a.k.a. *root infinitives* (Rizzi, 1993) – is most salient between the approximate ages of 1;07 and 2;06. However, as the target-like use of finiteness becomes more consistent by the age of 4;00, the production of root infinitives gradually ceases (Wexler et al., 2004:172; see also Phillips, 2010:77). In the English OI stage, not only are the past tense *-ed* (43a) and 3rd.SING.-*s* (43b) often omitted, but it is also common for the copula (43c) and auxiliary verbs (43d) to be dropped (Schütze & Wexler, 1996):

(43)   a.   what happen it? [+ IMT]   (Eve, age 1;10, CHAT file 01100b, line 139)

b.   he love me too.   (Sarah, age 2;07, CHAT file 020712, line 23)

c.   what new?   (Adam, age 2;10, CHAT file 021002, line 223)

d. we waiting (.) sir. (Adam, age 2;10, CHAT file 021002, line 197)

(CHILDES English Brown Corpus)

Moreover, it has been argued that the suppliance of OIs is not arbitrary. Schütze & Wexler (2000) identified a plausible correlation between root infinitives and null subjects. In their study, monolingual English children between the ages of 2;2 and 3;11 partook in a set of elicitation scenarios where they were asked to correct a puppet who did not pay attention to the scenes. What the authors found was that the children supplied root infinitives more frequently when subject NPs were dropped but tended to inflect verbs when the subject NP was present. This trend was especially noticeable in the oldest group (3;6 – 3;11) of children. Schütze & Wexler therefore concluded that the suppliance of OIs alongside null subjects was an instance of underspecification of tense and agreement in early child grammar. Another study conducted by Harris & Wexler (1996) showed higher occurrences of root infinitives following a negator (e.g., not, no). The authors examined naturalistic data (from CHILDES) produced by 10 English-speaking children between the ages of 1;6 and 4;1 (ibid., p.15). Focusing on negative sentences without auxiliary DO (a.k.a. medial-NEG sentences), main verbs that followed a negator were merely inflected 9.6% of the time (ibid., p.16). In their second, elicitation study, the authors observed that four children (out of 27) who managed to supply different question types and medial-NEG sentences (ibid., p.27) produced higher numbers of verbal inflections in affirmative sentences and in negative sentences with the adverb *never* than in sentences negated with *not* (ibid., pp.29-30). These findings led Harris & Wexler to conclude that children would not randomly inflect verbs for tense and agreement in non-finite contexts.

In terms of syntactic derivation, Wexler (1994) posits that the syntactic representation in child grammar is fully established under UG guidance but certain functional features are yet to be specified. Concerning finiteness, the TP layer hosting tense and agreement features is intact. What is lacking in the derivation is the full specification of the features themselves (Schütze & Wexler, 1996; Wexler et al., 1998). According to Wexler (1994, 1998, 2011; see also Borer & Wexler, 1992), such underspecification is due to maturational constraints. As children grow and mature, they eventually converge on the adult grammar. This is when tense and agreement and other functional features become fully specified.

Rizzi (1993/1994) is another proponent of Wexler's OI stage. However, he contends that the clausal structure in child grammar is underdeveloped. That is, root infinitives are "truncated clausal structures" projected below the TP layer (ibid., p.390). In other words, they are maximally VP projections. According to Rizzi, optionality of tense cannot go hand in hand

with the obligatory projection of TP simply because optionality is not allowed in the syntactic representation (ibid., pp.379-380). To wit, when root infinitives are still prevalent in early child speech, TP has yet to take root. That said, as children go through a "sudden growth of utterances" at about three years old (ibid., p.379), the functional projections above the VP such as TP and CP will become fully developed and operative, and root infinitives will eventually disappear. Rizzi's account of the OI stage suggests that the occasional suppliance of finite verbs in child speech is arbitrary. However, it does not suffice in accounting for the systematicity observed in children's production of root infinitives, for example, with null subjects and after a negator in the respective empirical studies of Schütze & Wexler (2000) and Harris & Wexler (1996).

What is lacking in Wexler (1994) and Rizzi's (1993/1994) coverage of child grammar is that both authors do not account for commission errors related to finiteness, which are attested in child speech (Brown, 1973; Marchman & Bates, 1994; Tesan & Thornton, 2004). The errors referred to here include the overgeneralisation of tense inflections such as the regular past *-ed* on irregular verbs (Marchman & Bates, 1994) and the overgeneration of suppletive morphemes such as BE (Brown, 1973; Tesan & Thornton, 2004). Marchman & Bates's (1994) study presented a parental report on the production of verb forms by English monolingual children ($n = 1,130$). Participants between the ages of 1;4 and 2;6 were involved in the study, and their speech was monitored and noted by their parents. In the authors' analysis of irregular verbs, over-regularisation of past *-ed* (e.g., **eated/ated* instead of *ate*) was attested. Although it was not as frequently supplied as bare/uninflected forms and correct past forms, it showed that the children were already able to abstract and produce past *-ed*. In Brown's (1973:265) longitudinal study on Adam, Eve, and Sarah, Brown noticed that it was only Adam who over-produced the contracted BE (i.e., *'s*; e.g., **It's fell*; Brown, 1973:392). As contracted BE was almost always suffixed to the subject pronoun *it*, Brown attributed it to an issue of phonological segmentation, suggesting that contracted BE was misanalysed as a nominative marker of the subject NP (ibid., pp. 265, 391-392). Similarly, Tesan & Thornton (2004) observed the superfluous suppliance of the *'s* morpheme in the elicited production of three two-year-old children. In particular, the authors observed that *'s* was suffixed to singular subject NPs about 11% of the time in affirmative sentences ($\frac{173}{1,553}$) and negative sentences ($\frac{21}{188}$), respectively (ibid., p.5). Tesan & Thornton did not claim that the displaced morpheme *'s* was a contracted form of BE overgeneration. However, similar to Brown's (1973) speculation, they maintained that *'s* was "taken to be the realization of singular verbal agreement" (Tesan & Thornton, 2004:5).

Although finiteness errors of commission (e.g., overgeneralisation) are indeed not as common as omission errors in child speech, they cannot be overlooked as it may shed light on children's knowledge of tense and agreement. For instance, the over-regularisation of past *-ed* on irregular verbs (Marchman & Bates, 1994) may attest to the presence of the tense feature in children's linguistic repertoire, whereas the overgeneration of BE (Brown, 1973; Tesan & Thornton, 2004) may be an overt form of syntactic agreement with singular subject NPs. Therefore, it follows that the TP layer must have already been established very early on in order to host even misanalysed finiteness features. This would lend further support to Wexler's (1994) claim of a fully established clausal structure in the linguistic repertoire of young L1 acquirers.

4.4.2. Acquisition of finiteness by L2-English learners

In bilingual and L2+ acquisition research, it is often argued that bilinguals and subsequent language learners – the latter being counted as bilinguals, too – are qualitatively different from their monolingual peers in the acquisition of a target language. Grammatical features have been shown to pose learnability issues for bilinguals and cause a lag (Austin, 2009; Blom, 2010; Müller & Hulk, 2001; Nicoladis et al., 2012) and/or representational deficits in interlanguage development (Hawkins & Chan, 1997; Hawkins & Hattori, 2006; Hawkins & Liszka, 2003; Liszka, 2015; Lozano, 2009; Tsimpli & Dimitrakopoulou, 2007; see also Tsimpli et al., in press). Even at advanced proficiency levels, there are researchers who argue that bilinguals will, at most, achieve near-native competence (Coppieters, 1987; Donaldson, 2016; Hopp, 2009; Montrul & Slabakova, 2003), which implicates that they can never be on par with their native-speaker peers. Accordingly, this subsection addresses the (persistent) types of finiteness errors supplied by bilinguals and explores how SLA theories account for the representation of finiteness in the bilingual repertoire. Other factors such as crosslinguistic influence, input, and language environment will also be discussed in addition to SLA theorising. This subsection concludes by presenting an empirical study conducted by Buschfeld (2020, 2021) on bilingual children acquiring English in Singapore to demonstrate the linguistic complexity (including crosslinguistic influence and the types of target-language input) of the country's multilingual environment and how native-speakerism can be a problematic ideology in envisaging successful acquisition in such ecologies.[9]

---

[9] Buschfeld's (2020, 2021) language acquisition study in the Singaporean context is worth bringing up in this thesis because Singapore shares many linguistic similarities with Malaysia in terms of their multilingual

As seen with young L1-English children, the omission of finiteness marking is prevalent amongst bilingual learners of L2-English. In particular, affixal morphemes bearing tense and agreement (e.g., past *-ed*, 3rd.SING.-*s*) are more susceptible to omission than suppletive morphemes (e.g., copula BE, aspectual auxiliaries, DO auxiliary) (García Mayo & Olaizola, 2010; Ionin & Wexler, 2002; Paradis et al., 2008). Ionin & Wexler (2002) examined speech samples from 20 young Russian learners of L2-English (between the ages of 3;9 to 13;10) residing in the US and found that the omission rates of verbal inflections (i.e., 3rd.SING.-*s*: 78%; past *-ed*: 58%) were notably higher than those of suppletive morphemes (i.e., auxiliary BE: 33%; copula BE: 16%). In a similar vein, Paradis and colleagues (2008) compared the suppliance of English finiteness marking amongst typically developing (TD) monolingual children (*n* = 20; mean age = 3;0), monolinguals with Specific Language Impairment (SLI; *n* = 24; mean age = 5;7), and child L2-English learners from various L1 backgrounds (*n* = 24; mean age = 5;7). Elicitation results from the administration of TEGI showed that the L2 learner group exhibited considerably lower rates of past and present tense inflections as compared to their TD (3rd.SING.-*s*: L2=16% vs. TD=42%; past *-ed*: L2=20% vs. TD=47%) and SLI (3rd.SING.-*s*: L2=16% vs. SLI=52%; past *-ed*: L2=20% vs. SLI=46%) monolingual peers but performed similarly to the monolinguals in producing the copula/auxiliary BE and DO-support (ibid., pp.703-704). In a more recent study on L3 acquisition, García Mayo & Olaizola (2010) administered a narrative task and showed that adolescent Basque-Spanish bilinguals, who started acquiring L3-English at the age of 8 years, had issues producing tense inflections but fared well in supplying the copula/auxiliary BE. When comparing between the learner group who undertook a Content and Language Integrated Learning (CLIL) programme and the other group who went through a mainstream (non-CLIL) programme, the authors found that, by the end of their three-year longitudinal study, there was a significant decrease in the omission of 3rd.SING.-*s* for the CLIL (from 71.16% to 21.84%) and non-CLIL groups (from 71.89% to 12.34%). However, the omission rates of past *-ed* remained high in both groups (CLIL: from 47.72% to 40.54%; non-CLIL: from 22.22% to 43.68%). This shows that the omission of finiteness marking can persist into the later stages of interlanguage development, contrariwise to the phasing-out of the OI stage observed amongst (TD) L1-English children.

There are also other finiteness errors apart omission. The overgeneration of BE is one that is attested in many L2 studies (Abdul Aziz & Mohd. Don, 2014: L1-Malay learners; García

---

environments (especially the types of languages being in contact with one another) and how people are exposed to different subvarieties of English growing up.

Mayo et al., 2005: Basque/Spanish learners; Hawkins & Casillas, 2008: Chinese and Spanish learners; Ionin & Wexler, 2002: Russian learners; Ntalli, 2021: Chinese and Russian learners; Paradis et al., 2008: different L1 backgrounds including Mandarin, Japanese, Arabic, and Ukrainian). BE-overgeneration usually takes an inflected form (e.g., *is*, *were*) and precedes a lexical verb that is not inflected with progressive *-ing* (e.g., "The cookie *is* belong-*ø* to the man."). Furthermore, lexical verbs following the superfluous BE commonly appear in the bare form (García Mayo et al., 2005:472; Hawkins & Casillas, 2008), though it is not uncommon to find them inflected for tense as well (Abdul Aziz & Mohd. Don, 2014; Ionin & Wexler, 2002; Ntalli, 2021). In García Mayo and colleagues' (2005) study of Basque/Spanish bilingual students (*n* = 58), the authors observed that while BE was appropriately supplied in progressive contexts (15%), it was also inappropriately produced with generic/habitual (73%) and stative (3%) verbs, in past tense contexts (2%), and in ambiguous instances (8%) (ibid., p.465; see also Ionin & Wexler (2002:112) for a similar report). Accordingly, García Mayo et al. (2005) suggested that BE-overgeneration was likely due to L1 transfer. Specifically, the placeholder *is* supplied by the Basque/Spanish bilinguals was seen as a reanalysis of the agreement morphemes of Basque and Spanish. On the other hand, Hawkins & Casillas (2008:610) dismissed the plausibility of L1 effects. Having administered a sentence completion task, the authors found that the L1-Chinese and L1-Spanish children in their study overgenerated BE, particularly in sentences with a pluralised complex subject NP (e.g., "The assistants of the math teacher…" + (*i*)*s* + V; ibid., pp.608-609). This led the authors to propose that this generic L2 error reflected syntactic deficits that would apply across the board. Notwithstanding, what García Mayo et al. (2005:464), Hawkins & Casillas (2008), and many others (Abdul Aziz & Mohd. Don, 2014:41; Gavruseva, 2008; Ionin & Wexler, 2002, Paradis et al., 2008) agree on is that BE-overgeneration is likely to be a functional morpheme that bears agreement and/or tense features in the interlanguage of L2-English learners, similar to what has been reported in a few L1 studies (Brown, 1973; Marchman & Bates, 1994; Tesan & Thornton, 2004). Other commission errors of finiteness include overgeneralisations of tense inflections (e.g., *breaked* instead of *broke*) and agreement errors (e.g., *they was* instead of *they were*). Notwithstanding, their occurrences have been reported to be lower than omission errors (García Mayo and Olaizola, 2010:141; Gutiérrez-Clellen et al., 2008:14; Hawkins & Casillas, 2008:610; Ionin & Wexler, 2002:107; Jia & Fuse, 2007:1294; Paradis, 2005:180).

Having looked at finiteness errors in SLA studies, what might the structural representation of finiteness look like in the interlanguage? As mentioned in Chapter 1 (**§1.2.**), two main SLA views frequently brought up by many SLA studies examining finiteness

(Chondrogianni & Marinis, 2011; García Mayo & Olaizola, 2010; Geçkin & Haznedar, 2008; Ionin & Wexler, 2002; Liszka, 2009, 2015; Leung, 2003; Wong, 2012) are the *Partial Access* (*PA*) (Beck, 1998; Eubank, 1993; Hawkins & Chan, 1997; Meisel, 1997; Tsimpli & Dimitrakopoulou, 2007) and *Full Access* (FA) views (Goad & White, 2006; Lardiere, 1998; Prévost & White, 2000; Schwartz & Sprouse, 1994, 1996). The PA view proposes that functional categories and features, such as those pertaining to finiteness, cannot be successfully acquired due to impairment in the interlanguage caused by maturational issues. On the other hand, the FA view entertains the plausibility of successful morphosyntactic acquisition, and that "learner errors" are likely due to a mapping problem between the target-like syntactic representation and the non-target-like morphophonological exponent.

An advocate of the PA view is Liszka (2009, 2015), who demonstrated that the inconsistent use between the English present simple versus progressive forms by adult Francophone learners of advanced L2-English was an indication of a syntactic deficit in the interlanguage. In a picture description task, the Francophone participants in Liszka's studies managed to produce target-like present simple constructions ($\geq 90\%$). According to the author, this seemingly successful acquisition of tense marking in L2-English could rebut the PA view. However, the author (2015) dismissed this possible interpretation by presenting results from two other tasks, namely video description and contextualised dialogue, where the L2 groups did not supply the English progressive construction ($55\% < 80\%$) as much as the native English controls ($>97\%$) (ibid., pp.76-77). Liszka argued that the L2 groups were overusing the present simple construction under the influence of L1-French (ibid., p.78), since French has only one *présent* form which can bear a generic/habitual or progressive reading (ibid., p.62). Another study close to home is that of Wong (2012). In a grammaticality judgement task, Malaysian university students of L1-Malay (lower intermediate: $n = 22$; higher intermediate: $n = 17$) and L1-Chinese (lower intermediate: $n = 12$; higher intermediate: $n = 9$) seemed to be accurate in their acceptance of correctly inflected verbs (including the copula and auxiliary BE) ($>90\%$; ibid., p.11). However, both L1 groups did not consistently reject omission and commission errors, even at the upper intermediate level of English proficiency. Furthermore, although Wong observed that the advanced learners performed better than those of lower proficiencies, she maintained that such apparent improvement was reflective of conscious learning and that the participants' general inconsistency in rejecting ungrammatical items alluded to a representational deficit (ibid., pp.12-13).

In defence against the PA view, Goad & White (2006) argue that apparent failure to supply grammatical features is not due to syntactic deficits per se but rather due to prosodic constraints from the L1 affecting the morphophonological expression of those features. In a combined sentence completion and elicited production task administered by the authors, Mandarin-speaking learners of L2-English were able to do well in selecting and producing the appropriate tense inflections in relation to the established context. Although the linguistic performance of the L1-Mandarin learners surpassed Goad & White's own prediction, it indicated that these learners had successfully acquired tense in L2-English. In another study, Chondrogianni & Marinis (2012) observed that the L1-Turkish children ($n = 39$; age range: 6;2 – 9;7) in their study performed significantly poorer than the native English controls in producing 3rd.SING.-s and past -ed in a TEGI production task. However, when it came to processing ungrammatical conditions (i.e., tense and non-tense omissions) in a word-monitoring task, the L1-Turkish learners showed longer reaction times, indicating sensitivity towards ungrammaticality. This led Chondrogianni & Marinis to suggest that the L1-Turkish learners' syntactic knowledge (as seen in the word-monitoring task) was intact but their failure to produce tense inflections (as seen in the TEGI task) was merely an issue of mapping from syntactic categories to their morphophonological forms (Prévost & White, 2000).

While there has been much debate on how SLA theories may account for real or apparent morphosyntactic deficits in the bilingual representation, there are other factors (e.g., crosslinguistic influence, input, language environment) that cannot be overlooked (Sridhar & Sridhar, 1986). Crosslinguistic influence has been shown in many studies (Goad & White, 2006; Liszka, 2009, 2015) to affect interlanguage development at various linguistic domains. To reiterate Liszka's (2009, 2015) studies, the overuse of the English present simple construction, instead of employing the progressive structure, by advanced Francophone learners was likely due to influence from the French *présent* form, which does not distinguish between generic/habitual and progressive interpretations.

The role of input also plays a vital part in bilingual acquisition (e.g., Buschfeld, 2021; Nicoladis et al., 2007). In Nicoladis and colleagues' (2007) study on the production of past tense morphemes, the authors argued that lower accuracies observed amongst French-English bilingual children in comparison to their French and English monolingual peers were likely due to less exposure to either French or English in the bilingual environment. Concomitantly, lower token frequency in the input would delay if not impede the bilinguals' acquisition of the past tense morphemes.

In addition, the language environment in which bilinguals are exposed to has been shown to affect their interlanguage repertoire (e.g., Jia & Fuse, 2007; Liszka, 2015). Liszka (2015) compared three L1-French groups of different linguistic experiences and observed that the group residing in the UK for an average of 15.5 years produced the highest accuracy score of the English progressive form in the appropriate contexts (79.5%). The group whose exposure to L2-English was restricted to the classroom context in France performed the poorest (43.1%) (ibid., p.79). In Jia & Fuse's (2007) longitudinal study, the authors showed that a richer target-language environment (e.g., number of hours of television/books to read, of friends speaking the L1/L2, of time speaking the L1/L2 at home; ibid., p.1285) was a strong predictor of higher accuracy in the suppliance of English finiteness by L1-Mandarin children and adolescents residing in the US.

However, acquisitional studies generally tend to focus on native-speaker input and environment (e.g., Jia & Fuse, 2007; Liszka, 2015; Nicoladis et al., 2007). In many multilingual contexts, such as those in New Englishes settings, bilingual acquirers of English do not usually have sufficient exposure to the native-speaker input (Sridhar & Sridhar, 1986), which further complicates matters. One such case is found in the Outer Circle country of Singapore, where Buschfeld (2020, 2021) examined the roles of ethnic language and input in the endeavour to unravel the complexity of Singapore's multilingual environment. In Buschfeld's investigation of the use of L1-English by bi/multilingual children, the author observed that the Chinese children (52.7%) produced considerably higher rates of past tense omission than their Indian peers (29.3%). The author ascribed this difference to the analytic versus synthetic effects of the respective ethnic languages of Chinese and India (Buschfeld, 2021:203-204). However, she also argued that "cross-linguistic influence is not always the most important factor" when explaining the suppliance of non-standard features (ibid., p.203). To demonstrate her point, the Chinese-influenced {lexV + *FINISH*} perfective construction was not only supplied by the children of Chinese descent (3.6%) but was also found in the production of the Indian children (1.5%). This led Buschfeld to suggest that this linguistic innovation was likely to be undergoing cross-linguistic *homogenisation* (ibid., pp.204-205). Moreover, Buschfeld flagged up the problem of trying to trace or assume the type of English input received by children in the Singaporean ecology. This is because, in addition to being exposed to native models of British English and American English (e.g., via media), the children were also exposed to the prevalent colloquial variety viz., Singapore English (ibid., pp.195-196), let alone the competing grammars of the local ethnic languages. What this all shows is that while linguistic mechanisms and crosslinguistic transfer are undeniably crucial in paving the acquisitional route towards the

ultimate attainment of a target language, the question as to what "successful" acquisition looks like with a native-speaker model in mind becomes problematic when i) there are indeed many other factors (e.g., the type of input, sociolinguistic environment) to consider (Buschfeld, 2020:266-267) and ii) native-speaker models are not usually the goal of successful acquisition, such as in Outer Circle contexts (Kachru, 1992a:358).

4.4.3. Summary

To sum up this section, there are similarities and differences between English monolinguals and bilinguals in the acquisition of finiteness. One similarity includes markedly higher occurrences of omission errors than commission errors. However, for the L1-English monolinguals, the optional marking of finiteness in their early years of acquisition is temporary, whereas for their bilingual counterparts, it can persist into the later stages of acquisition. Commission errors (e.g., BE-overgeneration, overgeneralisation of regular past on irregular verbs) are also attested albeit scant in the production of both cohorts. That said, the linguistic environments in which monolingual and bilingual acquirers are exposed to do generate different linguistic outcomes. Therefore, to claim that bilinguals will always be susceptible to representational deficits without accounting for other sociolinguistic factors (including the types of input in a language environment) will instigate a myopic view of how language acquisition works.

4.5. Learnability issues

This chapter has established the concept of finiteness (i.e., a linguistic property associated with grammatical tense and agreement) and explained how tense and aspect are intertwined as they encode temporality. It has also looked at studies that examine the acquisition of finiteness and how the linguistic property is represented in the linguistic repertoires of monolingual and bilingual acquirers of the target language, which is English in this case. As we prepare to delve into the main study, we now provide a summary of the parametric differences between the languages of interest and speculate the learnability issues participants in our study may face in the acquisition of finiteness in StE:-

StE (target language):

- *Finiteness*: Finiteness is obligatory in finite clauses, and it is marked across all verbal elements including lexical verbs, the copula BE, and auxiliary verbs.

- *Verbal elements including the copula*: All verbal elements must be overtly pronounced.

Malay and Chinese:

- *Finiteness*: The finiteness features of tense and agreement are not specified in Malay and Chinese. In the absence of GRAMMATICAL TENSE, aspectual markers, temporal adverbials, and other contextual information are some of the linguistic devices employed to convey time.
  - *The acquisition task*: L1-Malay and L1-Chinese learners need to acquire the morphological expression of grammatical tense and agreement of StE.
- *Copula*: Unlike in StE, the copula verbs in Malay and Chinese i) do not bear finiteness and ii) exhibit a certain degree of optionality.
  - *The acquisition task*: L1-Malay and L1-Chinese learners need to acquire the obligatory feature of the StE copula BE, on which tense and agreement need to be expressed.

CME:

- *Finiteness*: The structural representation of CME is similar to that of StE in that it has the finiteness features of tense and agreement, but they are optional.
  - *The acquisition task*: L1-MalE learners need to acquire the obligatory features of tense and agreement in StE.
- *Copula*: Similarly, the copula BE in CME is the same as in StE but its use is optional.
  - *The acquisition task*: L1-MalE learners need to acquire the obligatory feature of copula BE in StE.

5 Methodology

5.1. Research questions and hypotheses

The operationalisation of the main study is driven by four research questions (hereafter *RQ*s). The first three RQs deal with the acquisition of finiteness in StE, whereas RQ4 explores attitudinal responses to CME and StE. The RQs are listed below, and their respective hypotheses (H_n) are formulated, correspondingly:

RQ1. **To what extent does L1 play a role in the linguistic performance on StE?**

The learnability issues presented in the previous chapter are revisited with reference to the predictions posited by SLA theories.

H_1: Under the FA view, successful acquisition of functional features is possible under UG guidance. Malaysian learners of L1-Malay and L1-Chinese will successfully acquire the morphological expressions of grammatical tense and agreement of L2-StE, which do not exist in their L1 repertoires. They will also successfully acquire the obligatory feature of the StE copula, which is otherwise optional in their L1s. Learners of L1-MalE will successfully reset the optional features of finiteness marking allowed in CME to obligatory, as required in StE.

H_2: Under the PA view, successful acquisition of functional features, even at advanced levels of proficiency, is not possible. Malaysian learners of L1-Malay and L1-Chinese will not be able to acquire grammatical tense and agreement as well as the obligatory feature of copula BE in L2-StE. Learners of L1-MalE will not be able to reset the English finiteness features from optional to obligatory in StE. In these cases, greater variability in the linguistic performance of the StE finiteness will be observed across the board.

RQ2. **What is the contribution of CME on the error types in StE?**

Due to the nature of CME as an L2 variety, it can be difficult to identify which non-target features of finiteness are a result of cross-linguistic transfer (e.g., from CME, L1-Malay, or L1-Chinese) or a reflex of other L2 behaviours such as

overgeneralisation. Furthermore, by looking only into the linguistic performance of L1-MalE speakers with the assumption that they all acquired CME as the only L1 would be erroneous. This is because acquirers of L1-MalE might have been exposed to a variety of English input in the home environment (see Buschfeld, 2020), which makes it difficult to measure and single out an English variety for them.

H: That said, given the prevalence of CME in the Malaysian milieu, if a non-target feature recurs across all Malaysian L1 groups, then the feature shall be entertained as a plausible contribution of CME.

RQ3.   **Are there task effects on finiteness?**

The linguistic tasks employed to elicit the metalinguistic knowledge and use of finiteness are the grammaticality judgement and narrative tasks. This inquiry sets out to investigate if there are differences in linguistic performance due to the administration of these tasks.

H_1:   By administering the GJT and NT, no task effects will be observed in the introspection and production of finiteness. Two dichotomous outcomes are further predicted:

(a) Learners of StE who consistently exhibit target-like performances in both tasks will lend support to the FA view.

(b) Learners of StE who consistently exhibit non-target performances in both tasks will corroborate the PA view.

H_2:   By administering the GJT and NT, there will be task effects on the introspection and production of finiteness.

(a) Under the FA view, the MSIH anticipates that learners across different L1s may perform more poorly in the NT than in the GJT due to issues pertaining to morphophonological realisations of finiteness.

(b) Alternatively, learners may generally perform better in the NT than in the GJT. In this case, the MSIH cannot be supported but the FA view may still be substantiated.

RQ4.   **To what extent is CME perceived as functionally different from StE in Malaysia?**

H_1:   Under the null hypothesis, Malaysians are not aware of the different types of English spoken in Malaysia.

H[2]: Malaysians are able to discriminate CME and StE and attribute their communicative functions to different social contexts. This hypothesis is based on previous studies (Rubdy, 2007; Tan & Tan, 2008) which have demonstrated Singaporean children's ability to distinguish between the colloquial and standard varieties of Singapore English with respect to interpersonal closeness and formality of the social context.

Having laid out the RQs and their predictions, the remainder of this chapter focuses on the methodological design of the main study. The rationales and designs of the instruments employed are presented in the next section (**§5.2.**). In **§5.3.**, the procedure covers the data collection phases which took place before and during Covid. This is followed by an overview of participants' demographic and linguistic backgrounds (**§5.4.**). **§5.5.** then describes the organisation and tools used for the analyses of the data. The chapter wraps up by confirming the ethical approval granted by the University for the conduct of this study (**§5.6.**).

5.2. Instruments

This main study implemented a concurrent embedded methodological design comprising i) the linguistic tasks of grammaticality judgement and narrative-telling to examine the acquisition of finiteness in StE, and ii) a sociolinguistic questionnaire eliciting participants' demographic and linguistic backgrounds as well their attitudes towards CME and StE.[10]

5.2.1. Grammaticality judgement task

5.2.1.1. Rationale

Grammaticality judgement tasks (GJTs) are a popular methodological apparatus in linguistic research that is used to elicit participants' intuitions about the (un)grammaticality of sentence constructions. As one's mental grammar cannot be directly measured (García Mayo, 2003:97), GJTs may be employed to draw inferences about it (Schütze & Sprouse, 2013:28). The

---

[10] Concurrent embedded designs are a type of methodological approach that collects quantitative and qualitative data simultaneously. In this design, one of the types of data serves as a supplementary role to support the interpretation of the main findings (Mackey & Bryfonski, 2018).

methodological design of GJTs allows the researcher to construct stimulus sentences around the morphosyntactic phenomenon of interest. This is especially useful if the linguistic phenomenon is not often produced (Murphy, 1997:36) or is intentionally avoided due to its difficulty in speech (Ellis, 1991:163). In terms of feasibility, GJTs are easier to administer than oral production tasks because the former tasks can be distributed to many participants at once, whereas the latter tasks can only be done with one individual at a time (Ellis, 1991:162).

However, since GJTs are a performance task, one cannot be certain that the judgement data yielded can faithfully reflect linguistic competence (Murphy, 1997:36). This is because there are other extralinguistic factors, such as processing issues, boredom effects, and fatigue, which may affect the participant's perception of the sentences being presented (Ellis, 1991:164; Lust & Blume, 2017:159). Furthermore, it is difficult to determine whether the judgement data are indeed genuine responses to the linguistic stimuli or a result of guesswork (Hedgcock, 1993:4). What is more, when judging a sentence construction, the participant may not necessarily attend to its morphosyntactic well-formedness but instead take into account other linguistic criteria such as the semantic or pragmatic aspects of the sentence (Bard et al., 1996:33; Lust & Blume, 2017:158; Murphy, 1997:36).

Notwithstanding, some of the methodological issues that come with GJTs can be mitigated through careful design and execution of the task, such as running a pilot study with native speakers of the target language to check the stimulus sentences, encouraging participants to take breaks to minimise fatigue or boredom effects, and/or ensuring that the study's aims and instructions are straightforwardly simple to follow through.

Conclusively, GJTs are a versatile linguistic tool that allows researchers to zero in on certain linguistic features of interest by manipulating their (un)grammaticality and to elicit participants' perceptions of them. For the main study, we employed a GJT as one of the linguistic tasks in the endeavour to elicit our participants' metalinguistic knowledge of finiteness in StE. The next subsection describes its task design.

5.2.1.2. Task design

The GJT employed for the main study was a timed one. It contained a total of 150 written stimulus sentences (75 grammatical, 75 ungrammatical), which were further compartmentalised into:

i)   a practice set of 6 sentences (3 grammatical, 3 ungrammatical),

ii)   a filler set of 54 sentences (36 grammatical, 18 ungrammatical), and

iii)   an experimental set of 90 sentences (36 grammatical, 54 ungrammatical), targeting finiteness.

In what is to follow, we focus only on the construction of the experimental and filler items and stipulate the presentation of the experimental task.

Experimental items, fillers, and their conditions: Motivated by the Rice/Wexler Test of Early Grammatical Impairment (TEGI; Rice & Wexler, 2001), the finiteness features comprising the experimental items were the English copula BE, progressivity, DO-support, 3rd.SING.-*s*, and regular past -*ed*. All features under the experimental set contained the standard and omission conditions. That said, the features of progressivity, DO-support, and past -*ed* had an additional commission condition. Filler items encompassed features from other (morpho-)syntactic domains, such as plurality, adjective-noun word order, and *wh*-question construction. For the whole list of stimulus sentences, refer **Appendix A.2.1.2**..

Lexicalisations: Generally, each experimental condition is to be assigned between four (Weskott & Fanselow, 2009:231; Ionin, 2012:41) and eight (Schütze & Sprouse, 2013:39) lexical items (hereafter *lexicalisations*). In this study, six lexicalisations were assigned to each condition. This generated 54 grammatical sentences and 72 ungrammatical sentences. Altenberg & Vago (2004) warn of a response bias caused by an imbalance between grammatical and ungrammatical sentences. For instance, if participants notice a higher number of ungrammatical sentences in the GJT, they might be primed to reject sentences that are grammatical. Therefore, to avoid this kind of response bias, 18 more grammatical filler sentences were added in the endeavour to achieve the balance.

All sentences designed for the GJT were single-clause constructions. They were controlled for word frequency and sentence length to avoid confounding factors (e.g., semantic complexity, lexical frequency, and different sentence lengths) from affecting participants' rating responses (Aranuchalam, 2013:223-224; Cowart, 1997:45-46; Keller, 1998:10). The number of words generally ranged from 7 to 9 per sentence. Sentential and prenominal modifiers (e.g., adverbial phrases, adjectives, pre-modifying nouns) were added to achieve the target sentence length.

In terms of word choice, the construction of stimulus sentences started with the selection of transitive verbs. The verbs loosely fell within the midrange of Leech et al.'s (2001) frequency verb list, so that they were not too easy nor too difficult to comprehend. Subject NPs chosen for the lexical verbs were balanced for animacy (i.e., 72 animate subjects and 72

inanimate subjects) and number (i.e., 72 singular subjects and 72 plural subjects). Finally, all word selections were matched against Oxford 3,000, which is a vocabulary list supplied by the Oxford Learner's Dictionaries online. This was to ensure that all the words fell between CEFR's A1 (elementary) and B2 (upper-intermediate) levels of proficiency.

Presentation: The test sentences were pseudorandomised and ordered in six blocks of 24 items per block.[11] The list of blocked sentences was then counterbalanced so that half of the participants did not see the same order as the other (Cowart, 1997:93; Keller, 1998:10). This would help minimise the likelihood of participants' rating responses being affected by presentation order (Arunachalam, 2013:224).

The GJT was run on PyschoPy 3.0 (Pierce et al., 2019; before Covid) and Gorilla.sc (Anwyl-Irvine et al., 2019; during Covid). Sentences were presented one by one for 3.5 seconds (Bialystok, 1979; Han, 2000). After each display, participants were given an indefinite time to rate the sentence on a five-point Likert scale, with 1 indicating "least grammatical" and 5 indicating "completely grammatical". They were also given the choice to opt for the "Don't know" option, which was made available to prevent them from attributing the "don't know" value to the midpoint (i.e., 3) of the rating scale. Finally, a break was introduced in the middle of the experiment as a means to minimise task-induced fatigue (Arunachalam, 2013:224). To see the written instructions and presentation flow, refer **Appendix A.2.1.1.**.

5.2.2. Narrative task

5.2.2.1. Rationale

Narratives are a genre of discourse that taps into the skill of (re)constructing a sequence of real life or fictitious events by means of language. Not only are they practised across cultures (Berman, 2009:355; Langdon, 2016), but they are also acquired since childhood and develop into adulthood (Berman, 2004; Berman, 2009:355; Mäkinen et al., 2014:25).

---

[11] Pseudo-randomisation is a crucial step to avoid sentences of similar syntactic constructions from appearing successively during the presentation of the test stimuli (Arunachalam, 2013:224; Cowart, 1997:51; Sprouse et al., 2013:227). Blocking divides the whole sentence list into smaller lists/blocks that contain only one sentence from each condition (Cowart, 1997:94). This procedure helps to avoid sentences of the same syntactic structure from clustering at certain points in the presentation (Ionin, 2012:42).

The establishment of a successful story requires prior knowledge of the participating characters (whom the narrator needs to introduce and (re-)refer to throughout the narration) and the temporal succession of the plot (in which the narrator must retrospect in order to relay to the audience) (Berman, 2009). Navigation of these facets of information demands i) cognitive abilities to index references and temporality coherently, and ii) linguistic abilities to use the appropriate deictic expressions cohesively (Berman, 2009:359; see also Karmiloff-Smith, 1985; Cain, 2003). As such, the production of narratives can be analysed according to the parameters of macrostructure and microstructure. Macrostructure is concerned with cognitive-general (global) processes involved in the organisation of a story. Microstructure relates to language-related (local) domains in which linguistic elements are used to compose the story (Stein & Glenn, 1979, in Altman et al., 2022; see also Norbury & Bishop, 2003).

Since narratives naturally generate a wealth of production data, they have been utilised as an ecological instrument to assess language development and disorders (Altman et al., 2022; Botting, 2002; Justice et al., 2006; Mäkinen et al., 2014; Norbury et al., 2014). In the main study, one of the linguistic tasks involved narrative-telling, which was implemented to analyse microstructure, but, more specifically, the grammatical components of finiteness.

5.2.2.2. Task design

A non-verbal animated short film entitled "Snack Attack" (2012) was used for the narrative task (NT). It runs for 4 minutes and 38 seconds and is publicly accessible on the internet.

The short features two main characters: an old woman and a young man, who seem to be fighting over a packet of cookies at a train station. As they are sitting on a bench, the old woman thinks that the young man is eating her snack and thus becomes enraged by his insolence. This conjures up a tug of war between them. As they try to snatch the last cookie from each other, the young man wins but decides to share one half with the old woman. She does not appreciate his kindness and storms into her train, which has just arrived at the platform. It is not until she finds her own packet of cookies inside her handbag that she realises that the snack, which was already on the bench by the time she got there, was identical to hers and that it belonged to the young man. The film ends with the old woman having mixed feelings of remorse and gratitude as she looks out the train window to see the young man leaving the platform.

As narratives can be told in the past or historical present tense (Bonilla, 2011; Schriffrin, 1981; Wolfson, 1978), there was a likelihood that our participants would do so in the NT. As

such, the short was chosen because, as seen in the synopsis above, it contained a flashback scene which would require the narrator to describe it in past tense. Furthermore, in anticipating that the narrator would consistently use either past or present tense in their narration, different interview prompts were designed to elicit different tense types from the narrator (Table 5.1).

Table 5.1: Question prompts for the NT

Prompts	Expected use of tense
Q1. Tell me what happened in the video.	• Past tense / Historical present for the most part • Past tense in the flashback scene
Q2. If you were the young man, what would you do when the old lady took your packet of cookies?	• Hypothetical past
Q3. If you were the old lady, what would you do when you found out that you ate the young man's cookies?	• Hypothetical past
Q4. What is the moral of the story?	• Non-past

Q2 and Q3 might yield hypothetical past, in which case past tense modals (e.g., would, could) would be used. Indeed, the main study is not concerned with the use of modal verbs. However, what Q2 and Q3 might elicit in addition to hypothetical propositions was the recounting of events in the story, where participants might employ past tense constructions. Q4 was designed in the present tense to glean present-tense sentences while also anticipating imperatives (e.g., "Don't judge.") and non-past declaratives involving deontic modals (e.g., must, have to) in the participants' response. Accordingly, the question prompts in Table 5.1 were presented in that order, and the NT session was audio-recorded and transcribed for data analysis.

5.2.3. Language background and attitude questionnaire

5.2.3.1. Rationale

A self-administered questionnaire designed for the main study served a two-fold aim: i) to extract the biodata of participants; and ii) to tap into the participants' perceptions of, attitudes towards, and preference for the English varieties of interest viz., CME versus StE. Addressing the first aim, this was necessary to describe whom the sample populations of our study consisted of. In this case, the samples comprised the Malaysian and UK groups. With regards to the second aim, attitudinal questions were constructed to showcase the views of Malaysians in the endeavour to address RQ4 of the main study.

Unlike other linguistic tasks, (socio)linguistic questionnaires are not employed to assess respondents' performance in the survey (Dörnyei & Taguchi, 2010:4). Not only does this put the respondents at ease when filling out the questionnaire, but it also provides a safe space for them to respond to the task truthfully, provided that their anonymity is warranted (Dewaele, 2018:271). In terms of efficacy, running surveys digitally can save researchers' time and effort as the data collected can be processed easily and quickly by a compatible computer software (Dewaele, 2018:271-272; Dörnyei & Taguchi, 2010:6; Iwaniec, 2020:324-325).

However, the caveats of administering questionnaires are not few, as there can be problems of respondents misinterpreting the questions or statements being posed, inaccurately providing answers about themselves due to personality reasons, not completing the survey due to boredom or fatigue, or conforming to societal beliefs or researcher expectations instead of their own, just to list a few (Dörnyei & Taguchi, 2010:6-9; see also Dewaele, 2018:271-272; Iwaniec, 2020:326-327).

Indeed, there are ways to mitigate some of the concerns raised above, such as designing surveys that do not take too long to complete (Dollinger, 2015:231-233; Dörnyei & Taguchi, 2010:12-13; Schleef, 2014), constructing questions and/or statements that are unambiguous and simple to understand (Lietz, 2010), and including a variety of question types to minimise monotonicity of the survey (Iwaniec, 2020:327).

5.2.3.2. Task design

The questionnaire of the main study contained a total of 45 questions that bifurcated according to the nationality (i.e., Malaysia or UK) of respondents. The questions were further partitioned

into four main sections, which are briefly described as follows. For the entire list of questions, refer **Appendix A.2.2.**).

Section A: Language history and use. This section sought information about respondents' general linguistic background. This included *a list of languages/dialects learned* alongside their corresponding *age of onset of acquisition*, *frequency of language exposure and use* (in speech, writing, listening, and reading), *language use with different people* (e.g., family, friends, classmates/colleagues, strangers), and *language of instruction across different levels of formal education* (i.e., primary through tertiary education).

Section B: Language proficiency. In this section, respondents were asked to indicate their language proficiencies. This included *a self-report on the language skills of speaking, listening, writing, and reading* in the languages acquired, as well as the most recent *formal assessment in English* alongside the *grade* obtained.

Section C: Language perception and attitude. This section was only available in the Malaysian pipeline of survey as it aimed to elicit respondents' views on CME ("Manglish" in the survey) and StE. The section included overt questions on respondents' *awareness* and *description* of CME, *attitude-related statements on CME (n = 5) and StE (n = 5)* upon which the respondents had to rate their (dis)agreement, and *varietal choice across different social contexts* (i.e., with family, friends, non-/English Language teachers, colleagues, Malaysian strangers, foreigners).[12]

Section D: Demographic information. This section sought respondents' demographic information, which included *which part of the country respondents were from*, *highest education level achieved* (e.g., A-levels/sixth-form, Bachelor's), *types of primary and secondary education*, *details of the respondents' current occupation* (e.g., student, employee, retired), *monthly household income before taxation*, *ethnicity*, *year of birth*, and *biological sex*.

Items of the questionnaire were drafted with reference to Wigdorowitz's (2022, see also Wigdorowitz et al., 2020) Contextual and Individual Linguistic Diversity Questionnaire

---

[12] In Section C, the layman term "Manglish" was compared to another nomenclature "the Rojak language" or more widely known as *Bahasa Rojak* (hereafter *Rojak*) in Malay. Rojak is defined as "any mixture of two or more languages in communication, with any of the languages being the base language" (Abu Bakar, 2009:99). CME is an example of the Rojak phenomenon since it allows for codeswitching to occur (ibid., pp.99-100). In the MalE literature, CME is usually recognised as Manglish instead of Rojak. While there was not much theoretical grounding in our study to bring this comparison to the fore, it was interesting nonetheless to explore our respondents' views on these two terms, such as whether they were distinct, overlapping, or identical linguistic varieties. Unfortunately, due to space limitations, findings on the CME-Rojak comparison will not be presented.

(CILD-Q), Birdsong and colleagues' (2012) Bilingual Language Profile (BLP), Li et al.'s (2014) Language History Questionnaire (LHQ 2.0), and Rubdy (2007) and Harada's (2009) attitudinal studies on the colloquial and formal varieties of Singapore English. Having finalised the draft, the questionnaire was published online via Qualtrics. All data collected were then processed by the same online survey tool and exported into Excel files, accordingly.

5.2.4. English proficiency test

5.2.4.1. Rationale

The administration of standardised language proficiency or placement batteries is a common practice in experimental SLA studies which aims to identify and control for research participants' proficiency in the target language (Tremblay, 2011). As noted by Long et al. (2018), participants may be placed into different proficiency groups (Brown, 1989:65) or along a spectrum of proficiency scores, depending on the research objective.

There are several types of English standardised tests. The commercial ones, such as the International English Language Testing System (IELTS) and the Test of English as a Foreign Language (TOEFL), offer a comprehensive assessment across different language skills (i.e., reading, listening, speaking, writing). There are also quick placement tests targeting specific domains such as grammar and vocabulary (e.g., British Council English Level Test, Cambridge General English Test, Macmillan Straightforward Quick Placement and Diagnostic Test).

There are issues pertaining to the reliability and validity of standardised proficiency measures employed in SLA studies. For instance, standardised tests may fall short in providing fine-grained information on specific aspects of the learner's knowledge of the target language (Thomas, 1994:326). Concerning quick placement tests, the lack of comprehensibility in the assessment may fail to provide a holistic estimate of the learner's proficiency. Moreover, there is "a lack of uniformity among researchers" in choosing and justifying their choice of placement test (Tremblay, 2011:343).

In spite of the caveats that come with different placement tests, they serve the purpose of providing a standardised proficiency measure crucial for SLA studies because, in addition to indicating participants' proficiency scores or levels, they also help to control for variation which arises from the different language learning backgrounds of participants (Tremblay, 2011:340). Until there is unanimity on how proficiency could be methodically measured, it calls for wisdom for the researcher to select a placement test most appropriate for their study.

5.2.4.2. Task design

The standardised proficiency test chosen for the main study was the British Council Online English Level Test (hereafter *English Proficiency Test* (EPT)).

The EPT contained 25 questions with three choices of answers. The questions were designed to measure one's knowledge of the English grammar, vocabulary, and word collocations. Each question also came with a subsidiary question asking test takers to rate their confidence level (i.e., Not sure/Fairly sure/Certain) when choosing their answer. At the end of the test, an approximated proficiency score (in percentage), alongside its CEFR level and proficiency category (e.g., intermediate, upper intermediate), was displayed onscreen.

One of the benefits of using this EPT was that it was freely accessible online. Another interesting component was the level-of-confidence indicator which was taken into account in the grading system. Furthermore, that the EPT calculated the proficiency score upon completion of the test saved researcher time. However, one of the shortcomings was that the test did not explain how the weightage of accuracy scores was allocated to the main and subsidiary questions. Also, it was not clear how the range of scores was partitioned in terms of their categorical levels of proficiency. Even at 100%, the score was labelled as upper intermediate, not advanced. Moreover, the brevity and validity of the EPT raised concerns as test takers were not assessed on other skills such as listening, writing, or speaking.

Notwithstanding, for practicality reasons, we decided to utilise the present EPT to keep the entire experimental study under, maximally, 1.5 hours per session.

5.3. Procedure

5.3.1. Pilot

Before the main study was officially conducted, a pilot was run in order to test and improve the flow, feasibility, and materials of the linguistic tasks (Gass & Mackey, 2017:52; Mackey & Gass, 2005:43). A total of 24 volunteers studying at the University of Cambridge were recruited: 16 of them were Malaysians and 8 were English-dominant speakers from the UK, South Africa, and Singapore. Not only were their data examined to see if there were any issues

with the linguistic stimuli, but their feedback about the study was also collected to improve the study design. Due to limited space in this thesis, the data from the pilot study are not reported.

5.3.2. Pre-Covid

5.3.2.1. Kuala Lumpur, Malaysia

The first official phase of recruitment took place between November and December 2019 at the University of Malaya, Kuala Lumpur. The call for participants was made through several means. One of them was carried out by advertising the study in several English Language classes under the permission of the host faculty i.e., the Faculty of Languages and Linguistics. These compulsory classes were run by the faculty's Language Unit for university students from various degree courses and of different levels of English proficiency. The classes which I was arranged to go to consisted of students who scored Band 5 (corresponding to CEFR C1) in the Malaysian University English Test (MUET). Those who were interested in taking part were invited to leave their names and contact details on a sign-up sheet. Apart from that, social media such as Facebook, Instagram, and Twitter were also utilised to publicise the study. Twitter had the best reception as it was a preferred platform amongst the local students to disseminate campus and course-related information. In the final push for recruitment, posters were put up on noticeboards at faculties, bus stops, and canteens in the vicinity of the host faculty.

Correspondences were made via email to arrange one-on-one appointments with me. The invitation email contained not only a link to the Doodle poll where participants could anonymously choose their preferred time slots but also digital copies of the participant information sheet and consent form which they could read before making the appointment.

The experimental sessions were conducted face-to-face in a research room in the host faculty building. At the beginning of the session, participants were briefed about the study, covering its main aim, linguistic tasks, and the ethical conduct of the research. The audio-recording activity, which had to be carried out for the NT, was also flagged up in advance. Furthermore, participants were given a chance to read the participant information sheet and consent form and ask me any questions about the study. Once they were happy to proceed with the tasks, they were asked to sign two copies of the consent form, one for their personal record and the other for mine. As the main study aimed at eliciting the use of StE amongst Malaysians, I made sure that, at least during the conduct of the GJT and NT, I only used StE with the

participants in the endeavour to create an air of formality in the room. The participants were also made aware that they were expected to use StE as though they were in a formal setting.

The first task was the NT. Participants were instructed to watch the animated film, which was played on my laptop. They were then asked to narrate it to me as though I had never watched it before. They could only watch it once but were allowed to write down some notes. Once they were ready to begin the task, I started the audio recording session upon my participants' permission and began the interview, accordingly. During the narration, the participants were allowed to refer to their notes every now and then for quick references.[13] When the narrative was over, the sound recorder was turned off in front of the participants to assure them that the rest of our session was not recorded.

The second task was the GJT. Verbal instructions were first given to the participants before they read the more simplified, written version on the computer screen. Specifically, they were told, verbally, to imagine that the series of sentences they would be reading were randomly extracted from different manuscripts and that their role was to introspect the grammaticality of those sentences using their knowledge of StE. They were also strongly advised to take a break when they were halfway through the task, as instructed onscreen. This was to reduce any boredom effects which might affect their concentration. Moreover, they were encouraged to read every stimulus aloud in case the 3.5-second sentence display was too short for them. A unique identification number was assigned in order for the participants to begin the GJT. The same ID was used for the rest of the tasks. After completing the GJT, the participants were asked to notify me so that I could direct them to the next one.

Concerning the language background and attitude questionnaire, I sat next to the participants as they filled it out. Not only did this allow them to ask me questions about the survey, but it also gave me a chance to seek further information or clarification about their demographic and linguistic backgrounds, which were not already captured by the questionnaire.

The final task was the EPT. This was freely accessible online. It is worth mentioning that both Malaysian and British participants were required to undertake this test, as their scores would be used as standardised proficiency covariates when analysing their performances in the GJT and NT. As the test was straightforward, the participants were left alone to do the test. At the end of it, their scores were screenshot and saved in .jpg format using their unique ID.

---

[13] There was one incident where a participant kept reading from her note as she narrated the story. Consequently, I had to stop the audio recording and asked the participant to redo the task without relying too much on her note.

The experimental session lasted between 1 and 1.5 hours for the Malaysian group. At the end of each session, every participant had completed all the linguistic tasks and were reimbursed RM40 for their time.

5.3.2.2. Cambridge, UK

The second round of recruitment was carried out from February to March 2020 at the University of Cambridge. It was aimed at British university students, who were needed to form the control group. The study was advertised through JCR and MCR newsletters across the colleges. The structure of the experimental session was similar to the one with Malaysians, except that the British participants were not required to do the sociolinguistic survey (i.e., Section C in the questionnaire). The experimental sessions were therefore shorter, lasting about one hour per session. The interviews were all carried out in a couple of meeting rooms at Newnham College. Unfortunately, the recruitment had to be halted midway as the UK underwent its first Covid lockdown on 16 March 2020. At this point, only 15 participants were successfully interviewed.

5.3.3. Covid

By the new academic year in October 2020, movement was still restricted in the UK due to Covid. Consequently, the final data collection phase had to be moved online, as the number of participants needed for the control group was still lacking. Having piloted the online study with three volunteers, the call for participants was readvertised on the college newsletters. The data collection ran from October through November 2020 via Zoom. By the end of November, 15 more British participants were successfully added to the control group.

The British participants recruited before and during Covid completed all the tasks and were reimbursed £10 Amazon e-vouchers after each bout of data collection was completed.

5.4. Participants

5.4.1. General demographics

A total of 175 adult participants took part in the main study. Of them all, 145 were Malaysians and 30 were British. The average ages of the Malaysians and British were similar (**Malaysia:** $M = 20$ years, $SD = 1.212$, age range $= 18 - 24$ years; **UK:** $M = 21$ years, $SD = 2.54$, age range $= 18 - 28$ years). For both nationality groups, there were considerably more female participants than male participants (**Malaysia:** 61.38% females, 38.62% males; **UK:** 63.33% females, 36.67% males). Most of the Malaysian and British participants were students from the University of Malaya and the University of Cambridge, respectively. The two exceptions came from the Malaysian group. One of them was an employee who graduated with a bachelor's degree in Engineering at Universiti Teknologi PETRONAS (Malaysia) sixth month prior to the time of recruitment. Another participant just completed his sixth form at a national secondary school (in Malay: *Sekolah Menengah Kebangsaan*) and was preparing to embark on an undergraduate degree.[14]

5.4.2. Language background

5.4.2.1. First language(s)

Table 5.2 presents the L1(s) of participants. Most of the participants acquired only one language as their L1 (Malaysia = 71.03%; UK = 96.67%). There were also simultaneous bilinguals (Malaysia = 26.9%; UK = 3.33%) and trilinguals (Malaysia = 2.07%), but they were fewer in number than that of the "one L1" category.

  In the Malaysian group, the two L1 groups constituting the highest proportions of participants were the L1-Chinese (29.66%) and Malay (22.76%) groups. L1 groups with the least number of participants ($n = 1$ per group) included a MalE-Kelabit bilingual and three trilingual speakers of L1-MalE-Malay-Iban, L1-MalE-Malay-Tamil, and L1-MalE-Tamil-Telugu. Note that for all bilingual and trilingual groups, MalE was one of the L1s.

---

[14] Exceptions were made to these two individuals because L1-MalE speakers were still needed for the study. These participants fulfilled most of the criteria i.e., being at the minimal age of 18 years and highly proficient in English.

In the UK group, an overwhelming majority of the participants (96.67%) acquired BritE as their only L1. An exception was a bilingual who acquired Gujarati alongside BritE.

Table 5.2: Participants' L1(s)

L1(s)		Malaysia	UK
Malaysia:	*UK:*	N=145	N=30
One L1		**103 (71.03)**	**29 (96.67)**
*MalE	*BritE	19 (13.10)	29 (96.67)
*Malay		33 (22.76)	
*Chinese		43 (29.66)	
Tamil		8 (5.52)	
Two L1s		**39 (26.90)**	**1 (3.33)**
*MalE, Malay	BritE, Gujarati	21 (14.48)	1 (3.33)
*MalE, Chinese		13 (8.97)	
MalE, Tamil		4 (2.76)	
MalE, Kelabit		1 (0.69)	
Three L1s		**3 (2.07)**	
MalE, Malay, Iban		1 (0.69)	
MalE, Malay, Tamil		1 (0.69)	
MalE, Tamil, Telugu		1 (0.69)	

At this point, it is worth mentioning that only participants under the L1 groups marked with the asterisk (*) in Table 5.2 were included in the statistical analyses of the GJT and NT. This is because the rest of the L1 groups contained very few participants ($n < 10$), which would (further) undermine the power of the study.

5.4.2.2. The acquisition and use of English

5.4.2.2.1. Onset age of acquisition

Figure 5.1 charts the distribution of onset age of acquisition (AoA) in English amongst the Malaysian cohort. British participants were not included in the chart because they had all been exposed to English since birth. The AoA of MalE (hereafter *AoA-MalE*) ranged from ages 0 to 7. Participants who had been exposed to MalE since birth (age 0) were considered as L1

acquirers, whereas those acquiring MalE from two years old onwards were categorised as L2 learners.[15] Age 0 comprised the largest number of participants ($n = 61$, 42.07%). This included acquirers of L1-MalE as well as the simultaneous bilingual and trilingual speakers of L1-MalE. On the other hand, learners of L2-MalE generally started acquiring the target language at the ages of 4 ($n = 29$, 20%) and 5 ($n = 24$, 16.55%). This would potentially be the time when children were enrolled in kindergartens and were exposed to MalE for the first time. Meanwhile, the latest AoA-MalE (i.e., 6 and/or 7 years old) roughly coincided with the time children entered primary education. In this phase, Standard English would be introduced as a compulsory language subject in the classrooms.

Figure 5.1: Age of acquisition of MalE
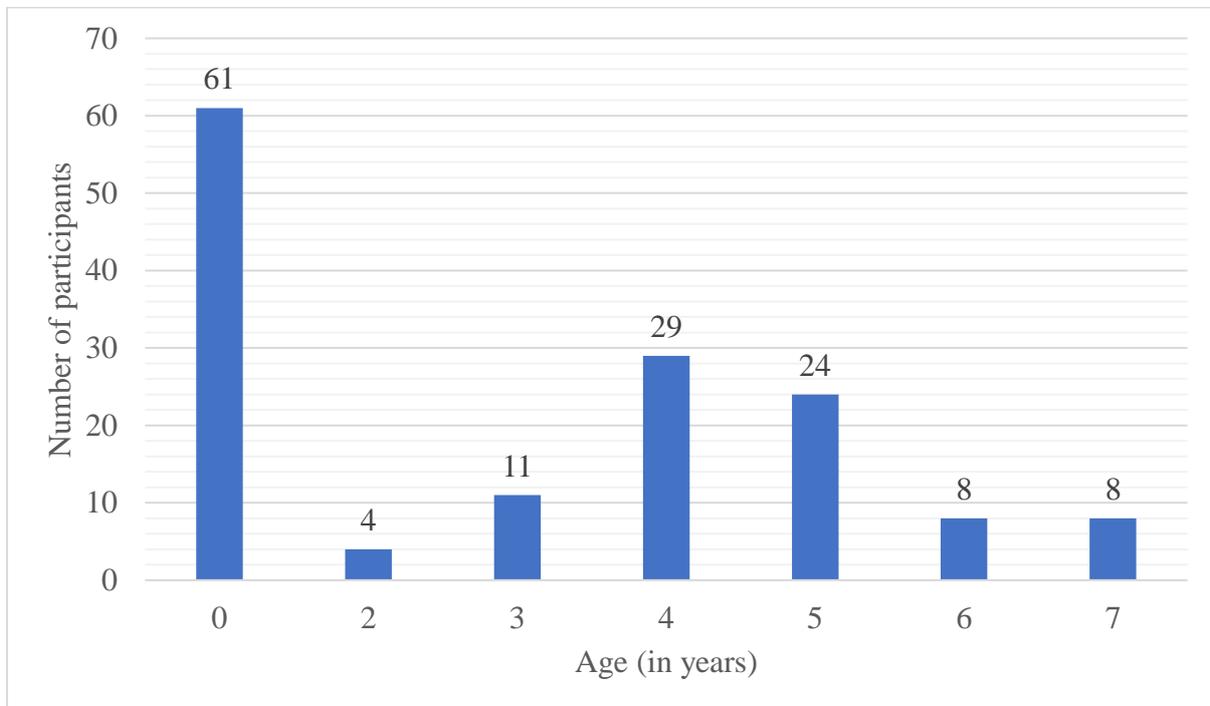
5.4.2.2.2. Current exposure to English

Regarding the current exposure to and use of English (see Table 5.3 for their distribution), we focus only on the Malaysian cohort, since the British participants were exposed to and used English every day. As shown in Table 5.3(A), most participants were exposed to (i.e., listening, reading) and used (i.e., speaking, writing) English on a daily basis. This indicates the

---

[15] Only one participant (MM002) indicated that she started acquiring English at one year old. Accordingly, we subsumed this datapoint under age 0.

prevalence of English in the Malaysian linguistic environment. In Table 5.3(B), it was fairly common for Malaysians to use English across different social settings. Amongst family members, over half of the Malaysian cohort claimed that they used English as one of their languages of choice with their parents (63.45%) and siblings (62.07%).[16] These percentages suggest that English was used as a home language, even if it was not initially acquired as an L1. Moving away from the home environment, English was reported to be used more frequently with friends, acquaintances, and strangers (Table 5.3(B)). More interestingly, almost all participants used English with people of different ethnicities (friends = 99.31%; classmates/colleagues = 100%; strangers = 100%), suggesting that English was employed to facilitate interethnic communication in the local milieu.

Table 5.3: Current exposure to and use of English amongst Malaysians, in frequency (and %)

	Malaysia N=145
(A) Exposure to and use of English	
Speaking	
Daily	123 (84.83)
Weekly	22 (15.17)
Writing	
Daily	142 (97.93)
Weekly	3 (2.07)
Listening	
Daily	143 (98.62)
Weekly	2 (1.38)
Reading	
Daily	145 (100)
Weekly	0
(B) Use of English in social contexts	
Parents	92 (63.45)
Siblings	90 (62.07)

---

[16] In the questionnaire, participants were asked to tick as many (or as little) language(s) of their choice with different people, as listed in Table 5.3(B). However, as the current subsection focuses on the use of English, I will not report the other language choices here.

Relatives	64 (44.14)
Friends of the same ethnicity	108 (74.48)
Friends of different ethnicities	144 (99.31)
Classmates/Colleagues of the same ethnicity	125 (86.21)
Classmates/Colleagues of different ethnicities	145 (100)
Strangers of the same ethnicity	102 (70.34)
Strangers of different ethnicities	145 (100)

5.4.2.2.3. English Proficiency

Table 5.4 shows the distribution of grades/bands obtained by Malaysian participants in their most recent English Language assessment. The three proficiency tests undertaken by the participants were MUET (*n* = 138), IELTS (*n* = 6), and "English 2" (n = *1*). Having matched the band scores against the CEFR levels, most participants fell on the C1 level, indicating proficient or advanced usage in English (CEFR, 2022). A relatively small proportion of participants (11.72%), who scored Band 4 in MUET, were mapped to the CEFR's B2 level indicating independent usage or upper intermediate proficiency. The "English 2" test was a local assessment set by Universiti Teknologi PETRONAS. There was no information about this test, so its grading system could not be mapped against the CEFR metric. Notwithstanding, based on the overall performance in the recent English assessments, one can say that the Malaysian participants were upper intermediate or advanced learners of English.

Table 5.4: Recent English assessments taken by Malaysian participants, in frequency (and %)

Test type	Band/Grade	CEFR	*N* (%)
MUET[a]	6	C1/C2	1 (0.69)
	5	C1	120 (82.76)
	4	B2	17 (11.72)
IELTS	8	C1	1 (0.69)
	7.5	C1	4 (2.76)
	7	C1	1 (0.69)
English 2	B+	?	1 (0.69)
Total			**145 (100)**

[a] Malaysian Examinations Council (2022)

Finally, as all participants were required to undertake the EPT published by the British Council, Figure 5.2 charts the distribution of their EPT scores according to nationality. The Malaysian group saw a wide range of EPT scores between 55% and 95% with an average proficiency score of 79.48% ($SD = 8.27$). Meanwhile, the proficiency range of the UK group was considerably smaller than that of the Malaysian cohort, with the lowest score starting at 83% and the highest landing at 99%; the proficiency scores averaged to 92.83% ($SD = 4.11$).

Zeroing in on the L1 groups for the GJT and NT studies, their EPT scores are displayed in Table 5.5. The L1 group with the highest average score was L1-BritE (93%, $SD = 4.08$), followed by the L1-MalE-Malay group which obtained an average score of 84.76% ($SD = 4.64$). The lowest mean score was found with the L1-Chinese group, which was at 72.51% ($SD = 8.92$). Altogether, the EPT proficiency scores of participants falling under the L1 groups listed in Table 5.5 were later entered as a covariate in the regression analyses of the GJT and NT.

Table 5.5: Mean, standard deviation, and range of EPT scores according to L1 experimental group (for the GJT and NT studies)

Country	L1 group	*M*	*SD*	Range
Malaysia	Malay	81.48	6.83	63 – 95
	Chinese	72.51	8.92	55 – 91
	MalE	82.63	5.81	71 – 93
	MalE-Malay	84.76	4.64	76 – 93
	MalE-Chinese	81.08	6.87	69 – 91
UK	BritE	93.00	4.08	83 – 99

Figure 5.2: Distribution of the British Council English Level Test scores according to nationality
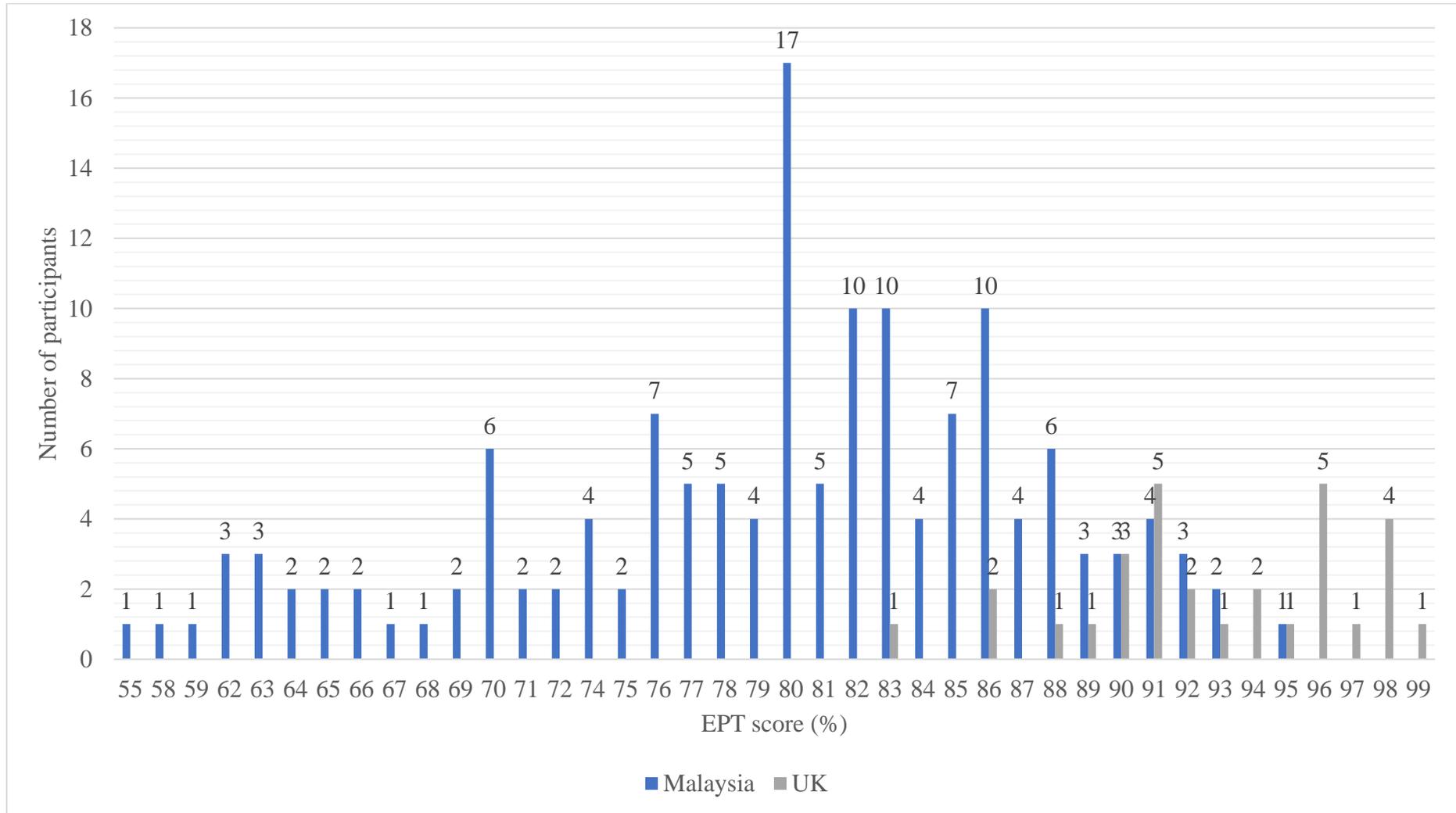
5.4.3. Summary

To sum up the characteristics of participants recruited for the main study:

- Participants were adults (Malaysia average age: 20 years; UK average age: 21 years old) who were mostly studying at university.
- The Malaysian cohort comprised participants who came from various L1 backgrounds. By contrast, participants in the UK group mostly spoke BritE as their only L1.
- Focusing on the Malaysian group:
  - The onset age of acquisition of English (MalE) ranged from 0 to 7 years old. However, almost half of the sample population (42.07%) were already exposed to English since birth.
  - During the time of testing, most Malaysians were exposed to and used English on a regular basis.
  - Based on their recent formal English assessment, the majority of participants (87.59%) were already at advanced proficiency in English, whereas a relatively small number (11.72%) were of upper intermediate levels.

Findings from the standardised EPT showed that the UK group (92.83%) scored markedly higher than the Malaysian group (79.48%).

5.5. Data management and analysis

5.5.1. Transcription

Participants' narratives were recorded using the Sony ICD-UX560F sound recorder, whereas those conducted online were done using the recording tab in Zoom. These audio recordings were all transcribed orthographically in CHAT (MacWhinney, 2000). In CHAT, every utterance was manually linked to its corresponding audio segment using the built-in facility called *Walker Controller*.

Every transcript was checked at least twice with the aid of the attached audio file running in the background. Any typographical errors detected were corrected immediately. After the whole data collection was completed, the transcripts were checked once more to ensure that the participants' narratives were transcribed as accurately as possible.

The CHAT files were duplicated into two folders. The original files are kept unadulterated for future publications. The duplicated files are for error coding and analyses.

With regards to data extractions from the transcripts, they were managed using the Computerized Language Analysis (CLAN) programme in CHAT. Key words, regular expressions, and other features in the transcripts were coded in the CLAN Commands window, which would then be extracted either onto the CLAN output window or an Excel spreadsheet.

5.5.2. Statistical analyses

Data from all the linguistic tasks were cleaned and organised separately in Microsoft Excel. For statistical analyses, the data files were saved in .csv format and analysed in RStudio (R Core Team, 2021). Data visualisations were generated in R and Microsoft Excel. Further details on the statistical tests for the linguistic tasks are provided in their respective Results chapters.

5.6. Ethical considerations

This research was approved by the University's Research Ethics Committee and was carried out in line with their ethical guidelines.

Prospective participants were given softcopies of the participant information sheet and consent form so that they were aware of what the study involved. When reserving an appointment via Doodle, the names they put on the poll were made visible only to me.

At the beginning of the appointment, the participants were briefed on the study and how their data would be anonymised and handled. They were also assured of their right to opt out during the experimental session without having to give any reasons. Having flagged up the possibility of uploading their anonymised narrative sound files on a public corpus, all but three participants gave their permission. The dissent of the three participants is respected and adhered to. Once the participants were happy to begin the study, they were asked to sign the consent forms. The physical copies for my record are kept safely in a locked folder.

Every participant was assigned a unique ID, which was used to label all their data files in order to retain their anonymity. All data gathered for this PhD project are hitherto stored with passcode security both in my laptop and portable hard drive for backup.

Finally, all reimbursements granted to the participants in Malaysia and the UK were funded by the University's Faculty of Modern and Medieval Languages and Linguistics.

6   Results of the grammaticality judgement task

6.1. Introduction

This chapter reports findings of the grammaticality judgement on Standard English (StE) finiteness, with the aim to respond to the following two research questions:

RQ1: To what extent does L1 play a role in the linguistic performance on StE?

The six L1 groups of interest are L1-British English (BritE) from the UK (control) and, from Malaysia, L1-Malay, L1-Chinese, L1-Malaysian English (MalE), L1-MalE-Malay, and L1-MalE-Chinese. Their rating performances in the grammaticality judgement task (GJT) will be examined and compared to one another. The L1 effects may be facilitative (H_1), adverse (H_2), or neutral (H_0), whose hypotheses are posited in the following:

H_1:   If the L1-MalE(+) groups perform better than their L1-Malay and L1-Chinese peers, whether or not their linguistic judgement of StE is markedly divergent from the L1-BritE control's, then having English in the L1 repertoire of Malaysians is argued to serve a facilitative/remedial role.

H_2:   If any of the Malaysian L1 groups diverges considerably from the L1-BritE control, then the plausibility that the group performance is affected by the L1 is not rejected.

H_0:   If all L1 groups perform similarly in their linguistic judgements across the board, then the argument for L1 effects cannot be supported.

The investigation is further carried out by clustering the L1 groups according to the areal-cum-acquisitional status of English, thereby forming three broader categories: L1-BritE, L1-MalE, and L2-MalE (Figure 6.1). The aforementioned hypotheses are retained, except that the L1 groups are now replaced by their corresponding English Status categories.

Figure 6.1: The hierarchical levels of English status and L1 groups

Languages and lectal varieties						
English status	L1-BritE		L1-MalE		L2-MalE	
L1 group	BritE	MalE	MalE-Malay	MalE-Chinese	Malay	Chinese
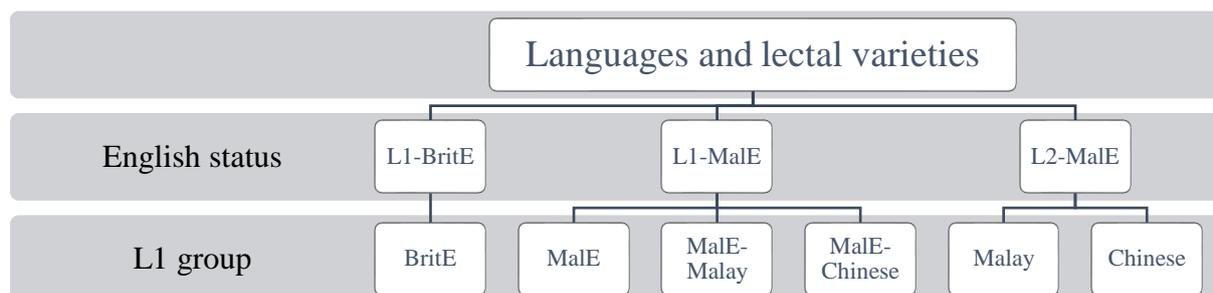
RQ2: What is the contribution of CME on the error types in StE?

As the morphosyntactic system of CME is dynamic and contingent upon several factors such as the effects of L1 or proficiency in StE, it is not straightforward to claim whether a non-target feature is indeed a feature of CME or, for instance, a reflex of L2 behaviour. Notwithstanding, in relation to H_2 above, if it happens that *all* the Malaysian L1 groups diverge from their British counterpart in their linguistic performances, then the pervasiveness of CME shall be entertained.

In what is to follow, the pre-processing and analytical procedures of the GJT data are explained in **§6.2.**. The data are then presented by looking into the predictors of L1 (**§6.3.1.**) and the status of English as an L1 or L2 (**§6.3.2.**). Finally, the key findings are summarised in response to the research inquiries in **§6.4.**.

6.2. Data management and analysis

The finiteness features examined in the GJT were copula BE, progressivity, 3rd.SING.-*s*, regular past tense, and auxiliary DO. The participants ($n = 175$) were asked to rate them on a five-point Likert scale (1 "least grammatical" to 5 "completely grammatical") and were given the option to click on the zero (0) "don't know" key.

In the data pre-processing phase, the rating scores were *z*-transformed per participant as a means to standardise the rating scale – which would have been interpreted differently – across the participants (Langsford et al., 2018; Schütze & Sprouse, 2013). The 0 values ($n = \frac{19}{15,750}$) were treated as missing data as they were not part of the rating scale. Several L1 groups were excluded (see Table 6.1(A)) because the lack of participants forming small group sample sizes ($n \leq 8$) would generate low statistical power, which attenuates the predictability of a true effect and reliability of the analysis, etc. (Button et al., 2013). Additionally, a participant

(MM003 from the L1-Chinese group) was dropped due to his inconsistent use of the scale. Altogether, 157 participants' GJT data were analysed, and the L1 groups being formed are listed in Table 6.1(B).

Table 6.1: L1 groups that were (A) excluded from and (B) included in the GJT analysis

	(A) Excluded		(B) Included	
Nationality	**L1 group**	**N**	**L1 group**	**N**
Malaysia	Tamil	8	Malay	33
	MalE, Kelabit	1	Chinese	42
	MalE, Tamil	4	MalE	19
	MalE, Malay, Iban	1	MalE, Malay	21
	MalE, Malay, Tamil	1	MalE, Chinese	13
	MalE, Tamil, Telugu	1		
UK	BritE, Gujarati	1	BritE	29
Total		**17**		**157**

General linear mixed-effects regression models were fitted using the `lmer()` function of the "lme4" R package (Bates et al., 2015). Their overall ANOVA (Type III Wald χ^2) results were procured from the `Anova()` function of the "car" package (Fox & Weisberg, 2019). The "emmeans" package was used to run Tukey-adjusted post hoc pairwise comparisons using `emmeans()`, as well as to generate plots of estimated marginal means with `emmip()` (Lenth, 2022). Regarding the presentation of statistical analyses, the results of the Type III Wald χ^2 tests and the plots of estimated marginal means are displayed in this chapter. Due to space limit, the complete summaries of the mixed model outputs and post hoc test results are attached in **Appendix C**.

With regards to model fitting, the rating responses (*z*-transformed) constituted the dependent variable. The fixed-effects structure varied depending on what was being examined. Specifically, in the investigation of L1 effects, L1 Group (levels: *BritE*, *Malay*, *Chinese*, *MalE*, *MalE-Malay*, *MalE-Chinese*) was entered as a predictor, whereas in the investigation of the role of English as an L1 or L2, English Status (levels: *L1-BritE*, *L1-MalE*, *L2-MalE*) was used instead. Both L1 Group and English Status were not computed into the fixed-effects structure at the same time to avoid multicollinearity issues, since L1 Group was subsumed under English

Status (Figure 6.1). The fixed effects retained across the models were Feature (levels: *copula*, *progressive*, DO, 3rd.SING.-s, *regular past*) and Grammaticality Type (levels: *standard*, *omission*, *commission*); their interactions with each other and with L1 Group/English Status were also computed therein. In addition, the covariate Proficiency (centred) was entered as a control variable. Concerning the random-effects structure, only (sentence) Items were included as random intercepts; Participants had to be removed because its inclusion gave rise to singularity issues in the model. The R codes of the regression models were roughly as follows:

- **For L1 effects**: …lmer(Judgement z-score ~ L1 Group * Feature * Grammaticality Type + Proficiency + (1|Items)…)
- **For English status effects**: …lmer(Judgement z-score ~ English Status * Feature Grammaticality Type + Proficiency + (1|Items)…)

Having established the models, their statistical results are reported in the subsequent sections.

6.3. The findings

6.3.1. L1 effects

Table 6.2 presents the results (Type III Wald χ^2) of the model fitted for L1 effects. The simple main effects of L1 Group ($\chi^2(5) = 63.65$, $p < 0.001$) and Grammaticality Type ($\chi^2(2) = 938.81$, $p < 0.001$) are statistically significant, as are the interaction terms of L1 Group × Feature ($\chi^2(20) = 68.61$, $p < 0.001$), L1 Group × Grammaticality Type ($\chi^2(10) = 78.5$, $p < 0.001$), and L1 Group × Feature × Grammaticality Type ($\chi^2(30) = 172.56$, $p < 0.001$).

Table 6.2: Results of the analysis of deviance table (Type III Wald χ^2 tests) with L1 Group being one of the predictors in the GJT analysis

	χ^2	DF	Pr ($>\chi^2$)	
(Intercept)	401.36	1	<0.001	***
L1 Group	63.65	5	<0.001	***
Feature	2.58	4	0.630	
Grammaticality Type	938.81	2	<0.001	***
Proficiency	0.01	1	0.932	

L1 Group × Feature	68.61	20	<0.001	***
L1 Group × Grammaticality Type	78.50	10	<0.001	***
Feature × Grammaticality Type	11.05	6	0.087	.
L1 Group × Feature × Grammaticality Type	172.56	30	<0.001	***

The estimated marginal means of the model are visualised in Figure 6.2. As anticipated, the standard constructions of the finiteness features were given markedly higher ratings than the ungrammatical constructions of omission and commission. Between omission and commission, the former condition received lower estimated ratings than the latter. Furthermore, the ratings against omission errors were fairly consistent across the board, whereas the ratings against commission errors, particularly in auxiliary DO and progressive BE, revealed fluctuating trends across the L1 groups.

Figure 6.2: Plot of estimated marginal means of the L1 groups according to feature and grammaticality type in the GJT analysis
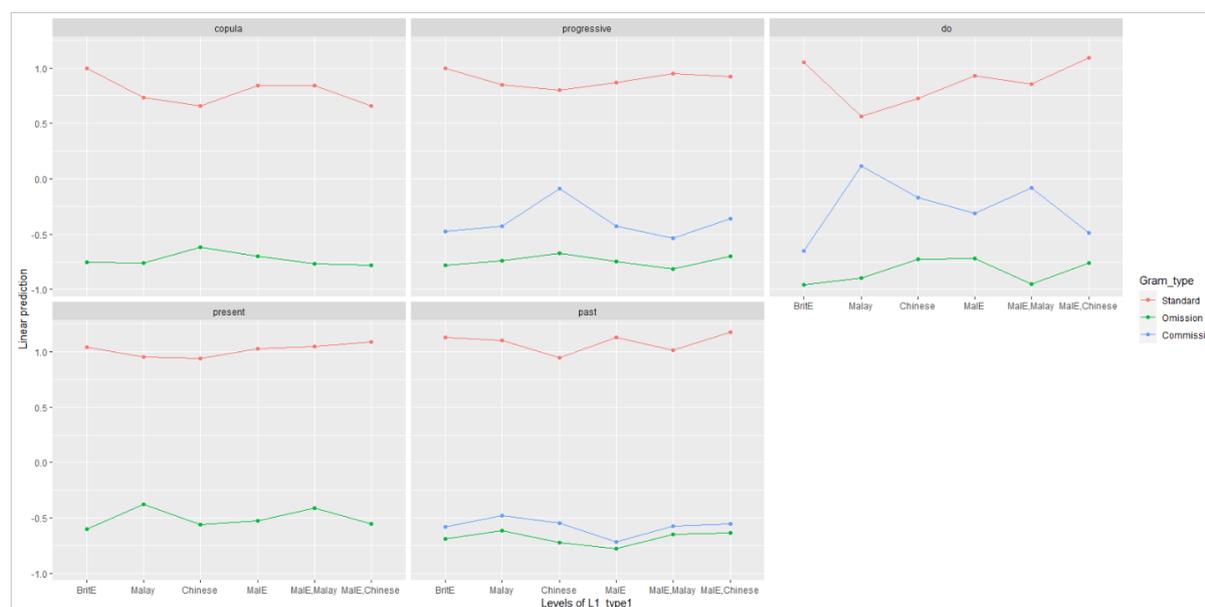
Comparing between the Malaysian L1 groups and the L1-British control group, the post hoc test revealed that the L1-Chinese group exhibited the most salient divergence in its rating performance across several grammaticality types of finiteness features. For instance, it behaved significantly differently from the L1-BritE control in that it assigned markedly higher ratings to the omission errors of copula BE ($\beta = 0.137$, $z = 2.86$, $p = 0.049$) and auxiliary DO ($\beta = 0.23$, $z = 3.52$, $p = 0.006$), as well as to the commission error of stative progressivity ($\beta = 0.385$, $z =$

5.89, $p < 0.001$). Moreover, its acceptance of the standard construction of progressive BE was significantly lower than that of L1-BritE ($\beta = -0.498$, $z = -3.02$, $p = 0.03$).

In comparison, the L1-Malay group performed comparatively better than its L1-Chinese counterpart. A couple of exceptions were its ratings of double-tense marking in auxiliary DO and omission of 3rd.SING.-s, which were significantly higher not only than those of L1-Chinese ($p < 0.05$) but also those of the L1-BritE control ($p < 0.05$).

The L1-MalE(+) cohorts generally performed better than their L1-Chinese and L1-Malay peers as they were being compared to their L1-BritE counterpart. Interestingly, a conspicuous contrast was observed between L1-MalE and L1-BritE in their judgements of the omission of auxiliary DO, viz. the ratings of the former group were significantly higher than the ratings of the latter ($\beta = 0.241$, $z = 3.09$, $p = 0.025$). Apart from that, the judgements of the L1-MalE(+) groups were rather similar to the L1-BritE control.

Notwithstanding, there were a couple of finiteness features that posed patent learnability issues to the Malaysian L1 groups, including those who acquired English as an L1. One of them concerned the acceptance of the standard construction of copula BE, which received significantly lower ratings from the Malaysian L1 groups in contrast to their L1-BritE cohort ($p < 0.05$). Another finiteness feature that witnessed a divergent performance between the two nationalities was auxiliary DO, especially when it concerned the standard versus commission (i.e., double-tense marking) conditions. In contrast to the L1-BritE group, the L1-Malay and L1-Chinese groups consistently rejected the standard condition in favour of double-tense marking ($p < 0.001$). While the L1-MalE(+) groups gave relatively high ratings to the standard condition, the L1-MalE and L1-MalE-Malay groups were considerably more accepting of double-tense marking, thereby yielding a significant contrast to the L1-BritE group, respectively ($p < 0.001$). Only the L1-MalE-Chinese group performed most similarly to its L1-BritE counterpart in accepting the standard construction ($\beta = 0.043$, $z = 0.49$, $p = 0.997$) and rejecting the commission condition ($\beta = 0.167$, $z = 1.89$, $p = 0.407$).

6.3.2. English status effects

To investigate whether acquiring English as an L1 or L2 had an effect in the rating performances of StE, Table 6.3 shows that the simple main effect of English Status ($\chi^2(2) = 53.88$, $p < 0.001$) was significant, so were Grammaticality Type ($\chi^2(2) = 935.12$, $p < 0.001$) and the interaction terms of English Status × Feature ($\chi^2(8) = 31.3$, $p = 0.001$), English Status

$\times$ Grammaticality Type ($\chi^2(4) = 56.2$, $p < 0.001$), and English Status $\times$ Feature $\times$ Grammaticality Type ($\chi^2(12) = 59.86$, $p < 0.001$).

Table 6.3: Results of the analysis of deviance table (Type III Wald χ^2 tests) with English Status being one of the predictors in the GJT analysis

	χ^2	DF	Pr ($>\chi^2$)	
(Intercept)	402.41	1	<0.001	***
English Status	53.88	2	<0.001	***
Feature	2.57	4	0.632	
Grammaticality Type	935.12	2	<0.001	***
Proficiency	0.00	1	0.947	
English Status $\times$ Feature	31.30	8	0.001	***
English Status $\times$ Grammaticality Type	56.20	4	<0.001	***
Feature $\times$ Grammaticality Type	11.01	6	0.088	.
English Status $\times$ Feature $\times$ Grammaticality Type	59.86	12	<0.001	***

Figure 6.3 illustrates the plot of estimated marginal means of the model. The standard constructions of the finiteness features witnessed a general downward slope whereas the ungrammatical constructions of omission and commission observed a rising trend, all of which were due to the status of English. As observed in the previous section, the post hoc test confirmed a significant divergence between the Malaysian (i.e., L1- and L2-MalE) and British (i.e., L1-BritE) cohorts concerning their ratings of the commission condition of auxiliary DO ($p < 0.001$) and the standard condition of copula BE ($p < 0.001$). Apart from that, the L1-MalE group did not diverge significantly from its L1-BritE counterpart in the rest of the finiteness features. On the other hand, a significant contrast was observed between L2-MalE and L1-BritE in their ratings of the standard ($\beta = -0.178$, $z = -3.02$, $p = 0.007$) and commission conditions ($\beta = 0.236$, $z = 4.01$, $p < 0.001$) of progressive BE, as well as in the omission condition of auxiliary DO ($\beta = 0.156$, $z = 2.65$, $p = 0.022$). From the previous subsection, these were due to the L1-Chinese group who diverged the most in their judgements of these features. Regarding the standard construction of auxiliary DO, the significant difference between L2-MalE and L1-BritE ($\beta = -0.397$, $z = 6.73$, $p < 0.001$) stemmed from the assignment of markedly low ratings by both L1-Chinese and L1-Malay groups in contrast to the control group.

Figure 6.3: Plot of estimated marginal means of the English Status groups according to feature and grammaticality type in the GJT
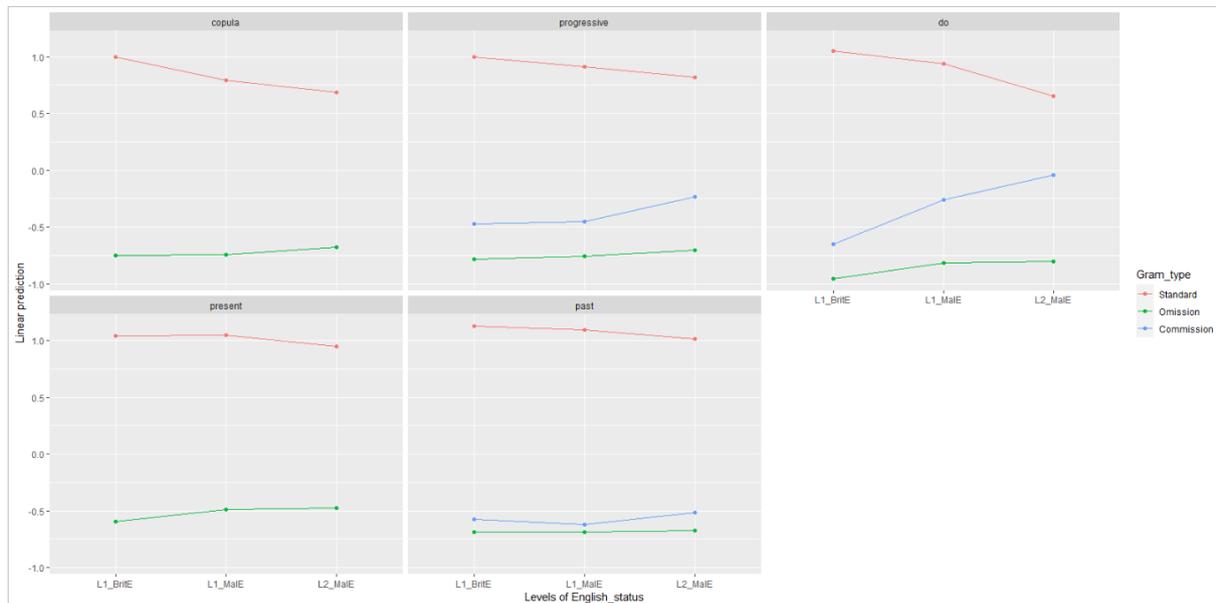
6.4. Summary of findings

L1 effects:

- In response to RQ1, we have seen that L1 was a prominent predictor of the rating responses, particularly coming from the L1-Chinese group (e.g., lower ratings in the standard construction of progressive BE, higher ratings in the omissions of copula BE and auxiliary DO, and higher ratings in the commission of progressive BE).

- Apart from the L1-Malay group's higher acceptance of the double-tense marking in auxiliary DO and omission of 3rd.SING.-*s*, it generally outperformed its L1-Chinese counterpart.

English status effects:

- Having English in the L1 repertoire of Malaysians seemed to facilitate performance in the GJT in comparison to acquiring it as an L2.

- However, the significantly higher acceptance of DO omission by the L1-MalE group as compared to its L1-BritE counterpart hints at the plausible influence from CME.

Vulnerable features:

- The syntactic construction of auxiliary DO pertaining to tense marking unearthed a vulnerable site for most Malaysian L1 groups, with L1-MalE-Chinese being the exception. For one thing, a strong preference for double-tense marking and consistent rejection of the standard construction was observed with the L1-Malay and L1-Chinese groups. For another, this non-target feature was also highly accepted by the L1-MalE and L1-MalE-Malay peers. Ergo, in response to RQ2, the prevalence of double-tense marking of DO across most of the Malaysian L1 groups, regardless of whether English was acquired as an L1 or L2, seems to allude to the pervasiveness of CME in the linguistic performance on StE, since double-tense marking has been attested in the colloquial variety (Pillai & Greig, 2020).

- Concerning the standard construction of copula BE, there was a clear distinction in the linguistic judgements between the L1-BritE group and the Malaysian L1 groups. While the ratings of the former group were generally high, those given by the latter cohort were significantly lower. As this divergent performance was observed across all Malaysian groups, this cannot be due to L1 effects. We will discuss this phenomenon, as well as the other key findings above, in more detail in **Chapter 9**.

7 Results of the narrative task

7.1. Introduction

This chapter presents the microstructural analysis of the narrative data, with a focus on finiteness marking. The research questions of the narrative task (NT) are the same as those stated in the previous chapter; they are:

RQ1: To what extent does L1 play a role in the linguistic performance on StE?

RQ2: What is the contribution of CME on the error types in StE?

RQ3: Are there task effects on finiteness?

As narratives generate semi-spontaneous production data, this chapter takes on an exploratory approach in investigating the production of finiteness features. The rest of the chapter is organised as follows. **§7.2.** recounts how the narratives were coded, filtered, and analysed. **§7.3.** reports findings of the narrative data. Here, the linguistic productivity and complexity of the narratives are briefly described (**§7.3.1.**). However, heavier emphasis is placed on findings pertaining to the use of finiteness (**§7.3.2.**) and other verbal errors not covered by the main features of interest (**§7.3.3.**). The key findings on finiteness are then summarised in **§7.3.4.**. Having conducted the analyses of the GJT (from the previous chapter) and NT, this chapter then addresses RQ3 by investigating potential task-based effects between GJT and NT in **§7.4.** before wrapping up in **§7.5.**.

7.2. Data management and analysis

7.2.1. Coding procedures

The narratives transcribed in CHAT (MacWhinney, 2000) were all duplicated in a new folder dedicated for manual error coding. Error codes targeting the verbal elements of interest followed the conventions laid out in the CHAT manual. To demonstrate, the uninflected verb

"look" produced by Participant MM125 in (44) was first coded with its target replacement [:looked], followed by the morphological error code specified for regular past tense marking [*m:0ed]:

(44) *MM125: he look [: looked] [* m:0ed] at that old lady.

The next round of coding was done in Microsoft Excel. To do this, all verb stems supplied by the participants were identified, extracted, and exported to Excel using the following code in the CLAN programme (MacWhinney, 2000):

```
kwal +t*M* +sm;*,|v,o% -s"[+ exc]" *.cha +d40
```

Component	Description
kwal	‣ Runs searches for linguistic elements at the speaker or morphological tier
+t*M*	‣ Selects speakers whose IDs start with M
+sm;*,|v,o%	‣ Selects only verb stems
-s"[+ exc]"	‣ Excludes utterances marked with the postcode [+ exc]
*.cha	‣ Selects every CHAT file in the working directory
+d40	‣ Exports output to Microsoft Excel

(CLAN manual, pp. 90-92; MacWhinney, 2000)

The output in the Excel spreadsheet contained, not exhaustively, the participant unique IDs and, crucially, the verb stems alongside their speaker and morphological tiers. From there, the spreadsheet was reorganised by manually entering the relevant demographic and linguistic backgrounds of the participants. Utterances from the speaker tier were coded manually for their verb type, feature type, verb regularity (applies to both auxiliary and lexical verbs), binary grammaticality type, specified grammaticality type, and the preceding context of the verbal element (Figure 7.1) – all of these were coded in their respective columns. For instance, if an utterance (in T-unit) contained five verbal elements, as underlined in (45), then four additional rows would be added so as to code all five verbal elements, respectively:

Figure 7.1. Components and levels in the verb coding scheme

Note: White textbox indicates code type; yellow textbox indicates features of interest; grey textbox indicates features not analysed in the NT study.
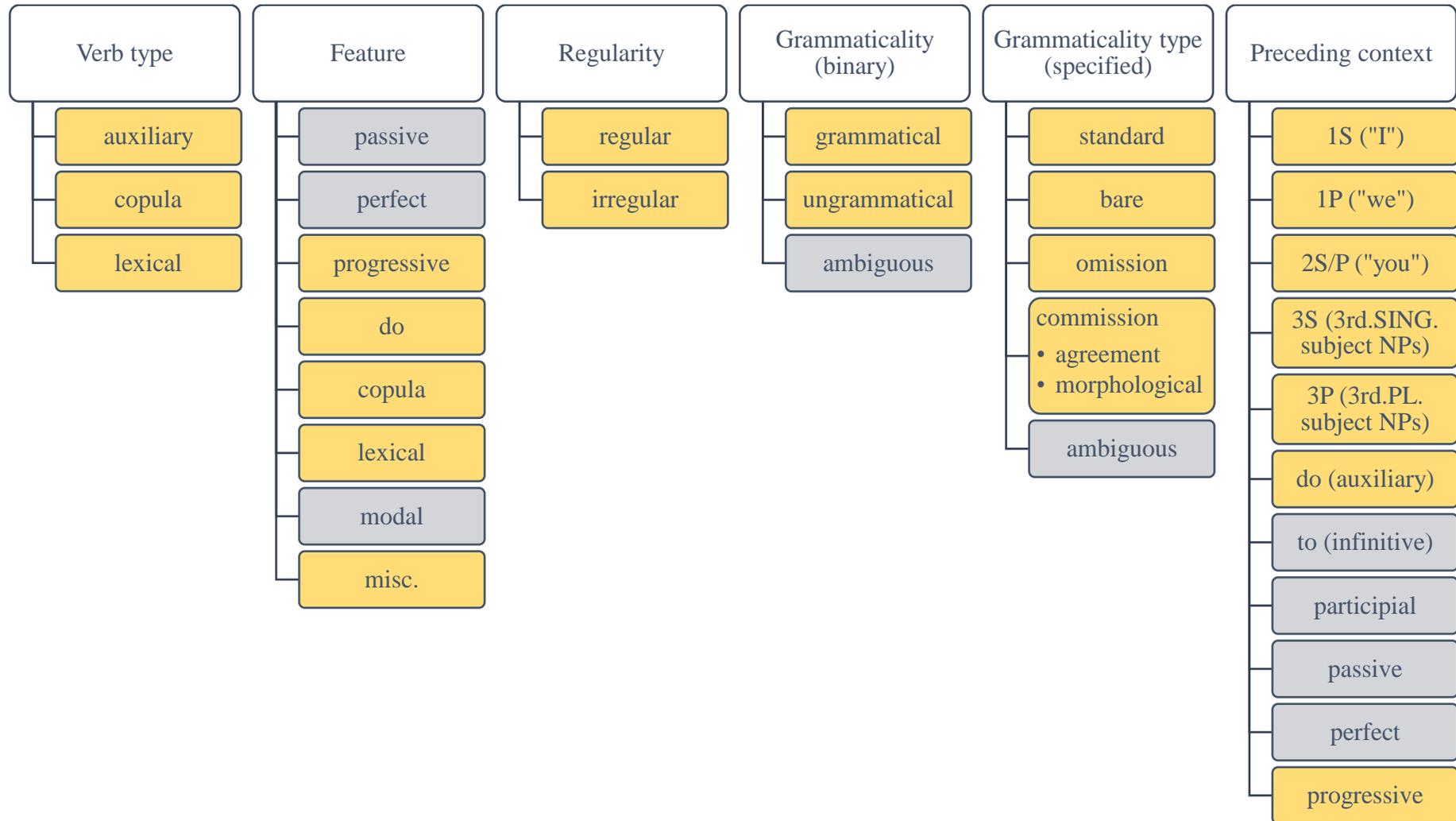
(45) *MM034: and then <he found out> [//] she found out that he was
     actually listening to some soft music instead of &-like not [/]
     not the rock music that she imagine [: imagined] [* m:0ed] it to
     be.

Although error coding was done twice (i.e., first time in CHAT; second time in Excel), this practice served me well for two main reasons. Firstly, any erroneous instances that had been bracketed by the error codes were easily identified during verb coding in the spreadsheet. Secondly, the error codes served as a backup option in case any specific error type needed to be retrieved for further analysis, which could be done instantaneously via CLAN.

There were altogether 9,696 utterances in the NT corpus. That said, excluded from the quantitative analysis were utterances marked with the [+ exc] postcode in CHAT ($n = 991$); these were single-(non)word utterances, frozen expressions (e.g., "thank you"; "I'm sorry"), self-repetitions, echo questions, simple clauses containing the quotative BE LIKE, and incomplete or elided utterances without the main verb. Only utterances containing at least a subject NP and a main verb, even if some were incomplete without their complements, were coded and included in the analysis.

7.2.2. Data pre-processing

RStudio (RStudio Team, 2022) was used to sort out the relevant participants and instances containing the finiteness features of interest before carrying out the quantitative analyses of the narrative data. Corresponding to the finiteness features investigated in the GJT, only instances of copula BE, progressivity, DO-support, and lexical verbs – as shaded in yellow in Figure 7.1 above – were kept for the analysis. Suppletive morphemes labelled with "misc." (miscellaneous) were separately examined because they did not fit into any of the features above. Grammatically ambiguous cases were removed from the quantification because they involved tense suffixations that could not be disambiguated due to their phonological assimilation with the onset of the proceeding word (e.g., belong**ed t**o; see**s s**ome). The L1 groups excluded from the GJT analysis were also excluded here due to their small sample sizes. That said, Participant MM003 from the L1-Chinese group, who was dropped from the GJT analysis, was included herein. Altogether, the narrative data of 158 participants (Table 7.1) were analysed.

Table 7.1: L1 groups that were (A) excluded from and (B) included in the NT analysis

	(A) Excluded		(B) Included	
Nationality	L1 group	*N*	L1 group	*N*
Malaysia	Tamil	8	Malay	33
	MalE, Kelabit	1	Chinese	43
	MalE, Tamil	4	MalE	19
	MalE, Malay, Iban	1	MalE, Malay	21
	MalE, Malay, Tamil	1	MalE, Chinese	13
	MalE, Tamil, Telugu	1		
UK	BritE, Gujarati	1	BritE	29
Total		**17**		**158**

7.2.3. Statistical tests

The statistical tests employed in the narrative analysis were simple linear regressions for the overall measures of linguistic productivity and complexity (**§7.3.1.**), binomial generalised linear mixed models for the examination of finiteness marking (**§7.3.2.**), and correlation analyses for the investigation of task effects (**§7.4.**). The implementations of these tests are further described in their respective sections.

7.3. Findings on finiteness

7.3.1. Productivity and complexity

Following Justice et al. (2006), the general microstructural properties of the narratives are described in terms of their linguistic productivity (i.e., *total numbers of T-units* [independent clauses, which may also include their embedded clauses] *and word tokens; lexical diversity* [type-token ratio]) and complexity (i.e., *mean length of utterance in morphemes; syntactic complexity* [number of verbs per T-unit]). The components bracketed under these two parameters were all automatically generated from the pre-error-coded CHAT transcripts using the following command in CLAN (MacWhinney, 2000):

```
eval +t*M* +u *.cha
```

Component	Description
eval	‣ Calculates the narrative microstructural properties of interlocutors
+t*M*	‣ Selects speakers whose IDs start with M
+u	‣ Combines a series of searches into one file
*.cha	‣ Selects every CHAT file in the working directory

(CLAN manual, pp. 131-136; MacWhinney, 2000)

The components constituted the dependent variables and were entered into separate simple regression models using the `lm()` function ("stats" R package; R Core Team, 2021). For all the models, the predictors were L1 Group and Proficiency (centred). Tukey-adjusted post hoc pairwise comparisons were carried out via `emmeans()` ("emmeans" R package; Lenth, 2022).

Figure 7.2 illustrates the distribution of the measures of productivity. Results from the models revealed that L1 Group was a significant predictor for Total T-units ($F(5) = 2.29$, $p = 0.049$) and Lexical Diversity ($F(5) = 3.08$, $p = 0.011$), but not for Total Words (in tokens) ($F(5) = 1.23$, $p = 0.297$). With regard to Total T-units, the post hoc results revealed that the L1-BritE group produced a marginally lower number of T-units than the respective groups of L1-Malay ($\beta = -17.4$, $t = -2.733$, $p = 0.075$) and L1-MalE ($\beta = -19.37$, $t = -2.752$, $p = 0.071$). With regard to lexical diversity, the post hoc test yielded a significant difference between L1-BritE and L1-Malay ($\beta = 0.041$, $t = 3.28$, $p = 0.016$), as the former group produced a considerably higher degree of lexical or vocabulary variation than the latter group.

Figure 7.2: Linguistic productivity of the narrative microstructure according to (A) the total number of T-units, (B) the total number of word tokens, and (C) lexical diversity measured in type-token ratio
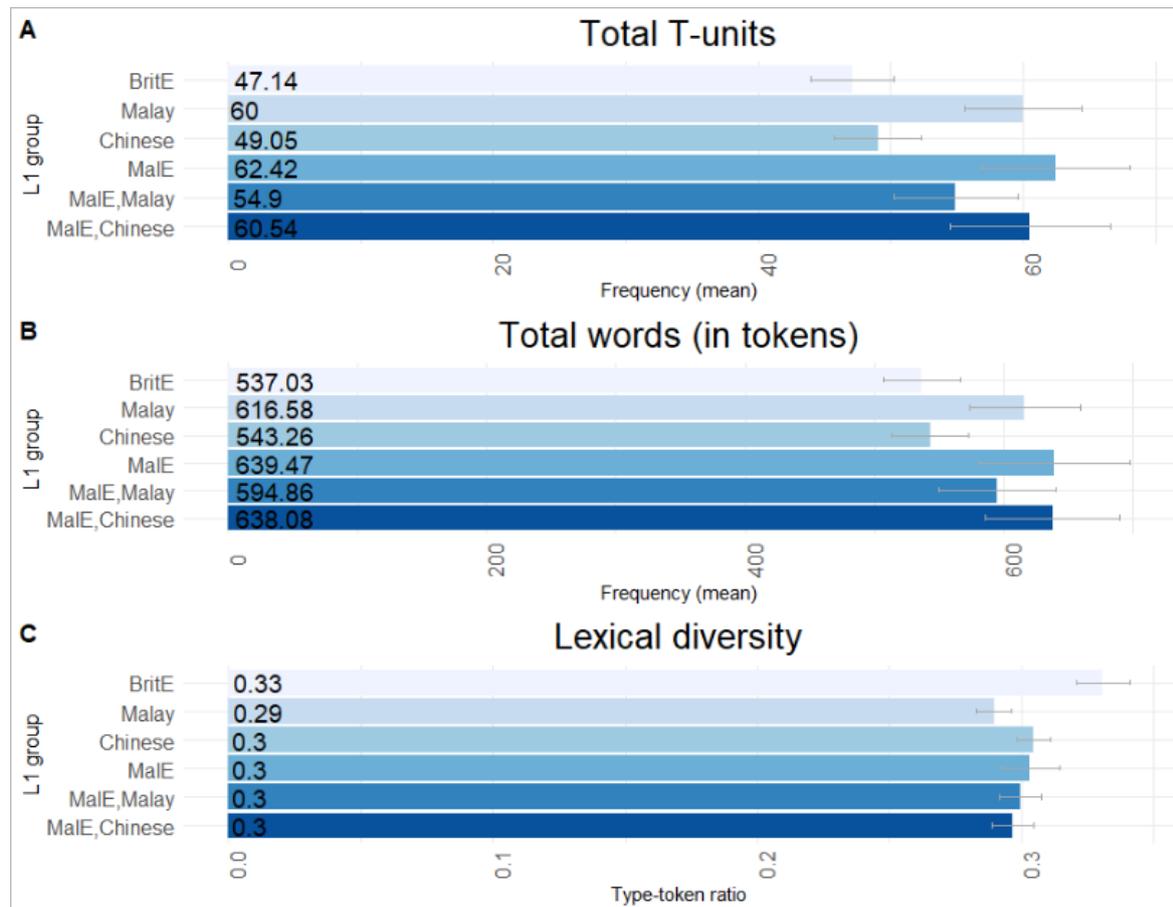
Turning to the measures of complexity (Figure 7.3), results of the regression models showed a significant main effect of L1 Group on MLU (in morphemes) ($F(5) = 3.06, p = 0.012$) and Syntactic Complexity (i.e., verbs per utterance) ($F(5) = 4.26, p = 0.0012$). Results from the post hoc pairwise comparison revealed that there were no significant differences between the L1 groups in the production of MLU but a significant difference between L1-BritE and L1-Malay in Syntactic Complexity ($\beta = -0.273, t = 3.125, p = 0.026$), indicating that the T-units produced by the former group were substantially more complex than the latter.

Figure 7.3: Linguistic complexity of the narrative microstructure according to (A) the mean length of utterance in morphemes and (B) syntactic complexity measured in number of verbs per T-unit
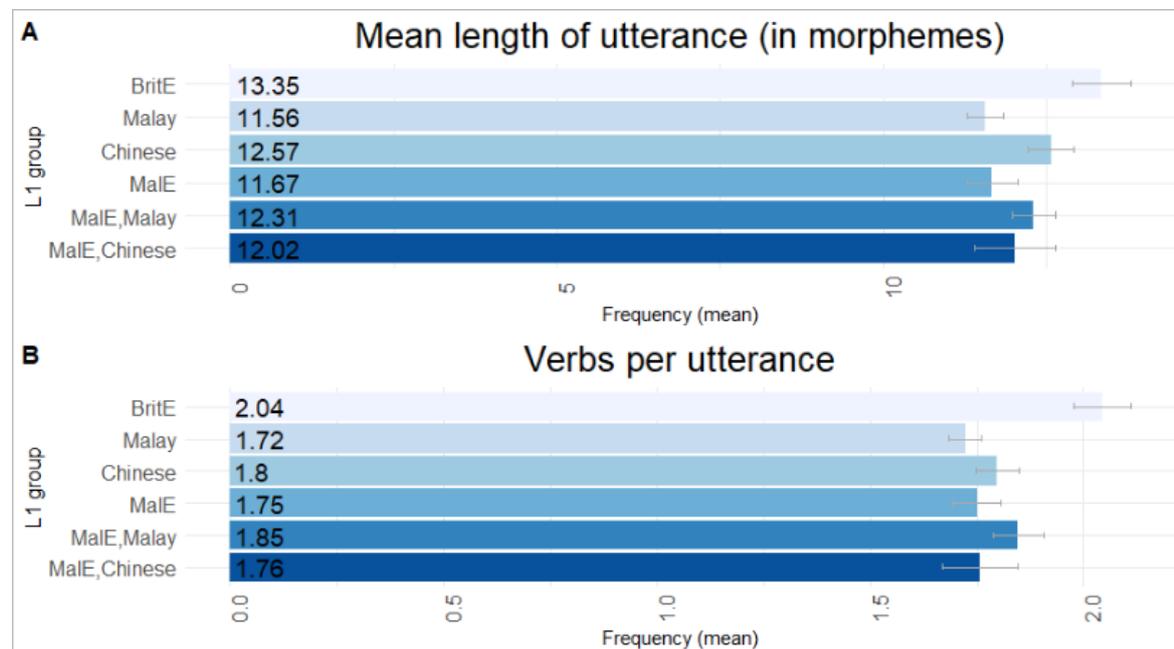
In all the regression analyses on productivity and complexity, Proficiency was not a significant predictor. Only for Total Words (in tokens) was this predictor marginally significant ($F(1) = 3.47$, $p = 0.064$). This suggests that higher proficiency in English was not necessarily associated with longer narratives in terms of the total number of utterances or vocabulary items, or of complex grammatical structures in terms of the number of morphemes per utterance.

7.3.2. Finiteness marking

7.3.2.1. L1 effects

The analysis of finiteness marking was conducted using binomial generalised linear mixed models. The analysis employed the "lme4" R package for model fitting [function: `glmer()`] (Bates et al., 2015), the "car" package for Type III Wald χ^2 test results [function: `Anova()`] (Fox & Weisberg, 2019), and the "emmeans" package for post hoc pairwise comparisons [function: `emmeans()`]. The construction of the model comprised the binary response variable with the "grammatical" and "ungrammatical" categories, L1 Group and Proficiency as predictors in the fixed-effects structure, as well as Participants and Verb Stems as random

intercepts.[17] Feature (levels: *copula, progressive, do, tenses*) was not included as a fixed effect in the current model because there were L1 groups that did not supply certain response types in certain features; including this predictor led to perfect separation in the model, thereby giving rise to convergence failures (Eager & Roy, 2017). The structure of the R code looked like the following: `…glmer(Grammaticality ~ L1 Group + Proficiency + (1|Participants) + (1|Items)…)`

According to the overall results (Type III Wald test χ^2) of the model, there were significant simple main effects of L1 Group ($\chi^2(5) = 65.3$, $p < 0.001$) and Proficiency ($\chi^2(1) = 21.3$, $p < 0.001$). Addressing the effect of proficiency first, Table 7.2 showed that the likelihood of producing ungrammatical finiteness decreased significantly as proficiency in English increased ($\beta = -0.069$, $z = -4.62$, $p < 0.001$).

Table 7.2: Summary of the binomial model output with L1 Group as one of the main predictors in the narrative task

	Estimate	**SE**	**z value**	**Pr(>|z|)**	
(Intercept)	-5.196	0.457	-11.370	<0.001	***
L1 Group (Malay)	3.014	0.490	6.160	<0.001	***
L1 Group (Chinese)	3.807	0.538	7.080	<0.001	***
L1 Group (MalE)	1.501	0.533	2.820	0.005	**
L1 Group (MalE-Malay)	2.310	0.511	4.520	<0.001	***
L1 Group (MalE-Chinese)	2.464	0.565	4.360	<0.001	***
Proficiency	-0.069	0.015	-4.620	<0.001	***

With regards to L1 effects, results from the Tukey-adjusted post hoc pairwise test (Table 7.3) showed that the differences between the L1-BritE control and most of the Malaysian L1 groups (i.e., Malay, Chinese, MalE-Malay, MalE-Chinese) were statistically significant ($p < 0.001$), indicating that the latter cohorts were much more likely to produce finiteness errors than the control group. The likelihood of L1-MalE supplying more

---

[17] A multinomial regression model was attempted using the `mblogit()` function of the "mclogit" R package (Elff, 2022). The fixed- and random-effects structures were the same as those in the binomial model. The main difference, apart from the R packages used, was that the outcome variable was trichotomous: *Standard*; *Omission*; *Commission*. However, the multinomial regression analysis was abandoned due to nonconvergence after 100 iterations. Therefore, this thesis settled on binomial regression for the overall analysis of the narrative data.

ungrammatical instances was slightly higher than L1-BritE, yielding a marginally significant difference ($\beta = 1.5$, $z = 2.82$, $p = 0.055$). Turning to within-nationality comparisons, the L1-Chinese group was significantly more likely to produce finiteness errors than all the L1-MalE(+) groups ($p < 0.05$). Meanwhile, the L1-Malay group only diverged significantly from the L1-MalE group ($\beta = 1.51$, $z = 3.93$, $p = 0.001$), but not from the L1-MalE-Malay and L1-MalE-Chinese cohorts.

Table 7.3: Results of the post hoc comparisons between L1 groups from the narrative task

Contrast	Estimate	SE	z ratio	p value
BritE - Malay	-3.01	0.49	-6.16	<0.001
BritE - Chinese	-3.81	0.538	-7.08	<0.001
BritE - MalE	-1.5	0.533	-2.82	0.055
BritE - MalE,Malay	-2.31	0.511	-4.52	<0.001
BritE - MalE,Chinese	-2.46	0.565	-4.36	<0.001
Malay - Chinese	-0.79	0.31	-2.55	0.109
Malay - MalE	1.51	0.385	3.93	0.001
Malay - MalE,Malay	0.7	0.366	1.92	0.387
Malay - MalE,Chinese	0.55	0.412	1.34	0.765
Chinese - MalE	2.31	0.396	5.82	<0.001
Chinese - MalE,Malay	1.5	0.387	3.86	0.002
Chinese - MalE,Chinese	1.34	0.412	3.26	0.014
MalE - MalE,Malay	-0.81	0.432	-1.87	0.42
MalE - MalE,Chinese	-0.96	0.474	-2.03	0.325
MalE,Malay - MalE,Chinese	-0.15	0.459	-0.34	0.999

As we now delve into the descriptive analysis of narrative performance, Figure 7.4 presents the proportions of the trichotomous (un)grammatical responses (i.e., *standard*, *omission*, *commission*) of finiteness produced by the L1 groups. The omission category includes the omissions of copula BE, progressive BE and -*ing*, auxiliary DO, and tense inflections. The commission category encompasses subject-verb agreement errors, morphological errors, and overgeneralisations (e.g., double-tense marking in DO, overuse of -*ing*). In Figure 7.4, the proportions of ungrammatical finiteness produced by the L1 groups were generally low. Between the omission and commission errors, the rates of the former type

were considerably higher than the rates of the latter type. Amongst the L1 groups, the L1-British control exhibited the highest proportion of grammatical finiteness (99.46%). Contrariwise, the L1-Chinese group produced the most errors (21.98%). This was followed by the L1-Malay group, although its suppliance of errors was 11.42% less than that of the L1-Chinese group. Meanwhile, the L1-MalE(+) groups all produced grammatical finiteness above 90% of the time, with L1-MalE exhibiting the highest rate (95.88%), and L1-MalE-Chinese exhibiting the lowest amongst the three groups (93.2%). Notwithstanding, their differences were negligible.

Figure 7.4: Overall suppliance of (un)grammatical finiteness (%) according to L1 group in the narrative task
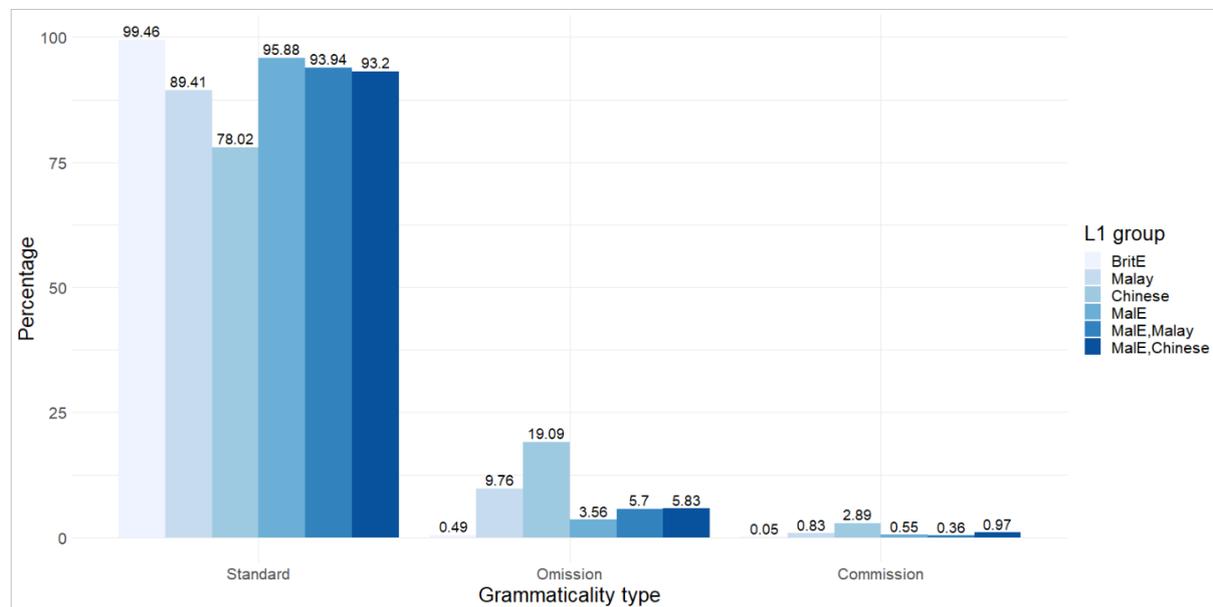
Figure 7.5 illustrates the proportions of the (un)grammatical features supplied by the L1 groups, and Tables 7.4(A-D) detail the frequencies and percentages not indicated in the bar graphs. But before delving into the descriptive statistics, a few notes about the presentation of the data should be mentioned. Firstly, regarding the tabulated data in Table 7.4, the omission and commission errors of some finiteness features were further specified into subcategories, such as the omissions of progressive BE and -ing in Table 7.4(B) which were not captured in the corresponding graph in Figure 7.5(B). Secondly, in relation to the first point, unless their distinctions were important (i.e., see auxiliary DO in Table 7.4(C) and Figure 7.5(C)), the specified ungrammaticality types in the tables were collapsed into the broader categories of omission and commission in Figure 7.5. This is because the low frequencies of some

subcategories could not be captured visibly in the graphs. Thirdly, as subject-verb agreement is related to the syntactic category of Tense, agreement errors on the copula and auxiliaries BE and DO were not quantified in their respective categories but subsumed under the Tenses category. Finally, as mentioned earlier, due to the absence of certain ungrammatical response types in some L1 groups, regression analyses could not be conducted and thus the data to be presented could only be interpreted descriptively.

Figure 7.5: (Un)grammatical instances (%) produced by the L1 groups across the finiteness features of (A) copula BE, (B) progressivity, (C) auxiliary DO, and (D) tenses
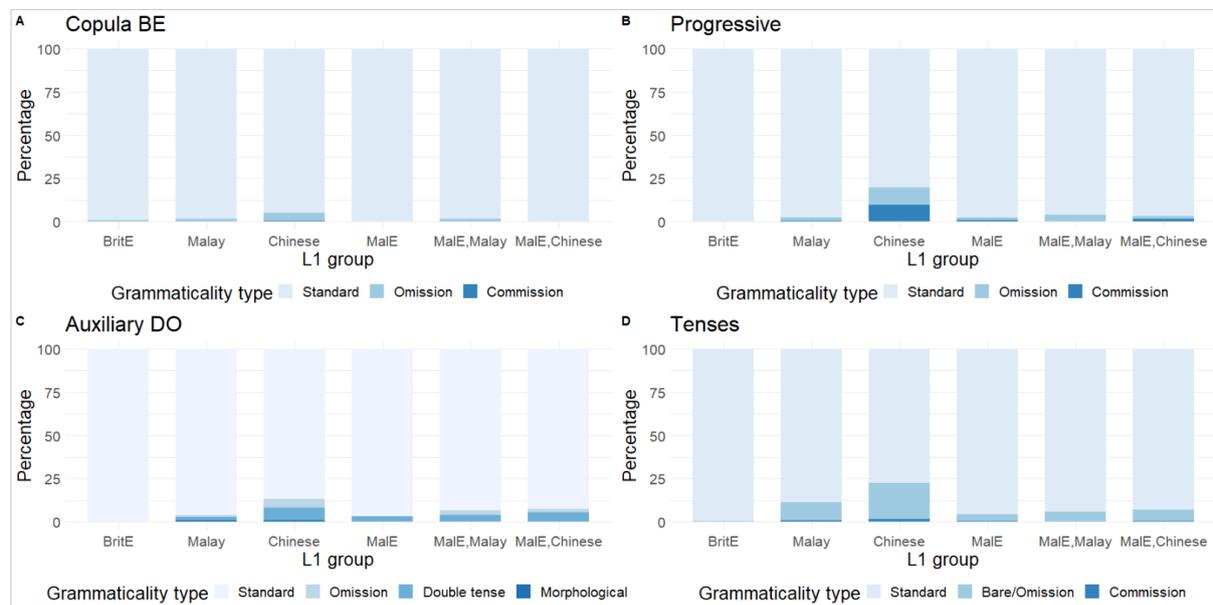
Table 7.4: (Un)grammatical instances (%) produced by the L1 groups across the finiteness features of (A) copula BE, (B) progressivity, (C) auxiliary DO, and (D) tenses

(A): Copula BE

	Grammatical	Ungrammatical		Total (100)
	Standard	Omission	Commission *Morphological*	
BritE	449 (99.30)	3 (0.66)	-	452
Malay	685 (98.40)	11 (1.58)	-	696
Chinese	649 (95.00)	32 (4.69)	2 (0.29)	683
MalE	450 (100.00)	-	-	450
MalE-Malay	434 (98.20)	8 (1.81)	-	442
MalE-Chinese	290 (100.00)	-	-	290

(B): Progressivity

	Grammatical	Ungrammatical			Total (100)
	Standard	**Omission**		**Commission**	
		BE	*-ing*	*Morphological*	
BritE	178 (100.00)	-	-	-	178
Malay	199 (98.00)	3 (1.48)	-	1 (0.49)	203
Chinese	166 (80.20)	17 (8.21)	4 (1.93)	20 (9.66)	207
MalE	132 (97.80)	2 (1.48)	-	1 (0.74)	135
MalE-Malay	127 (96.20)	5 (3.79)	-	-	132
MalE-Chinese	62 (96.90)	1 (1.56)	-	1 (1.56)	64

(C): Auxiliary DO

	Grammatical	Ungrammatical			Total (100)
	Standard	**Omission**	**Commission**		
			Double-tense	*Morphological*	
BritE	75 (100.00)	-	-	-	75
Malay	142 (95.90)	2 (1.35)	3 (2.03)	1 (0.68)	148
Chinese	119 (86.90)	7 (5.11)	10 (7.30)	1 (0.73)	137
MalE	65 (97.00)	-	2 (2.99)	-	67
MalE-Malay	72 (93.50)	2 (2.60)	3 (3.90)	-	77
MalE-Chinese	51 (92.70)	1 (1.82)	3 (5.45)	-	55

(D): Tenses

	Grammatical	Ungrammatical			Total (100)
	Standard	**Omission**	**Commission**		
			Agreement	*Morphological*	
BritE	1765 (99.50)	7 (0.40)	1 (0.06)	-	1773
Malay	2005 (88.60)	242 (10.70)	16 (0.71)	-	2263

Chinese	1825 (77.70)	486 (20.70)	36 (1.53)	2 (0.09)	2349
MalE	1361 (95.80)	56 (3.94)	4 (0.28)	-	1421
MalE-Malay	1340 (94.00)	82 (5.75)	1 (0.07)	2 (0.14)	1425
MalE-Chinese	843 (93.10)	58 (6.41)	3 (0.33)	1 (0.11)	905

Referring to Table 7.4/Figure 7.5 above, the L1-Chinese group produced the highest proportions of omission and commission errors across the board. Starting off with the use of tenses, L1-Chinese supplied the highest omission rate (20.70%). Although L1-Malay came next, its omission rate (10.70%) was half as much. Amongst the L1-English groups, the suppliance of tense omission was below 7% for the Malaysian cohorts and below 1% for the British counterpart. While subject-verb agreement errors and morphological errors (i.e., over-regularisation of tense/aspectual inflections) were attested in the narrative data, their suppliance was miniscule across the L1 groups.

Regarding the use of the progressive feature, the highest proportion of errors was also found amongst the L1-Chinese group (19.8%). In contrast, the rest of the L1 groups – including L1-Malay – produce errors under 4%. The omission and commission errors supplied by the L1-Chinese group were of more or less equal proportions, which were at 10.14% and 9.66%, respectively. The L1-Chinese group was also the group that produced the highest number of commission errors i.e., (over)use of *-ing* ($n = 20$) amongst the L1 peers.[18] Table 7.5 presents a complete list of the overuse of *-ing*, whose investigation could not be carried out without examining the lexical aspect of the verbs being inflected as well. Of the 20 instances produced by L1-Chinese, more than half of them involved activity ($n = 10$) and accomplishment ($n = 3$) verbs, which usually allow for the use of *-ing*. There were also verbs of achievement ($n = 3$), semelfactive ($n = 3$), and stative ($n = 1$) denotations that received *-ing*, which is not normally expected. However, in 19 of these 20 cases, the inflected verbs were not preceded by progressive BE, but either by subject NPs ($n = 10$), conjunctions ($n = 4$), or even modal auxiliaries ($n = 5$). The only ungrammatical instance with the BE…*-ing* construction is found in #4 of Table 7.5, which involved a stative verb.

---

[18] Apart from the L1-Chinese group, the Malaysian L1 groups of Malay, MalE, and MalE-Chinese also seemed to over-generate *-ing* (see Table 7.5). However, each of these groups merely supplied one such instance, so the errors they produced were not explained further in this chapter.

Looking at the use of auxiliary DO, the L1-Chinese group once again exhibited the highest rate of errors amongst the L1 groups (13.14%) (see Table 7.4/Figure 7.5(C) above). This time, its suppliance of commission errors (8.03%) – especially double-tense marking (7.3%) – was slightly higher than that of DO omission (5.11%). A similar pattern was observed in the L1-Malay group, but the ungrammatical instances it produced (4.06%) were three times less than those produced by its L1-Chinese counterpart (13.14%). Furthermore, the L1-Malay and L1-MalE(+) groups were not substantially dissimilar to each other in their use of auxiliary DO, except for the L1-MalE-Chinese group, whose suppliance of double-tense marking revealed a slightly higher percentage (5.45%) by comparison.

Concerning copula BE, the highest cumulative error rate was a mere 4.98%, which came from the L1-Chinese group. As it is the least problematic feature, it will not be discussed further in the current analysis.

Looking at the performance of L1-Malay participants, the group behaved similarly to the L1-MalE(+) groups in that it generally supplied at least 95% of the standard uses of finiteness. The only exception where it dropped slightly below the 90% threshold (i.e., 88.6%) was in the production of tenses. Notwithstanding, as we compare between L1-Malay and L1-Chinese, the former group performed substantially better than the latter group in the production of standard finiteness.

Table 7.5: Overgeneration of *-ing*

No.	Example	Verb stem	Lexical aspect	Prog BE	NP subj.	Conjunct.	Modal	Infinitive
				Component preceding the inflected lexical verb				
1.	"And he just sit back, plug in back the earphone and then just, like, **hearing** the music." <MM014, L1-Chinese>	hear	activity			✓		
2.	"And that what's, what **happening** in the film." <MM014, L1-Chinese>	happen	accomplishment		✓			
3.	"So I will, like, **giving** her the whole packet of cookies." <MM014, L1-Chinese>	give	achievement				✓	
4.	"Sometimes, even though people has done something which you are not quite, um, um, not quite **understanding** or not quite knowing what they are doing, but, um, it is not actually wrong to offer them something, just like the man who actually, um…" <MM015, L1-Chinese>	understand	stative	✓				
5.	"And before he, he's, before she's going to, um, **waiting** for the train outside, she	wait	activity					✓

#		Verb	Aspect		
	went to the vending machine." <MM017, L1-MalE,Chinese>				
6.	"Next, um, she go to the um train station and **waiting** for the train." <MM021, L1-Malay>	wait	activity	✓	
7.	"Yeah, and with, we must think the consequences before we **doing** anything." <MM022, L1-Chinese>	do	activity	✓	
8.	"And we must, like, uh, **trying** not to repeat the mistake again." <MM022, L1-Chinese>	try	activity		✓
9.	"Or we need to figure out first then only we start angry or not just, uh, at first straightaway mad or **doing** anything." <MM030, L1-Chinese>	do	activity	✓	
10.	"The lady was infuriated and, as such, **beginning** to, to fight with the teenage boy for her packet of cookie back." <MM031, L1-Chinese>	begin	semelfactive	✓	
11.	"So maybe just xxx, I mean, kind of remorse in myself and then to remind	take	semelfactive	✓	

	myself not to do it again to make sure everything like just as I thought before I **taking** another move." <MM033, L1-Chinese>			
12.	"You should, like, **speaking** to her or even talking to, explaining to her she's taking, she's taking his food and even like chitchatting would, like, improve the situation more better." <MM039, L1-Chinese>	speak	activity	✓
13.	"Mmph maybe, uh, you should've check your things before you **trying** to scold other people." <MM077, L1-Chinese>	try	activity	✓	
14.	"Like, mm, if apply to my life, it's like, um, I have to know where am I first before I **scolding** at other else." <MM077, L1-Chinese>	scold	activity	✓	
15.	"Uh, after she **putting** the coins into the machine, the pack of cookies' stuck at	put	achievement	✓	

	the, in the vending machine." <MM088, L1-Chinese>			
16.	"So after that the old lady, like, plug out the earphone from the, from the handset and then, uh, like, **scolding** the young man." <MM096, L1-Chinese>	scold	activity	✓
17.	"And the young man, like, **seeing** in the woman, seeing the old lady and then seeing on the biscuit and, uh, break the last piece into half, so, uh, giving half to the old lady and another half eat by himself." <MM096, L1-Chinese>	see	semelfactive	✓
18.	"So after that the train **coming**." <MM096, L1-Chinese>	come	accomplishment	✓
19.	"So she, like, **looking** out of, through her windows and seeing the young man, like, and then feel like a bit guilty but still very grateful to the teenage for still giving her to ate the biscuits." <MM096, L1-Chinese>	look	activity	✓

No.	Example	Verb	Category			
20.	"But then I will still, like, **acting** nothing or just give her, share, sharing with, uh, with me, because I don't think, like, mm, stopping her from eating is quite manner lah." <MM096, L1-Chinese>	act	activity			✓
21.	"Or don't so fast, like, **making** a decision base(d) on your judging." <MM096, L1-Chinese>	make	accomplishment			✓
22.	"And also, uh, even after she **pressing** it, it was stuck." <MM113, L1-Chinese>	press	achievement	✓		
23.	"Well if I was already on the train and the train was about to leave, then I would have bang real hard and **trying** to apologise to him." <MM125, L1-MalE>	try	activity		✓	

7.3.2.2. English status effects

Turning to the investigation of the English status (levels: *L1-BritE, L1-MalE, L2-MalE*) effects, the syntax of the binomial regression model was similar to that for L1 effects, with the exception that the fixed effect English Status now substituted L1 Group: …glmer(Grammaticality ~ English Status + Proficiency + (1|Participants) + (1|Items)…). The overall results (Type III Wald χ^2 test) of the binomial regression model indicated that English Status ($\chi^2(2) = 52.2$, $p < 0.001$) and Proficiency ($\chi^2(1) = 36.7$, $p < 0.001$) were significant predictors, and Table 7.6 presents the summary of the model output. Results showed that the Malaysian L1- and L2-MalE groups were significantly more likely to produce finiteness errors than the L1-BritE control ($p \leq 0.001$; see also Table 7.7). However, as proficiency increased, that likelihood to produce finiteness errors decreased significantly ($\beta = -0.085$, $z = -6.06$, $p < 0.001$).

Table 7.6: Summary of the binomial model output with English Status as one of the main predictors in the narrative task

	Estimate	SE	*z* value	Pr(>|z|)	
(Intercept)	-5.053	0.460	-11.000	<0.001	***
English Status (L1 MalE)	1.931	0.472	4.090	<0.001	***
English Status (L2 MalE)	3.237	0.494	6.550	<0.001	***
Proficiency	-0.085	0.014	-6.060	<0.001	***

Furthermore, the post hoc (Tukey-adjusted) test revealed a significant contrast between the Malaysian groups of L1-MalE and L2-MalE ($\beta = -1.31$, $z = -5.11$, $p < 0.001$; see Table 7.7). This indicates that the likelihood of the former group committing ungrammatical finiteness was significantly lower than the latter group, thereby suggesting that having MalE as an L1 played a facilitative if not remedial role in the production of finiteness.

Table 7.7: Results of the post hoc comparisons between English Status groups from the narrative task

Contrast	Estimate	SE	*z* ratio	*p* value
L1-BritE - L1-MalE	-1.93	0.472	-4.09	0.001
L1-BritE - L2-MalE	-3.24	0.494	-6.55	<0.001

| L1-MalE - L2-MalE | -1.31 | 0.256 | -5.11 | <0.001 |

Delving into the descriptive analysis of the features, Figure 7.6 (for values, refer Table 7.8) shows a clear increment of finiteness errors produced by the English Status groups: L1-BritE < L1-MalE < L2-MalE. The feature with the most errors was tense. The cumulative error rate of L2-MalE (16.96%) was three times higher than that of L1-MalE (5.52%) (Table 7.8(D)). The progressive feature also witnessed a considerably high cumulative suppliance of errors coming from L2-MalE (10.97%), which was almost four times the ungrammatical suppliance of the L1-MalE counterpart (3.02%) (Table 7.8(B)). In comparison with the other features, auxiliary DO had the fewest constructions. Yet its ungrammatical suppliance was mostly concerned with the commission error of double-tense marking. These error rates coming from the L1- and L2-MalE groups were about the same, which were at 4.02% and 5.26%, respectively (Table 7.8(C)). Albeit low, they offer a glimpse of the pervasiveness of double-tense marking amongst Malaysians. Notwithstanding, looking at the overall picture, the L1-MalE group still outperformed the L2-MalE group and, as mentioned in the previous section, much of the ungrammatical finiteness came from the L1-Chinese group.

Figure 7.6: (Un)grammatical instances (%) produced by the English Status groups across the finiteness features of (A) copula BE, (B) progressivity, (C) auxiliary DO, and (D) tenses
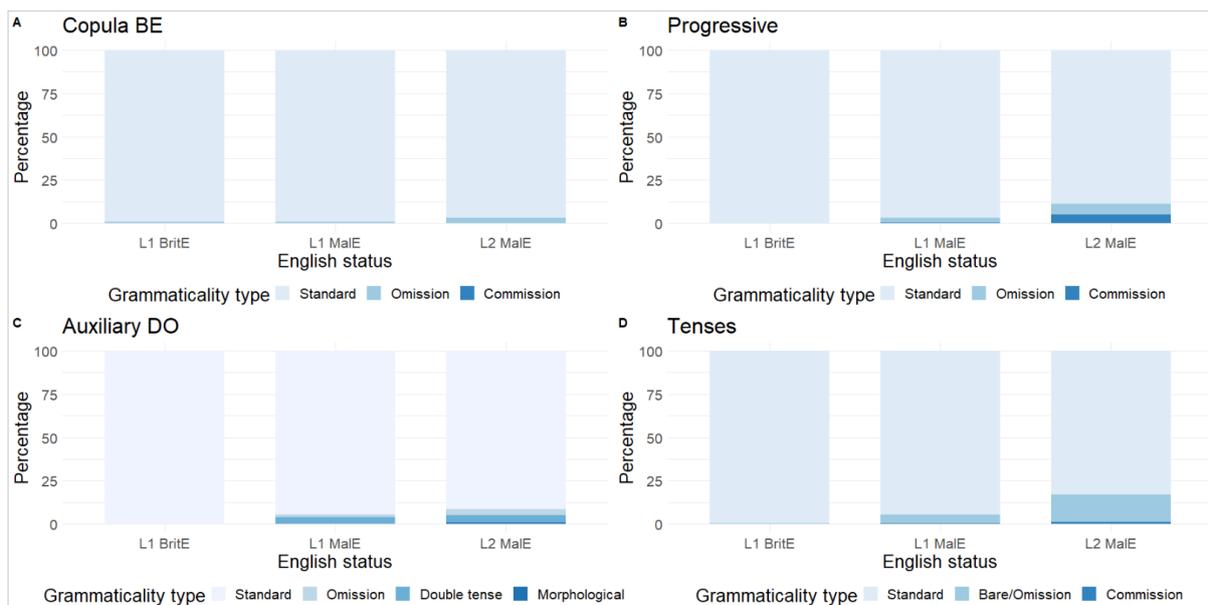
Table 7.8: (Un)grammatical instances (%) produced by the English Status groups across the finiteness features of (A) copula BE, (B) progressivity, (C) auxiliary DO, and (D) tenses

| | | Grammatical | Ungrammatical | | Total |
		Standard	Omission	Commission	(100)
(A) Copula BE	L1 BritE	449 (99.34)	3 (0.66)	-	452
	L1 MalE	1174 (99.32)	8 (0.68)	-	1182
	L2 MalE	1334 (96.74)	43 (3.12)	2 (0.15)	1379
(B) Progressive BE	L1 BritE	178 (100.00)	-	-	178
	L1 MalE	321 (96.98)	8 (2.42)	2 (0.60)	331
	L2 MalE	365 (89.02)	24 (5.85)	21 (5.12)	410
(C) Auxiliary DO	L1 BritE	75 (100.00)	-	-	75
	L1 MalE	188 (94.47)	3 (1.51)	8 (4.02)	199
	L2 MalE	261 (91.58)	9 (3.16)	15 (5.26)	285
(D) Tenses	L1 BritE	1765 (99.55)	7 (0.39)	1 (0.06)	1773
	L1 MalE	3544 (94.48)	196 (5.23)	11 (0.29)	3751
	L2 MalE	3830 (83.04)	728 (15.78)	54 (1.17)	4612

7.3.3. Miscellaneous finiteness errors

This section covers non-target suppletive morphemes (i.e., copula and auxiliary verbs) pertaining to finiteness that do not fit into any of the features of interest above. As we shall see below, the verbal element that is most frequently supplied is BE, but there are also other suppletive morphemes worth bringing to the fore. In what is to follow, the BE + BELONG structure, a pattern recurrent in the narrative data is first explored (**§7.3.3.1**). The remaining part covers the other morphological errors attested in the narratives (**§7.3.3.2.**).

7.3.3.1. BE + BELONG

In StE, the lexical verb BELONG denotes the state of being possessed by a possessor. It is always used as a phrasal verb with the preposition TO. Its argument structure comprises a subject NP,

which is assigned the theta role of a theme (in this case, it is the possessee), and an object NP, which takes the role of a recipient (in this case, the possessor) (see 46a). In the current data, there are attested instances of the phrasal verb additionally receiving a preceding morphological marker, BE (46b):

(46) a. "And then she realizes that that packet of cookies actually **belonged** to the teenager." <MM089, L1-MalE>

b. "And that packet of cookies ***was belonged** to that teenager, actually." <MM014, L1-Chinese>

To retrieve all instances containing BELONG, the following code was implemented via CLAN:

```
kwal + sbelong* +t*M* *.cha +d40
```

Component	Description
kwal	‣ Runs searches for linguistic elements at the speaker or morphological tier
+sbelong	‣ Specifies the word search: BELONG
+t*M*	‣ Selects speakers whose IDs start with M
*.cha	‣ Selects every CHAT file in the working directory
+d40	‣ Exports output to Microsoft Excel

(CLAN manual, pp. 90-92; MacWhinney, 2000)

There were altogether 48 tokens, 44 being verbs and 4 being nouns (i.e., belonging(s)). For the purpose of the main study, only verbal constructions ($n = 44$) were included. Additionally, four instances supplied by speakers of L1-Tamil were excluded from the quantification. This resulted in 40 instances for the analysis.

Table 7.9 presents the distribution of BELONG across the L1 groups. The highest suppliance of its target and non-target constructions came from the L1-Chinese group ($n = 24$), whereas the rest of its cohort produced a handful or less of it. More interestingly, the non-target construction of BELONG was slightly higher in number ($n = 22$) than the standard construction ($n = 18$), and an overwhelming majority of the former type involved BE + BELONG ($n = 20$), which was once again supplied mostly by the L1-Chinese group ($n = 13$). Concerning the two ungrammatical instances under the miscellaneous (misc.) category, BELONG was not accurately realised in its participial form "belonging" in reduced relative clause constructions (e.g., "The

166

biscuit ***belongs** to the old lady* is still in her own handbag," <MM029, L1-Chinese>). In cases where BELONG was used as a two-place predicate without periphrastic BE, it was always accurately inflected for tense and agreement ($n = 18$).

Table 7.9: The distribution of target and non-target constructions of the verb BELONG by the L1 groups

L1 group	Target	Non-target		Total
		BE + BELONG	*Misc.*	
BritE	-	1	-	1
Malay	1	3	-	4
Chinese	9	13	2	24
MalE	4	1	-	5
MalE-Malay	3	1	-	4
MalE-Chinese	1	1	-	-
Total	18	20	2	**40**

Zeroing in on the BE + BELONG structure, Figure 7.7 presents its manifestations in environments where third person subject NPs were either in the singular (3SG) or plural (3PL) forms. The most common manifestation was "is…belong" ($n = 10$), followed by "is…belongs" ($n = 5$). Not only did they both appear in 3SG contexts ($n = 7$), but they were also used in 3PL contexts ($n = 8$). What this tells us is that although BE was periphrastically attached to the verb BELONG, it did not obligatorily carry an agreement system, just as how it did not agree – at least in number – with 3PL subject NPs. A similar instance was found via the "was…belong" manifestation in the 3PL context, where BE was overtly inflected for past tense but not for agreement in number. Furthermore, the lack of agreement in BE was mostly found amongst the L1-Chinese participants – apart from one instance coming from a L1-Malay participant – which might unveil a representational issue of BE in their L2-StE interlanguage.

In addition, the verb BELONG was occasionally inflected with 3rd.SING.-*s* ($n = 6$). However, with the remaining 14 instances of BE + BELONG, it is unclear whether the lexical verb appeared in its bare form or was inflected for past tense. In the four cases where "belong-ϕ/*ed*" was preceded by "was" (in 3S contexts) and "were" (in 3PL contexts), it is possible that the lexical verb was doubly inflected with -*ed*, in which case its phonological realisation ([d]) would assimilate to the onset of the proceeding preposition TO ([t]) and thus become audibly

undetectable. However, if periphrastic BE was a main bearer of tense, then there is also a possibility for BELONG to remain uninflected. For instance, the fact that the "is…belong" pairing was twice the number of its doubly inflected counterpart "is…belongs" suggests that BELONG was optionally marked with 3rd.SING.-*s*. Leaving the subject NPs aside, -*s* might be an optional reflex of the present tense form of "is", which means the past tense -*ed* should not be expected when BELONG was supposedly c-commanded by "is". Contradicting this proposal of tense concord is the attestation of "was…belongs" (*n* = 1). Here, BE was realised in its past tense form whereas BELONG was inflected with the present tense -*s*. Notwithstanding, "was…belongs" merely appeared once, and thus the idea of tense concord cannot be rejected.

Figure 7.7: Manifestations of the non-target BE + BELONG structure produced by the L1 groups in third person singular (3SG) and plural (3PL) subject NP environments
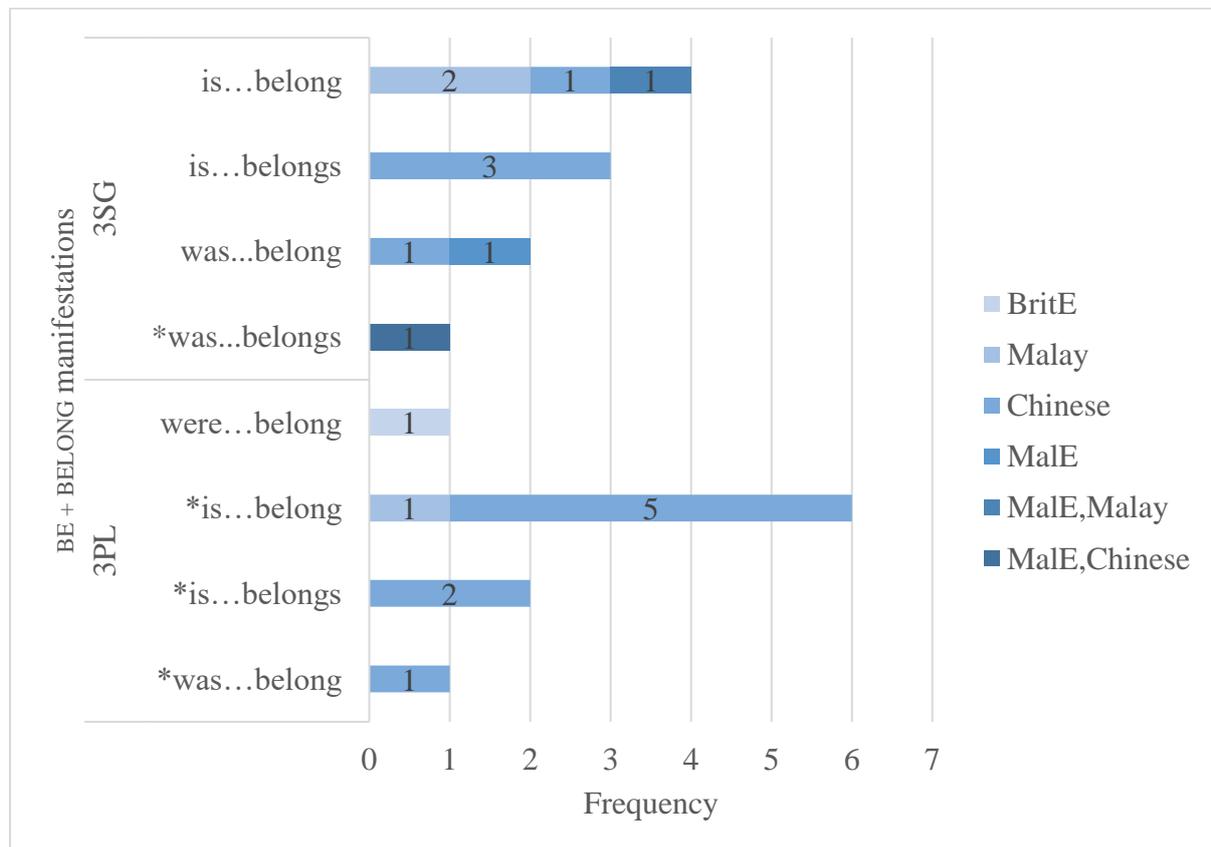
7.3.3.2. Other suppletive errors

Apart from the periphrastic suppliance of BE with BELONG, there were other errors of suppletive morphology. Of the 26 instances listed in Table 7.10, BE remained as the most supplied auxiliary ($n = 12$), which was followed by the contracted ′s ($n = 8$) and HAVE ($n = 6$).

In some instances, it is uncertain whether the suppletive morphemes produced were superfluous errors or a result of language processing issues. The examples in #1 and #2 in Table 7.10 demonstrate such ambiguous cases because it could either be that the manifestations of BE (i.e., "is") proceeded the missing expletive "it" in #1 and missing conjunction "if" in #2, or that "is" was a case of mispronunciation of those missing elements. Moreover, it is difficult to determine whether the contracted ′s was merely an error of phonological processing or indeed the contracted form of "is", "was", or "has".

On the other hand, there were morphological errors of BE that were clearly instances of overgeneration, such as the periphrastic insertions of BE before auxiliary DO (#5 and #10) and before lexical verbs in non-progressive contexts (#8, #12, #19, and #24). Furthermore, BE was also used in place of other auxiliary verbs such as the perfect HAVE (#6, #7, #25).

Concerning the existential construction in #21, it is difficult to decide if BE was a case of overgeneration since it was supplied in the presence of the CME GOT. For one thing, it could be that BE was not a superfluous marker because it would be expected to serve as a copula to link the existential quantifier THERE to its predicate. However, THERE had been dropped in #21. For another, GOT served as an existential verb in this context. As GOT usually suffices in replacing the StE THERE BE, plus given the absence of THERE, it is more likely that BE *was* a superfluous marker. That said, due to the lack of empirical investigation on GOT, it remains unclear if BE + GOT is an attested pattern of existential constructions amongst Malaysians. However, what remains clear is that GOT is a feature of CME and its use, to a certain extent, attests to the influence of the colloquial variety in the speaker's (MM096) StE interlanguage.

Concerning to the non-target uses of HAVE, it is worth noting that these instances ($n = 6$) all came from one participant (MM055). Not only was HAVE used in place of copula BE (#15), but it also seemed to bear readings of habituality (#13), perfectivity (#14), and existentiality (#16, #17, #18). While HAVE has not been documented as a feature of CME, I argue that it is an alternative to GOT, since it shares similar functions with it.

Finally, looking at the performances of the L1 groups, it was the L1-Chinese group that produced the highest number of morphological errors ($n = 19$) in contrast to the rest of its peers, who supplied less than four instances per group.

169

Table 7.10: List of suppletive errors produced by participants of the NT study

No.	Participant	L1 Group	Proficiency	Verb	Transcript	Notes
1	MM001	Chinese	78	aux|be	&-uh in the film, **is** basically started with a cartoon series <of it> [//] &-uh with a [//] an old lady there going for the cookie in a vending machine.	superfluous/ phonological
2	MM005	Chinese	59	aux|be	and I don't know **is** &-like he is [: was] [* m] not focus or alert there, she realized that there's a pack of cookies that already open <beside him> [//] &-ei beside her.	superfluous/ phonological
3	MM005	Chinese	59	's	yeah, that**'s** is the whole story.	contracted 's
4	MM027	MalE	93	's	so what happens is &-um she**'s** takes up her cookies or +...	contracted 's
5	MM029	Chinese	71	aux|be	<and then> [//] but the woman **is** [* m] still don't [: didn't] [* m:base:ed] appreciate it because she's still angry because &-like the guy took her biscuit.	superfluous
6	MM029	Chinese	71	aux|be	but then &-uh the scene end [: ends] [* m:03s] with the public transport **is** [*] already started moving.	auxiliary
7	MM029	Chinese	71	aux|be	but then in the clip, the public transport **is** [*] already started moving.	auxiliary
8	MM030	Chinese	76	aux|be	the cookies **is** [*] consist of four cookie in one tray.	superfluous
9	MM039	Chinese	74	's	&-uh after several times bang [//] banging the machines, which it**'s** happen [: happens] [* m:03s] quite frequently	contracted 's

					in my faculty also, &*INV:&=laughs the food finally drop out.	
10	MM041	MalE	77	aux|be	+" why **is** this kid &-like does not have any manners and didn't ask for permission while taking the food?	superfluous
11	MM042	Chinese	77	's	that**'s** means that &-uh the packet of biscuit does not belong to her.	contracted 's
12	MM045	Mal.E,Malay	77	aux|be	and finally, when she **was** [* m] actually come [: came] [* m:base:ed] to the point of frustrated, she apparently &-um pushed the vending machine and sit [: sat] [* m:base:ed] down &-um with her back towards the vending machine.	superfluous
13	MM055	Chinese	55	aux|have	what happen [: happened] [* m:0ed] is [//] <by is> [//] <what &+happe> [//] at the beginning is an old woman go [: going] [* m:0ing] to a somethings [: something] [* m:+s] like a train station and go [: went] [* m:base:ed] to the vending machine, which **have** [*] sell the cookies .	morphosyntactic
14	MM055	Chinese	55	aux|have	and so she **have** [*] take out her coin to buy the cookies.	morphosyntactic
15	MM055	Chinese	55	aux|have	and so she keep [: kept] [* m:base:ed] on &+s &-like 0cop angry and knocking the vending machine and finally **have** [*] successful get the cookies.	lexical

16	MM055	Chinese	55	aux|have	and the bench there **have** [*] another young male adult where wearing earphone.	lexical
17	MM055	Chinese	55	aux|have	<and when left> [///] total **have** [*] four biscuit inside the packing.	lexical
18	MM055	Chinese	55	aux|have	and then <when he> [//] after she sat down and saw her bag inside **have** [*] a new &-uh, is &nn like didn't open before one@s$co with original packing cookies.	lexical
19	MM067	Chinese	80	aux|be	but the young man **is** [*] just &-like seems to be on his own world while he's listening to his own phones and texting messages.	superfluous
20	MM067	Chinese	80	's	so maybe what I can do is &-uh I can see that the old woman**'s**, &-uh is not that convenient <to walk &-like so &+man> [//] &-uh to walk so much.	contracted 's
21	MM096	Chinese	65	aux|be	so on the bench **is** [*] actually **got** one young man, or teenage [: teenager], sitting there, listen to the &+mu music, using the earphones.	expletive
22	MM110	Malay	95	's	and then she realized that**'s** it was her mistake to through and through <at first> [//] at the very first.	contracted 's
23	MM118	Chinese	91	's	and [/] and she [/] she**'s** just &-uh looks very unbelievable.	contracted 's

24	MM134	MalE	71	aux|be	&-um he **was** [* m] just keep [: kept] [* m:base:ed] on smiling.	superfluous
25	MM140	Malay	63	aux|be	and then finally, the train **is** [*] arrive [: arrived] [* m:0en].	auxiliary
26	MU011	Brit.E	98	′s	&-um and she**'s** gets angry.	contracted ′s

7.3.4. Summary of findings

L1 effects:

- The proportions of finiteness errors in the NT were generally relatively low. The narrative performance of the L1-BritE control was barely touched on as the finiteness errors committed by the group were miniscule (0.54%).

- The error rates of the L1-MalE(+) groups, albeit relatively higher than that of the control group, were well below 10%, with the error rate of L1-MalE being the lowest (4.12%).

- Contrariwise, the L1-Chinese group performed the most divergently amongst its L1 cohorts. Not only did it commit the highest number of omission errors across the board, but it also produced the most commission errors such as the overgeneration (e.g., progressive *-ing*, BE + BELONG, etc.) and non-target uses (e.g., the use of auxiliary BE instead of perfect HAVE, etc.) of verbal morphology.

- In spite of its relatively low suppliance of obligatory tense marking (88.6%), the L1-Malay group performed markedly better than the L1-Chinese group in its suppliance of grammatical finiteness, making it almost on par with its L1-MalE(+) counterparts.

English status effects:

- As the L1 groups were collapsed into their respective English Status groups, an increase in the production of finiteness errors was observed, with the L1-BritE group supplying the least errors and the L2-MalE group supplying the most.

- However, it is clear that, within the L2-MalE group of L1-Chinese and L1-Malay, the main contributor of ungrammatical finiteness was the L1-Chinese group.

Vulnerable features:

- The feature most susceptible to errors was tense. This was especially so for the L2-MalE cohort (16.95%), in contrast to its L1-MalE(+) counterparts who cumulatively produced three times less errors (5.52%) than the former group.

- The ungrammatical constructions involving DO-support were low. That said, the proportions of double-tense marking produced by the L1-MalE(+) (4.02%) and L2-MalE (5.26%) groups were about equal, thereby hinting at the pervasiveness of double-tense marking as a feature of CME amongst the Malaysian participants.

- The BE + BELONG structure is a recurrent pattern attested across all L1 groups. However, a stark contrast was observed between the L1-Chinese group and the rest of its peers in that the former group produced the highest number ($n = 13$) of this ungrammatical construction as compared to the rest who supplied less than a handful, each.

7.4. Task effects

In response to RQ3, this section compares the overall findings of the GJT and NT to investigate if there are any effects of task types affecting the linguistic performances of participants.

7.4.1. Data management and analysis

Prior to conducting the statistical analysis, the results of each linguistic task were bifurcated into the grammatical and ungrammatical categories and collated into an Excel spreadsheet. Regarding the GJT ratings (i.e., 1 = least grammatical, 5 = completely grammatical), they were converted into accuracy scores (i.e., 1 = least accurate, 5 = most accurate). That is, while the rating scores of grammatical test sentences were retained, those of ungrammatical sentences were inverted to correspond to their respective accuracy scores (See Table 7.11).

Table 7.11: Accuracy score conversion schema for the GJT ratings

	Rating score →	Accuracy score
Grammatical items	5 (expected)	5 (most accurate)
	4	4
	3	3
	2	2
	1 (least expected)	1 (least accurate)
	0/NA	NA
Ungrammatical items	1 (expected)	5 (most accurate)
	2	4
	3	3
	4	2
	5 (least expected)	1 (least accurate)
	0/NA	NA

Meanwhile, the grammatical and ungrammatical instances of finiteness in the NT were measured in proportions, which altogether added up to 1. Following this, correlation analyses using Kendall's tau (τ) were carried out using the `cor.test()` function in RStudio (R Core Team, 2021). Kendall's τ was selected because in theory it deals with rank data (Field et al., 2012:225; Levshina, 2015:132-133). In the investigation of task effects according to L1 Group and English Status, the main dataset was split into the respective levels of these factors.

7.4.2. Results

The overall statistical results revealed significant positive and negative correlations between the performances of the GJT and NT for the grammatical ($\tau = 0.314$, $z = 5.594$, $p < 0.001$) and ungrammatical ($\tau = -0.172$, $z = -3.078$, $p = 0.002$) constructions, respectively. This indicates that participants who received higher accuracy scores in the GJT also produced higher proportions of grammatical finiteness and lower proportions of ungrammatical finiteness in the NT. Looking into Kendall's τ by English Status (Table 7.12(A)), significant correlations between the task types were found in the L2-MalE group for both grammatical and ungrammatical constructions ($p < 0.05$). Upon closer inspection on the L1 groups (Table 7.12(B)), it was the performances of both L1-Malay and L1-Chinese groups that contributed a significant negative correlation in the ungrammatical context (L1-Malay: $\tau = -0.518$, $z = -4.167$, $p < 0.001$) and a significant positive correlation in the grammatical context (L1-Chinese: $\tau = 0.236$, $z = 2.162$, $p = 0.031$), respectively. On the other hand, the overall performance of the L1-English status groups did not yield any significant correlation between the task types. Apart from the L1-MalE-Malay group, which showed a significant negative correlation in the ungrammatical context ($\tau = -0.3.28$, $z = -2.01$, $p = 0.044$), the results indicate that the L1-BritE, L1-MalE, and L1-MalE-Chinese groups consistently produced high accuracy scores in the GJT and high proportions of grammatical finiteness, thus rendering the task effects insignificant. Altogether, the effect of task type was significant, but it was mainly observed amongst the L2-MalE groups and not so much amongst the L1-English-speaking groups.

Table 7.12: Kendall's correlation between the linguistic performances on finiteness in the GJT and NT by grammaticality type, (A) English status, and (B) L1 group

		Grammaticality type	τ	z	p
(A) English status	L1-BritE	Grammatical	-0.218	-1.434	0.152
		Ungrammatical	0.080	0.531	0.595
	L1-MalE	Grammatical	0.129	1.301	0.191
		Ungrammatical	-0.179	-1.817	0.069
	L2-MalE	Grammatical	0.243	3.016	0.003
		Ungrammatical	-0.219	-2.743	0.006
(B) L1 group	BritE	Grammatical	-0.218	-1.434	0.152
		Ungrammatical	0.080	0.531	0.595
	Malay	Grammatical	0.175	1.388	0.165
		Ungrammatical	-0.518	-4.167	<0.001
	Chinese	Grammatical	0.236	2.162	0.031
		Ungrammatical	-0.058	-0.532	0.595
	MalE	Grammatical	0.180	1.013	0.311
		Ungrammatical	-0.206	-1.158	0.247
	MalE-Malay	Grammatical	0.223	1.371	0.170
		Ungrammatical	-0.328	-2.01	0.044
	MalE-Chinese	Grammatical	0.146	0.677	0.498
		Ungrammatical	0.144	0.675	0.500

7.5. Conclusion

In this chapter, we have looked at the production of finiteness in the NT and found a clear effect of the L1/L2 status of English in predicting the suppliance of (un)grammatical finiteness. Although the production of finiteness errors was generally low, the highest proportion was found amongst the L2-MalE cohort, which predominantly came from the L1-Chinese group. The L1-Malay group, on the other hand, performed similarly to its L1-MalE(+) peers in producing higher rates of grammatical finiteness, apart from exhibiting a weaker performance in the suppliance of tense marking. In response to RQ1, therefore, the effects of L1 were clear in that the effect of L1-Chinese was adverse, whereas the effects of L1-Malay and L1-MalE(+),

to a greater extent, were positive. In relation to RQ2, the influence of CME on the use of StE finiteness was implied through the suppliance of double-tense marking by both L1- and L2-MalE cohorts. Notwithstanding, the proportions of this commission error were relatively low. Finally, regarding RQ3, there were clear effects of task types on the use of finiteness, but the effects were mainly found in the L2-MalE status cohort. The main findings highlighted in this chapter are discussed further in **Chapter 9**. In the meantime, the following chapter, which presents a sociolinguistic study, reports the perceptions and attitudes of the Malaysian participants on CME and StE.

8   Results of the sociolinguistic survey

8.1. Introduction

This chapter reports findings from the sociolinguistic section (Section C) of the language background and attitude questionnaire. The overarching research question motivating the current study is as follows:

RQ4: To what extent is CME perceived as functionally different from StE in Malaysia?

This study was designed only for Malaysian participants, and it sought to investigate i) to what extent they were aware of the indigenised English variety spoken in Malaysia, ii) what their attitudes were towards it and the standard variety, and iii) what their linguistic preferences between these two varieties would be in different social contexts.[19] Accordingly, this chapter takes on an exploratory trajectory in addressing the inquiries above. Pre-processing and analysis of the data are first described in **§8.2.**. Findings on lectal awareness (**§8.3.1.**), attitude (**§8.3.2.**), and preference (**§8.3.3.**) are then reported in the sections after. Finally, the key findings are summarised in **§8.4.**.

8.2. Data management and analysis

The data of the sociolinguistic survey were automatically generated from Qualtrics and exported into an Excel spreadsheet. Responses elicited from the questions of interest were selected, formatted, and saved in separate Excel files. The demographic and linguistic details of the Malaysian participants were also copied therein.

Regarding the statistical analyses, a simple linear regression was conducted on the data for lectal attitude (**§8.3.2.**), and a binomial mixed regression was fitted for lectal preference (**§8.3.3.**). All 145 Malaysian participants, regardless of their L1 backgrounds, were included. This was because inclusion of L1 Group (only with the levels of Malay, Chinese, MalE, MalE-Malay, MalE-Chinese) as a fixed effect did not significantly improve the model fit in the

---

[19] Due to time constraints, this thesis will not be able to cover the written responses of participants describing the local English variety.

regression analyses for both lectal attitude and preference. Other covariates such as proficiency and biological sex were also entertained, but they too did not improve the models. In other words, the linguistic and demographic variables did not significantly predict the outcomes of lectal attitude and preference. Therefore, it was safe to include to all participants in the analyses. The structures of the best fitted models are later specified in their corresponding sections.

8.3. The findings

8.3.1. Lectal awareness

Out of the 145 participants, an overwhelming majority of them ($n = 140$) stated that they were aware of a local variety of English being spoken in Malaysia. The five exceptions came from different demographic and linguistic backgrounds as well as English proficiencies. In other words, there were no particular traits that stood out amongst these participants. Their profiles can be found in Table 8.1.

Table 8.1: The linguistic and demographic profiles of Malaysian participants who were not aware of a local English variety in Malaysia

Participant	Sex	State	L1 group	Proficiency
MM003	Male	Penang	Chinese	63
MM012	Male	Selangor	Malay	76
MM133	Female	Selangor	Malay	81
MM140	Female	Sabah	Malay	63
MM144	Male	Kedah	Tamil	82

Turning to those who were aware of the local variety, they were asked to provide as little or as many names as they could think of. As presented in Table 8.2, not all of them were informed about the name(s) associated with the variety ($n = 16$). However, amongst those that were, almost 50% of the participants labelled it as *Manglish*. Within this subgroup of respondents, they were further requested to specify the portmanteau of Manglish, were they aware of it. Subsequently, "Malaysian English" comprised the highest frequency ($n = 44$), followed by language-specific combinations such as "Malay & English" ($n = 23$) and "Mandarin & English" ($n = 7$). There were also cases where no specifications had been given

to Manglish ($n = 24$). The generic label receiving the second highest proportion of recognition – albeit 30.1% less popular than Manglish – was Rojak (18.2%). It encompassed the following specific terms stated by the participants: "Rojak", "Rojak English", and "Bahasa Rojak" [Malay: the Rojak Language]. The terms Malaysian English (independent from the label Manglish), Broken English, English, and other miscellaneous neologisms were also attested, though they were considerably less frequent than Manglish and Rojak.

Table 8.2: Names of the local English variety given by Malaysian participants

Variety name	N (%)
Manglish	**101 (48.30)**
• *Malaysian English*	*44 (43.60)*
• *Malay & English*	*23 (22.80)*
• *Mandarin & English*	*7 (6.93)*
• *Mandarin & Malay & English*	*2 (1.98)*
• *Malay & English & Mandarin & Tamil*	*1 (0.99)*
• *NA*	*24 (23.80)*
Malaysian English	**16 (7.66)**
English	**4 (1.91)**
Broken English	**14 (6.70)**
Rojak	**38 (18.20)**
• *Rojak*	*18 (47.40)*
• *Rojak English*	*11 (28.90)*
• *Bahasa Rojak*	*9 (23.70)*
Misc.	**20 (9.57)**
Singlish; Grammatically error English but understandable; Bahasa pasar; Taglish; KL English; Wichet English; Bad English; Speaking; Bahasa London; Gwai Lor Language; White Man Language; Chinglish; Minglish; Kong Angmoh; Malay-English; Broken; Mat Salleh Celup; Speaking; Easy English; Cincai English	
Don't know	**16 (7.66)**

8.3.2. Lectal attitude

Next, the perceptions of all Malaysian participants about CME versus StE were evaluated.[20] The participants were presented with 10 statements about CME and StE, each of which was designed along five themes: Nationalism, Friendliness, Educatedness, Purism, and Pedagogy. As shown in Table 8.3, the paired statements in each theme were designed to contrast each other. This was done to evaluate the extent to which the participants would (dis)agree with them across the themes.

Table 8.3: Theme-based item pairs for the language attitude survey

	CME	StE
Nationalism	Speaking Manglish reveals one's identity as a Malaysian.	Speaking Standard English covers up one's national identity.
Friendliness	People who speak Manglish seem friendly.	People who speak Standard English seem less approachable.
Educatedness	People who speak Manglish are thought to be less well-educated.	People who speak Standard English are thought to be well-educated.
Purism	Manglish is bad English.	Standard English is the only correct English.
Pedagogy	Using Manglish can be beneficial in helping Malaysians learn Standard English better.	In order to improve on Standard English, one must avoid using Manglish.

The rating responses of the participants were indicated using a seven-point Likert scale, with 1 being "Completely disagree" and 7 being "Completely agree". Figure 8.1 presents the participants' mean ratings on the variety-related statements. Under the sociocentric themes of Nationalism and Friendliness, the ratings of the CME-related statements were considerably higher than those of the StE ones. CME was perceived to be associated with national identity and friendliness. However, statements where StE would conceal one's national identity or create social distance were not rejected nor accepted by the participants, thereby indicating the participants' neutral attitude towards StE in those regards. Moving on to themes of

---

[20] In the survey, the layman term *Manglish* was used instead of CME because Manglish was assumed to be a familiar term to Malaysians. CME is usually used in academic contexts (e.g., World Englishes research).

Educatedness, Purism, and Pedagogy, the trends were flipped such that the mean ratings were now higher for the StE statements than for the CME ones. Participants generally agreed that the use of StE would give an impression of the speaker's well-educatedness. In terms of language purism, while there was a consensus on the correctness of StE, the bastardization of CME was not readily accepted by the participants. That said, the mean rating of the CME statement was nevertheless slightly on the agreeableness side of the scale. Finally, concerning pedagogical prospects, the notion of limiting the use of CME when learning StE was agreed upon to a relative degree but, at the same time, the idea of deploying CME as a pedagogical tool was positively welcomed by the participants.

Figure 8.1: Mean ratings of statements about CME and StE along the themes of nationalism, friendliness, educatedness, purism, and pedagogy
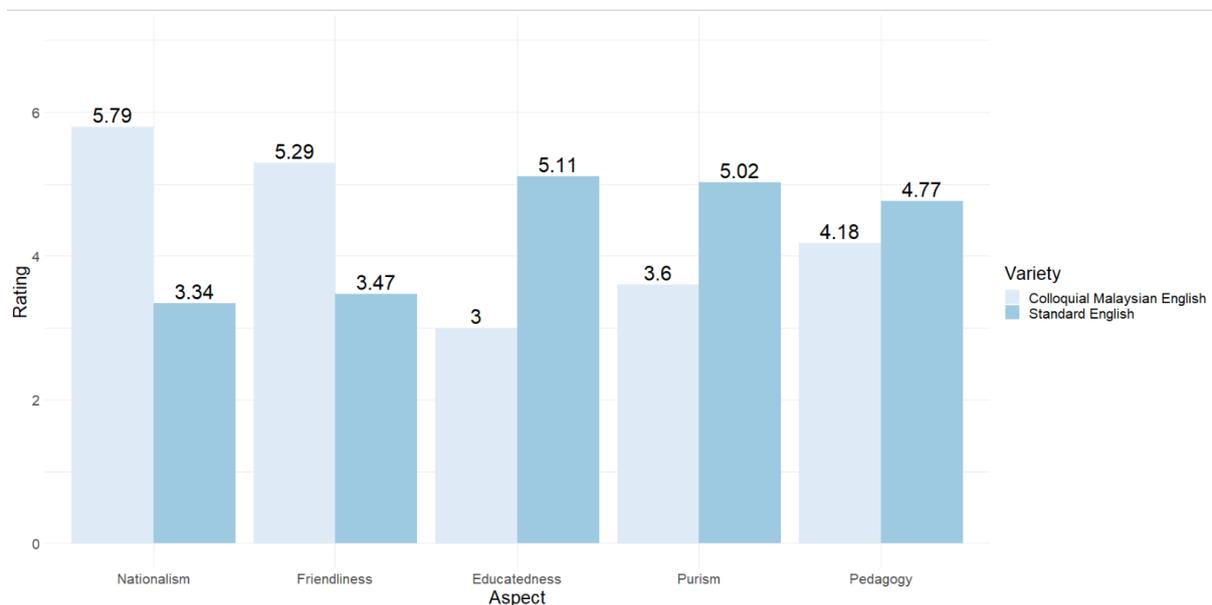
To test the aforementioned rating differences further, a simple linear regression model was fitted (R Core Team, 2021). The rating responses were *z*-transformed and constituted the dependent variable. The variables Theme and English Variety and their interaction formed the fixed-effects structure. The R code of the model was roughly constructed as follows: …`lm(Rating.z ~ Theme * Variety…)`

Results of the model revealed a significant simple main effect of Theme ($F(4) = 5.02$, $p < 0.001$) and a strong interaction between Theme and English Variety ($F(4) = 165.93$, $p < 0.001$) (Table 8.4).

Table 8.4: Summary of the model output for language attitude

Estimate	Est.	SE	t value	Pr(>|t|)	
(Intercept)	0.780	0.061	12.710	<0.001	***
Theme (Friendliness)	-0.282	0.087	-3.250	0.001	**
Theme (Educatedness)	-1.533	0.087	-17.670	<0.001	***
Theme (Purism)	-1.166	0.087	-13.430	<0.001	***
Theme (Pedagogy)	-0.920	0.087	-10.590	<0.001	***
Variety (StE)	-1.328	0.087	-15.300	<0.001	***
Theme (Friendliness): Variety (StE)	0.305	0.123	2.480	0.013	*
Theme (Educatedness): Variety (StE)	2.540	0.123	20.700	<0.001	***
Theme (Purism): Variety (StE)	2.088	0.123	17.010	<0.001	***
Theme (Pedagogy): Variety (StE)	1.706	0.123	13.900	<0.001	***

The post hoc pairwise comparison by theme (Tukey adjusted) showed that the rating differences between the StE and CME statements were statistically significant across the themes (Table 8.5). This indicates that for every contrastive pair of statements, one of them received a substantially higher rating than the other.

Table 8.5: Results of the post hoc test for language attitude

Theme	Pair	Est.	SE	df	LCL	UCL	t ratio	p val.
Nationalism	CME - StE	1.330	0.087	1440	1.160	1.500	15.30	<.001
Friendliness	CME - StE	1.020	0.087	1440	0.850	1.190	11.80	<.001
Educatedness	CME - StE	-1.210	0.087	1440	-1.380	-1.040	-14.00	<.001
Purism	CME - StE	-0.760	0.087	1440	-0.930	-0.590	-8.800	<.001
Pedagogy	CME - StE	-0.380	0.087	1440	-0.550	-0.210	-4.400	<.001

8.3.3. Lectal preference in different social contexts

Finally, the Malaysian participants' lectal preferences between CME and StE in different social contexts were examined. The hypothetical social contexts included interactions with family, friends, colleagues, English Language teachers, teachers who did not teach English, Malaysian strangers, and foreigners. Accordingly, Figure 8.2 visualises the proportions of lectal preference across these contexts.

Figure 8.2: Proportions (%) of lectal preferences according to social context
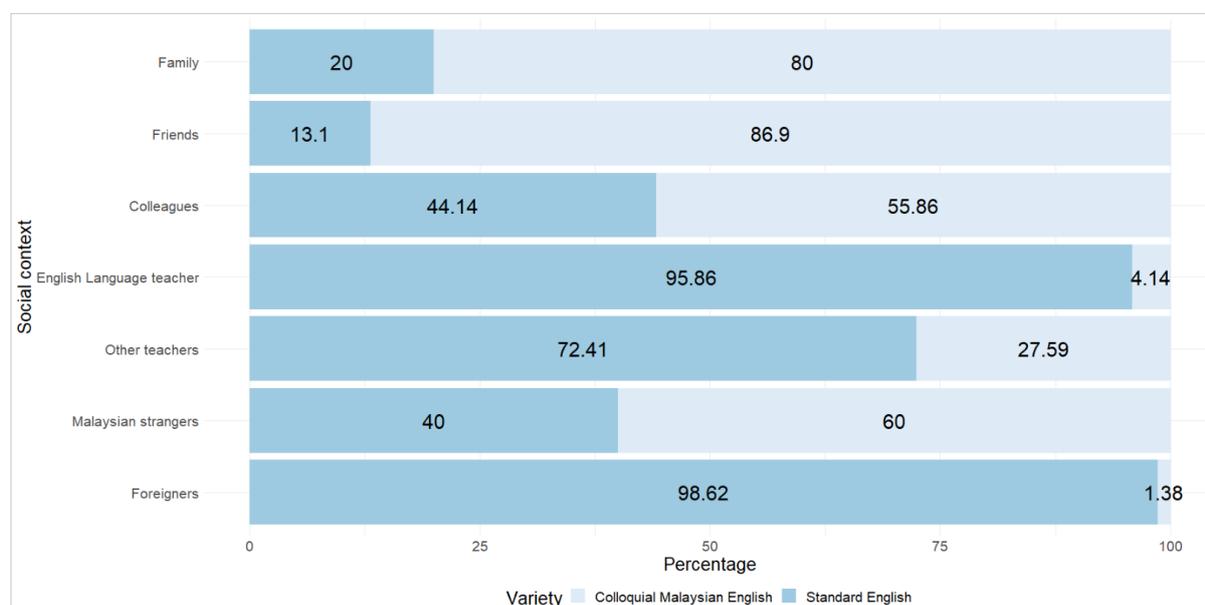
A binomial mixed model was fitted using the `glmer()` function of the "lme4" R package (Bates et al., 2015). English Variety was entered as a binomial response variable, whereas Social Context was entered as a fixed effect. Furthermore, by-participants were included as random intercepts. The R code therefore looked like the following:

…`glmer(Variety ~ Context + (1|Participants)`…

The overall results of the binomial model (Type III Wald χ^2 test) revealed that social context was a significantly strong predictor for lectal preference ($\chi^2(6) = 181.7$, $p < 0.001$; Table 8.6).

Table 8.6: Summary of the model output for lectal preference

	Estimate	SE	z value	Pr(>|z|)	
(Intercept)	-1.967	0.292	-6.740	<0.001	***
People (Friends)	-0.660	0.369	-1.790	0.074	.
People (Colleagues)	1.615	0.324	4.980	<0.001	***
People (English Language teachers)	6.109	0.597	10.240	<0.001	***
People (Other teachers)	3.352	0.369	9.070	<0.001	***
People (Malaysian strangers)	1.369	0.323	4.240	<0.001	***
People (Foreigners)	7.347	0.855	8.590	<0.001	***

185

In line with Figure 8.2 above, the Tukey adjusted post hoc test in Table 8.7 showed that CME was a significantly stronger choice than StE for the Family ($\beta = -2$, $z = -6.7$, $p < 0.001$, 95% CI[-2.5, -1.4]) and Friends ($\beta = -2.6$, $z = -7.9$, $p < 0.001$, 95% CI[-3.3, -2.0]) categories. StE was preferred significantly more than CME in interactions with teachers ($p < 0.001$), foreigners ($\beta = 5.4$, $z = 6.8$, $p < 0.001$, 95%CI[3.8, 6.9]), and Malaysian strangers ($\beta = -0.6$, $z = -2.4$, $p = 0.015$, 95% CI[-1.1, -0.1]). Regarding interactions with colleagues, the preference for CME was merely slightly more than for StE, and results from the post hoc test demonstrated no statistical significance between these choices.

Table 8.7: Results of the post hoc test for language preference

People	Emmean	SE	LCL	UCL	*z* ratio	*p* value
Family	-2.000	0.290	-2.500	-1.400	-6.700	<.0001
Friends	-2.600	0.330	-3.300	-2.000	-7.900	<.0001
Colleagues	-0.400	0.240	-0.800	0.100	-1.500	0.147
English Language teachers	4.100	0.510	3.100	5.100	8.100	<.0001
Other teachers	1.400	0.270	0.900	1.900	5.200	<.0001
Malaysian strangers	-0.600	0.250	-1.100	-0.100	-2.400	0.015
Foreigners	5.400	0.790	3.800	6.900	6.800	<.0001

What these findings suggest is that while CME was preferred in intimate and informal settings (as seen with families and friends), the choice for CME decreased and StE became a more favoured option as the socio-cultural distance and/or formality increased (as seen with teachers, local strangers, and foreigners). More importantly, the preferences given to both CME and StE in different social contexts not only indicated the Malaysians' sensitivity of when to employ the English varieties, but that the participants did not discriminate any of them.

8.4. Summary of findings

Language awareness:

- From the sociolinguistic survey, most of the Malaysian participants were aware that there was a local variety of English spoken in their home country and recognised it as Manglish.

Language attitude:

- The participants perceived the use of CME to express national identity and friendliness.
- However, there was a stronger inclination to attribute StE to a higher educational status than CME.
- Although the participants generally agreed to the practice of limiting the use of CME when learning StE, they entertained the prospect of using it to facilitate the learning experience.

Language preference:

- Finally, when asked for their preference between CME and StE in different social interactions, the participants indicated that they would use CME in informal and/or intimate settings and StE in formal and less intimate ones.

To sum up, not only were the participants aware of the different social functions of CME and StE, but they also did not show clear signs of discrimination against any of them.

9 Discussion

9.1. Introduction

The main study adopted a concurrent embedded design which comprised i) the linguistic tasks of grammaticality judgement (GJT) and narrative-telling (NT) to examine the linguistic performance of Malaysians on StE, and ii) a sociolinguistic survey exploring Malaysians' attitudes towards CME and StE. This chapter is divided into two parts: **§9.2.** addresses the first three questions that are related to the acquisition of StE. **§9.3.** addresses the final question that is concerned with sociolinguistic matters. The chapter then concludes by presenting the study's limitations and recommendations for future research (**§9.4.**).

9.2. Discussion of findings from the acquisition tasks

9.2.1. RQ1: To what extent does L1 play a role in the linguistic performance on StE?

To address the first research question, the participants of the main study were categorised into six L1 groups: L1-BritE being the control group and the five Malaysian groups of L1-Malay, L1-Chinese, L1-MalE, L1-MalE-Malay, and L1-MalE-Chinese. These groups were further clustered into three broader categories according to English Status: L1-BritE, L1-MalE (i.e., L1-MalE, L1-MalE-Malay, L1-MalE-Chinese; hereafter *L1-MalE*(+)), and L2-MalE (i.e., L1-Malay, L1-Chinese). Accordingly, we sought to investigate if any of the (clusters of) Malaysian L1 groups would exhibit divergent linguistic performance from the L1-BritE control group in their judgements (via the GJT) and production (via the NT) of finiteness in StE. If so, it would implicate negative transfer from the L1. The L1-MalE(+) groups were further scrutinized to see if having English in their L1 would facilitate or interfere with their performance in the linguistic tasks.

  The primary findings from the GJT and NT revealed that performance in finiteness marking in StE was overtly influenced by the L1. Although there were statistically significant differences between the Malaysian cohorts and the British control group in their judgements and suppliance of certain finiteness features, the results nevertheless showed clear facilitation

effects when English was acquired as an L1 or one of the L1s in the Malaysian context. Contrariwise, the L1 group exhibiting the greatest linguistic divergence from all other groups across the experimental tasks was the L1-Chinese group. By comparison, the L1-Malay participants performed considerably better and were almost on par with their L1-MalE(+) peers. In what is to follow, the main findings of the Malaysian L1 groups are explicated, starting with the L1-MalE(+) cohorts, followed by the L1-Chinese group and finally the L1-Malay group. Particular attention shall also be drawn towards non-target features in the production data to examine the extent to which L1 effects could sufficiently account for those features.

Addressing first the linguistic performance of the three **L1-MalE(+) groups**, we compared them with their L1-BritE counterpart in the **GJT**. In many occasions, the linguistic judgements of the L1-MalE(+) peers were similar to those of the L1-BritE controls. Few exceptions include: i) the significantly lower ratings of standard copular constructions by all L1-MalE(+) groups, ii) the significantly higher acceptance of double-tense marking in DO constructions by the L1-MalE and L1-MalE-Malay groups, and iii) the significantly higher acceptance of DO omission by the L1-MalE group. As the first two conditions have also been observed with the L2-MalE cohorts, we will discuss them in **§9.2.2.**. We turn to the third condition, i.e., DO omission. In CME, auxiliary verbs including DO-support are commonly omitted in direct questions (Hashim, 2020:387-388; Pillai & Greig, 2020). According to Pillai & Greig (2020), this feature is considered to be pervasive. For Malaysians who acquired English as (one of) their L1(s), it is unclear as to what types of – and how much of each type of – input in English they have been exposed to since early childhood (see Buschfeld, 2020 on child L1 SingE). However, given the pervasiveness and informal use of CME in Malaysia, it should come as no surprise if CME is found to be spoken in the homes of English-speaking Malaysians (see Gupta, 1994 on CSE). Should this be the case, then the fact that the L1-MalE group generously accepted DO omission in direct questions in the GJT hints at the influence of CME. Notwithstanding, there were many other instances where the L1-MalE group was on par in their linguistic judgement with their L1-BritE counterpart. More interestingly, in the **NT**, not once did the L1-MalE group omit auxiliary DO from its obligatory contexts (i.e., yes/no-questions, negated declaratives). Looking at the performance of the L1-MalE(+) cohorts as a whole in the NT, these groups supplied grammatical finiteness more than 90% of the time for all the features of interest. If we follow the 90% accuracy threshold (Brown, 1973), then this means that the L1-MalE(+) groups all successfully acquired grammatical finiteness despite being statistically different from their L1-BritE peers. Therefore, the current research reveals that the plausibility of having a colloquial variety of English in the L1 does not severely

interfere with the acquisition of StE, although there is some influence from CME which seems to mainly be affecting perception/judgements.

Turning to the **L1-Chinese participants**, their linguistic performance generally diverged from the other L1 groups across the board. In the **GJT**, this group accepted the omission of copula BE and auxiliary DO significantly more than the L1-BritE control group, thereby suggesting influence from Chinese. Regarding the omission of copula BE, the lack of rejection of this condition indicates transfer from the optional use of the Chinese copula *shì* (Li & Thompson, 1981) (see **§4.3.3.**). Regarding the omission of auxiliary DO, not only does Chinese not have a counterpart for this dummy auxiliary, it also necessitates *in-situ* question constructions, though with the help of intonation, question particles, and/or A-not-A constructions (Li & Thompson, 2009:718-719). However, with respect to the omission of 3rd.SING.-*s* and regular past -*ed*, the L1-Chinese group behaved similarly to their L1-BritE counterpart by consistently rejecting these ungrammatical conditions. Such findings came as a surprise, given that the absence of tense inflections in Chinese would lead one to anticipate low rejection rates by the L1-Chinese participants. However, in the **NT**, the most vulnerable finiteness features faced by this group were indeed the past/present tense inflections, which were omitted 20.7% of the time. At first glance, the discrepancy between the L1-Chinese's performance in the NT (non-target-like) and GJT (target-like) in relation to tense inflections might indicate a failure to pronounce such inflections due to constraints imposed by the prosodic structure of L1-Chinese (Goad & White, 2006). However, looking beyond tense markings, the L1-Chinese group was also found to commit not only omission errors but also commission errors in other finiteness features (i.e., progressive and DO-support). Quantitatively, these commission errors brought down the accuracy rates below 90% in the NT (i.e., progressive: 80.2%; DO-support: 86.9%). More broadly, the amount of omission and commission errors produced by the L1-Chinese participants across the board, and how this group's linguistic performance generally diverged from the rest of L1 groups (including the L1-Malay group), altogether do not support the prosodic transfer approach (Goad & White, 2006). Instead, these observations seem to reflect lack of target language attainment in the interlanguage amongst the L1-Chinese group.

Looking into the commission errors of the L1-Chinese group, one of the highest rates of this type of error is the overgeneration of the progressive -*ing*. In the GJT, the L1-Chinese group was amongst the L1 cohorts that gave the highest average rating to stative progressive constructions. In the NT, although the L1-Chinese group oversupplied -*ing* less than 10% (i.e., 9.66%) of the time, it was nevertheless the L1 group with the highest overuse of -*ing*. The

acceptance of stative progressivity in the GJT could be due to influence from the Chinese -*zhe*, which is one of the durative aspect markers that denotes the "continuance of a state" (Kwan-Terry, 1979:220). However, in the NT, the overgeneration of -*ing* – apart from one instance (see #4 in Table 7.5 under **§7.3.2.1.**) – did not involve stative verbs. More specifically, out of the 20 instances, 19 of them were all dynamic verbs (including activity, achievement, accomplishment, and semelfactive), and they were not preceded by the progressive BE. In Bardovi-Harlig & Reynolds's (1995) study on the acquisition of the simple past tense, the authors demonstrated that L2 learners of different L1 backgrounds were influenced by the inherent semantic aspect of verbs in their use of simple past and non-target tense/aspect markings. A couple of patterns observed by the authors included the frequent use of the progressive marking with activity verbs and non-past (i.e., simple present; uninflected/bare) forms with stative verbs. In the current study, while we did not analyse tense alternations on stative verbs, we noticed that most ($n = 10$) of the 19 dynamic verbs which were overly inflected with -*ing* in non-progressive contexts were indeed activity verbs. This pattern was in some ways similar to that observed by Bardovi-Harlig & Reynolds (ibid.). However, if "the influence of lexical aspectual class [is] an acquisitional universal" (ibid., p.120), then what is left to be explained is this: why was the overuse of -*ing* in the NT mostly found amongst the L1-Chinese group and not the other L1 groups as well? Also, if L1 influence was mainly at play, we would not expect to find the overuse of -*ing* predominantly supplied by the L1-Chinese group in the NT, since progressive marking is optional in Chinese (see Smith, 1997:277 on viewpoint morphemes; see also Zeng et al., 2019, pp.271, 282). An alternative explanation offered by Bardovi-Harlig & Reynolds (1995) is the potential role of input suggested by Andersen (1990). That is, if a particular grammatical form X is used more frequently than another form Y in the input, or if certain verbs occur more frequently with a particular form than with other forms, then the L2 learner would be more inclined to use the form with the higher frequency (Andersen, 1990:58-59). Therefore, the overuse of -*ing* in the current study might possibly reflect the saliency and frequency of the form in the input received by our L1-Chinese group. It could also be due to factors pertaining to lexical aspect, since -*ing* was over-supplied most frequently with activity verbs. Even though L1 influence could not account for the L1-Chinese's overuse of -*ing* in the NT, it cannot be dismissed entirely as a factor because, as observed in the GJT, it might to a certain extent be able to account for the acceptance of stative progressivity by the L1-Chinese learners.

Another type of commission error supplied mostly by the L1-Chinese group is the overuse of the auxiliary BE in the NT. One particular pattern observed is the use of BE with the

verb BELONG, which was produced more than half of the time (i.e., $n = \frac{13}{24}$; 54.17%) when BELONG was used by this L1 group. The over-suppliance of BE in this BE + BELONG structure is likely due to the influence of the Chinese copula *shì*, which can be used, optionally, as a marker to emphasise a statement (Li & Thompson, 1981:154):

(47) English: (BE)   BELONG

　　　Chinese: (是) 属于

　　　　　shì　*shǔyú*

As the overuse of BE has been argued to be a syntactic carrier of agreement in relation to the subject NP (**L1 studies:** Brown, 1973; Tesan & Thornton, 2004; **L2 studies:** Abdul Aziz & Mohd Don, 2014; García Mayo et al., 2005; Gavruseva, 2008; Hawkins & Casillas, 2008; Ionin & Wexler, 2002; Paradis et al., 2008), the next question to pose is whether it serves such a function – in addition to its role as an emphatic marker – in the BE + BELONG structure. From the findings in the NT, although BE was accurately manifested in its singular forms (i.e., *is/was*) in 3rd person singular contexts, these forms were inaccurately extended to 3rd person plural ones. What this suggests is that the singular forms of BE were not used to bear agreement but mostly as a default marker to indicate emphasis on the statement.[21] Apart from the BE + BELONG structure, BE also occurred with other verbs in non-progressive contexts (e.g., #1, #2, #5, #6, #8, #19, and #21 in Table 7.10 in **§7.3.3.2.**). This superfluous suppliance might be a reflex of the pragmatic marking of *shì* that possibly served as "a focus marker in an emphatic sentence" (Lee & Huang, 2004:213).

　　　Apart from the overgeneration of BE, non-target suppliance of auxiliary verbs by the L1-Chinese group was not prevalent in the NT data. That said, the non-target uses of HAVE (*n* = 6) by one L1-Chinese speaker (MM055) is worth discussing in the light of L1 influence. In Chinese, the verb 有 *yǒu* bears the same lexical meaning as the English verb HAVE in that they both denote "to possess". When used as a grammatical item, the Chinese *yǒu* serves as an existential verb (Li & Thompson, 1981:510-516; Her, 1991). This is unlike English, in which the existential counterpart is the copula BE. In the colloquial English varieties of Malaysia and Singapore, *yǒu* has undergone relexification, and its semantic and syntactic properties are

---

[21] At this stage, it remains unclear if BE was correctly inflected for tense or if its present form *is* was represented as a default/base form in the interlingual repertoire of the L1-Chinese.

masked under the English derivative GOT (Bao, 2005, 2014; Lee et al., 2009). However, the phonetic string of the English-lexifier adopted in #16 through #18 in Table 7.10 (**§7.3.3.2.**; repeated in (48) below) was not GOT but its competitor HAVE, which bears the existential reading:

(48) a.   and the bench there **have** [*] another young male adult where wearing earphone.

b.   <and when left> [///] total **have** [*] four biscuit inside the packing.

c.   and then <when he> [//] after she sat down and saw her bag inside **have** [*] a new &-uh, is &nn like didn't open before one@s$co with original packing cookies.

For one thing, it remains unclear as to why GOT was chosen over HAVE in relexifying *yǒu* in CSE (Bao, 2014:150) and CME. But in the interlanguage repertoire of Participant MM055, the Chinese-influenced lexical entry of HAVE might have been a stronger competitor and thus undergone a similar process of relexification, overlapping with some functions of GOT, to say the least. In example #13 in Table 7.10 (**§7.3.3.2.**; repeated in (49) below), the functional HAVE once again seemed be used in place of GOT in that, this time, it gave an emphatic reading (see Lee et al., 2009:298-299 on GOT) like the English emphatic DO:

(49) what happen [: happened] [* m:0ed] is [//] <by is> [//] <what &+happe> [//] at the beginning is an old woman go [: going] [* m:0ing] to a somethings [: something] [* m:+s] like a train station and go [: went] [* m:base:ed] to the vending machine, which **have** [*] sell the cookies.

However, concerning #14 (Table 7.10 under **§7.3.3.2.**; repeated in (50) below), it is difficult to determine whether the suppliance of HAVE was intended to mark emphasis like GOT or was simply a speech error:

(50) and so she have [*] take out her coin to buy the cookies.

Meanwhile, #15 (Table 7.10 under **§7.3.3.2.**; repeated in (51) below) presents a clear case of substitution error where HAVE was used instead of the copula BE. In this instance, it was no longer a matter of L1 influence but more generally an L2 learner error.

(51) and so she keep [: kept] [* m:base:ed] on &+s &-like 0cop angry and knocking the
vending machine and finally **have** [*] successful get the cookies.

Lee & Huang (2004), in their investigation of the acquisition of BE in L2-English, observed that 61% ($n = 20$) of substitution errors ($n = 33$) produced by Hong Kong Chinese children did involve the use of HAVE, as well as DO, in place of BE (ibid., pp.220-221). The authors explained that influence from L1-Chinese was an unlikely factor in this case because Chinese "does not have the equivalent auxiliaries of *have* and *do*". Hence, having to resort to other language learning strategies, the Chinese children in Lee & Huang might have misanalysed HAVE and DO and used them inappropriately instead of BE (ibid., p.224). The substitution error in (51) above might reflect such a case as well, but it is difficult to justify this based on just one instance. Notwithstanding, the non-target uses of HAVE by Participant MM055 demonstrate that while crosslinguistic influence might have played a main role in many of the instances, other factors such as general L2 learning strategies should not be overlooked, too.

Having discussed the linguistic performance of the L1-Chinese group, the next group to examine is the **L1-Malay** group. One of the finiteness features that posed learnability issues to these learners of L2-StE was DO-support, specifically in relation to the commission error of double-tense marking, as revealed in the **GJT**. Participants under this group were observed to consistently reject the standard constructions of DO-support in favour of the ungrammatical ones where both DO and the lexical verb were doubly inflected for tense. Under these conditions, the rating performance of the L1-Malay group was significantly divergent from that of the L1-BritE group.[22] As Malay does not have a counterpart of DO-support in its morphosyntactic inventory, crosslinguistic influence from L1-Malay cannot account for the participants' high acceptance of double-tense marking. This calls for the need to look into other factors, such as overgeneralisation as a L2 learning strategy, which will be elucidated in **§9.2.2.**.

Another finiteness feature that the L1-Malay group seemed to struggle with is the StE tense marking. In the **GJT**, participants under this group did not consistently reject the omission of 3rd.SING.-*s*, thereby yielding a significant difference from the L1-BritE controls. However, when it came to the omission of past -*ed*, the L1-Malay group had no issue rejecting this ungrammatical feature. As Malay does not inflect for past or present tense (Goddard,

---

[22] This behavioural pattern was also observed with the L1-Chinese group. However, as double-tense marking is a problematic feature that also affects the L1-MalE(+) cohorts, this issue will be addressed as a pervasive feature in response to RQ2 in the next subsection (**§9.2.2.**).

1996:431; Wong, 2012:7-8), the inconsistency in the L1-Malay group's rating performance between 3rd.SING.-*s* and past -*ed* was an unexpected finding. In the **NT**, although the suppliance of tense omissions by the L1-Malay group (10.7%) was relatively higher than that of the L1-MalE(+) cohorts, it was, impressively, half of that of the L1-Chinese group (20.7%). In this regard, L1 influence may not be as relevant as a main factor that can sufficiently account for the linguistic behaviour of the L1-Malay group.

Zeroing in on the behavioural discrepancies between the L1-Malay and L1-Chinese groups, it is clear from both GJT and NT that the former cohort outperformed the latter counterpart across the board. These group differences pose an interesting inquiry, given that both Malay and Chinese do not have the grammatical equivalents of finiteness in StE, and that both languages of the respective Austronesian and Sino-Tibetan families are typologically distinct from each other and from English (Germanic). A likely explanation regards the quantity of input in L2-English which L1-Malay and L1-Chinese participants had received since childhood. Under the national education system in Malaysia, Malay and English are compulsory language subjects to be learned from primary education onwards. At national (Malay-medium) primary schools (in Malay: *Sekolah Kebangsaan*), children are merely – if not minimally – required to take Malay and English lessons. However, at national-type (Chinese/Tamil-medium) primary schools (in Malay: *Sekolah Jenis Kebangsaan*), children are additionally required to learn the school's vernacular language viz., Mandarin Chinese or Tamil. This means that the instructional hours for the Malay and English subjects have to be reduced in order to incorporate lessons on the vernacular language into the classroom timetable. Table 9.1 presents the time (in minutes per week) allocated for the English Language subject imposed by the former and current primary school curricula.

Table 9.1: A comparison of time (in minutes per week) allocated for the English Language subject in the former KBSR and current KSSR curricula

Curriculum	KBSR (former)		KSSR (current)	
	National school	National-type school	National school	National-type school
Level 1 (*Primary 1 – 3*)	240 mins	60 mins	300 mins	150 mins
Level 2 (*Primary 4 – 6*)	210 mins	90 mins	300 mins	180 mins

In the former Integrated Curriculum for Primary Schools (in Malay: *Kurikulum Bersepadu Sekolah Rendah*, hereafter *KBSR*), students at national schools received more than half of the time allocated for English than their national-type school peers across the school levels. In the current Standards-Based Primary School Curriculum (in Malay: *Kurikulum Standard Sekolah Rendah*, hereafter *KSSR*), although classroom time for English has relatively increased for both school types, students at national schools still receive more exposure to English as their peers at national-type schools. As pointed out in many studies (e.g., Blom, 2010; Kuo et al., 2020; Paradis et al., 2011; Tsimpli, 2014; Unsworth, 2014), linguistic input and exposure, especially during childhood, play a crucial role in the development of the target language. Therefore, given the higher number of hours of input in StE received in the classroom setting, participants who attended national schools – most of whom were L1-Malay learners – might be at a greater advantage than those who went to national-type schools, such as the L1-Chinese learners. This increased amount of input would have therefore helped accelerate the development of L2-StE amongst L1-Malay speakers. Learners of L1-Chinese, on the other hand, might have faced an acquisitional delay due to the restricted amount of input received in class, which would in turn affect their development into the steady-state in L2-StE. That said, we have only focused on classroom input. As the participants recruited in this study were all adults, it was beyond the scope of this thesis to elicit the amount and type of input they had received outside the classroom in their younger years.

To sum up, we have observed how L1 transfer clearly affected the linguistic judgement and production of finiteness in StE. The divergent performance of the L1-Chinese group implicated negative transfer from the L1. Examples included the acceptance of stative progressivity and the omissions of copula BE and auxiliary DO (a.k.a. in-situ yes/no-questions) in the GJT, as well as the overuse of BE and non-target uses of HAVE (by one speaker) in the NT. Amongst the L1-MalE(+) groups, there were certain non-target features that seemed to persist in their performance on finiteness. For instance, the lack of rejection of DO omission (by the L1-MalE group) and double-tense marking in DO-constructions (by the L1-MalE and L1-MalE-Malay groups) in the GJT were speculated to be influenced by CME. Notwithstanding, that the L1-MalE(+) cohorts generally outperformed their L2-MalE peers across the board strongly indicates that their linguistic performances were more likely facilitated than interfered by the presence of English (potentially including CME) in their L1. Furthermore, it has also been demonstrated that crosslinguistic influence alone does not suffice

196

in accounting for the overall performance of the L1 groups. Other factors, such as input and L2 learning strategies, have also been implicated to be at play. For example, that the L1-Malay group generally outperformed its L1-Chinese counterpart might have been partially due to the role of input, as illustrated by the number of instructional hours allocated for the English Language subject in primary education. In another instance, the high acceptance of double-tense marking by the L1-Malay group in the GJT seemed to reflect L2 learning strategies, which will be discussed further in the following subsection. All in all, L1 transfer has been showed to serve both facilitative and adverse roles in the linguistic performances of the Malaysian L1 cohorts. However, there are also other factors such as input and L2 learning strategies that cannot be dismissed because they too have a part to play in the development of the target language.

9.2.2. RQ2: What is the contribution of CME on the error types of StE?

This inquiry sought to investigate the extent to which CME would affect the acquisition of StE. For one thing, it is not easy to distinguish between CME features and non-target-like features that resemble L2 errors in the Malaysian context. This is because CME itself has, for the most part, undergone restructuring as a result of substrate transfer and other L2 learning mechanisms (as demonstrated in **Chapter 2**). For another, given the prevalence of CME in the local milieu, non-target features that seem to persist across the English-speaking communities regardless of their L1s might allude to the pervasiveness of the nativized variety. In relation to the main study, morphosyntactic features exclusive to CME (and/or CSE), such as GOT, would faithfully serve as indicators of CME influence if they were used extensively across the Malaysian L1 groups. However, such instances are scant in the data. Accordingly, our endeavour to approach the current research inquiry was an exploratory one. Having gleaned through the linguistic data, a couple of linguistic patterns shall be brought to the fore. They are: i) the attestation and/or lack of rejection of double-tense marking in DO-constructions (mainly in the GJT), and ii) the low acceptance of the standard copular constructions in the GJT. In both instances, *all* Malaysian L1 groups, including those who acquired English as their L1, performed divergently from their L1-BritE counterpart. These instances shall be expounded in relation to the current inquiry. Before wrapping up this subsection, a final topic to be discussed concerns the plausibility of having CME as (part of) one's L1 and how the nativized variety might be able to answer the better linguistic performances of the L1-MalE(+) groups. Coming from an

acquisitional standpoint, this will challenge the perceived linguistic deficiency of CME in the endeavour to vindicate its utility in facilitating the acquisition of StE.

Double-tense marking: Double-tense marking in DO-support is a non-target feature that seemed to persist across the Malaysian L1 cohorts. In the **GJT**, the L1-Malay and L1-Chinese groups consistently rejected the standard condition of DO-support in strong favour of the ungrammatical condition of double-tense marking, thereby yielding statistically significant differences from the L1-BritE cohort. A facilitative effect was observed amongst the L1-MalE(+) counterparts but only to the extent of the grammatical condition of DO. Unlike the L1-BritE group, the L1-MalE and L1-MalE-Malay cohorts showed significantly greater acceptance towards double-tense marking, placing themselves more or less on par with their L2-MalE peers. Only the L1-MalE-Chinese group behaved like L1-BritE in that it consistently rejected double-tense marking in favour of the standard condition. Amongst the Malaysian L1 groups that accepted double-tense marking, it was the L1-Malay group that gave the highest average rating to this ungrammatical condition. This pattern was indeed unexpected because Malay does not inflect for tense and does not have an equivalent of the StE DO-support. However, the L1-Malay speakers' preference for double-tense marking in the GJT was not reflected in the **NT**. In this latter task, their suppliance of double-tense marking was, surprisingly, the lowest (2.03%, $n = \frac{3}{148}$) amongst the Malaysian L1 cohort. And while the L1-MalE and L1-MalE-Malay groups exhibited greater leniency towards double-tense marking than their L1-MalE-Chinese peers in the GJT, it was now the L1-MalE-Chinese's turn in the NT to produce the highest proportion of double-tense marking (5.45%, $n = \frac{3}{55}$), albeit relative, as compared to their L1-MalE (2.99%, $n = \frac{2}{67}$) and L1-MalE-Malay (3.9%, $n = \frac{3}{77}$) peers. The L1 group supplying the highest rate of double-tense marking was the L1-Chinese group, which was at 7.3% ($n = \frac{10}{137}$). This was another unexpected finding because, like Malay, Chinese does not have the counterparts of DO-support and tense inflections, and hence L1 speakers of this language were not expected to produce the most instances of double-tense marking. Indeed, one may remark that the proportions of this commission error produced by the Malaysian participants in the NT were generally low (i.e., below 10%). However, it could be that most of the standard constructions of DO-support were formulaic chunks that involved high-frequency

verbs such as *know* and *have* (e.g., did not know; does not have) (Hilbert, 2008:279).[23] Notwithstanding, given their significantly divergent performances in the GJT, the strong preference for double-tense marking, as shown by most of the Malaysian L1 groups, is telling about how DO-support might be represented in their mental grammar.

Following Adger (2003), the Pronounce Tense Rule (PTR), which is a "language-specific pronunciation rule" in English (ibid., p.194), states that tense is pronounced on the verbal element in *v* only if *v* heads the sister of T. However, if the (T[tense]… *v*[uINFL: _]) chain is broken by an intervening element that is the negation *not*, then the PTR can no longer apply (ibid., pp.192-193). Consequently, DO-support kicks in and gets inserted into T, bearing its tense feature on behalf of the verbal element in *v*. Hence, for Malaysians who accepted and/or supplied double-tense marking in the current study, it could be that they had not fully acquired the PTR in relation to DO-support. This would be especially challenging for L2 learners (e.g., the L1-Malay and L1-Chinese speakers). This is because, in addition to learning the L2-English auxiliary DO from scratch, these learners were also faced with the additional task of learning the PTR, failing which they might overgeneralise tense marking on the lexical verb in DO-support constructions. Meanwhile, Malaysians who acquired English as their L1 might be faced with both standard and non-standard (i.e., double-tense marking) options and thus would have acquired the PTR as an optional rule.

Could CME have contributed to the acceptance and use of double-tense marking in the main study? Since double-tense marking has been attested in this local variety (Pillai & Greig, 2020), it seems likely. Given the prevalence of CME, not only would L1-MalE speakers have potentially been exposed to this non-standard feature, but those acquiring English as an L2 would, too. This explains why double-tense marking was accepted in the GJT and used in the NT by both L1- and L2-MalE Malaysian cohorts. However, double-tense marking is not exclusive to CME. It has also been attested in other English varieties such as Hong Kong English, Pure Fiji English, Aboriginal English, Tristan da Cunha English (Kortmann et al., 2020), Black South African English (Mesthrie, 2014:139), and Indian English (Stringer, 2015:125), as well as amongst L1-English children (Hurford, 1975; Rowland, 2007; with SLI: van Der Lely & Battell, 2003; van Der Lely et al., 2011) and L2-English learners from Expanding Circle countries (Al-Makatrah et al., 2017; Platt, 1989:10). To wit, double-tense marking in DO-constructions is an instance of overgeneralisation error, where acquirers of L1-

---

[23] Ineptly, the current research did not take such formulaic chunks into account when quantifying instances of DO-support. This has formed a limitation which this study has to bear.

and L2-English productively mark tense on the lexical verb in addition to the auxiliary DO. This pattern occurs even if it is not necessarily found in the input, just as in the case of monolingual English children. Notwithstanding, while child errors often "show a gradual decline over the acquisition process" (Rowland, 2007:130), errors made by learners of L2-English might persist into the steady-state grammar of the interlanguage. And if errors like double-tense marking have fossilised in the interlanguage and are used productively at a societal level, then they become a feature of the language variety spoken by the local speech community (see Hilbert, 2008 on a similar explanation on interrogative inversion as a product of second language acquisition in Indian and Singapore English).

As double-tense marking in DO-constructions has paved its way into the morphosyntactic inventory of CME, it likely explains its pervasiveness as a CME feature, as indicated in the acceptance and production of both L1- and L2-MalE cohorts in the main study. At the same time, it is worth bearing in mind that double-tense marking is a fossilised overgeneralisation error that reflects the status of English as an L2 in the Malaysian milieu.

Standard copular constructions: In the GJT, all Malaysian L1 groups gave significantly lower ratings to the grammatical sentences of copula BE than their L1-BritE counterpart. This was an unexpected finding, given that most of the Malaysian groups – except for that of L1-Chinese – had no issue rejecting the omission of copula BE as being an ungrammatical feature in StE. A likely source of the rating differences concerns the use of adverbials in copular sentences. In terms of the experimental design of the GJT, the targeted length for each stimulus sentence was between 7 and 9 words. Therefore, when constructing copular sentences for the experimental task, adverbials and other modifying words such as attributive adjectives were added to achieve the minimum sentence length. For example, for copular constructions taking PP complements, adverbs were placed at the beginning of the sentence (e.g., "*Usually*, all photo albums are in the glass cupboard."). For copular constructions taking AdjP complements, two adverbs were used consecutively mid-sentence to modify the adjectival predicate (e.g., "All washing liquids are *actually quite* effective."). In Belletti and colleagues' (2007) study on L2-Italian, the authors found that adverbials tended to be produced at the end of sentences by both native (i.e., L1-Italian) and near-native learners of L1-English, whereas adverbials at sentence-initial and sentence-medial positions were hardly supplied by either group (ibid., p.671). In a different study, Hinkel (2003) observed that the adverbials supplied in academic writing by advanced learners of L2-English were usually those that were frequently used in conversations (ibid., pp.1065-1066). Based on these two studies, we speculate that the position (Belletti et al.,

2007) and usage frequency of adverbials (Hinkel, 2003) were potential factors affecting the Malaysians' ratings of the grammatical copular constructions in the GJT study. In other words, the lower ratings could be due to i) the placement of adverbs at sentence-initial positions (as in the case of copular sentences taking PP complements), or ii) the stacking of adverbs (as in the case of copular sentences taking AdjP complements), both of which are plausibly dispreferred if not infrequent in the CME discourse. That said, more empirical work is needed to investigate the acquisition of English adverbs by Malaysians to (dis)confirm these speculations.

CME in the L1: Finally, let us explore the likelihood of CME constituting (part of) the linguistic repertoire of L1-MalE(+) speakers and, accordingly, address the extent to which CME may affect (i.e., facilitate or interfere with) the acquisition of StE.

First of all, the assumption that CME comprises the *entire* L1 system of English-speaking Malaysians is ostensible. This is because it is likely that these speakers have been exposed to more than one type of input variety in English since early childhood. In Buschfeld's (2020, 2021) investigation of Singaporean children acquiring L1-English, the author argues that the intra-speaker variability observed in the children's use of subject pronouns and past tense inflections was partly due to the quality of input they received in their social surroundings. According to the author, the input was not necessarily restricted to the home environment but also determined by other sources such as input provided by peers, the materials children read, watch, and/or listen to, and the input received in the classroom and, more broadly, the wider community (ibid., 2020:48-49; 2021:206). Accordingly, Buschfeld "[does] not expect the children to copy their parental input one-to-one" (ibid., 2020:67), but that they would "pick and mix" features from different linguistic varieties that are available in the input environment (ibid., 2020:262; see also Schreier, 2014). Therefore, due to the diversity of input, Buschfeld (2020:67) refuses to classify which English variety was acquired by the Singaporean children in her study. While I concur with Buschfeld that it is difficult to pin down one specific variety that comprises the linguistic repertoire of L1-MalE(+) speakers, I argue that it is not impossible for CME to form *part of* their mental grammar growing up. This is because CME is inextricably extensive in the Malaysian milieu. Even if it is not used in the home environment (e.g., due to attitudinal reasons; see Buschfeld, 2020, 2021), it is actively used in the wider, L1/L2-English-speaking communities in Malaysia. Moreover, according to the findings on lectal preference from the sociolinguistic survey (see **§8.3.3.**) – we will come back to this again in **§9.3.1.** – the majority (i.e., 80%) of the Malaysian participants, regardless of whether they spoke English as their L1 or L2, chose CME over StE were they to speak English with their family. What this

pattern suggests is that i) it is likely for CME to be used in the home of Malaysians irrespective of the status of English being an L1 or L2; and ii) it potentially alludes to the speech of caregivers who choose to speak English to their children (see Schreier (2014) and Gupta (1994) on the role of caregivers in shaping child speech).

Having argued for the plausibility of CME being part of the L1 system of L1-MalE(+) speakers, the next question to address is to what extent CME plays a role in the acquisition of StE for these speakers. As discussed in RQ1, although the L1-MalE(+) groups were statistically different from their L1-BritE peers in their linguistic judgement and production of (certain) finiteness features, they generally outperformed their L2-MalE peers. Assuming that CME was one of the input varieties that L1-MalE(+) participants grew up with, it could have potentially facilitated their learning experience with StE, especially when they started receiving explicit instruction on StE during their formal education. In an intervention study conducted by Yiakoumetti (2006) on the acquisition of Standard Modern Greek (SMG), the author observed that Cypriot-speaking learners who undertook a bidialectal language programme exhibited significant improvement in their use of SMG over a series of oral and written tests. Contrariwise, those who did not receive the treatment did not show much change in their performance by the end of the study. Although Yiakoumetti did also notice that certain features of the Cypriot dialect [e.g., the use of /-n/ in the accusative singular in students' speech and writing (ibid., pp.309-310)] persisted throughout the study, she argued that "students' greater awareness of what constitutes dialectal interference may have simply assisted their efforts to reduce it" (ibid., p. 307). In other words, the improved performance shown by the Cypriot participants in Yiakoumetti's study might be facilitated by their metalinguistic awareness between the Cypriot dialect and SMG. In the current study, there, too, were certain finiteness features (e.g., DO-support) that posed greater learnability issues to the L1-MalE(+) speakers. However, looking at the bigger picture, the L1-MalE(+) groups were ahead of their L2-MalE counterparts in their linguistic performance on finiteness in StE. Unlike in Yiakoumetti's study, the L1-MalE(+) cohorts – alongside other L2-MalE peers – might not have undertaken any bidialectal programme throughout their formal education. Nonetheless, they might have some linguistic awareness of the differences between the colloquial and standard varieties of English, since StE is explicitly taught in the classroom. Subsequently, these learners might know when (not) to use non-standard features of English in the appropriate context.[24]

---

[24] Assuming that the L1-MalE(+) speakers had implicit knowledge of CME, it would not be surprising if their linguistic awareness was subconsciously heightened when StE was explicitly taught in class.

To sum up, we have looked at a couple of non-target-like instances where most, if not all, Malaysian L1 groups behaved significantly differently from the L1-BritE group. These anomalous linguistic behaviours were speculated to be associated with CME. However, given how tricky it can be to distinguish between CME features and L2 errors, such conjectures could not be justified with confidence. Concurrently, given the extensive use of CME, we have speculated the plausibility of CME constituting (part of) the linguistic repertoire of L1-MalE(+) speakers. If this were true, then the better linguistic performances exhibited by the L1-MalE(+) cohorts, in comparison to their L2-MalE peers, would shed light on the role of CME in facilitating – rather than impeding – the acquisition of StE.

9.2.3. RQ3: Are there task effects on finiteness?

The linguistic performances between the GJT and NT were compared for task effects, with the rationale being that these two tasks might tap into explicit and implicit linguistic knowledge (see Bowles, 2011; Ellis, 2005; van Osch et al., 2018), respectively. This is because the GJT is a perception comprehension task whereas the NT is a production task. Therefore, any divergent performances observed would signify the effects of task modality. To recapitulate, the overall findings in **§7.4.2.** revealed a significant positive correlation between the GJT and NT under the grammatical condition ($\tau = 0.314$, $p < 0.001$), and a significant negative correlation between these two tasks in the ungrammatical condition ($\tau = -0.172$, $p = 0.002$). Upon closer scrutiny, it was the L2-MalE cohort through whom the significant correlations were exhibited. In the grammatical condition, a significant positive correlation between the two tasks was observed for the L1-Chinese group ($\tau = 0.236$, $p = 0.031$) whereas in the ungrammatical condition, a significant negative correlation was found for the L1-Malay group ($\tau = -0.518$, $p < 0.001$). While the L1-MalE-Malay group shared a similar pattern (i.e., significant negative correlation: $\tau = -0.328$, $p = 0.044$) with its L1-Malay counterpart, it was the only group in the L1-MalE English Status cohort that yielded a significant correlation. Altogether, these findings reveal two points. Firstly, higher accuracy in the GJT was associated with higher suppliance of grammatical finiteness (this positive correlation was significant amongst the L1-Chinese group). At the same time (second point), higher accuracy in the GJT was associated with lower suppliance of ungrammatical finiteness (this negative correlation was significant amongst the L1-Malay and L1-MalE-Malay groups). In other words, the improved linguistic performance of the L1 groups in the GJT and in the NT corroborated each other, indicating no task effects

in this study. However, why there were significant correlations in certain L1 groups under different conditions of grammaticality is something that this thesis cannot account for. Therefore, we remain agnostic about these observations.

9.2.4. Revisiting the SLA theories

Having discussed findings from the GJT and NT, this subsection explicates them with reference to SLA theories, which postulate different degrees of accessibility to UG in L2 acquisition. Accordingly, this subsection reviews the Full versus Partial Access views – alongside their respective theories – in light of the findings and lends support to the former view.

The Full Access (FA) view proposes that UG is fully accessible in L2 acquisition and envisages the possibility of achieving successful or target-like acquisition. Advocating this view, Schwartz & Sprouse's (1994; 1996) Full Transfer Full Access Hypothesis (FTFA) anticipates L1 transfer in the initial stages of L2 acquisition but maintains that restructuring of the interlanguage is fully governed by UG. Notwithstanding, it does not warrant native-like attainment in the steady state of the L2 grammar due to influence from different L1 grammars at the starting point of L2 acquisition. In the current study, the L1-MalE(+) and L1-Malay cohorts in many instances exhibited performances similar to their L1-BritE counterpart, and their overall performances seem to corroborate FTFA. The L1-MalE(+) cohorts generally – apart from three instances – did not struggle with their introspection of grammatical and ungrammatical finiteness in the GJT. Although they were all statistically different from the L1-BritE group in the NT, the suppliance of grammatical finiteness was nevertheless above 90%, indicating successful acquisition. What this further suggests is that if CME comprised part of these speakers' L1, then its optional finiteness features had already been successfully reset to obligatory by the steady state of the StE interlanguage grammar. Meanwhile, the fact that the L1-Malay group outperformed its L1-Chinese counterpart and was about on par with its L1-MalE(+) peers was a remarkable finding, considering the fact that both Malay and Chinese do not have the grammatical counterparts of finiteness as English has. Although the L1-Malay participants were observed to be struggling with tense inflections in the GJT (i.e., 3rd.SING.-*s*) and NT (88.6% accuracy rate), they generally did not have issues with the suppletive morphemes of finiteness in StE. This indicates that they had successfully acquired grammatical morphemes (i.e., progressive BE, auxiliary DO) not instantiated in their L1, and that this success could be fully sanctioned by UG.

Two other theories corroborating the FA view are the Missing Surface Inflection Hypothesis (MSIH) by Prévost & White (2000) and the Prosodic Transfer Hypothesis (PTH) by Goad & White (2006). According to MSIH, failure to supply target-like morphological exponents is not a reflection of syntactic deficits underlying the L2 grammar but of issues pertaining to "processing reasons" or "communication pressures" (Prévost & White, 2000:129). PTH differs slightly from MSIH in that the issue specifically pertains to phonological constraints imposed by the L1. Following these hypotheses therefore means that there will be differences in linguistic performance depending on the modality of the experimental tasks employed (Prévost & White, 2000:129; Slabakova, 2016:193). One instance supporting these hypotheses concerns the L1-Chinese learners' linguistic performance on tense marking in L2-StE. On the one hand, this group exhibited target-like judgement in the GJT. On the other hand, it supplied the highest proportion of tense omissions in the NT. As Chinese only allows for sonorant codas in its phonological representation (Goad & White, 2006:259), acquiring tense inflections in L2-StE would pose a learnability issue for L1-Chinese learners at the phonological level. At first blush, this could be the case for our L1-Chinese speakers in the NT. However, to say that they successfully acquired finiteness in L2-StE based on their target-like performance in the GJT would be impetuous. This is because there are several other instances where the L1-Chinese group performed divergently from the rest of their L1 peers, which MSIH and PTH cannot account for.

As an alternative to the FA view, the Partial Access (PA) view postulates that morphosyntactic features not instantiated in the L1 can never be successfully acquired due to the inaccessibility of UG in certain linguistic (e.g., syntactic) domains. Corroborating this view include the Failed Functional Features Hypothesis (FFFH) by Hawkins & Chan (1997) and the Interpretability Hypothesis (IH) by Tsimpli & Dimitrakopoulou (2007). Although both theories expect L1 transfer to take place throughout L2 development, they also anticipate restructuring in the interlanguage. However, if the linguistic feature in question is morphosyntactic in nature, then a deficit in its representation is predicted due to the partial availability of UG. L2 learners will consequently have to resort to other acquisitional strategies, such as relying on interpretable features (Tsimpli & Dimitrakopoulou, 2007), in the attempt to learn the grammatical component. The divergent performances of the L1-Chinese group in the GJT and NT provide a strong case for the PA view. For one thing, negative transfer from Chinese seemed to persist in the linguistic judgement and suppliance of these speakers. Instances in the GJT include leniency in the ratings of omissions of copula BE and auxiliary DO (a.k.a., in-situ yes/no-question constructions) and the commission error of stative progressivity, whereas

instances in the NT include the overgeneration of BE (especially in BE + BELONG structures) and the non-target uses of HAVE (as supplied by one participant). For another, errors not instigated by crosslinguistic influence, such as the overgeneration of progressive *-ing* (9.66%), were supplied the most by the L1-Chinese group. These observations not only implicate failure to reset parameters (e.g., from optional to obligatory for copula BE; from [+stative] to [-stative] in progressive BE), but also failure to acquire new morphosyntactic features (e.g., progressive *-ing*, auxiliary DO) successfully in L2-StE, thereby lending support to the PA view.

As we compare between the FA and PA views, the latter view seems to be inadequate in accounting for the different linguistic performances observed across the Malaysian L1 cohorts. Addressing first the performances between the L1-Malay and L1-Chinese cohorts, the PA view predicts that both groups would behave similarly in their performances on grammatical finiteness in L2-StE because this linguistic phenomenon, alongside the features associated with it, is not instantiated in the Malay and Chinese L1s. However, that the L1-Malay group outperformed its L1-Chinese counterpart across the board and was about on par with the L1-MalE(+) cohorts counters the predictions made by the PA view. Following this, the FA counterpart, particularly FTFA, serves as a better theory in accounting for the different linguistic behaviours exhibited by these L2-MalE groups. For one thing, FTFA anticipates that different acquisitional trajectories are driven by different L1s that constitute the initial stage of the L2 grammar. For another, it also acknowledges the importance of input in facilitating L2 development (Schwartz & Sprouse, 1996). As explained in **§9.2.1.**, there is a difference in the instructional hours allocated to StE as a language subject between national (i.e., Malay-medium) and national-type (e.g., Chinese-medium) primary schools. Given that these schools are usually attended by the Malays and Chinese, respectively, this means that the higher amount of input received by the L1-Malay participants might have accelerated their learning of StE in comparison to their L1-Chinese peers.

Now, there are other anomalous linguistic behaviours – observed with the L1-MalE group (e.g., acceptance of DO-omission) and across all Malaysian L1 groups (e.g., double-tense marking) – that might at first blush signal representational deficits in the finiteness features involved. However, it seems that these learners from different L1s might have developed a learner-cum-indigenised variety where there could be more than one way to express finiteness in StE. Therefore, it does not necessarily reflect linguistic deficiency per se but rather combines both standard and non-standard features such as double-tense marking which are borrowed and tolerated by CME which is in itself an L2 variety.

To conclude, although the linguistic performance of the Malaysian L1 groups can be accounted for by both FA and PA views, they generally tend to favour the former view, and, in particular, FTFA.

9.3. Discussion of findings from the sociolinguistic study

9.3.1. RQ4: To what extent is CME perceived as functionally different from StE in Malay

Finally, the current study sought to inquire the roles of the indigenised and standard varieties of English in the Malaysian ecology. To address this research inquiry, three subsidiary questions were further presented:

(i)   To what extent were Malaysians aware of the indigenised variety (i.e., CME) spoken in Malaysia?

(ii)   What were the attitudes of Malaysians towards CME and StE in relation to different themes (i.e., nationalism, friendliness, educatedness, purism, pedagogy)?

(iii)   What were the Malaysians' preferences between CME and StE across different social contexts (i.e., family, friends, colleagues, English Language teacher, other teachers, Malaysian strangers, foreigners)?

Following this, data from the study's sociolinguistic survey are discussed in relation to the subsidiary questions above.

Addressing the question in (i), the study saw an overwhelming majority of the Malaysian participants ($n = 140$; 96.55%) attesting to the presence of an indigenised variety in Malaysia, which was most commonly known *by* them as "Manglish" ($n = 101$; 48.3%). This shows that the participants were already aware of the different varieties of English spoken locally.

Turning to question (ii), the attitudes of the Malaysian participants towards CME and StE were elicited using five pairs of contrastive statements that touched on the themes of national identity, friendliness, educatedness, language purism, and language pedagogy.

To reiterate the findings in **§8.3.2.**, CME received relatively high (averaged) ratings of agreeableness in relation to national identity and friendliness. StE, on the other hand, was perceived to be associated with well-educatedness, purity as a language variety, and appropriateness as a language model in the learning of English. These findings seem to lend support to the notion that the colloquial and standard varieties bear the sociocentric traits of solidarity (indicated by *national identity* and *friendliness*) and status (indicated by *well-educatedness*), respectively (Dragojevic, 2017). However, the average scores of contrastive statements (except for that on *pedagogy*) leaned slightly towards disagreeableness. For instance, the participants tended to disagree – albeit slightly – with the contrastive statements where StE would conceal one's national identity and/or create social distance. This indicates that they did not unanimously view StE in a bad light when it was stated negatively on the solidarity dimension. Also, slight disagreement was shown towards the statements where CME was thought to be used by less educated people or was "bad English", thereby suggesting that the participants did not generally associate CME with poor education and, importantly, were not as accepting of the idea that CME was deficient as a language variety. More interestingly, concerning the paired statements on pedagogy, although the participants showed a tendency – albeit weak – towards avoiding the use of CME in the endeavour to improve on StE, they were neutral to the idea of using the local variety as a facilitative linguistic tool in the learning of the standard counterpart.

In Ng & Diskin-Holdaway's (2021) attitudinal study on CME and StE, findings from the authors' matched-guise test revealed that, on the solidarity dimension, CME was indeed strongly associated with Malaysian-ness and slightly so with sociability. However, in terms of friendliness and likeability, it was rated as being almost equal with StE. Under the status category, StE was assigned higher ratings than CME in relation to the themes of educatedness, professionalism, and clarity, and significantly more so with the former two themes.

The findings of Ng & Diskin-Holdaway were, to a certain extent, in line with those from the current attitudinal survey, especially when it touched on the themes of Malaysian-ness (in relation to CME) and educatedness (in relation to StE). However, what our study has additionally shown is that the negative statements about CME (i.e., in relation to lower educational status, deficiency as a language variety, and avoidance of use in the learning of StE) and StE (i.e., in relation to social and identity distancing) were not overwhelmingly embraced by our participants, suggesting that the participants were not strongly prejudiced against CME or StE. What is more, with regard to language pedagogy, although there was no consensus as to whether or not CME could be used to facilitate the learning of StE, it

nevertheless implicates the participants' openness to entertain the plausibility of using CME as an additive tool in language learning.

Finally, addressing question (iii), which looked at the Malaysians' preferences between CME and StE in different social contexts, three key findings were observed:

- CME was a significantly popular choice for interactions with family (80%), friends (86.9%), and Malaysian strangers (60%).

- On the flipside, StE was significantly favoured in interactions with teachers (English Language: 95.86%; non-English subjects: 72.41%) and foreigners (98.62%).

- There was almost a balance between the two lectal varieties when it came to interactions with colleagues, though CME was slightly preferred over StE at 55.86%.

These findings were generally in line with those in Tan & Tan (2008). One of the questions asked in the authors' survey to secondary school students in Singapore was the degree to which Singlish (i.e., CSE) would be used in different situations. The social settings posed under this question were constructed according to the types of audience (i.e., friends/classmates, family, teachers), location/setting (i.e., outside school, outside or during lesson time in school), and school subject (i.e., Mathematics, English). Accordingly, the authors observed that Singlish was strongly preferred if it was used with friends/classmates and family. Between Mathematics and English teachers, it was markedly more likely for Singlish to be used with the former group of teachers. With regards to location, the use of the local variety would be more appropriate if used outside school whereas within the school boundaries, outside lesson time would be a better time to use it than in class. Following this, Tan & Tan showed that the Singaporean students in their study "[had] a sense of when Singlish might be appropriate and when it might be less so, and therefore [showed] an awareness and sensitivity to audience and settings" (ibid., pp. 473-474). Referring back to the current sociolinguistic study, our results revealed that CME was generally favoured in intimate (i.e., family, friends) and/or informal (e.g., Malaysian strangers) settings, whereas StE was chosen for formal and/or more socially distanced situations (e.g., teachers, foreigners). What is more, the overwhelming preference for StE over CME in interactions with foreigners indicates that the use of the former variety was better suited to foster mutual intelligibility with those who might not be familiar with the indigenised variety.

To sum up, we have observed that the majority of the Malaysian participants were aware of the indigenised and standard varieties of English spoken in Malaysia. While both varieties were generally ascribed to the corresponding traits of solidarity and status, neither of them was unanimously discriminated by the participants. This is because CME and StE were rightly understood to serve different functions in the local milieu, whose choice was contingent upon the formality of the social setting, social distance, and sociocultural familiarity shared between interlocutors.

9.4. Study limitations and future research

The main study of this PhD research employed a concurrent embedded design to investigate the acquisitional and sociolinguistic aspects pertaining to the linguistically diverse ecology in Malaysia. Notwithstanding, there are several limitations that are worth bringing to the fore in the hope that they will inform and create avenues for future research.

The scope of the acquisitional study zeroed in on finiteness in StE. Accordingly, implications were drawn about the acquisitional status of the Malaysian L1 groups based on their performance on this morphosyntactic phenomenon. There are, however, other features beyond finiteness or beyond morphosyntax that are worth investigating to (dis)confirm the claims made in this thesis. For instance, interface phenomena that tap into language-internal (e.g., syntactic) and language-external (e.g., discoursal) domains require much input to warrant successful acquisition, and hence have been argued to pose greater learnability issues to L2 learners and even early bilinguals (Sorace, 2005; Sorace & Serratrice, 2009; Yuan & Dugarova, 2012; see also Tsimpli, 2014). Given the highly variable input in World Englishes contexts like Malaysia, research into interface phenomena would therefore provide further insights into the acquisition of English in these settings.

Another drawback is that the main study only considered Malaysia as the only site to investigate L2 acquisition. It would be good for future studies to compare between two (or more) New Englishes or even L2 varieties in Expanding Circle countries to examine the extent to which learners are similar or divergent in their linguistic performances in (L2-)StE. Indeed, the linguistic mechanisms involved (e.g., crosslinguistic transfer and other language learning strategies) in paving the acquisitional trajectory of the StE interlanguage might be similar across varieties in the Outer and Expanding Circles. However, other sociolinguistic dimensions which lie at the crux of the evolution of indigenised varieties may shed light on the linguistic

complexity of these varieties which L2 varieties in Expanding Circle contexts might not be able to account for.

Additionally, the sample sizes of the Malaysian L1 groups posed a statistical limitation in the study. The size of a sample has been argued and proven to affect the statistical outcomes of the research (Button et al., 2013; Donner & Eliasziw, 1987; Field et al., 2012; Porte, 2002). Correspondingly, small sample size would generate low statistical power, which would in turn lead to a misinterpretation of the statistical findings (i.e., Type I error) (Button et al., 2013:365). In the current study, the L1-MalE ($n = 19$), L1-MalE-Malay ($n = 21$), and L1-MalE-Chinese ($n = 13$) groups comprised relatively small sample sizes. L1-MalE speakers were included in the analysis because they were needed for the main study. Furthermore, as it was difficult to find L1-MalE-only speakers, simultaneous bilinguals were recruited as well. Nevertheless, this justification does not diminish the fact that the low powers yielded by these small sample sizes would give rise to the misinterpretation that any statistical outcomes generated would reflect a true effect (see Button et al., 2013).

Finally, concerning the demographics of the study's participants, these participants were all adults and predominantly students enrolled at elite universities. For the Malaysian recruits, they were already of upper intermediate or advanced proficiencies in English by the time they partook in the study. Although it was indeed the aim of this thesis to recruit participants of these proficiency levels, it should be mentioned that the linguistic performance of our participants is not representative of the wider Malaysian population. Wider ranges of English proficiency, age, and/or other sociolinguistic factors not mentioned herein are variables worth considering in future work to account for a better representation of the Malaysian population.

10 Conclusion

This thesis set out to address the paradigm gap between SLA and WE by adopting an acquisitional-cum-sociolinguistic approach to investigate language acquisition and sociolinguistic sentiments that revolved around the Englishes in Malaysia (viz., CME and StE). The linguistically diverse ecology of Malaysia has served as an interesting site of study because it not only hosts an eclectic mix of local and heritage languages but has also provided fertile ground for the germination of English to take place. The emergence of CME as an indigenised variety may reflect the L2 status of English in Malaysia. However, it also unravels the identity rewritings of Malaysians as an endeavour to decolonise itself from the ex-colonial variety (viz., BritE). In this process, Malaysians have been granted full liberty to exercise their bilingual creativity to further propel the restructuring of English, making it their own. Concurrently, as the official – albeit restricted – status of English continues to be upheld, many attempts have been made at the ministerial level to reverse the declining proficiency in English as observed across the country.

Accordingly, this thesis examined the roles of L1 and CME in contributing to the ultimate attainment of StE. As the morphosyntactic phenomenon of interest was finiteness, a grammaticality judgement task (GJT) and a narrative task (NT) were employed to elicit the metalinguistic knowledge and use of features (i.e., copula BE, progressivity, DO-support, tense inflections) related to this phenomenon. The Malaysian participants who took part in the main study encompassed those who spoke Malay, Chinese, and MalE(+) as (one of) their L1(s). British (L1-BritE) monolinguals were also recruited to serve as controls. Although the overall results from the linguistic tasks showed that the Malaysian cohort were statistically different from their BritE counterpart, the L1-MalE(+) groups outperformed their L1-Malay and L1-Chinese peers across the board. Not only does this implicate their success in resetting the optionality feature in CME finiteness to obligatory in StE (corroborating the *Full Transfer Full Access Hypothesis* under the *Full Access* view), it also indicates the additive role of L1-MalE (under which CME is subsumed) in facilitating the acquisition of StE. On the other hand, the L1-Chinese participants performed poorly than the rest of their L1 peers on several occasions. Negative transfer from the L1 was salient amongst this cohort (e.g., tense omissions, overgeneration of BE, acceptance of stative progressivity, etc.), so were overgeneralisation errors resulting from other L2 learning behaviours (e.g., overuse of *-ing*). Therefore, the divergent performance of the L1-Chinese group suggests deficits in the syntactic representation

of finiteness in L2-StE (corroborating the *Partial Access* view). Meanwhile, the L1-Malay participants outperformed their L1-Chinese peers and were about on par with the L1-MalE(+) cohort. Their general success could lend support to the *Full Access* view, but it could also be due to the higher amount of instructional hours on StE they received as compared to their L1-Chinese peers in the (primary) classroom setting. In other words, for L1s that are typologically more distant than the target L2, the role of input might be an especially important factor in facilitating interlanguage development.

This thesis also conducted a sociolinguistic study to explore the attitudinal behaviours of Malaysians towards CME and StE. An overwhelming majority of participants in this study demonstrated awareness of these varieties being spoken in Malaysia. Furthermore, they showed no strong discrimination against either variety when being presented with statements on their assumed social prestige. Their acceptance of the two varieties was further supported by their preferences for these varieties in different social contexts. That is, CME was a preferred code of interaction in intimate and/or informal settings, whereas StE was deemed as a more suitable choice in more socio-culturally distanced and/or formal settings.

Altogether, what this thesis hopes to have achieved is that it has i) destigmatised any preconceived notion about CME being a sociolinguistically inferior variety of indigenised English, ii) demonstrated the additive role of CME in the acquisition of StE, and iii) vindicated and equally embraced the functional importance of both CME and StE as legitimate varieties in their own right.

References

Abdul Aziz, R. and Mohd Don, Z. (2014). The overgeneration of *be* + verb constructions in the writing of L1-Malay ESL learners in Malaysia. *Research in Corpus Linguistics*, 2, pp.35-44. Available at: https://ricl.aelinco.es/index.php/ricl/article/view/28 [Accessed 31-December-2022]

Abu Bakar, H. (2009). Code-switching in Kuala Lumpur Malay: The "Rojak" Phenomenon. *Explorations*, 9, pp.99-107. Available at: http://hdl.handle.net/10125/10712 [Accessed 31-December-2022]

Adger, D. (2003). *Core syntax: A Minimalist Approach*. Oxford: Oxford University Press.

Adger, D. (2007). Three domains of finiteness: a Minimalist perspective. In Nikolaeva, I. (Ed.). *Finiteness: Theoretical and Empirical Foundations*. Oxford: Oxford University Press, pp.23-58.

Adger, D. and Smith, J. (2005). Variation and the minimalist program. In Cornips, L. and Corrigan, K. P. (Eds.). *Syntax and variation: reconciling the biological and the social*. Amsterdam: John Benjamins Publishing Company, pp.149-178.

Adi Badiozaman, I. F. (2019). Rethinking English Language Education in Malaysia. *The Journal of Asia TEFL*, 16(1), pp.349-359. DOI: 10.18823/asiatefl.2019.16.1.24.349

Al-Makatrah, E. S., Mohd Yasin, M. S., Sulaiman, M. Z., and Al-Khawaldeh, M. (2017). The Markedness in Acquiring Do-support in Negation and Inversion by Adult Native-Arabic Speakers Learning English. *3L: The Southeast Asian Journal of English Language Studies*, 23(4), pp.143-156.

Alsagoff, L. and Ho, C. L. (1998). The relative clause in colloquial Singapore English. *World Englishes*, 17(2), pp.127-138. DOI: 10.1111/1467-971X.00087

Altenberg, E. P. and Vago, R. M. (2004). The role of grammaticality judgments in investigating first language attrition: A cross-disciplinary perspective. In Schmid, M. S., Köpke, B., Keijzer, M., and Weilemar, L. (Eds.). *First Language Attrition: Interdisciplinary perspectives on methodological issues*. Amsterdam: John Benjamins Publishing Company, pp.105-129.

Altman, C., Avraham, I., Meirovich, S. S., and Lifshitz, H. (2022). How do students with intellectual disabilities tell stories? An investigation of narrative macrostructure and microstructure. *Journal of Applied Research in Intellectual Disabilities*, 35(5), pp.1119-1130. DOI: 10.1111/jar.12997

Andersen, R. W. (1990). Models, Processes, Principles and Strategies: Second Language Acquisition Inside and Outside the Classroom. In VanPatten, B. and Lee, J. F. (Eds.). *Second Language Acquisition/Foreign Language Learning*. Multilingual Matters Ltd., pp.45-68.

Andersen, R. W. (1991). Developmental sequences: the emergence of aspect marking in second language acquisition. In Huebner, T. and Ferguson, C. A. (Eds.). *Crosscurrents in second language acquisition and linguistic theories*. Amsterdam: John Benjamins Publishing Company, pp.305-324.

Anwyl-Irvine, A. L., Massonié, J., Flitton, A., Kirkham, N. Z., and Evershed, J. K. (2019). *Gorilla in our midst: an online behavioural experimenter builder*. Behavior Research Methods. DOI: 10.3758/s13428-019-01237-x

Arunachalam, S. (2013). Experimental Methods for Linguists. *Language and Linguistics Compass*, 7 (4), pp.221-232. DOI: 10.1111/lnc3.12021

Austin, J. (2009). Delay, interference and bilingual development: The acquisition of verbal morphology in children learning Basque and Spanish. *International Journal of Bilingualism*, 13(4), pp.447-479. DOI: 10.1177/1367006909353234

Awang, H. W. (1981). *Some aspects of Malay relativization: A Transformational Account by Deletion*. PhD thesis: The School of Oriental and African Studies, University of London. Available at: https://eprints.soas.ac.uk/28602/1/10672761.pdf [Accessed 11-December-2022]

Bamgboṣe, A. (1998). Torn between the norms: innovations in world Englishes. *World Englishes*, 17(1), pp.1-14. DOI: 10.1111/1467-971X.00078

Bao, Z. (2001). The origins of empty categories in Singapore English. *Journal of Pidgin and Creole Languages*, 16(2), pp.275-319. DOI: 10.1075/jpcl.16.2.03zhi

Bao, Z. (2005). The aspectual system of Singapore English and the systemic substratist explanation. *Journal of Linguistics*, 41, pp.237-267. DOI: 10.1017/S0022226705003269

Bao, Z. (2014). *Got* in Singapore English. In Green, E. and Meyer, C. F. (Eds.). *The Variability of Current World Englishes*. De Gruyter Mouton, pp.147-167. DOI: 10.1515/9783110352108.147

Bar-Shalom, E. and Snyder, W. (1998). Root Infinitives in Child Russian: A Comparison with Italian and Polish. In Shillcock, R., Sorace, A., and Heycock, C. (Eds.). *Proceedings of the Gala '97 Conference on Language Acquisition: Knowledge Representation and Processing*. Edinburgh: Edinburgh University Press.

Barber, C. (2000). *The English language: a historical introduction*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139106894

Bard, E. G., Robertson, D., and Sorace, A. (1996). Magnitude estimation of linguistic acceptability. *Language*, 72(1), pp.32-68. Available at: https://www.jstor.org/stable/416793 [Accessed 31-December-2022]

Bardovi-Harlig, K. and Reynolds, D. W. (1995). The Role of Lexical Aspect in the Acquisition of Tense and Aspect. *TESOL Quarterly*, 29(1), pp.107-131. DOI: 10.2307/3587807

Bardovi-Harlig, K. and Sprouse, R. A. (2018). Negative Versus Positive Transfer. In Liontas, J. I. (Ed.). *The TESOL Encyclopedia of English Language Teaching*. John Wiley & Sons, Inc., pp.1-6. DOI: 10.1002/9781118784235.eelt0084

Baskaran, L. (1987). *Aspects of Malaysian English syntax*. PhD thesis: University College. University of London, UK. Available at: https://discovery.ucl.ac.uk/id/eprint/1317756/1/283105.pdf [Accessed 09-December-2022]

Baskaran, L. (1994). The Malaysian English mosaic. *English Today 37*, 10(1), pp.27-32. DOI: 10.1017/S0266078400000857

Baskaran, L. (2008a). Malaysian English: phonology. In Mesthrie, R. (Ed.). *Varieties of English 4: Africa, South and Southeast Asia*. Berlin: Mouton de Gruyter, pp.278-291. DOI: 10.1515/9783110208429.1.278

Baskaran, L. (2008b). Malaysian English: morphology and syntax. In Mesthrie, R. (Ed.). *Varieties of English 4: Africa, South and Southeast Asia*. Berlin: Mouton de Gruyter, pp.610-623. DOI: 10.1515/9783110208429.2.610

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), pp.1-48. DOI: 10.18637/jss.v067.i01

Beck, M.-L. (1998). L2 acquisition and obligatory head movement. *Studies in Second Language Acquisition*, 20(3), pp.311-348. DOI: 10.1017/S0272263198003027

Belletti, A., Bennati, E., and Sorace, A. (2007). Theoretical and developmental issues in the syntax of subjects: Evidence from near-native Italian. *Natural Language & Linguistic Theory*, 25, pp.657-689. DOI: 10.1007/s11049-007-9026-9

Berman, R. A. (2004). Between emergence and mastery: The long developmental route of language acquisition. In Berman, R. A. (Ed.). *Language Development across Childhood and Adolescence*. Amsterdam: John Benjamins Publishing Company, pp.9-34.

Berman, R. A. (2009). Language development in narrative contexts. In Bavin, E. L. (Ed.). *The Cambridge Handbook of Child Language*. Cambridge: Cambridge University Press, pp.355-375.

Bhatt, R. M. (2001). World Englishes. *Annual Review of Anthropology*, 30(2001), pp.527-550. Available at: https://www.jstor.org/stable/3069227 [Accessed 09-December-2022]

Bhatt, R. M. (2004). Indian English: syntax. In Kormann, B., Burridge, K., Mesthrie, R., Schneider, E. W., and Upton, C. (Eds.). *A Handbook of Varieties of English. Volume 2: Morphology and Syntax.* Mouton de Gruyter, pp.1016-1030. DOI: 10.1515/9783110197181-133

Bialystok, E. (1979). Explicit and implicit judgements of L2 grammaticality. *Language Learning*, 29(1), pp.81-103. DOI: 10.1111/j.1467-1770.1979.tb01053.x

Biberauer, T. and Richards, M. (2006). True optionality: When the grammar doesn't mind. In Boecks, C. (Ed.). *Minimalist Essays*. Amsterdam: John Benjamins Publishing Company, pp.35-67.

Bickerton, D. (1973). The Nature of a Creole Continuum. *Language*, 49(3), pp.640-669. Available at: https://www.jstor.org/stable/412355 [Accessed 09-December-2022]

Bickerton, D. (1975). *Dynamics of a Creole System*. Cambridge: Cambridge University Press.

Birdsong, D. (1989). *Metalinguistic Performance and Interlinguistic Competence*. London: Springer.

Birdsong, D., Gertken, L.M., and Amengual, M. (2012). *Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism*. COERLL, University of Texas at Austin. Available at: https://sites.la.utexas.edu/bilingual/ [Accessed 31-December-2022]

Blom, E. (2010). Effects of input on the early grammatical development of bilingual children. *International Journal of Bilingualism*, 14(4), pp.422-446. DOI: 10.1177/1367006910370917

Bolton, K. (2010). Creativity and world Englishes. *World Englishes*, 29(4), pp.455-466. DOI: 10.1111/j.1467-971X.2010.01674.x

Bonilla, C. L. (2011). The Conversational Historical Present in Oral Spanish Narratives. *Hispania*, 94(3), pp.429-442. Available at: https://www.jstor.org/stable/23032119 [Accessed 31-December-2022]

Borer, H. and Wexler, K. (1992). Bi-unique relations and the maturation of grammatical principles. *Natural Language and Linguistic Theory*, 10, pp.147-189. DOI: 10.1007/BF00133811

Botting, N. (2002). Narrative as a tool for the assessment of linguistic and pragmatic impairments. *Child Language Teaching and Therapy*, 18(1), pp.1-21. DOI: 10.1191/0265659002ct224oa

Bowles, M. A. (2011). Measuring implicit and explicit linguistic knowledge. *Studies in Second Language Acquisition*, 33(2), pp.247-271. DOI: 10.1017/S0272263110000756

Britannica (2022). agglutination. *Encyclopaedia Britannica*. Available at: https://www.britannica.com/topic/agglutination-grammar [Accessed 24-December-2022]

British Council (2022). *Online English level test*. Available at: https://learnenglish.britishcouncil.org/english-levels/online-english-level-test [Accessed 31-December-2022]

Brown, J. D. (1989). Improving ESL Placement Tests Using Two Perspectives. *TESOL Quarterly*, 23(1), pp.65-83. Available at: https://www.jstor.org/stable/3587508 [Accessed 31-December-2022]

Brown, R. (1973). *A First Language: The Early Stages*. Cambridge, Massachusetts: Harvard University Press.

Buschfeld, S. (2020). *Children's English in Singapore: Acquisition, Properties, and Use*. London: Routledge.

Buschfeld, S. (2021). L1 Singapore English: The Influence of Ethnicity and Input. In Peters, P. and Burridge, K. (Eds.). *Exploring the Ecology of World Englishes in the Twenty-first Century*. Edinburgh: Edinburgh University Press, pp.193-214. DOI: 10.1515/9781474462877-013

Buschfeld, S. and Kautzsch, A. (2016). Towards an integrated approach to postcolonial and non-postcolonial Englishes. *World Englishes*, 36(1), pp.104-126. DOI: 10.1111/weng.12203

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., and Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, 14, pp.365-376. DOI: 10.1038/nrn3475

Cain, K. (2003). Text comprehension and its relation to coherence and cohesion in children's fictional narratives. *British Journal of Development Psychology*, 21(3), pp.335-351. DOI: 10.1348/026151003322277739

Cambridge University Press & Assessment (2022). *Test Your English: General English*. Available at: https://www.cambridgeenglish.org/test-your-english/ [Accessed 31-December-2022]

Chambers, J. K. (2004). Dynamic typology and vernacular universals. In Kortmann, B. (Ed.). *Dialectology meets Typology: Dialect Grammar from a Cross-Linguistic Perspective*. Mouton de Gruyter, pp.127-145.

Chen, C.-H. (2008). *Chinese relativization: Ordering at the syntax-phonology interface*. PhD thesis: University of Ottawa. DOI: http://dx.doi.org/10.20381/ruor-19873

Chen, J. and Shirai, Y. (2010). The development of aspectual marking in child Mandarin Chinese. *Applied Psycholinguistics*, 31(1), pp.1-28. DOI: 10.1017/S0142716409990257

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris Publications.

Chomsky, N. (1986). *Barriers*. The MIT Press.

Chomsky, N. (1993). A Minimalist Program for Linguistic Theory. In Hale, K. and Keyser, S. J. (Eds.). *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*. The MIT Press, pp.1-52.

Chomsky, N. (1995). *The Minimalist Program*. The MIT Press

Chomsky, N. (2015). *The Minimalist Program: 20th Anniversary Edition*. The MIT Press.

Chondrogianni, V. and Marinis, T. (2012). Production and processing asymmetries in the acquisition of tense morphology by sequential bilingual children. *Bilingualism: Language and Cognition*, 15(1), pp.5-21. DOI: 10.1017/S1366728911000368

Chong, S. L. (2004). Distribution of Chinese dialects. In Haji Omar, A. (Ed.). *The Encyclopedia of Malaysia. Volume 9: Languages and Literature*. Archipelago Press, pp.52-53.

Common European Framework of References for Languages (CEFR) (2022). *Global scale – Table 1 (CEFR 3.3): Common Reference levels*. Available at: https://www.coe.int/en/web/common-european-framework-reference-languages/table-1-cefr-3.3-common-reference-levels-global-scale# [Accessed 31-December-2022]

Coppieters, R. (1987). Competence Differences between Native and Near-Native Speakers. *Language*, 63(3), pp.544-573. Available at: https://www.jstor.org/stable/415005 [Accessed 25-December-2022]

Cowart, W. (1997). *Experimental Syntax: Applying Objective Methods to Sentence Judgments*. SAGE Publications.

Crystal, D. (2003). *The Cambridge encyclopedia of the English language*. 2nd edition. Cambridge: Cambridge University Press.

Crystal, D. (2012). A global language. In Seargeant, P. and Swann, J. (Eds.). *English in the World: History, Diversity, Change.* Routledge, pp.151-196.

Danchev, A. (1992). The evidence for analytic and synthetic developments in English. In Rissanen, M., Ihalainen, O., Nevalainen, T., and Taavitsainen, I. (Eds.). *History of Englishes: New Methods and Interpretations in Historical Linguistics*. Mouton de Gruyter, pp.25-41.

David, M. K. (2006). Role and functions of code-switching in Malaysian courtrooms. *Multilingua*, 22, pp.5-20. DOI: 10.1515/mult.2003.005

De Costa, P. I. and Crowther, D. (2018). Second Language Acquisition and world Englishes: Dialogue and engagement. *World Englishes*, 37(1), pp.19-33. DOI: 10.1111/weng.12300

De Klerk, V. (2003). Towards a norm in South African Englishes: the case for Xhosa English. *World Englishes*, 22(4), pp.463-481. DOI: 10.1111/j.1467-971X.2003.00313.x

De Swart, H. (2012). Verbal Aspect. In Binnick, R. I. (Ed.). *The Oxford Handbook of Tense and Aspect*. Oxford: Oxford University Press, pp.752-780. DOI: 10.1093/oxfordhb/9780195381979.013.0026

DeKeyser, R., Alfi-Shabtay, I., and Ravid, D. (2010). Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics*, 31(3), pp.413-438. DOI: 10.1017/S0142716410000056

Department of Statistics Malaysia (2020). *Education Statistics*. Putrajaya: Department of Statistics, Malaysia. Available from: https://www.dosm.gov.my/v1/index.php?r=column/cthree&menu_id=Z1hCMUVLQWVOL2 tScVlhSmo5eEd3QT09 [Accessed 04-December-2022]

Department of Statistics Malaysia (2021). *Demographic Statistics Second Quarter 2021, Malaysia*. Putrajaya: Department of Statistics, Malaysia. Available from: https://www.dosm.gov.my/v1/index.php?r=column/cthemeByCat&cat=430&bul_id=eGtwdjd 4amZJb1JmcFFkYXBKNHg3dz09&menu_id=L0pheU43NWJwRWVSZklWdzQ4TlhUUT0 9 [Accessed 04-December-2022]

Dewaele, J.-M. (2018). Online Questionnaires. In Phakiti, A., De Costa, P., Plonsky, L., and Starfield, S. (Eds.). *The Palgrave Handbook of Applied Linguistics Research Methodology. Palgrave Macmillan*, pp.269-286. DOI: 10.1057/978-1-137-59900-1_13

Dollinger, S. (2015). *The Written Questionnaire in Social Dialectology: History, theory, practice.* Amsterdam: John Benjamins Publishing Company.

Donaldson, B. (2017). Negation in Near-Native French: Variation and Sociolinguistic Competence. *Language Learning*, 67(1), pp.141-170. DOI: 10.1111/lang.12201

Donner, A. and Eliasziw, M. (1987). Sample size requirements for reliability studies. *Statistics in Medicine*, 6(4), pp.441-448. DOI: 10.1002/sim.4780060404

Dörnyei, Z. and Taguchi, T. (2010). *Questionnaires in Second Language Research: Construction, Administration, and Processing*. 2nd edition. Routledge.

Dragojevic, M. (2017). Language Attitudes. *Oxford Research Encyclopedias: Communication*. Oxford: Oxford University Press. DOI: 10.1093/acrefore/9780190228613.013.437

Eager, C. and Roy, J. (2017). Mixed Effects Models are sometimes Terrible. *arXiv*. DOI: 10.48550/arXiv.1701.04858

Eberhard, D. M., Gary, F. S., and Fennig, C. D. (Eds.). (2022). *Ethnologue: Languages of the World*. 25th edition. Dallas, Texas: SIL International. Available from: http://www.ethnologue.com/ [Accessed 04-December-2022]

Elff, M. (2022). *mclogit: Multinomial Logit Models, with or without Random Effects or Overdispersion*. R package version 0.9.4.2. Available at: https://CRAN.R-project.org/package=mclogit [Accessed 07-December-2022]

Ellis, N. C. and Sagarra, N. (2010). The bounds of adult language acquisition: Blocking and Learned Attention. *Studies in Second Language Acquisition*, 32(4), pp.553-580. Available at: https://www.jstor.org/stable/44485954 [Accessed 20-December-2022]

Ellis, R. (1991). Grammaticality judgments and second language acquisition. *Studies in Second Language Acquisition*, 13(2), pp.161-186. DOI: 10.1017/S0272263100009931

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language. *Studies in Second Language Acquisition*, 27(2), pp.141-172. DOI: 10.1017/S0272263105050096

Ernst, T. (2009). Functional categories and the Chinese Infl. *Linguistics*, 32, pp.191-212. DOI: 10.1515/ling.1994.32.2.191

Eubank, L. (1993). On the Transfer of Parametric Values in L2 Development. *Language Acquisition*, 3(3), pp.183-208. DOI: 10.1207/s15327817la0303_1

Fabricius-Hansen, C. (2006). Tense. In Brown, K. (Ed.). *Encyclopedia of Language & Linguistics*. 2nd edition. Elsevier Ltd., pp.566-573. DOI: 10.1016/B0-08-044854-2/00248-0

Federation of Malaya (1951). *Report on* the Barnes report on Malay education *and* the Fenn-Wu report on Chinese education. Kuala Lumpur: Government Press.

Field, A., Miles, J., and Field, Z. (2012). *Discovering statistics using R*. London: SAGE Publications Ltd.

Filip, H. (2012). Lexical Aspect. In Binnick, R. I. (Ed.). The Oxford Handbook of Tense and Aspect. *Oxford: Oxford University Press*, pp.721-751. DOI: 10.1093/oxfordhb/9780195381979.013.0025

Fox, J. and Weisberg, S. (2019). *An R Companion to Applied Regression*. 3rd edition. Sage Publications. Available at: https://socialsciences.mcmaster.ca/jfox/Books/Companion/ [Accessed 07-December-2022]

García Mayo, M. D. P. (2003). Age, Length of Exposure and Grammaticality Judgements in the Acquisition of English as a Foreign Language. In García Mayo, M. D. P. and Lecumberri, M. L. G. (Eds.). *Age and the Acquisition of English as a Foreign Language*. Multilingual Matters Ltd., pp.94-114.

García Mayo, M. D. P. and Olaizola, I. V. (2010). The development of suppletive and affixal tense and agreement morphemes in the L3 English of Basque-Spanish bilinguals. *Second Language Research*, 27(1), pp.129-149. DOI: 10.1177/0267658310386523

García Mayo, M. D. P., Ibarrola, A. L., and Liceras, J. M. (2005), Placeholders in the English Interlanguage of Bilingual (Basque/Spanish) Children. *Language Learning*, 55(3), pp.445-489. DOI: 10.1111/j.0023-8333.2005.00312.x

Gass, S. (2018). Coda 1 World Englishes and Second Language Acquisition: Working towards a common goal. *World Englishes*, 37(1), pp.120-126. DOI: 10.1111/weng.12307

Gass, S. M. and Mackey, A. (2017). *Stimulated Recall Method in Applied Linguistics and L2 Research*. Routledge. DOI: 10.4324/9781315813349

Gavruseva, E. (2008). The development of copula and auxiliary *be* and overgeneration of *be* in child L2 English. In Haznedar, B. and Gavruseva, E. (Eds.). *Current Trends in Child Second Language Acquisition: A generative perspective*. Amsterdam: John Benjamins Publishing Company, pp.145-175.

Geçkin, V. and Haznedar, B. (2008). The morphology/syntax interface in child L2 acquisition: Evidence from verbal morphology. In Haznedar, B. and Gavruseva, E. (Eds.). *Current Trends in Child Second Language Acquisition: A generative perspective*. Amsterdam: John Benjamins Publishing Company, pp.237-267.

Gill, S. K. (1999). Standards and emerging linguistic realities in the Malaysian workplace. *World Englishes*, 18(2), pp.215-231. DOI: 10.1111/1467-971X.00135

Gill, S. K. (2012). The Complexities of Re-reversal of Language-in-Education Policy in Malaysia. In Kirkpatrick, A. and Sussex, R. (Eds.). *English as an International Language in Asia: Implications for Language Education*. Springer, pp.45-61. DOI: 10.1007/978-94-007-4578-0_4

Goad, H. and White, L. (2006). Ultimate attainment in interlanguage grammars: a prosodic approach. *Second Language Research*, 22(3), pp.243-268. DOI: 10.1191/0267658306sr268oa

Goddard, C. (1996). The "Social Emotions" of Malay (Bahasa Melayu). *Ethos*, 24(3), pp.426-464. Available at: https://www.jstor.org/stable/640479 [Accessed 24-December-2022]

Goddard, C. (2002). Semantic Primes and Universal Grammar in Malay (Bahasa Melayu). In Goddard, C. and Wierzbicka, A. (Eds.). *Meaning and Universal Grammar: Theory and empirical findings. Volume I*. Amsterdam: John Benjamins Publishing Company, pp.87-172.

Gooch, L. (2009). Malaysia Ends Use of English in Science and Math Teaching. *The New York Times*. 08 July. Available at: https://www.nytimes.com/2009/07/09/world/asia/09iht-malay.html [Accessed 08-December-2022]

Görlach, M. (1990). *Studies in the history of the English language*. Heidelberg: Carl Winter.

Government of Malaysia (1976). *Third Malaysia Plan, 1976-1980.* Kuala Lumpur: Government Press. Available at: https://www.epu.gov.my/en/economic-developments/development-plans/rmk/third-malaysia-plan-1976-1980 [Accessed 08-December-2022]

Govindan, I. V. P. and Pillai, S. (2009). English question forms used by young Malaysian Indians. *The English Teacher*, XXXVIII, pp.74-79.

Grangé, P. (2013). Aspect in Indonesian: free markers vs bound markers. In Bowden, J. (Ed.). *Tense, aspect, mood and evidentiality in languages of Indonesia*. NUSA: Linguistic studies of languages in and around Indonesia, 55, pp.57-79. Available at: http://hdl.handle.net/10108/74326 [Accessed 24-December-2022]

Gupta, A. F. (1992). The pragmatic particles of Singapore colloquial English. *Journal of Pragmatics*, 18(1), pp.31-57. DOI: 10.1016/0378-2166(92)90106-L

Gupta, A. F. (1994). *The Step-tongue: Children's English in Singapore*. Clevedon: Multilingual Matters Ltd.

Gutiérrez-Clellen, V. F., Simon-Cereijido, G., and Wagner, C. (2008). Bilingual children with language impairment: A comparison with monolinguals and second language learners. *Applied Psycholinguistics*, 29(1), pp.3-19. DOI: 10.1017/S0142716408080016

Gvozdanović, J. (2012). Perfective and Imperfective Aspect. In Binnick, R. I. (Ed.). *The Oxford Handbook of Tense and Aspect*. Oxford: Oxford University Press, pp.781-802. DOI: 10.1093/oxfordhb/9780195381979.013.0027

Haegeman, L. (1997). Elements of Grammar. In Haegeman, L. (Ed.). *Elements of Grammar: Handbook in Generative Syntax*. Springer Science+Business Media, B. V., pp.1-72.

Haji Omar, A. (1982). *Language and society in Malaysia*. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Haji Omar, A. (1985). The language policy of Malaysia: A formula for balanced pluralism. *Papers in South-East Asian linguistics*, 9, pp.39-49.

Haji Omar, A. (1996). Post-imperial English in Malaysia. In Conrad, A. W. and Rubal-Lopez, A. (Eds.). *Post-Imperial English: Status Change in Former British and American Colonies, 1940 – 1990.* De Gruyter Mouton, pp.513-534. DOI: 10.1515/9783110872187.513

Haji Omar, A. (Ed.) (2004). *The Encyclopedia of Malaysia. Volume 9: Languages and Literature*. Archipelago Press.

Haji Omar, A. (2004). Social varieties of Malay. In Haji Omar, A. (Ed.). *The Encyclopedia of Malaysia. Volume 9: Languages and Literature*. Archipelago Press, pp.28-29.

Haji Omar, A. (2012). Pragmatics of maintaining English in Malaysia's education system. In Low, E.-L., and Hashim, A. (Eds.). *English in Southeast Asia: Features, policy and language in use*. Amsterdam: John Benjamins Publishing Company, pp.155-174.

Haji Omar, A. (2015). *Languages in the Malaysian Education System: Monolingual strands in multilingual settings*. London: Routledge. DOI: 10.4324/9781315669472

Han, Y. (2000). Grammaticality Judgment Tests: How Reliable and Valid Are They? *Applied Language Learning*, 11(1), pp.177-204.

Harada, S. (2009). The Roles of Singapore Standard English and Singlish. *Information Research*, 40, pp.69-82.

Harris, T. and Wexler, K. (1996). The Optional-Infinitive Stage in Child English: Evidence from Negation. In Clahsen, H. (Ed.). *Generative Perspectives on Language Acquisition: Empirical findings, theoretical considerations and crosslinguistic comparisons*. Amsterdam: John Benjamins Publishing Company, pp.1-42.

Hashim, A. (2007). The Use of Malaysian English in Creative Writing. *Asian Englishes*, 10(2), pp.30-44. DOI: 10.1080/13488678.2007.10801211

Hashim, A. (2014). English and the linguistic ecology of Malaysia. *World Englishes*, 33(4), pp.458-471. DOI: 10.1111/weng.12107

Hashim, A. (2020). Malaysian English. In Bolton, K., Botha, W., and Kirkpatrick, A. (Eds.). *The Handbook of Asian Englishes*. Wiley Blackwell. DOI: 10.1002/9781118791882.ch16

Hashim, A. and Leitner, G. (2015). English in language education policies and planning in Malaysia. In Haji Omar, A. (Ed.). *Languages in the Malaysian Education System: Monolingual strands in multilingual settings*. London: Routledge.

Hashim, A. and Leitner, G. (2016). Arabic in contact with English and Malay in Malaysia. In Leitner, G., Hashim, A., and Wolf, H.-G. (Eds.). *Communicating with Asia: The Future of English as a Global Language*. Cambridge: Cambridge University Press, pp.85-101. DOI: 10.1017/CBO9781107477186.007

Hashim, A., Leitner, G., and Al Aqad, M. (2017). Arabic in contact with English in Asia: Linguistic, social and political influences of Arabic in the region. *English Today*, 33(1), pp.25-32. DOI: 10.1017/S0266078416000377

Hawkins, R. and Casillas, G. (2008). Explaining frequency of verb morphology in early L2 speech. *Lingua*, 118(4), pp.595-612. DOI: 10.1016/j.lingua.2007.01.009

Hawkins, R. and Chan, C. Y.-H. (1997). The partial availability of Universal Grammar in second language acquisition: the 'failed functional features hypothesis'. *Second Language Research*, 13(3), pp.187-226. DOI: 10.1191/026765897671476153

Hawkins, R. and Hattori, H. (2006). Interpretation of English multiple *wh*-questions by Japanese speakers: a missing uninterpretable feature account. *Second Language Research*, 22(3), pp.269-301. DOI: 10.1191/0267658306sr269oa

Hawkins, R. and Liszka, S. (2003). Locating the source of defective past tense marking in advanced L2 English speakers. In van Hout, R., Hulk, A., Kuiken, F., and Towell, R. (Eds.). *The Lexicon-Syntax Interface in Second Language Acquisition*. Amsterdam: John Benjamins Publishing Company, pp.21-44.

Hedgecock, J. (1993). Well-formed vs. ill-formed strings in L2 metalingual tasks: specifying features of grammaticality judgements. *Second Language Research*, 9(1), pp.1-21. Available at: http://www.jstor.com/stable/43104460 [Accessed 31-December-2022]

Her, O.-S. (1991). On the Mandarin Possessive and Existential Verb *You* and its Idiomatic Expressions. *Language Sciences*, 13(3-4), pp.381-398. DOI: 10.1016/0388-0001(91)90023-T

Hermas, A. (2019). Genericity in third language English: Acquisition pattern and transfer in ultimate attainment. *International Journal of Bilingualism*, 24(2), pp.266-288. DOI: 10.1177/1367006919826865

Hess, C. W., Ritchie, K. P., and Landry, R. G. (1984). The type-token ratio and vocabulary performance. *Psychological Reports*, pp.51-57. DOI: 10.2466/pr0.1984.55.1.51

Hewson, J. (2012). Tense. In Binnick, R. I. (Ed.). *The Oxford Handbook of Tense and Aspect*. Oxford: Oxford University Press, pp.507-535. DOI: https://doi.org/10.1093/oxfordhb/9780195381979.013.0017

Hilbert, M. (2008). Interrogative inversion in non-standard varieties of English. In Siemund, P. and Kintana, N. (Eds.). *Language Contact and Contact Languages*. Amsterdam: John Benjamin Publishing Company, pp.261-289.

Hilbert, M. (2011). Interrogative inversion as a learner phenomenon in English contact varieties: A case of Angloversals? In Mukherjee, J. and Hundt, M. (Eds.). *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a paradigm gap*. Amsterdam: John Benjamins Publishing Company, pp.125-143.

Hinkel., E. (2003). Adverbial markers and tone in L1 and L2 students' writing. *Journal of Pragmatics*, 35(7), pp.1049-1068. DOI: 10.1016/S0378-2166(02)00133-9

Hiramoto, M. and Sato, Y. (2012). *Got*-interrogatives and answers in Colloquial Singapore English. *World Englishes*, 31(2), pp.198-207. DOI: 10.1111/j.1467-971X.2012.01750.x

Hirschman, C. (1972). Educational Patterns in Colonial Malaya. *Comparative Education Review*, 16(3), pp.486-502. Available at: http://www.jstor.org/stable/1186779 [Accessed 04-December-2022]

Hoffman, M. (2014). Sociolinguistic Interviews. In Holmes, J. and Hazen, K. (Eds.). *Research Methods in Sociolinguistics: A Practical Guide*. West Sussex: John Wiley & Sons, Inc, pp.25-41.

Hogg, R. M. (Ed.) (1992). *The Cambridge history of the English language. Volume I: The Beginnings to 1066*. Cambridge: Cambridge University Press. DOI: 10.1017/CHOL9780521264747

Hopp, H. (2009). The syntax-discourse interface in near-native L2 acquisition: Off-line and on-line performance. *Bilingualism: Language and Cognition*, 12(4), pp. 463-483. DOI: 10.1017/S1366728909990253

Hopp, H. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. *Lingua*, 20(4), pp.901-931. DOI: 10.1016/j.lingua.2009.06.004

Huang, C.-T. J. (1984). On the Distribution and Reference of Empty Pronouns. *Linguistic Inquiry*, 15(4), pp. 531-574. Available at: https://www.jstor.org/stable/4178404 [Accessed 11-December-2022]

Huang, J. C.-T. (1998*). Logical Relations in Chinese and the Theory of Grammar*. New York: Garland Publishing, Inc.

Huddleston, R. (1988). *English Grammar: An Outline*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9781139166003

Hulk, A. and Müller, N. (2000). Bilingual first language acquisition at the interface between syntax and pragmatics. *Bilingualism: Language and Cognition*, 3(3), pp.227-244. DOI: 10.1017/S1366728900000353

Hurford, J. R. (1975). A child and the English question formation rule. *Journal of Child Language*, 2(2), pp.299-301. DOI: 10.1017/S0305000900001148

IELTS (2022). *IELTS in CEFR scale*. Available at: https://www.ielts.org/about-ielts/ielts-in-cefr-scale [Accessed 31-December-2022]

Ionin, T. (2012). Formal Theory-Based Methodologies. In Mackey, A. and Gass, S. M. (Eds.). *Research Methods in Second Language Acquisition: A Practical Guide*. Wiley-Backwell, pp.30-52.

Ionin, T. and Wexler, K. (2002). Why is 'is' easier than '-s'?: acquisition of tense/agreement morphology by child second language learners of English. *Second Language Research*, 18(2), pp.95-136. DOI: 10.1191/0267658302sr195oa

Iwaniec, J. (2020). Questionnaires: Implications for effective implementation. In McKinley, J. and Rose, H. (Eds.). *The Routledge Handbook of Research Methods in Applied Linguistics*. Routledge, pp.324-335.

Jakobson, R. (1971). *Selected writings. Volume II: Word and Language*. De Gruyter Mouton. DOI: 10.1515/9783110873269

Jia, G. and Fuse, A. (2007). Acquisition of English Grammatical Morphology by Native Mandarin-Speaking Children and Adolescents: Age-Related Differences. *Journal of Speech, Language, and Hearing Research*, 50(5), pp.1280-1299. DOI: 10.1044/1092-4388(2007/090)

Johnson, J. S. and Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21(2), pp.60-99. DOI: 10.1016/0010-0285(89)90003-0

Johnson, W. (1944). Studies in language behavior. I. A program of research. *Psychological Monographs*, 56(2), pp.1-15. DOI: 10.1037/h0093508

Justice, L. M., Bowles, R. P., Kaderavek, J. N., Ukrainetz, T. A., Eisenberg, S. L., and Gillam, R. B. (2006). The Index of Narrative Microstructure: A Clinical Tool for Analyzing School-Age Children's Narrative Performances. *American Journal of Speech-Language Pathology*, 15, pp.177-191. DOI: 10.1044/1058-0360(2006/017)

Kachru, B. B. (1985). Bilingual's creativity. *Annual Review of Applied Linguistics*, 6, pp.20-33. DOI: 10.1017/S0267190500003032

Kachru, B. B. (Ed.) (1992a). *The Other Tongue: English across Cultures.* 2nd edition. Urbana: University of Illinois Press.

Kachru, B. B. (1992b). World Englishes: approaches, issues and resources. *Language Teaching*, 25(1), pp.1-14. DOI: 10.1017/S0261444800006583

Kärchner-Ober, R. (2013). Linguistic quandary in multilingual Malaysia: Socio-political issues, language policy, educational changes. In Singleton, D., Fishman, J. A., Aronin, L., and Ó Laoire, M. (Eds.). *Current Multilingualism: A New Linguistic Dispensation*. De Gruyter Mouton, pp.297-315. DOI: 10.1515/9781614512813.297

Karim, N. S., M., M. Onn, Haji Musa, and Mahmood, A. H. (2008). *Tatabahasa Dewan*. Edisi Ketiga. Kuala Lumpur: Dewan Bahasa dan Pustaka.

Karmiloff-Smith, A. (1985). Language and cognitive processes from a developmental perspective. *Language and Cognitive Processes*, 1(1), pp.61-85. DOI: 10.1080/01690968508402071

Keller, F. (1998). Grammaticality Judgments and Linguistic Methodology. *Research Paper EUCCS-RP-1998-3*, pp.1-15.

Klein, W. (1995). Frame of analysis. In Dietrich, R., Klein, W., and Noyao, C. (Eds.). *The acquisition of temporality in a second language*. Amsterdam: John Benjamins Publishing Company, pp.17-29.

Kortmann, B. and Szmrecsanyi, B. (2004). Global synopsis: morphological and syntactic variation in English. In Kortmann, B., Burridge, K., Mesthrie, R., Schneider, E. W., and Upton, C. (Eds.). *A Handbook of Varieties of English. Volume 2: Morphology and Syntax*. Mouton de Gruyter, pp.1142-1202. DOI: 10.1515/9783110197181-142

Kortmann, B. and Szmrecsanyi, B. (2009). World Englishes between simplification and complexification. In Hoffmann, T. and Siebers, L. (Eds.), *World Englishes – Problems, Properties and Prospects: Selected papers from the 13th IAWE conference*. Amsterdam: John Benjamins Publishing Company, pp.265-285.

Kortmann, B. and Szmrecsanyi, B. (2011). Parameters of morphosyntactic variation in World Englishes: prospects and limitations of searching for universals. In Siemund, P. (Ed.). *Linguistic Universals and Language Variation*. De Gruyter Mouton, pp.264-290.

Kortmann, B., Lunkenheimer, K., and Ehret, K. (Eds.) (2020). *The Electronic World Atlas of Varieties of English*. Zendono. DOI: 10.5281/zenodo.3712132. Available at: http://ewave-atlas.org [Accessed 29-December-2022]

Kuo, L.-J., Ku, Y.-M., and Gezer, M. Ü. (2020). The relationship between input and literacy and metalinguistic development: A study with Chinese-English bilinguals. *International Journal of Bilingualism*, 24(1), pp.26-45. DOI: 10.1177/1367006918768312

Kwan-Terry, A. (1979). Two progressive markers in Chinese. *Southeast Asian Linguistic Studies*, 4, pp.213-232.

Labov, W. (1972). Some Principles of Linguistic Methodology. *Language in Society*, 1(1), pp.97-120. Available at: https://www.jstor.org/stable/4166672 [Accessed 20-December-2022]

Labov, W. (1984). Field Methods of the Project on Linguistic Change and Variation. In Baugh, J. and Sherzer, J. (Eds.). *Language in Use*. Englewood Cliffs: Prentice Hall, pp.28-53.

Langdon, H. W. (2016). Communication Issues in a Multilingual Society. In Langdon, H. W. and Saenz, T. I. (Eds.). *Working With Interpreters and Translators: A Guide for Speech-Language Pathologists and Audiologists*. Plural Publishing Inc., pp.29-54.

Langsford, S., Perfors, A., Hendrickson, A. T., Kennedy, L. A., and Navarro, D. J. (2018). Quantifying sentence acceptability measures: Reliability, bias, and variability. *Glossa: a journal of general linguistics*, 3(1), pp.1-34. DOI: 10.5334/gjgl.396

Lardiere, D. (1998). Dissociating syntax from morphology in a divergent L2 end-state grammar. *Second Language Research*, 14(4), pp.359-375. DOI: 10.1191/026765898672500216

Larsen-Freeman, D. (2018). Second Language Acquisition, WE, and language as a complex adaptive system (CAS). *World Englishes*, 37(1), pp.80-92. DOI: 10.1111/weng.12304

Lee, N. and Huang, Y. Y. (2004). To Be or Not To Be – The Variable Use of the Verb BE in the Interlanguage of Hong Kong Chinese Children. *RELC Journal*, 35(2), pp.211-228. DOI: 10.1177/003368820403500208

Lee, N. H., Ping, L. A., and Nomoto, H. (2009). Colloquial Singapore English *got*: functions and substratal influences. *World Englishes*, 28(3), pp.293-318. DOI: 10.1111/j.1467-971X.2009.01594.x

Leech, G., Rayson, P., and Wilson, A. (2001). *Word Frequencies in Written and Spoken English: Based on the British National Corpus*. London: Routledge. DOI: 10.4324/9781315840161

Lefebvre, C. (1986). Relexification in Creole Genesis Revisited: the Case of Haitian Creole. In Muysken, P. and Smith, N. (Eds.). *Substrata versus universals in creole genesis: Papers from the Amsterdam Creole Workshop, April 1985*. Amsterdam: John Benjamins Publishing Company, pp.279-300.

Lefebvre, C. (1999). *Creole genesis and the acquisition of grammar: The case of Haitian creole*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511519826

Lenth, R. V. (2022). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.7.3. Available at: https://CRAN.R-project.org/package=emmeans [Accessed 07-December-2022]

Ler, S. L. V. (2006). A relevance-theoretic approach to discourse particles in Singapore English. In Fischer, K. (Ed.). *Approaches to discourse particles*. Oxford: Elsevier, pp.149-166.

Leung, Y.-k. I. (2003). Failed Features versus Full Transfer Full Access in the Acquisition of a Third Language: Evidence from Tense and Agreement. *Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference (GASLA 2000)*, pp.199-207.

Levshina, N. (2015). *How to do Linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins Company.

Li, A. Y. (1990). *Order and Constituency in Mandarin Chinese*. Dordrecht: Kluwer Academic Publishers.

Li, C. N. and Thompson, S. A. (1981). *Mandarin Chinese: A Functional Reference Grammar*. University of California Press.

Li, C. N. and Thompson, S. A. (2009). Chinese. In Comrie, B. (Ed.). *The World's Major Languages*. 2nd edition. London: Routledge, pp.703-723.

Li, L. and Siemund, P. (2001). From phasal polarity expression to aspectual marker: Grammaticalization of *already* in Asian and African varieties of English. In Kramer, R. L.

(Ed.). *The Expression of Phasal Polarity in African Languages*. De Gruyter Mouton, pp.515-543. DOI: 10.1515/9783110646290-021

Li, P., Zhang, F., Tsai, E., and Puls, B. (2014). Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition*, 17(3), pp.673-680. DOI: 10.1017/S1366728913000606

Lietz, P. (2010). Research into questionnaire design: A summary of the literature. *International Journal of Market Research*, 52(2), pp.249-272. DOI: 10.2501/S147078530920120X

Lim, H.-P. (1976). Errors and Error Analysis in TESL: The Malaysian Experience. *RELC Journal*, 7(2), pp.23-29. DOI: 10.1177/003368827600700203

Lim, L. (2007). Mergers and acquisition: on the ages and origins of Singapore English particles. *World Englishes*, 26(4), pp.446-473. DOI: 10.1111/j.1467-971X.2007.00522.x

Lin, J.-W. (2006a). Aspectual selection and negation in Mandarin Chinese. *Linguistics*, 41-3, pp.425-459. DOI: https://doi.org/10.1515/ling.2003.015

Lin, J.-W. (2006b). Time in a Language Without Tense: The Case of Chinese. *Journal of Semantics*, 23(1), pp.1-53. DOI: 10.1093/jos/ffh033

Liszka, S. A. (2009). Associating meaning to form in advanced L2 speakers: An investigation into the acquisition of the English present simple and present progressive. In Snape, N., Leung, Y.-k. I., and Sharwood Smith, M. (Eds.). *Representational Deficits in SLA: Studies in honor of Roger Hawkins*. Amsterdam: John Benjamins Publishing company, pp.229-246.

Liszka, S. A. (2015). The L2 acquisition of the English present simple – present progressive distinction: Verb-raising revisited. In Ayoun, D. (Ed.). *The Acquisition of the Present*. Amsterdam: John Benjamins Publishing Company, pp.57-86.

Liu, M. (2015). Tense and Aspect in Mandarin Chinese. In Wang, W. S.-Y., Sun, C. (Eds). *The Oxford Handbook of Chinese Linguistics*. Oxford: Oxford University Press, pp.274-289. DOI: 10.1093/oxfordhb/9780199856336.013.0070

Long, A. Y., Shin, S.-Y., Geeslin, K., and Willis, E. W. (2018). Does the test work? Evaluating a web-based language placement test. *Language Learning & Technology*, 22(1), pp.137-156.

Lowenberg, P. H. (1986). Non-native varieties of English: Nativization, norms, and implications. *Studies in Second Language Acquisition*, 8(1), pp.1-18. Available at: https://www.jstor.org/stable/44486848 [Accessed 09-December-2022]

Lowenberg, P. H. (1993). Language transfer and levels of meaning potential in Malaysian English. In Alatis, J. E. (Ed.). *Georgetown University Round Table on Languages and Linguistics 1992: Language, Communication, and Social Meaning*. Washington, D.C.: Georgetown University Press, pp.41-55. Available at: https://repository.library.georgetown.edu/bitstream/handle/10822/555485/GURT_1992.pdf?sequence=1#page=53 [Accessed 09-December-2022]

Lozano, C. (2009). Selective deficits at the syntax-discourse interface: Evidence from the CEDEL2 corpus. In Snape, N., Leung, Y.-k. I., and Sharwood Smith, M. (Eds.). *Representational Deficits in SLA: Studies in honor of Roger Hawkins*. Amsterdam: John Benjamins Publishing Company, pp.127-166.

Lust, B. and Blume, M. (2017). The Grammaticality Judgment Task. In Blume, M. and Lust, B. C. (Eds.). *Research Methods in Language Acquisition: Principles, Procedures, and Practices*. De Gruyter Mouton, pp.155-163. DOI: 10.1515/9783110415339-009

MacArthur, T. (1987). The English languages? *English Today*, 3(3), pp.9-13. DOI: 10.1017/S0266078400013511

Mackey, A. and Gass, S. M. (2005). *Second language research: Methodology and Design*. Routledge.

Mackey, B. and Bryfonski, L. (2018). Mixed Methodology. In Phakiti, A., De Costa, P., Plonsky, L., and Starfield, S. (Eds.). *The Palgrave Handbook of Applied Linguistics Research Methodology*. Palgrave Macmillan, pp.103-121. DOI: https://doi.org/10.1057/978-1-137-59900-1_5

Macmillan Education Limited (2022). *Placement Tests:* Straightforward *Quick Placement & Diagnostic test*. Available at: https://www.macmillanstraightforward.com/resources/tests/ [Accessed 31-December-2022]

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk*. 3rd edition. Mahwah, NJ: Lawrence Erlbaum Associates.

Mäkinen, L., Loukusa, S., Nieminen, L., Leinonen, E., and Kunnari, S. (2014). The development of narrative productivity, syntactic complexity, referential cohesion and event content in four- to eight-year-old Finnish children. *First Language*, 34(1), pp.24-42. DOI: 10.1177/0142723713511000

Malaysian Examinations Council (2022). *Band description*.
    Reference in Malay:
    Majlis Peperiksaan Malaysia (2022). *Keterangan BAND*.
    Available at: https://www.mpm.edu.my/muet/keterangan-band [Accessed 20-December-2022]

Marchman, V. A. and Bates, E. (1994). Continuity in lexical and morphological development: a test of the critical mass hypothesis. *Journal of Child Language*, 21(2), pp.339-366. DOI: 10.1017/S0305000900009302

McWhorter, J. (2001). The world's simplest grammars are creole grammars. *Linguistic Typology*, 5, pp.125-166. DOI: 10.1515/lity.2001.001

Meisel, J. M. (1997). The acquisition of the syntax of negation in French and German: contrasting first and second language development. *Second Language Research*, 13(3), pp.227-263. DOI: 10.1191/026765897666180760

Mesthrie, R. (2004). Synopsis: morphological and syntactic variation in Africa and South and Southeast Asia. In Kortmann, B., Burridge, K., Mesthrie, R., Schneider, E. W., and Upton, C. (Eds.). *A Handbook of Varieties of English. Volume 2: Morphology and Syntax*. Mouton de Gruyter, pp.1132-1141. DOI: https://doi.org/10.1515/9783110197181-141

Mesthrie, R. (2014). A robust, living substratum: contact and sociolinguistic factors in the evolution of a variety of Black English in Kimberley, South Africa. In Green, E. and Meyer, C. F. (Eds.). *The Variability of Current World Englishes*. De Gruyter Mouton, pp.127-144.

Mesthrie, R. and Bhatt, R. M. (2008). *World Englishes: The Study of New Linguistic Varieties*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511791321

Ministry of Education (1956). *Report of the Education Committee*. Kuala Lumpur: The Government Press.

Ministry of Education Malaysia (2014). *Upholding the Malay Language and Strengthening the English Language*. Putrajaya: Kementerian Pendidikan Malaysia.

      Reference in Malay:

      Kementerian Pendidikan Malaysia (2014). *Dasar Memartabatkan Bahasa Malaysia Memperkukuh Bahasa Inggeris (MBMMBI)*. Putrajaya: Kementerian Pendidikan Malaysia.

      Available at: https://www.moe.gov.my/dasar/1198-dasar-memartabatkan-bahasa-malaysia-memperkukuh-bahasa-inggeris-mbmmbi/file [Accessed 08-December-2022]

Ministry of Education Malaysia (2013). *Malaysia Education Blueprint 2013 – 2025: Preschool to post-secondary education*. Putrajaya: Kementerian Pendidikan Malaysia. Available at: https://www.moe.gov.my/menumedia/media-cetak/penerbitan/dasar/1207-malaysia-education-blueprint-2013-2025/file [Accessed 08-December-2022]

Ministry of Education Malaysia (2015). *English Language Education Reform in Malaysia: The Roadmap 2015 – 2025*. Ministry of Education Malaysia.

Ministry of Education Malaysia (2019). *Implementation of the Teaching and Learning of Science and Mathematics for Primary and Secondary Schools.*

      Reference in Malay:

      Kementerian Pendidikan Malaysia (2019). *Pelaksanaan Pengajaran dan Pembelajaran (P&P) Sains dan Matematik Bagi Sekolah Rendah dan Sekolah Menengah.*

Available at: https://www.moe.gov.my/?view=article&id=4211:pelaksanaan-pengajaran-dan-pembelajaran-p-p-sains-dan-matematik-bagi-sekolah-rendah-dan-sekolah-menengah&catid=2 [Accessed 08-December-2022]

Mohd. Don, Z. and Abdullah, M. H. (2019). The reform of English Language education in Malaysia. *Free Malaysia Today*. 23 May. Available at: https://www.moe.gov.my/en/menumedia/printed-media/newspaper-clippings/the-reform-of-english-language-education-in-malaysia-free-malaysia-today-22-mei-2019 [Accessed 08-December-2022]

Montrul, S. and Slabakova, R. (2003). Competence similarities between native and near-native speakers. *Studies in Second Language Acquisition*, 25(3), pp.351-398. DOI: 10.1017/S0272263103000159

Morais, E. (1998). Language choice in a Malaysian car-assembly plant. *International Journal of the Sociology of Language*, 130, pp.89-105. DOI: 10.1515/ijsl.1998.130.89

Mougeon, R., Nadasdi T., and Rehner, K. (2005). Contact-induced linguistic innovations on the continuum of language use: The case of French in Ontario. *Bilingualism: Language and Cognition*, 8(2), pp.99-115. DOI: 10.1017/S1366728905002142

Müller, N. and Hulk, A. (2001). Crosslinguistic influence in bilingual language acquisition: Italian and French as recipient languages. *Bilingualism: Language and Cognition*, 4(1), pp.1-21. DOI: 10.1017/S1366728901000116

Munalim, L. O. (2019). Subject-Auxiliary Inversion in Embedded Questions in Spoken Professional Discourses: A Comparison of Philippine English between 1999 and 2016-2019. *Journal of English as an International Language*, 14(1), pp.40-57. Available at: https://eric.ed.gov/?id=EJ1244513 [Accessed 20-December-2022]

Murphy, V. A. (1997). The effect of modality on a grammaticality judgement task. *Second Language Research*, 13(1), pp.34-65. DOI: 10.1191/026765897671676818

Mustaffa, A. R. (2018). Silent Syntactic Structures in Malay: Copular Clauses. *Journal of Modern Languages*, 28, pp.20-43. Available at: https://mojes.um.edu.my/index.php/JML/article/view/13172 [Accessed 11-December-2022]

Mustaffa, A. R. (2020). Null Expletives in Malay. *GEMA Online Journal of Language Studies*, 20(2), pp.224-243. DOI: 10.17576/gema-2020-2002-13

Myers-Scotton, C. (1992). Comparing codeswitching and borrowing. *Journal of Multilingual & Multicultural Development*, 13(1-2), pp.19-39. DOI: 10.1080/01434632.1992.9994481

Myers-Scotton, C. (1993). *Duelling Languages: Grammatical Structure in Codeswitching*. Oxford: Clarendon Press.

Myers-Scotton, C. (2000). The matrix language frame model: Development and responses. In Jacobson, R. (Ed.). *Codeswitching Worldwide II*. Mouton de Gruyter, pp.23-58.

Myers-Scotton, C. (2002). *Contact Linguistics: Bilingual Encounters and Grammatical Outcomes.* Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780198299530.001.0001

Nair-Venugopal, S. (2000). English, identity and the Malaysian workplace. *World Englishes*, 19(2), pp.205-213. DOI: 10.1111/1467-971X.00169

Nair-Venugopal, S. (2003). Malaysian English, normativity and workplace interactions. *World Englishes*, 22(1), pp.15-29. DOI: 10.1111/1467-971X.00269

National Language Act 1963/67. Kuala Lumpur: Malayan Law Journal Sdn. Bhd. and Percetakan Nasional Malaysia Bhd.

Newbrook, M. (1997). Malaysian English: Status, norms, some grammatical and lexical features. In Schneider, E. W. (Ed.). *Englishes around the world. Volume 2: Caribbean, Africa, Asia, Australia. Studies in honour of Manfred Görlach*. Amsterdam: John Benjamins Publishing Company, pp.229-256.

Newbrook, M (2006). Malaysian English: Status, norms, some grammatical and lexical features. In Bolton, K. and Kachru, B. B. (Eds.). *World Englishes: critical concepts in linguistics*. London: Routledge, pp.390-417.

Ng, J. C. and Diskin-Holdaway, C. (2021). Attitudes to English in contemporary Malaysia. *World Englishes*, 1, pp.1-17. DOI: 10.1111/weng.12562

Nicoladis, E. (2006). Cross-linguistic transfer in adjective-noun strings by preschool bilingual children. *Bilingualism: Language and Cognition*, 9(1), pp.15-32. DOI: 10.1017/S136672890500235X

Nicoladis, E., Palmer, A., and Marentette, P. (2007). The role of type and token frequency in using past tense morphemes correctly. *Developmental Science*, 10(2), pp.237-254. DOI: 10.1111/j.1467-7687.2007.00582.x

Nicoladis, E., Song, J., and Marentette, P. (2012). Do young bilinguals acquire past tense morphology like monolinguals, only later? Evidence form French-English and Chinese-English bilinguals. *Applied Psycholinguistics*, 33, pp.457-479. DOI: 10.1017/S0142716411000439

Nikolaeva, I. (2007). Introduction. In Nikolaeva, I. (Ed.). *Finiteness: Theoretical and Empirical Foundations*. Oxford: Oxford University Press, pp.1-19.

Nikolaeva, I. (2010). Typology in Finiteness. *Language and Linguistics Compass*, 4(12), pp.1176-1189. DOI: 10.1111/j.1749-818X.2010.00253.x

Nomoto, H. and Soh, H. L. (2019). Malay. In Vittrant, A. and Watkins, J. (Eds.). *The Mainland Southeast Asia Linguistic Area*. De Gruyter Mouton, pp.475-522. DOI: 10.1515/9783110401981-011

Norbury, C. F. and Bishop, D. V. M. (2003). Narrative skills of children with communication impairments. *International Journal of Language & Communication Disorders*, 38(3), pp.287-313. DOI: 10.1080/136820310000108133

Norbury, C. F., Gemmell, T., and Paul, R. (2014). Pragmatic abilities in narrative production: a cross-disorder comparison. *Journal of Child Language*, 41(3), pp.485-510. DOI: 10.1017/S030500091300007X

Nothofer, B. (2006). Malay. In Brown, K. (Ed.). *Encyclopedia of Language & Linguistics*. 2nd edition. Elsevier, pp.450-453. DOI: 10.1016/B0-08-044854-2/02130-1

Ntalli, N. (2021). *The acquisition of finiteness in English by child second language learners in intstructed contexts: age of onset and L1 effects.* PhD thesis. University of Cambridge. DOI: 10.17863/CAM.71247

Ortega, L. (2018). Ontologies of language, Second Language Acquisition, and world Englishes. *World Englishes*, 37(1), pp.64-79. DOI: 10.1111/weng.12303

Oxford University Press (2022). *The Oxford 3000*. Oxford: Oxford University Press. Available at: https://www.oxfordlearnersdictionaries.com/wordlist/american_english/oxford3000/ [Accessed-31 December-2022]

Packard, J. L. (2006). Chinese as an Isolating Language. In Brown, K. (Ed.). *Encyclopedia of Language & Linguistics*. 2nd edition. Elsevier, pp.355-358. DOI: 10.1016/B0-08-044854-2/00149-8

Papp, S. (2000). Stable and developmental optionality in native and non-native Hungarian grammars. *Second Language Research*, 16(2), pp.173-200. DOI: 10.1191/026765800666966395

Paradis, J. (2005). Grammatical Morphology in Children Learning English as a Second Language: Implications of Similarities With Specific Language Impairment. *Language, Speech, and Hearing Services in Schools*, 36, pp.172-187. DOI: 10.1044/0161-1461(2005/019)

Paradis, J. and Crago, M. (2001). The Morphosyntax of Specific Language Impairment in French: An Extended Optional Default Account. *Language Acquisition*, 9(4), pp.269-300. DOI: 10.1207/S15327817LA0904_01

Paradis., J., Nicoladis, E., Crago, M., and Genesee, F. (2011). Bilingual children's acquisition of the past tense: a usage-based approach. *Journal of Child Language*, 38(3), pp.554-578. DOI: 10.1017/S0305000910000218

Paradis, J., Rice, M. L., Crago, M., and Marquis, J. (2008). The acquisition of tense in English: Distinguishing child second language from first language and specific language impairment. *Applied Psycholinguistics*, 29, pp.689-722. DOI: 10.1017/S0142716408080296

Paradis, J., Rusk, B., Duncan, T. S., and Govindarajan, K. (2017). Children's Second Language Acquisition of English Complex Syntax: The Role of Age, Input, and Cognitive Factors. *Annual Review of Applied Linguistics*, 37, pp.148-167. DOI: 10.1017/S0267190517000022

Parodi, T. (2006). Functional Categories. In Brown, K. (Ed.). *Encyclopedia of Language & Linguistics*. 2nd edition. Elsevier Ltd., pp.666-668. DOI: 10.1016/B0-08-044854-2/01966-0

Parodi, T. and Tsimpli, I.-M. (2005). 'Real' and apparent optionality in second language grammars: finiteness and pronouns in null operator structures. *Second Language Research*, 21(3), pp.250-285. DOI: 10.1191/0267658305sr248oa

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., and Lindeløv, J. (2019). *PsychoPy 2: experiments in behavior made easy*. Behavior Research Methods. DOI: 10.3758/s13428-018-01193-y

Percillier, M. (2016a). Postcolonial and learner English in Southeast Asia: implications for international communication. In Leitner, G., Hashim, A., and Wolf, H.-G. (Eds.). *Communicating with Asia: The Future of English as a Global Language*. Cambridge: Cambridge University Press, pp.135-152. DOI: 10.1017/CBO9781107477186.010

Percillier, M. (2016b). *World Englishes and Second Language Acquisition: Insights from Southeast Asian Englishes*. Amsterdam: John Benjamins Publishing Company.

Pfaff, C. W. (1979). Constraints on Language Mixing: Intrasentential Code-Switching and Borrowing in Spanish/English. *Language*, 55(2), pp.291-318. Available at: https://www.jstor.org/stable/412586 [Accessed 20-December-2022]

Phillips, C. (2010). Syntax at Age Two: Cross-Linguistic Differences. *Language Acquisition*, 17(1-2), pp.70-120. DOI: 10.1080/10489221003621167

Pillai, S. (2013). Colloquial Malaysian English. In Kortmann, B. and Lunkenheimer, K. (Eds.). *The Mouton World Atlas of Variation in English*. De Gruyter Mouton, pp.573-581. DOI: 10.1515/9783110280128.573

Pillai, S. and Greig, A. (2020). Malaysian English. In Kortmann, B., Lunkenheimer, K., and Ehret, K. (Eds.). *The Electronic World Atlas of Varieties of English*. Available at: http://ewave-atlas.org/languages/58 [Accessed 07-December-2022]

Pillai, S. and Ong, L. T. (2018). English(es) in Malaysia. *Asian Englishes*, 20(2), pp.145-157. DOI: 10.1080/13488678.2018.1459073

Pillai, S., Mohd Don, Z., Knowles, G., and Tang, J. (2010). Malaysian English: an instrumental analysis of vowel contrasts. *World Englishes*, 29(2), pp.159-172. DOI: 10.1111/j.1467-971X.2010.01636.x

Platt, E. J. (1989). *Testing the null subject parameter in adult second language acquisition: a study of Vietnamese and Spanish speakers learning English*. PhD thesis. University of Illinois at Urbana-Champaign.

Platt, J. and Weber, H. (1980). *English in Singapore and Malaysia. Status: Features: Functions*. Oxford: Oxford University Press.

Platt, J., Weber, H., and Ho, M. L. (1983). *Singapore and Malaysia*. Amsterdam: John Benjamins Publishing Company.

Poeppel, D. and Wexler, K. (1993). The Full Competence Hypothesis of Clause Structure in Early German. *Language*, 69(1), pp.1-33. Available at: https://www.jstor.org/stable/416414 [Accessed 24-December-2022]

Poplack, S. (2018). *Borrowing: Loanwords in the Speech Community and in the Grammar*. Oxford: Oxford University Press. DOI: 10.1093/oso/9780190256388.001.0001

Porte, G. K. (2002). *Appraising Research in Second Language Learning: A practical approach to critical analysis of quantitative research*. Amsterdam: John Benjamins Publishing Company.

Pozzan, L. and Quirk, E. (2014). Second language acquisition of English questions: An elicited production study. *Applied Psycholinguistics*, 35(6), pp.1055-1086. DOI: 10.1017/S0142716412000690

Pratt, A. and Grinstead, J. (2007). Optional Infinitives in Child Spanish. *Proceedings of the 2nd Conference on Generative Approaches to Language Acquisition North America (GALANA)*, pp.351-362.

Prévost, P. and White, L. (2000). Missing Surface Inflection or Impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research*, 16(2), pp.103-133. DOI: 10.1191/026765800677556046

Qualtrics (2022). *Qualtrics*. Provo, Utah: Qualtrics. Available at: https://www.qualtrics.com/ [Accessed 31-December-2022]

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/ [Accessed 07-December-2022]

Radford, A. (2004). *English Syntax: An introduction*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511841675

Rajadurai, J. (2007). Sociolinguistic perspectives on variation in non-native varieties of English: The case of Malaysian English. *Multilingua*, 26, pp.409-426. DOI: 10.1515/MULTI.2007.019

Ramli, N. and Ab. Hamid, N. (2020). The Development of qualifiers 'Adalah' and 'Ialah'. *E-jurnal Bahasa dan Linguistik*, 2(2), pp.34-51. Available at:

https://www.ejbl.kuisjournal.com/index.php/e-jbl/article/view/48/17 [Accessed 24-December-2022]

Rice, M. L. and Wexler, K. (1996). Toward Tense as a Clinical Marker of Specific Language Impairment in English-Speaking Children. *Journal of Speech, Language, and Hearing Research*, pp.1239-1257. DOI: 10.1044/jshr.3906.1239

Rice, M. L. and Wexler, K. (2001). *Test of Early Language Impairment*. Pearson Publishing. Available at: http://hdl.handle.net/1808/32569 [Accessed 31-December-2022]

Rice, M. L., Noll, K. R., and Grimm, H. (1997). An Extended Optional Infinitive Stage in German-Speaking Children with Specific Language Impairment. *Language Acquisition*, 6(4), pp.255-295. DOI: 10.1207/s15327817la0604_1

Rice, M. L., Wexler, K., and Cleave, P. L. (1995). Specific Language Impairment as a Period of Extended Optional Infinitive. *Journal of Speech, Language, and Hearing Research*, 38(4), pp.850-863. DOI: 10.1044/jshr.3804.850

Ritter, E. and Wiltschko, M. (2014). The composition of INFL: An exploration of "tense, tenseless" languages, and "tenseless" constructions. *Natural Language & Linguistic Theory*, 32(4), pp.1331-1386. Available at: https://www.jstor.org/stable/43697757 [Accessed 24-December-2022]

Rizzi, L. (1993/1994). Some Notes on Linguistic Theory and Language Development: The Case of Root Infinitives. *Language Acquisition*, 3(4), pp.371-393. DOI: 10.1207/s15327817la0304_2

Robertson, D. (2000). Variability in the use of the English article system by Chinese learners of English. *Second Language Research*, 16(2), pp.135-172. DOI: 10.1191/026765800672262975

Rowland, C. F. (2007). Explaining errors in children's questions. *Cognition*, 104(1), pp.106-134. DOI: 10.1016/j.cognition.2006.05.011

RStudio Team (2020). *RStudio: Integrated Development for R*. Boston, MA: RStudio, PBC. Available at: http://www.rstudio.com/ [Accessed 07-December-2022]

Rubdy, R. (2007). Singlish in the School: An Impediment or a Resource? *Journal of Multilingual and Multicultural Development*, 28(4), pp.308-324. DOI: 10.2167/jmmd459.0

Sagarra, N. and Ellis, N. C. (2013). From seeing adverbs to seeing verbal morphology. *Studies in Second Language Acquisition*, 35(2), pp.261-290. DOI: 10.1017/S0272263112000885

Samuel, M. and Khan, M. H. (2013). Construction of nationhood through education in Malaya: revisiting the Barnes and Fenn-Wu Reports. *Asia Pacific Journal of Education*, 33(3), pp.249-260. DOI: 10.1080/02188791.2013.786678

Samuel, M., Khan, M. H., Ng, L. L., and Cheang, K. W. (2014). Articulation of medium of instruction politics in the Malaysian Chinese press. *Discourse: Studies in the Cultural Politics of Education*, 35(2), pp.206-218. DOI: 10.1080/01596306.2012.745731

Samuel, M., Tee, M. Y., and Pe Symaco, L. (2017). *Education in Malaysia: Developments and Challenges*. Singapore: Springer. DOI: 10.1007/978-981-10-4427-4

Sand, A. (2004). Shared morpho-syntactic features in contact varieties of English: article use. *World Englishes*, 23(2), pp.281-298. DOI: 10.1111/j.0883-2919.2004.00352.x

Sasse, H.-J. (2006). Aspect and Aktionsart. In Brown, K. (Ed.). *Encyclopedia of Language & Linguistics*. 2nd edition. Elsevier Ltd., pp.535-538. DOI: 10.1016/B0-08-044854-2/00261-3

Sato. Y. (2013). *Wh*-Questions in Colloquial Singapore English: Adaptive traits from vernacular Malay and typological congruence. *Journal of Pidgin and Creole Languages*, 28(2), pp.299-322. DOI: 10.1075/jpcl.28.2.02sat

Sato, Y. and Kim, C. (2012). Radical pro drop and the role of syntactic agreement in Colloquial Singapore English. *Lingua*, 122(8), pp.858-873. DOI: 10.1016/j.lingua.2012.02.006

Schiffrin, D. (1981). Tense variation in narrative. *Language*, 57(1), pp.45-62. DOI: 10.1353/lan.1981.0011

Schilling, N. (2013). *Sociolinguistic Fieldwork*. Cambridge: Cambridge University Press.

Schleef, E. (2014). Written Surveys and Questionnaires in Sociolinguistics. In Holmes, J. And Hazen, K. (Eds.). *Research Methods in Sociolinguistics: A Practical Guide*. West Sussex: John Wiley & Sons, Inc., pp.42-57.

Schneider, E. W. (2003). Evolutionary Patterns of New Englishes and the Special Case of Malaysian English. *Asian Englishes*, 6(2), pp.44-63. DOI: 10.1080/13488678.2003.10801118

Schneider, E. W. (2007). *Postcolonial English: Varieties around the world*. Cambridge: Cambridge University Press.

Schneider, E. W. (2008). Accommodation versus identity? A response to Trudgill. *Language in Society*, 37(2), pp.262-267. DOI: 10.1017/S0047404508080317

Schneider, E. W. (2014). New reflections on the evolutionary dynamics of world Englishes. *World Englishes*, 33(1), pp.9-32. DOI: 10.1111/weng.12069

Schreier, D. (2014). On cafeterias and new dialects: The role of primary transmitters. In Buschfeld, S., Hoffmann, T., Huber, M., and Kautzsch, A. (Eds.). *The Evolution of Englishes: The Dynamic Model and beyond*. Amsterdam: John Benjamins Publishing Company, pp.231-248.

Schütze, C. T. and Sprouse, J. (2013). Judgment data. In Podesva, R. J. and Sharma, D. (Eds.). *Research Methods in Linguistics*. Cambridge: Cambridge University Press, pp.27-50. DOI: 10.1017/CBO9781139013734.004

Schütze, C. T. and Wexler, K. (1996). Subject Case Licensing and English Root Infinitives. *BUCLD 20 Proceedings,* pp.670-681.

Schütze, C. T. and Wexler, K. (2000). An Elicitation Study of Young English Children's Knowledge of Tense: Semantic and Syntactic Properties of Optional Infinitives. *Proceedings of the 24th Annual Boston University Conference on Language Development*, pp.1-15. Available at: https://escholarship.org/uc/item/7kb2v2gg [Accessed 24-December-2022]

Schwartz, B. D. and Sprouse, R. A. (1994). Word Order and Nominative Case in Non-Native Language Acquisition: A Longitudinal Study of (L1 Turkish) German Interlanguage. In Hoekstra, T. and Schwartz, B. D. (Eds.). *Language acquisition studies in generative grammar: Papers in honor of Kenneth Wexler from the 1991 GLOW Workshop*. Amsterdam: John Benjamins Publishing Company, pp.317-368.

Schwartz, B. D. and Sprouse, R. A. (1996). L2 cognitive states and the Full Transfer/Full Access model. *Second Language Research*, 12(1), pp.40-72. Available at: https://www.jstor.org/stable/43104504 [Accessed 31-December-2022]

Sharma, D. (2009). Typological diversity in New Englishes. *English World-Wide*, 30(2), pp.170-195. DOI: 10.1075/eww.30.2.04sha

Shi, Y. and Li, C. N. (2002). The establishment of the classifier system and the grammaticalization of the morphosyntactic particle *de* in Chinese. *Language Science*, 24(1), pp.1-15. DOI: 10.1016/S0388-0001(00)00048-6

Shively, R. (2018). Naturalistic data in L2 pragmatics research: Challenges and opportunities. In Gudmestad, A. and Edmonds, A. (Eds.). *Critical Reflections on Data in Second Language Acquisition*. Amsterdam: John Benjamins Publishing Company, pp.197-218.

Siegel, J. (2003). Social Context. In Doughty, C. J. and Long, M. H. (Eds.). *The Handbook of Second Language Acquisition*. Oxford: Blackwell Publishing, pp.178-223.

Slabakova, R. (2016). *Second Language Acquisition*. Oxford: Oxford University Press.

Smakman, D. and Wagenaar, S. (2013). Discourse particles in Colloquial Singapore English. *World Englishes*, 32(3), pp.308-324. DOI: 10.1111/weng.12033

Smith, C. S. (1997). *The Parameter of Aspect*. 2nd edition. Dordrecht: Kluwer Academic Publishers.

Smith, C. S. (2019). Tense and aspect: Time across languages. In von Heusinger, K., Maienborn, C., and Portner, P. (Eds.). *Semantics: Typology, Diachrony and Processing*. De Gruyter Mouton, pp.57-91.

Snack Attack (2012). Cadelago, A. Metanoia Films and Arc Productions. Available at: https://snackattackmovie.com/ [Accessed 31-December-2022]

Sprouse, J., Schütze, C. T., and Almeida, D. (2013). A comparison of informal and formal acceptability judgments using a random sample from *Linguistic Inquiry* 2001-2010. *Lingua*, 134, pp.219-248. DOI: 10.1016/j.lingua.2013.07.002

Soh, H. L. and Nomoto, H. (2009). Progressive Aspect, the Verbal Prefix *meN*-, and Stative Sentences in Malay. *Oceanic Linguistics*, 48(1), pp.148-171. DOI: 10.1353/ol.0.0026

Sorace, A. (2000). Introduction: Syntactic optionality in non-native grammars. *Second Language Research*, 16(2), pp.93-102. Available at: https://www.jstor.org/stable/43103572 [Accessed 20-December-2022]

Sorace, A. (2003). Near-Nativeness. In Doughty, C. J. and Long, M. H. (Eds.). *The Handbook of Second Language Acquisition*. Blackwell Publishing, pp.130-151. DOI: 10.1002/9780470756492.ch6

Sorace, A. (2005). Selective optionality in language development. In Cornips, L. and Corrigan, K. P. (Eds.). *Syntax and variation: reconciling the biological and the social*. Amsterdam: John Benjamin Publishing Company, pp.55-80.

Sorace, A. and Serratrice, L. (2009). Internal and external interfaces in bilingual language development: Beyond structural overlap. *International Journal of Bilingualism*, 13(2), pp.195-210. DOI: 10.1177/1367006909339810

Sridhar, K. K. and Sridhar, S. N. (1986). Bridging the paradigm gap: second language acquisition theory and indigenized varieties of English. *World Englishes*, 5(1), pp.3-14. DOI: 10.1111/j.1467-971X.1986.tb00636.x

Stringer, D. (2015). Embedded *wh*-questions in L2 English in India: Inversion as a main clause phenomenon. *Studies in Second Language Acquisition*, 37(1), pp.101-133. DOI: 10.1017/S0272263114000357

Sun, C. (2006). *Chinese: A Linguistic Introduction.* Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511755019

Sybesma, R. (2004). Exploring Cantonese tense. *Linguistics in the Netherlands*, 21, pp.169-180.

Sybesma, R. P. E. (2007). Whether We Tense-Agree Overtly or Not. *Linguistic Inquiry*, 38(3), pp. 580 – 587. Available at: https://muse.jhu.edu/article/219273 [Accessed 24-December-2022]

Szmrecsanyi, B. and Kortmann, B. (2009a). The morphosyntax of varieties of English worldwide: A quantitative perspective. *Lingua*, 119(11), pp.1643-1663. DOI: 10.1016/j.lingua.2007.09.016

Szmrecsanyi, B. and Kortmann, B. (2009b). Vernacular Universals and Angloversals in a Typological Perspective. In Filppula, M., Klemola, J., and Paulasto, H. (Eds.). *Vernacular Universals and Language Contacts: Evidence from Varieties of English and Beyond*. Routledge, pp.33-53.

Szmrecsanyi, B. and Kortmann, B. (2011). Typological profiling: Learner Englishes versus indigenized L2 varieties of English. In Mukherjee, J. and Hundt, M. (Eds.). *Exploring Second-Language Varieties of English and Learner Englishes: Bridging a paradigm gap*. Amsterdam: John Benjamins Publishing Company, pp.167-187.

Tan, M. (2011). Mathematics and science teachers' beliefs and practices regarding the teaching of language in content learning. *Language Teaching Research*, 15(3), pp.325-342. DOI: 10.1177/1362168811401153

Tan, P. K. W. and Tan, D. H. (2008). Attitudes towards non-standard English in Singapore. *World Englishes*, 27(3/4), pp.465-479. DOI: 10.1111/j.1467-971X.2008.00578.x

Tan, Y. S. (2009). Educational development in Malaysia: The roles of Lim Lian Geok and Aminuddin Baki. *CenPRIS Working Paper* 108/09, pp.1-23.

Tan, Y. Y. (2003). Reading the Census: Language Use in Asia. In Lindsay, J. and Tan, Y. Y. (Eds.). *Babel or Behemoth: language trends in Asia*. Singapore: Asia Research Institute, National University of Singapore, pp.175-210.

Tay, L. C., Chan, M. Y., Yap, N. T., and Wong, B. E. (2016). Discourse Particles in Malaysian English. *Brill*, 172(4), pp.479-509. DOI: 10.1163/22134379-17204002

Taylor, B. P. (1975). Adult Language Learning Strategies and Their Pedagogical Implications. *TESOL Quarterly*, 9(4), pp.391-399. Available at: https://www.jstor.org/stable/3585623 [Accessed 17-January-2023]

Tesan, G. and Thornton, R. (2004). Overspecification in Early English. *LOT Occasional Series*, 3, pp.469-480. Available at: https://dspace.library.uu.nl/handle/1874/296254 [Accessed 25-December-2022]

Thilagawathi, K. and Ramasamy, K. (2004). Indian languages. In Haji Omar, A. (Ed.). *The Encyclopedia of Malaysia. Volume 9: Languages and Literature*. Archipelago Press, pp.56-57.

Thirusanku, J. and Md. Yunus, M. (2012). Review Article: The Many Faces of Malaysian English. *International Scholarly Research Network*, 2012, pp.1-14. DOI: 10.5402/2012/138928

Thirusanku, J. and Md. Yunus, M. (2013). Malaysian English – A Distinct Variety. *Internet Journal of Language, Culture and Society*, 37, pp.18-29.

Thomas, M. (1994). Review Article: Assessment of L2 Proficiency in Second Language Acquisition Research. *Language Learning*, 44(2), pp.307-336.

Tija, J. (2004). The development of possessive and passive-like constructions in Ambonese language. *NUSA*, 53, pp.51-62.

Tremblay, A. (2011). Proficiency assessment standards in second language acquisition research: "Clozing" the Gap. *Studies in Second Language Acquisition*, 33(3), pp.339-372. DOI: 10.1017/S0272263111000015

Tsimpli, I. M. (2014). Early, late or very late? Timing acquisition and bilingualism. *Linguistic Approaches to Bilingualism*, 4(3), pp.283-313. DOI: 10.1075/lab.4.3.01tsi

Tsimpli, I. M. and Dimitrakopoulou, D. (2007). The Interpretability Hypothesis: evidence from *wh*-interrogatives in second language acquisition. *Second Language Research*, 23(2), pp.215-242. DOI: 10.1177/0267658307076546

Tsimpli, I. M., Mukhopadhyay, L., Balasubramanian, A., and Treffers-Daller, J. (in press). *Microstructural properties in the narrative retellings of young English learners in EMI schools in India: the role of L2 literacy, minority languages and English input in the classroom*.

Unsworth, S. (2014). Comparing the role of input in bilingual acquisition across domains. In Grüter, T. and Paradis, J. (Eds.). *Input and Experience in Bilingual Development*. Amsterdam: John Benjamins Publishing Company, pp.181-201.

Unsworth, S., Argyri, F., Cornips, L., Hulk, A., Sorace, A., and Tsimpli, I. (2014). The role of age of onset and input in early child bilingualism in Greek and Dutch. *Applied Psycholinguistics*, 35(4), pp.765-805. DOI: 10.1017/S0142716412000574

Uzawa, H. (2009). A Study on the Pragmatic Functions of *ialah* and *adalah* in Malay. In Kawaguchi, Y., Minegishi, M., and Durand, J. (Eds.). *Corpus Analysis and Variation in Linguistics*. Amsterdam: John Benjamins Publishing Company, pp.315-338.

van Der Lely, H. K. J. and Battell, J. (2003). WH-Movement in Children with Grammatical SLI: A Test of the RDDR Hypothesis. *Language*, 79(1), pp.153-181. Available at: https://www.jstor.org/stable/4489388 [Accessed 29-December-2022]

van Der Lely, H. K. J., Jones, M., and Marshall, C. R. (2011). Who did Buzz see someone? Grammaticality judgement of wh-questions in typically developing children and children with Grammatical-SLI. *Lingua*, 121(3), pp.408-422. DOI: 10.1016/j.lingua.2010.10.007

van Osch, B., Hulk, A. H., Aalberse, S., and Sleeman, P. (2018). Implicit and Explicit Knowledge of a Multiple Interface Phenomenon: Differential Task Effects in Heritage Speakers and L2 Speakers of Spanish in The Netherlands. *Languages*, 3(25), pp.1-20. DOI: 10.3390/languages3030025

Vendler, Z. (1967). *Linguistics in Philosophy*. Cornell University Press, pp.97-121. DOI: 10.7591/9781501743726

Wang, J. (2016). Resumptive pronouns are not optional: evidence from topic constructions of the possessor in Mandarin Chinese. *Coyote Papers*, 14, pp.160-195. Available at: http://hdl.handle.net/10150/126612 [Accessed 11-December-2022]

Wee, L. (2018). World Englishes, Second Language Acquisition, and the linguistic system conundrum. *World Englishes*, 37(1), pp.51-63. DOI: 10.1111/weng.12302

Weskott, T. and Fanselow, G. (2009). Scaling issues in the measurement of linguistic acceptability. In Featherston, S. and Winkler, S. (Eds.). *The Fruits of Empirical Linguistics*. Volume 1: Process. Berlin: Mouton de Gruyter, pp.229-245.

Wexler, K. (1994). Finiteness and head movement in early child grammars. In Lightfoot, D. and Hornstein, N. (Eds.). *Verb movement*. Cambridge: Cambridge University Press, pp.305-350. DOI: 10.1017/CBO9780511627705.016

Wexler, K. (1998). Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive. *Lingua*, 106, pp.23-79. DOI: 10.1016/S0024-3841(98)00029-1

Wexler, K. (2011). Grammatical Computation in the Optional Infinitive Stage. In de Villiers, J. and Roeper, T. (Eds.). *Handbook of Generative Approaches to Language Acquisition*. Springer, pp.53-118. DOI: 10.1007/978-94-007-1688-9_3

Wexler, K., Schaeffer, J., and Bol, G. (2004). Verbal syntax and morphology in typically developing Dutch children and children with SLI: How developmental data can play an important role in morphological theory. *Syntax*, 7(2), pp.148-198. DOI: 10.1111/j.1467-9612.2004.00006.x

Wexler, K., Schütze, C. T., and Rice, M. (1998). Subject Case in Children With SLI and Unaffected Controls: Evidence for the Agr/Tns Omission Model. *Language Acquisition*, 7(2-4), pp. 317-344. DOI: 10.1207/s15327817la0702-4_8

White, L. (2003). *Second Language Acquisition and Universal Grammar*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511815065

Wigdorowitz, M. (2022). *The Case for Contextual Linguistic Diversity: Language Profiling, Multilingual Identity, and High-Level Listening Comprehension Ability in South African University Students*. PhD thesis. University of Cambridge. DOI: 10.17863/CAM.89625

Wigdorowitz, M., Pérez, A. I., and Tsimpli, I. M. (2020). A holistic measure of contextual and individual linguistic diversity. *International Journal of Multilingualism*, pp.1-19. DOI: 10.1080/14790718.2020.1835921

Wijnen, F., Kempen, M., and Gillis, S. (2001). Root infinitives in Dutch early child language: an effect of input? *Journal of Child Language*, 28, pp.629-660. DOI: 10.1017/S0305000901004809

Wolfson, N. (1978). A feature of performed narrative: the conversational historical present. *Language in Society*, 7(2), pp.215-237. Available at: https://www.jstor.org/stable/4166998 [Accessed 31-December-2022]

Wong, B. E. (2008). *Wh*-Questions in Malay: An Explanation for the Restriction of Extraction to Subject Position with *Yang*. *3L: Language, Linguistics, Literature*, 14, pp.109-126.

Available at: http://journalarticle.ukm.my/1081/1/6-Wong_Bee_Eng_6.pdf [Accessed 11-December-2022]

Wong, B. E. (2012). Acquisition of English Tense and Agreement Morphology by L1 Malay and L1 Chinese Speakers. *3L: The Southeast Asian Journal of English Language Studies*, 18(3), pp.5-14. Available at: https://ejournal.ukm.my/3l/article/viewFile/1105/997 [Accessed 11-December-2022]

Wong, B. E. and Quek, S. T. (2007). Acquisition of the English Definite Article by Chinese and Malay ESL Learners. *Electronic Journal of Foreign Language Teaching*, 4(2), pp.210-234.

Wong, J. (2004). The particles of Singapore English: a semantic and cultural interpretation. *Journal of Pragmatics*, 36(4), pp.739-793. DOI: 10.1016/S0378-2166(03)00070-5

Wu, J.-S. (2009). Tense as a discourse feature: rethinking temporal location in Mandarin Chinese. *Journal of East Asian Linguistics*, 18, pp.145-165. DOI: 10.1007/s10831-009-9042-3

Yiakoumetti, A. (2006). A Bidialectal Programme for the Learning of Standard Modern Greek in Cyprus. *Applied Linguistics*, 27(2), pp.295-317. DOI: 10.1093/applin/aml012

Yuan, B. (1998). Interpretation of binding and orientation of the Chinese reflexive *ziji* by English and Japanese speakers. *Second Language Research*, 14(4), pp.324-340. DOI: 10.1191/026765898670904111

Yuan, B. and Dugarova, E. (2012). *Wh*-topicalization at the syntax-discourse interface in English speakers' L2 Chinese grammars. *Studies in Second Language Acquisition*, 34(3), pp.533-560. DOI: 10.1017/S0272263112000332

Zeng, X., Shirai, Y., and Chen, X. (2019). Universals and transfer in the acquisition of the progressive aspect: Evidence from L1 Chinese, German, and Spanish learners' use of the progressive *-ing* in spoken English. *International Review of Applied Linguistics in Language Teaching*, 59(2), pp.267-292. DOI: 10.1515/iral-2017-0078

Zhan, F. and Sun, C. (2013). A Copula Analysis of *shì* in the Chinese Cleft Construction. *Language and Linguistics*, 14(4), pp.755-789.

Zhang, Y. (2005). Processing and formal instruction in the L2 acquisition of five Chinese grammatical morphemes. In Pienemann, M. (Ed.). *Cross-Linguistic Aspect of Processability Theory*. Amsterdam: John Benjamins Publishing company, pp.155-177.

Ziegeler, D. (2020). Changes in the functions of *already* in Singapore English. *Journal of Pidgin and Creole Languages*, 35(2), pp.293-331. DOI: 10.1075/jpcl.00062.zie

Ziegeler, D. and Lee, S. (2019). Lexical Retention in Contact Grammaticalisation: *Already* in Southeast Asian Englishes. *Journal of Language Contact*, 12(3), pp.737-783. DOI: 10.1163/19552629-01203006

Appendices

A. Instruments

A.1. CME case study: Questionnaire

Multilingualism Questionnaire

General information				
Age in 2018				
Nationality				
Ethnicity				
Familial information				
Father's nationality				
Father's ethnicity				
Mother's nationality				
Mother's ethnicity				
Education background				
Education level				
Medium of instruction	Primary school	Secondary school	Pre-university (e.g. STPM; A-levels)	University	Others:
Have you been abroad for further studies? If yes, where?				
General language profile				
What is your first language?				
What languages do you speak?	Malay ☐	English ☐	Chinese ☐	Tamil ☐	Others:

At what age were you first exposed to these languages?	Malay	English	Chinese	Tamil	Others:

At what age did you start using these languages?	Malay	English	Chinese	Tamil	Others:

How often do you use the languages?	Malay	English	Chinese	Tamil	Others:

0: Never (0%); **1**: Rarely (25%); **2**: Sometimes (50%); **3**: Often (75%); **4**: All the time (100%)

How well do you speak these languages?	Malay	English	Chinese	Tamil	Others:

0: Zero knowledge: *I can't speak this language at all.*
1: Not fluent: *I can only use a few basic words.*
2: Not so fluent: *I have enough words to form a complete sentence, but I struggle with the grammar.*
3: Quite fluent: *I can use this language quite comfortably, but I don't sound like a native speaker.*
4: Very fluent: *I am a native speaker of this language.*

Language use

Language(s) used at home

With father	
With mother	
With sibling 1	
With sibling 2	
With sibling 3	
With sibling 4	
With sibling 5	
With sibling 6	
With sibling 7	
With sibling 8	

With sibling 9	
With sibling 10	

Language(s) used with non-family members	
With friends of the same ethnicity	
With friends of different ethnicities	
With strangers	

Language(s) used in school / college / university					
With classmates of the same ethnicity	Primary school	Secondary school	Pre-university (e.g. STPM; A-levels)	University	Others:
With classmates of different ethnicities	Primary school	Secondary school	Pre-university (e.g. STPM; A-levels)	University	Others:
With educators with the same ethnicity	Primary school	Secondary school	Pre-university (e.g. STPM; A-levels)	University	Others:
With educators with different ethnicity	Primary school	Secondary school	Pre-university (e.g. STPM; A-levels)	University	Others:

English language profile				
Is English your first or second language?				
What is your English proficiency, based on your English assessments?	SPM	MUET	IELTS	Others

How much are you exposed to English, based on the following activities?				
Reading (e.g. books, newspapers, news, Internet articles)	Never	Rarely	Once a week	Every day
Watching/listening (e.g. TV, movies, dramas, radio)	Never	Rarely	Once a week	Every day

Thank you for your participation!

A.2. Main study

A.2.1. Grammaticality judgement task

A.2.1.1. Presentation flow

Step	Presentation
1	**Welcome**
	Welcome! You are about to rate a series of sentences in Standard English. This task is expected to last between 20 and 30 minutes, with a short break in between.
	Each sentence will be presented for 3.5 seconds. Use the MOUSE to rate how grammatical it is from 1 (least grammatical) to 5 (completely grammatical). If you are unsure about the sentence, press the 0 (zero) key.
	You may read the sentences aloud if it helps you focus better.
	Please note that there are no errors in spelling and punctuation.
	Press SPACEBAR to have a quick practice.

2	**Practice trial**
	a. Present stimulus one at a time (3.5 seconds).
	b. After the stimulus disappears, display 5-point Likert scale.
3	**Main trial instructions**
	Ready to begin the main experiment?
	Reminder: Use the MOUSE to click on the rating scale. If you are unsure about the sentence, press the 0 (zero) key. Also, note that there are no errors in spelling and punctuation.
	Press SPACEBAR to begin.
4	**Main trial: Block A**
	Same as 2a-b
5	**Break**
	Yay! You're halfway through the experiment.
	Time to refresh yourself with a short break.
	Once you're done, press SPACEBAR to resume the experiment.
6	**Main trial: Block B**
	Same as 2a-b

7	**End**
	Thank you for taking part in the sentence judgement task!

A.2.1.2. Sentence stimuli

Practice

1	The beautiful garden is hidden behind the trees.
2	How many opponents were defeated by the tennis player?
3	What does the rich man do for a living?
4	Some kitchens old were renovated last month.
5	How many lesson are taught by the professor?
6	Many international flights were cancelled this morning why?

Experimental

Copula

Copula+PP

Grammatical

7	Usually, all photo albums are in the glass cupboard.
8	Obviously, the old printer is on the wooden desk.
9	Apparently, some little cottages are near a freshwater lake.
10	As always, the international airport is by the sea.
11	Apparently, the mysterious recording is with the security guard.
12	Surprisingly, the wool carpet is from a local supermarket.

Omission

13   As usual, the old laboratory by the polluted river.

14   Normally, several silver buttons with the fashion designer.

15   Usually, several leather jackets from a charity shop.

16   Surprisingly, the new accommodation near the railway station.

17   Normally, some academic journals on the coffee table.

18   As always, all family portraits in the living room.

Copula+AdjP

Grammatical

19   All washing liquids are actually quite effective.

20   The market square is normally very packed.

21   A lot of language classes are apparently rather boring.

22   The beach resort is usually fairly popular.

23   Some root vegetables are surprisingly very healthy.

24   The sports centre is always very crowded.

Omission

25   The civil service uniform normally very colourful.

26   Quite a few free art exhibitions apparently rather interesting.

27   The luxury hotel reception always super busy.

28   Almost all summer music festivals actually quite exciting.

29   The traditional wedding ceremony usually fairly lively.

30   A lot of deep sea creatures surprisingly really dangerous.

Progressivity

Grammatical

31   Today, all doctors are submitting their medical reports.

32   At this time, the authorities are assessing the conflict.

33   At the moment, the secretary is copying the document.

34   Tonight, some artists are performing their latest songs.

35   Currently, some critics are reviewing the wildlife documentary.

36   Right now, the machine is processing cocoa beans.

Omission

37   Currently, the computer installing its software updates.

38   This year, all students studying a language subject.

39   At this time, the engineer modifying the car engine.

40   At the moment, certain researchers evaluating their study's findings.

41   Right now, some detectives investigating the crime scene.

42   Today, the psychologist examining his depressed patient.

Commission

43   All hardworking teachers are deserving everyone's respect.

44   The study room is containing many rare objects.

45   The foreign visitor is preferring winter over summer.

46   Some fitness instructors are believing in Buddhism.

47   Certain historical sites are seeming frightening.

48   The university library is possessing eight million books.

3rd.SING.-*s*

Grammatical

49   The cancer survivor truly appreciates people's support.

50   The expensive chocolate bar tastes very bitter.

51   The enthusiastic traveller loves all kinds of landscape photography.

52   The current president dislikes his political party.

53   The young public speaker looks particularly confident.

54   The ancient musical instrument sounds incredibly beautiful.

Omission

55   Every afternoon, the pupil practise his guitar.

56   Every month, the volunteer collect donated clothes.

57   Every year, the institution employ fresh graduates.

58   Every morning, the scientist conduct her experiments.

59   Every week, the banker analyse the stock market.

60   Every day, the driver worry about road accidents.

Past *-ed*

Grammatical

61   Last weekend, the celebrity purchased a few famous paintings.

62   A few years ago, the flood destroyed many properties.

63   Three days ago, the criminal murdered four innocent children.

64   Last July, the tourist escaped the forest fire.

65   Yesterday morning, the employer cancelled the job interview.

66   Last night, the farmer buried his favourite horse.

Omission

67   Last year, the industry generate a lot of income.

68   A week ago, the noise interrupt most nearby residents.

69   Years ago, the hero defeat his greatest enemy.

70   Last Monday, the school issue twenty warning letters.

71   Last month, the athlete participate in a regional competition.

72   Two days ago, the officer arrest three thieves.

Commission

73   Yesterday evening, the gardener encounters an unusual insect.

74   Last night, the device detects the stolen diamond.

75   Last Friday, the university celebrates its tenth anniversary.

76   Three weeks ago, the sailor locates five hidden islands.

77   Last April, the teacher contacts the pupil's father.

78   A few weeks ago, the council rejects the proposed scheme.

DO-support

Grammatical

79   Did any musicians recognise the background music?

80   Did all leaders attend the international conference?

81   Did the lady adopt a teenage orphan?

82   Did certain employees demand a pay rise?

83   Did the court release a revised policy?

84   Did the organisation supply free monthly calendars?

Omission

85   Last year, all products satisfy the client's expectation?

86   Last night, the programme repeat the political debate?

87   Last summer, any events attract many foreign tourists?

88   Last week, certain companies alter the original contract?

89   Last month, the training benefit most junior nurses?

90   Last weekend, the construction bother the nearby neighbourhood?

Commission

91   Did the manager replied to the angry customer?

92   Did any passengers survived the plane crash?

93   Did all members selected a new agent?

94   Did certain producers organised a committee meeting?

95   Did the campaign featured a few leading speakers?

96   Did the chemical enhanced the reaction time?

Fillers

Plural -*s*

Grammatical

97   Quite a few issues were highlighted in the magazine.

98   Almost all flights were delayed because of the storm.

99   Dozens of proposals were criticised by several experts.

100 A couple of towers were transformed into fancy resorts.

101 A few suspects were captured by the armed forces.

102 A lot of containers were lowered into the ground.

Omission

103 A lot of book were distributed to rural schools.

104 A couple of building were constructed by the architect.

105 Quite a few novel were translated into different languages.

106 A few state are governed by a military group.

107 Almost all solution are outlined in the article.

108 Dozens of house were damaged because of the earthquake.

Adj-NP word order

Grammatical

109 Most negative outcomes were associated with poor judgement.

110 Some famous celebrities were connected to a secret society.

111 Certain religious ceremonies were observed by native communities.

112 Many abandoned castles were converted into art museums.

113 Several ancient palaces were surrounded by electrical fences.

114 All music files were arranged in alphabetical order.

Commission

115 All topics complex were separated into ten lectures.

116 Most tales classical were adapted into successful films.

117 Many theories famous were proposed by the lecturer.

118 The cake delicious was ordered for a birthday celebration.

119 Some plants delicate were exposed to too much sunlight.

120 Certain strategies business were declined by the government.

Wh-movement

Grammatical

121 Where will the librarian store all personal information?

122 Where will most chefs demonstrate their cooking techniques?

123 Where will the assistant register the company car?

124 Where will the mother deliver her firstborn son?

125 Where will all winners display their gold trophies?

126 Where will the captain abandon the pirate ship?

Commission

127 Most representatives will address the general audience at where?

128 Some historians will acquire a few ancient scrolls from where?

129 The minister will welcome the royal family to where?

130 All soldiers will carry the military rockets to where?

131 The professor will interview all research candidates at where?

132 Some investigators will recover the misused funds from where?

Misc.

Grammatical

133 How many investors are protected against financial losses?

134 How many postgraduates are accused of academic misconduct?

135 How many victims were injured from the nuclear explosion?

136 How many colleagues are threatened by the senior manager?

137 How many painters are acknowledged for their brilliant artwork?

138 How many books were recommended by the young librarian?

139 How many participants were excluded from the pilot study?

140 How many princesses were married into foreign families?

141 How many offenders were defended by the old lawyer?

142 How many politicians were approached by news reporters?

143 How many immigrants were transferred to the detention centre?

144 How many parents were informed about the school regulations?

145 How many players were gathered at the football stadium?

146 How many dentists were assisted by their junior assistants?

147	How many citizens were attacked during the civil war?
148	How many members were ignored during the annual meeting?
149	How many students were measured for their school uniforms?
150	How many teenagers were handled by the strict teacher?

A.2.2. Language background and attitude questionnaire

Participant information sheet

Language Background Questionnaire

You are about to take part in a questionnaire, which is part of a PhD study run by Samantha Sie at the University of Cambridge. The purpose of this task is to collect demographic and language-related information which may represent the language backgrounds of Malaysia and the United Kingdom (UK). Individuals who wish to participate must meet all the following criteria:

- Either Malaysian or UK nationality
- 18 years or above
- Learned English at any point in life

There are altogether 60 questions in this questionnaire, which should take less than 15 minutes to complete. Please be assured that your data are anonymised, non-identifiable, and confidential. They are only accessible to researchers involved in this project and will be stored safely in compliance with the Data Protection Act.

There are no right or wrong answers. Please be as truthful as you can when answering all questions. Should there be any queries or concerns, please do not hesitate to get in touch with me. Thank you for your support!

Yours sincerely,
Samantha Sie ([email redacted])
PhD student,
Faculty of Modern and Medieval Languages and Linguistics,
University of Cambridge.

Consent form

Please confirm that you have read the information sheet and agree with **all** the following statements:
- ☐ I have read and understood the participation information sheet for this study.
- ☐ I understand that my participation is voluntary. I have the right to withdraw from the questionnaire at any time up until the point of submission without needing to provide any reason.

☐ I understand that the data collected in this study will be anonymised, made non-identifiable, and stored securely in compliance with the Data Protection Act.

☐ I understand that the anonymised data may be used in analyses, publications, and conference presentations by the researcher and her research collaborators from the University of Cambridge and other research institutions. I grant these individuals permission to use these data.

☐ By complying to the aforementioned statements, I agree to take part in this study.

Participant unique ID

Nationality

Where are you from?
o Malaysia
o United Kingdom

Section A: Language history and use

A1 What language(s)/dialect(s) have you learned?
You may choose more than one language.
☐ English
☐ Malay
☐ Mandarin
☐ Tamil
☐ Other (x6): _____

Carry Forward Selected Choices from A1

A2 At roughly what age (in years) did you start learning the language(s)?
If the language is your first language, state 0.
☐ English: _____
☐ Malay: _____
☐ Mandarin: _____
☐ Tamil: _____
☐ Other: _____

Carry Forward Selected Choices from A1

A3 How often do you **speak** the following language(s)?
This may include face-to-face conversations and phone calls, but excluding text-messaging and emails.

	Daily	Weekly	Monthly	Yearly	Never
English	○	○	○	○	○
Malay	○	○	○	○	○
Mandarin	○	○	○	○	○
Tamil	○	○	○	○	○
Other	○	○	○	○	○

A4   How often do you **write/type** in the following language(s)? *This may include academic writing, journaling, drafting emails, and texting on social media.*

	Daily	Weekly	Monthly	Yearly	Never
English	○	○	○	○	○
Malay	○	○	○	○	○
Mandarin	○	○	○	○	○
Tamil	○	○	○	○	○
Other	○	○	○	○	○

A5   How often do you **listen** to the following language(s)? *This may include listening to the radio, lectures, podcasts, songs, and videos without subtitles.*

	Daily	Weekly	Monthly	Yearly	Never
English	○	○	○	○	○

Malay	○	○	○	○	○
Mandarin	○	○	○	○	○
Tamil	○	○	○	○	○
Other	○	○	○	○	○

A6   How often do you **read** in the following language(s)?
This may include reading books, articles, news stories (online), feeds on social media, and video subtitles.

	Daily	Weekly	Monthly	Yearly	Never
English	○	○	○	○	○
Malay	○	○	○	○	○
Mandarin	○	○	○	○	○
Tamil	○	○	○	○	○
Other	○	○	○	○	○

A7   What language(s) do you use with the following people, if applicable?
You may choose more than one language.

	English	Malay	Mandarin	Tamil	Other (5)
Your parents	☐	☐	☐	☐	☐
Your siblings	☐	☐	☐	☐	☐
Your relatives	☐	☐	☐	☐	☐

Friends of the same ethnicity	☐	☐	☐	☐	☐
Friends of different ethnicities	☐	☐	☐	☐	☐
Classmates and/or colleagues of the same ethnicity	☐	☐	☐	☐	☐
Classmates and/or colleagues of different ethnicities	☐	☐	☐	☐	☐
Strangers of the same ethnicity	☐	☐	☐	☐	☐
Strangers of different ethnicities	☐	☐	☐	☐	☐

A8   Which of the following school languages are/were used during your education?
*You may choose more than one language per school level. Language subjects do **not** count.*

	Malay	Mandarin Chinese	Tamil	English
Primary school	☐	☐	☐	☐
Secondary school	☐	☐	☐	☐
Pre-university	☐	☐	☐	☐
University	☐	☐	☐	☐

Section B: Language proficiency

B1   How well do you **speak** the following language(s)?

	Not well at all	Slightly well	Moderately well	Very well	Extremely well
	1	2	3	4	5
English					
Malay					
Mandarin					
Tamil					
Other					

B2   How well do you **listen to and understand** the following language(s)?

	Not well at all	Slightly well	Moderately well	Very well	Extremely well
	1	2	3	4	5
English					
Malay					
Mandarin					
Tamil					
Other					

B3   How well do you **write** in the following language(s)?

	Not well at all	Slightly well	Moderately well	Very well	Extremely well

276

	1	2	3	4	5
English			●		
Malay			●		
Mandarin			●		
Tamil			●		
Other			●		

B4   How well do you **read and understand** in the following language(s)?

	Not well at all	Slightly well	Moderately well	Very well	Extremely well
	1	2	3	4	5
English			●		
Malay			●		
Mandarin			●		
Tamil			●		
Other			●		

B5MY   What is the most recent English proficiency test that you have **already** taken?
- o   SPM English
- o   IELTS
- o   MUET
- o   Other: _____

B6MY   State your grade for the English proficiency test.

Grade

SPM English	
IELTS	
MUET	
Other	

Section C: Language attitude

Section C1: Language perception of English in Malaysia

This section invites you to express **your own perception** of English in Malaysia. There are no right or wrong answers, so please do not let your judgement be influenced by others.

*You are strongly advised **not** to go back to the previous questions to change your answers.*

C1.i   Do you agree that there is a certain type of English that is spoken and understood in Malaysia (and, to a certain extent, Singapore) but not in other English-speaking countries (e.g. United Kingdom, United States, Australia)?
- Yes
- No

C1.ii   In your own terms, define the English variety that is only spoken and understood in Malaysia (and, to a certain extent, Singapore). *You may describe as many language properties of the spoken variety as possible.*

[Insert text]

C1.iii   List as many given names as possible of this specific type of English that is spoken and understood in Malaysia.

[Insert text]

C1.iv.a   Have you heard of the term "Manglish" before?
- Yes, just as I said earlier!
- Oh yeah, now that you've mentioned it.
- Not sure, but it sounds familiar.
- Not at all, this is my first time hearing it.

C1.iv.b   Have you heard of the term "Bahasa Rojak" / the "Rojak" language before?
- o   Yes, just as I said earlier!
- o   Oh yeah, now that you've mentioned it.
- o   Not sure, but it sounds familiar.
- o   Not at all, this is my first time hearing it.

C1.iv.c   Do   you   think   Manglish   and   Rojak   are   the   same   thing? *In full sentences, explain how they are the same/different.*
- o   Yes: _____
- o   Yes and no: _____
- o   No: _____

C1.v.d   Which of the following would you prefer calling the local type of English spoken in Malaysia?
- o   Manglish
- o   Rojak
- o   English
- o   None of the above. I've got a better name: _____

Transition

This PhD study recognises a particular type of English spoken amongst Malaysians (and, to a certain extent, Singaporeans) which might not be fully understood by other English-speaking people from different parts of the world. It is NOT this study's intention to impose a label on this local English variety, which is why we wanted to know from *you* what name this variety goes by. Until the data have all been collected, this study shall refer to it as "Manglish".

Section C2: Language attitude

This section invites you to indicate **your own attitude** towards Standard (or formal) English and what you consider as "Manglish" in the Malaysian context. There are no right or wrong answers, so please do not let your judgement be influenced by others.

*You are strongly advised **not** to go back to the previous questions to change your answers.*

To what extent do you agree with the following statement?

Completely disagree	Strongly disagree	Somewhat disagree	Neutral	Somewhat agree	Strongly agree	Completely agree
○	○	○	○	○	○	○

C2.i   Standard English is the only correct English.
C2.ii   People who speak Manglish are thought to be less well-educated.
C2.iii People who speak Standard English seem less approachable.
C2.iv   Speaking Manglish reveals one's identity as a Malaysian.
C2.v   In order to improve on Standard English, one must avoid using Manglish.
C2.vi   Manglish is bad English.
C2.vii People who speak Standard English are thought to be well-educated.
C2.viii People who speak Manglish seem friendly.
C2.ix   Speaking Standard English covers up one's national identity.
C2.x   Using Manglish can be beneficial in helping Malaysians learn Standard English better.

C11   If you have to speak English with the following people in Malaysia, which English variety would you use?

	Manglish	Standard English
With family	○	○
With friends	○	○
With your English Language teacher	○	○
With your teacher/lecturer who teaches subjects other than English	○	○
With colleagues at work	○	○
With Malaysian strangers	○	○
With foreigners	○	○

Section D: Demographic information

D1

Which state or federal territory of Malaysia are you from?	Which part of the UK are you from?
○ Johor ○ Kedah ○ Kelantan ○ Kuala Lumpur ○ Labuan ○ Malacca ○ Negeri Sembilan	○ England ○ Northern Ireland ○ Scotland ○ Wales

o Pahang
o Penang
o Perak
o Perlis
o Putrajaya
o Sabah
o Sarawak
o Selangor
o Terengganu

D2 What is the highest education level you have **already** achieved?

Malaysia	UK
o Doctoral degree (e.g. PhD, DPhil) o Master's degree/Postgraduate certificate/PGCE o Bachelor's degree (e.g. BA, BSc) o STPM/A-level/IB/Diploma/Foundation degree/DipHE o Up to SPM/GCSE/O-level o Other: _____	o Doctoral degree (e.g. PhD, DPhil) o Master's degree/Postgraduate certificate/PGCE o Bachelor's degree (e.g. BA, BSc) o A-level/Sixth form/BTEC's/GNVQ's/IB/Diploma/Foundation degree/DipHE o Up to GCSE/O-level o Other: _____

D3 Which type of school did you attend during your primary education?

Malaysia	UK
o Sekolah Kebangsaan o Sekolah Jenis Kebangsaan Cina o Sekolah Jenis Kebagsaan Tamil o Sekolah Rendah Agama o International/private/independent school o Home school o Other: _____	o State school o Faith school o Boarding school o Free school o Academy o International/private/independent school o Home school o Other: _____

D4 Which type of school did you attend during your secondary education?

Malaysia	UK
o Sekolah Menengah Kebangsaan o Sekolah Menengah (Persendirian)	o State school o Faith school o Boarding school o Grammar school

o Sekolah Menengah Jenis Kebangsaan (English)	o City technology college
o Sekolah Menengah Jenis Kebangsaan (Cina)	o International/private/independent school
o Maktab Rendah Sains MARA (MRSM)	o Other: _____
o Sekolah Berasrama Penuh	
o Sekolah Menengah Agama	
o Sekolah Menengah Teknik/Vokasional	
o International/private/independent school	
o Home school	
o Other: _____	

D5   What is your current occupation?
- o Student
- o Employee
- o Employer/Self-employed
- o Retired
- o Unemployed

D6.1   Please specify your current degree level, course, and university.
- Degree level: _____
- Course name: _____
- University: _____

D6.2   Please specify your current occupation: _____

D6.3   Please specify your previous occupation: _____

D7   What is your monthly household income (before tax)? *If you are a student who does not receive any income, state the total income of your parent(s) or caregiver(s).*

Malaysia	UK
o Less than RM3,000	o Less than £1,000
o RM3,000 – RM5,999	o £1,000 - £2,999
o RM6,000 – RM8,999	o £3,000 - £4,999
o RM9,000 – RM11,999	o £5,000 - £6,999
o RM12,000 and above	o £7,000 and above

D8 State your ethnicity.

Malaysia	UK
o Malay	o White
o Chinese	o Black
o Indian	o Asian
o Other Bumiputera: _____	o Mixed: _____
o Other non-Bumiputera: _____	o Other: _____

D9 Which year were you born? _____

D10 What is your sex?
- o Male
- o Female
- o Prefer not to say

End of survey

You have reached the end of this questionnaire. Thank you very much for completing it! Should you have further inquiries or points for discussion, please do not hesitate to get in touch with me at [email redacted].

With all best wishes,
Samantha Sie

B. Ethics

B.1. Participant information sheet

B.1.1. CME case study

UNIVERSITY OF CAMBRIDGE
Department of Theoretical
and Applied Linguistics

Faculty of Modern and Medieval Languages
Sidgwick Avenue
Cambridge, CB3 9DA

www.mml.cam.ac.uk/dtal

Cambridge, 13-02-2018

PARTICIPATION INFORMATION LETTER
DAILY CONVERSATIONS IN MANGLISH:
A DESCRIPTION FROM A SOCIOLINGUISTIC PERSPECTIVE

Dear participant,

I am a PhD student investigating the ways in which Manglish is used in daily conversations. Please read the following information on how the study will be conducted and how you will participate.

Who is participating?
This study aims to recruit participants currently in Tawau, Malaysia. They must be Manglish speakers of Malaysian nationality, whose age is between 18 and 35 years.

Where will it take place?
The testing will take place in Tawau, Malaysia.

What are the tasks involved in this study?
You are invited to take part in a casual conversation with another Malaysian speaker of Manglish. You are asked to interact using Manglish. The conversation will be audio-recorded and is expected to last for 1 hour. You will be provided some topics and questions if you need some ideas for the conversation.

What will happen with the data?
Data collection is entirely anonymous and confidential. The consent form in which you are to fill in is only accessible to the researchers involved and will be used and stored in compliance with the Data Protection Act.

Do you have to take part?
It is entirely up to you whether or not to take part in this study. You may withdraw at any time without having to provide any reason.

Why is your participation important?
Your participation is highly valuable as the use of Manglish, which is not an official codified language, is best captured through daily interactions.

Should there be any questions or concerns, please contact me for further clarification.

Thank you for your support.

Yours sincerely,
Samantha Sie

Email: [redacted]
Tel: [redacted]

B.1.2. Main study

UNIVERSITY OF CAMBRIDGE
Faculty of Modern and Medieval Languages and Linguistics
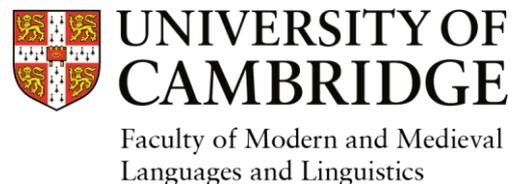
PARTICIPANT INFORMATION SHEET

Title

Investigating the use of English in Malaysia and the United Kingdom (UK)

Invitation paragraph

Before you decide to take part in this study, it is important for you to understand why the research is being done and what it will involve. Please take your time to read the following information carefully and discuss it with others if you wish. A member of the team can be contacted if there is anything that is not clear or if you would like more information. Take your time to decide whether or not you wish to take part.

Purpose of the study

This linguistic study investigates the use of English in a formal setting. Its interest lies particularly in the use of Standard English in a multilingual environment as compared to a monolingual one. Therefore, Malaysia and the UK have been chosen for the study.

Why have I been chosen?

You have been invited as you fulfil all the participation criteria required for this study:

- Born and raised either in Malaysia or the UK
- Between 18 and 30 years old
- Learned English at any point in life
- Studying at a university

Do I have to take part?

Participation is completely voluntary. You may withdraw from the study at any point without having to provide any explanation.

What am I asked to do if I take part?

You will be asked to complete **all** the following tasks using a computer/laptop:

- watch a short, animated film and narrate the story in English
- rate English sentences in terms of how grammatical they seem to you
- fill in a language background questionnaire
- complete an online English placement test

The whole session takes around 1 hour to complete, including short breaks. It will be done one-on-one with the researcher *online*. Links to the tasks will be provided during the session. The narrative task will be audio-recorded by the researcher for transcription purposes, but please rest assured that these recordings will be anonymised and identified only by a code; they will only be used for research purposes.

Furthermore, you will be asked to submit the following files (anonymised) to the researcher via an online form:

- your sound file (*.mp3*), as a backup in case of poor internet connectivity
- a screenshot (*.jpg*/*.jpeg*) of your English placement test score

Before COVID, the narrative task and the English test score were recorded by the researcher herself. However, as the study is now being conducted remotely, your cooperation and understanding in assisting the researcher record the data from your end is very much needed and appreciated.

Are there possible disadvantages and/or risks in taking part?

The nature of this study does not cause any discomfort or risks that would affect your health and well-being. However, if you experience fatigue during the session, please let the investigator know to re-schedule the session.

Will my taking part in this project be kept confidential?

In compliance with the Data Protection Act, all data will be anonymised and identified only by a code, with personal details kept in password-protected files in the investigator's personal computer. Access to the data is restricted to the immediate research team.

How long will my data be retained?

In line with University policy, data will generally be kept until the completion of the project + 10 years. After this, they may be destroyed or sent (in anonymised form) to the University's digital Institutional Repository.

What will happen to the results of the research project?

The anonymised data may be used in analyses, publications, and conference presentations by the investigator and her research collaborators from the University of Cambridge and other research institutions. Results are normally presented in terms of groups of individuals. If any individual data are presented, they will be given a unique identification code and made completely anonymous, without any means of identifying the individuals involved.

Anonymised and coded audio-recordings may be made accessible to academics and educators through open-access language repositories – e.g. TalkBank (MacWhinney, 2000) – which are used for studying conversational interaction. Please decide if you would like to contribute to the wider academic community by indicating in the consent form whether or not to grant the researcher permission to upload your audio-recorded data on an online public repository, on the condition that the data will be anonymised and non-identifiable.

Ethical review of the study

This project has received ethical approval from the Research Ethics Committee of the Faculty of Modern and Medieval Languages and Linguistics at the University of Cambridge.

Contact for further information

Should there be any queries or concerns, please do not hesitate to get in touch with Samantha Sie (PhD student) at [email redacted] or Professor Ianthi Tsimpli (Supervisor) at [email redacted].

Reference: MacWhinney, B. (2000). *The CHILDES Project: Tools for Analysing Talk*. 3rd Edition. Mahwah: NJ: Lawrence Erlbaum Associates.

B.2. Consent form

B.2.1. CME case study

UNIVERSITY OF CAMBRIDGE

Theoretical and Applied Linguistics Section

Raised Faculty Building
Sidgwick Avenue
Cambridge CB3 9DA

CONSENT FORM

Project title:
Daily conversations in Manglish: A description from a sociolinguistic perspective

Research team:
Samantha Sie (PhD student); Professor Ianthi Tsimpli (supervisor)

If you have any questions, please contact Samantha Sie, [email redacted].

☐ I confirm that I have read and understand the information sheet dated 13 February 2018 for the above mentioned study and have had the opportunity to ask questions.

☐ I understand that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, and without my rights being affected.

☐ I understand that any data that are collected will be used and stored anonymously, in accordance with the Data Protection Act. Results are normally presented in terms of groups of individuals. If any individual data were presented, the data would be completely anonymous, without any means of identifying the individuals involved.

☐ I understand that these data may be used in analyses, publications, conference presentations or by researchers at the University of Cambridge. I give permission for these individuals to have access to these data.

☐ I have been given a copy of this form to keep.

☐ I agree to take part in the study.

I _____ agree to participate in the abovementioned study run by Samantha Sie, a PhD student in the Theoretical and Applied Linguistics Section at the University of Cambridge.

_____   _____
Date   Signature of participant

_____   _____   _____
Name of researcher   Date   Signature   of
researcher

If you have any questions or complaints about the ethical aspects of this study, please contact [email redacted].

B.2.2. Main study

UNIVERSITY OF CAMBRIDGE
Faculty of Modern and Medieval
Languages and Linguistics
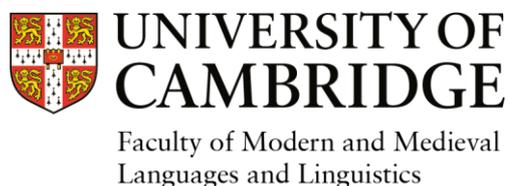
CONSENT FORM

Project title: **Investigating the use of English in Malaysia and the United Kingdom (UK)**

Research team: Samantha Sie (PhD student), Professor Ianthi Tsimpli (Supervisor)

If you have any questions, please contact Samantha Sie at [email redacted].

☐ I have read and understood the participant information sheet for this study and have had the opportunity to ask questions.

☐ I understand that my participation is voluntary. I have the right to withdraw from the study at any time without having to give any reason.

☐ I understand that all data collected in this study will be anonymised, made non-identifiable, and stored securely in compliance with the Data Protection Act.

☐ I understand that the anonymised data may be used in analyses, publications, and conference presentations by the researcher and her research collaborators from the University of Cambridge and other research institutions. I grant these individuals permission to use these data.

☐ I understand that anonymised and coded audio-recordings may be made accessible to academics and educators through open-access language repositories used for studying conversational interaction such as TalkBank (MacWhinney, 2000). I would like to contribute to the wider academic community, and thus grant the researcher permission to upload my audio-recorded data on an online public repository, on the condition that the data will be anonymised and non-identifiable.

☐ I have been given a copy of this form to keep.

I _____ agree to participate in the above-mentioned study run by Samantha Sie, a PhD student of the Faculty of MMLL at the University of Cambridge.

_____   _____
Date   Signature of participant

_____   _____   _____
Name of researcher   Date   Signature of researcher

The project has received ethical approval from the Faculty's Research Ethics Committee
If you have any questions or complaints about the ethical aspects of this study, please contact [email redacted].

C. Statistical data (Grammaticality judgement task)

Table C.1: Summary of the mixed model output with L1 Group as one of the predictors

	Est.	SE	df	*t* value	Pr(>|t|)	
(Intercept)	1.000	0.050	229.107	20.030	<0.0001	***
L1_Group(Malay)	-0.270	0.048	13952.035	-5.600	0.000	***
L1_Group(Chinese)	-0.343	0.048	13952.006	-7.200	0.000	***
L1_Group(MalE)	-0.161	0.055	13952.072	-2.900	0.004	**
L1_Group(MalE-Malay)	-0.161	0.054	13952.002	-3.000	0.003	**
L1_Group(MalE-Chinese)	-0.340	0.063	13952.003	-5.440	0.000	***
Proficiency	0.000	0.001	13952.032	0.090	0.932	
Feature(progressive)	-0.001	0.085	215.514	-0.010	0.993	
Feature(do)	0.052	0.085	214.695	0.610	0.542	
Feature(present)	0.039	0.085	213.887	0.450	0.651	

Feature(past)	0.127	0.085	213.887	1.490	0.137	
Grammaticality_Type(omission)	-1.753	0.069	214.493	-25.240	<0.0001	***
Grammaticality_Type(commission)	-1.705	0.098	214.493	-17.370	<0.0001	***
L1_Group(Malay):Feature(progressive)	0.117	0.082	13952.087	1.430	0.153	
L1_Group(Chinese):Feature(progressive)	0.145	0.078	13952.086	1.870	0.061	.
L1_Group(MalE):Feature(progressive)	0.027	0.095	13952.076	0.280	0.779	
L1_Group(MalE-Malay):Feature(progressive)	0.113	0.092	13952.062	1.230	0.219	
L1_Group(MalE-Chinese):Feature(progressive)	0.260	0.107	13952.046	2.430	0.015	*
L1_Group(Malay):Feature(do)	-0.221	0.082	13952.058	-2.700	0.007	**
L1_Group(Chinese):Feature(do)	0.020	0.078	13952.054	0.260	0.795	
L1_Group(MalE):Feature(do)	0.036	0.095	13952.059	0.380	0.706	

L1_Group(MalE-Malay):Feature(do)	-0.039	0.092	13952.039	-0.420	0.674	
L1_Group(MalE-Chinese):Feature(do)	0.383	0.107	13952.029	3.580	0.000	***
L1_Group(Malay):Feature(present)	0.186	0.082	13952.012	2.280	0.023	*
L1_Group(Chinese):Feature(present)	0.245	0.077	13952.002	3.160	0.002	**
L1_Group(MalE):Feature(present)	0.150	0.095	13952.024	1.580	0.113	
L1_Group(MalE-Malay):Feature(present)	0.167	0.092	13952.002	1.810	0.070	.
L1_Group(MalE-Chinese):Feature(present)	0.392	0.107	13952.002	3.660	0.000	***
L1_Group(Malay):Feature(past)	0.242	0.082	13952.012	2.960	0.003	**
L1_Group(Chinese):Feature(past)	0.160	0.078	13952.027	2.070	0.039	*
L1_Group(MalE):Feature(past)	0.165	0.095	13952.105	1.740	0.082	.

L1_Group(MalE-Malay):Feature(past)	0.047	0.092	13952.002	0.510	0.611	
L1_Group(MalE-Chinese):Feature(past)	0.390	0.107	13952.002	3.640	0.000	***
L1_Group(Malay):Grammaticality_Type(omission)	0.263	0.067	13952.073	3.940	0.000	***
L1_Group(Chinese):Grammaticality_Type(omission)	0.480	0.063	13952.056	7.580	0.000	***
L1_Group(MalE):Grammaticality_Type(omission)	0.213	0.077	13952.103	2.750	0.006	**
L1_Group(MalE-Malay):Grammaticality_Type(omission)	0.147	0.075	13952.030	1.960	0.050	.
L1_Group(MalE-Chinese):Grammaticality_Type(omission)	0.312	0.087	13952.022	3.570	0.000	***
L1_Group(Malay):Grammaticality_Type(commission)	0.127	0.094	13952.037	1.350	0.177	
L1_Group(Chinese):Grammaticality_Type(commission)	0.214	0.090	13952.059	2.380	0.017	*
L1_Group(MalE):Grammaticality_Type(commission)	-0.145	0.110	13952.166	-1.320	0.186	

L1_Group(MalE-Malay):Grammaticality_Type(commission)	0.116	0.106	13952.030	1.090	0.274	
L1_Group(MalE-Chinese):Grammaticality_Type(commission)	-0.024	0.124	13952.144	-0.200	0.844	
Feature(progressive):Grammaticality_Type(omission)	-0.028	0.120	214.902	-0.240	0.813	
Feature(do):Grammaticality_Type(omission)	-0.256	0.120	214.898	-2.130	0.034	*
Feature(present):Grammaticality_Type(omission)	0.116	0.120	214.089	0.970	0.335	
Feature(past):Grammaticality_Type(omission)	-0.059	0.120	214.089	-0.490	0.622	
Feature(progressive):Grammaticality_Type(commission)	0.233	0.139	214.800	1.680	0.095	.
Feature(do):Grammaticality_Type (commission)	0.001	0.139	214.493	0.010	0.996	
L1_Group(Malay):Feature(progressive):Grammaticality_Type(omission)	-0.069	0.116	13952.063	-0.600	0.549	
L1_Group(Chinese):Feature(progressive):Grammaticality_Type(omission)	-0.170	0.110	13952.062	-1.550	0.122	
L1_Group(MalE):Feature (progressive): Grammaticality_Type(omission)	-0.041	0.134	13952.062	-0.300	0.763	
L1_Group(MalE-Malay):Feature(progressive):Grammaticality_Type(omission)	-0.131	0.130	13952.041	-1.000	0.316	
L1_Group(MalE-Chinese):Feature(progressive):Grammaticality_Type(omission)	-0.146	0.152	13952.031	-0.960	0.336	

L1_Group(Malay):Feature(do):Grammaticality_Type(omission)	0.289	0.116	13952.118	2.500	0.012	*
L1_Group(Chinese):Feature(do):Grammaticality_Type(omission)	0.074	0.110	13952.072	0.670	0.502	
L1_Group(MalE):Feature(do):Grammaticality_Type(omission)	0.152	0.134	13952.070	1.140	0.256	
L1_Group(MalE-Malay):Feature(do):Grammaticality_Type(omission)	0.063	0.130	13952.048	0.480	0.629	
L1_Group(MalE-Chinese):Feature(do):Grammaticality_Type(omission)	-0.159	0.152	13952.036	-1.050	0.296	
L1_Group(Malay):Feature(present):Grammaticality_Type(omission)	0.044	0.116	13952.044	0.380	0.701	
L1_Group(Chinese):Feature(present):Grammaticality_Type(omission)	-0.344	0.110	13952.020	-3.140	0.002	**
L1_Group(MalE):Feature(present):Grammaticality_Type(omission)	-0.131	0.134	13952.036	-0.980	0.328	
L1_Group(MalE-Malay):Feature(present):Grammaticality_Type(omission)	0.033	0.130	13952.011	0.260	0.797	
L1_Group(MalE-Chinese):Feature(present):Grammaticality_Type(omission)	-0.316	0.151	13952.009	-2.090	0.037	*

L1_Group(Malay):Feature(past):Grammaticality_Type(omission)	-0.166	0.116	13952.026	-1.440	0.150	
L1_Group(Chinese):Feature(past):Grammaticality_Type(omission)	-0.335	0.110	13952.045	-3.050	0.002	**
L1_Group(MalE):Feature(past):Grammaticality_Type(omission)	-0.306	0.134	13952.076	-2.280	0.022	*
L1_Group(MalE-Malay):Feature(past):Grammaticality_Type(omission)	0.001	0.130	13952.011	0.010	0.994	
L1_Group(MalE-Chinese):Feature(past):Grammaticality_Type(omission)	-0.311	0.151	13952.009	-2.050	0.040	*
L1_Group(Malay):Feature(progressive):Grammaticality_Type(commission)	0.072	0.134	13952.048	0.540	0.587	
L1_Group(Chinese):Feature(progressive):Grammaticality_Type(commission)	0.370	0.127	13952.062	2.920	0.004	**
L1_Group(MalE):Feature(progressive):Grammaticality_Type(commission)	0.326	0.155	13952.104	2.100	0.036	*
L1_Group(MalE-Malay):Feature(progressive):Grammaticality_Type(commission)	-0.131	0.150	13952.038	-0.870	0.382	
L1_Group(MalE-Chinese):Feature(progressive):Grammaticality_Type(commission)	0.215	0.175	13952.090	1.230	0.219	

L1_Group(Malay):Feature(do):Grammaticality_Type(commission)	1.130	0.133	13952.037	8.460	<0.0001	***
L1_Group(Chinese):Feature(do):Grammaticality_Type(commission)	0.592	0.127	13952.059	4.670	0.000	***
L1_Group(MalE):Feature(do):Grammaticality_Type(commission)	0.610	0.155	13952.097	3.940	0.000	***
L1_Group(MalE-Malay):Feature(do):Grammaticality_Type(commission)	0.658	0.150	13952.030	4.380	0.000	***
L1_Group(MalE-Chinese):Feature(do):Grammaticality_Type(commission)	0.148	0.175	13952.083	0.850	0.398	

Table C2: Results of the post hoc comparisons between L1 groups by feature and grammaticality type

Copula: Standard

Contrast	Est.	SE	z	p
BritE-Malay	0.27	0.0481	5.6	<.0001
BritE-Chinese	0.343	0.0477	7.2	<.0001
BritE-MalE	0.161	0.0554	2.9	0.043
BritE-MalE,Malay	0.161	0.0535	3	0.032
BritE-MalE,Chinese	0.34	0.0626	5.44	<.0001
Malay-Chinese	0.074	0.0437	1.69	0.539
Malay-MalE	-0.109	0.0535	-2.03	0.324
Malay-MalE,Malay	-0.109	0.0518	-2.1	0.287
Malay-MalE,Chinese	0.071	0.0607	1.17	0.852
Chinese-MalE	-0.183	0.0519	-3.51	0.006
Chinese-MalE,Malay	-0.183	0.0505	-3.62	0.004
Chinese-MalE,Chinese	-0.003	0.0592	-0.05	1
MalE-MalE,Malay	0	0.0588	0	1
MalE-MalE,Chinese	0.18	0.0668	2.69	0.077
MalE,Malay-MalE,Chinese	0.18	0.0655	2.74	0.067

Copula: Omission

Contrast	Est.	SE	z	p
BritE-Malay	0.006	0.0482	0.13	1
BritE-Chinese	-0.137	0.0478	-2.86	0.049
BritE-MalE	-0.052	0.0555	-0.94	0.935
BritE-MalE,Malay	0.014	0.0536	0.25	1
BritE-MalE,Chinese	0.028	0.0627	0.45	0.998
Malay-Chinese	-0.143	0.0437	-3.27	0.014
Malay-MalE	-0.059	0.0535	-1.1	0.882
Malay-MalE,Malay	0.007	0.0518	0.14	1
Malay-MalE,Chinese	0.022	0.0607	0.36	0.999
Chinese-MalE	0.084	0.052	1.62	0.586

Contrast	Est.	SE	z	p
Chinese-MalE,Malay	0.15	0.0505	2.97	0.035
Chinese-MalE,Chinese	0.165	0.0592	2.78	0.06
MalE-MalE,Malay	0.066	0.0588	1.12	0.872
MalE-MalE,Chinese	0.081	0.0668	1.21	0.833
MalE,Malay-MalE,Chinese	0.015	0.0655	0.22	1

Progressive: Standard

Contrast	Est.	SE	z	p
BritE-Malay	0.152	0.0676	2.26	0.212
BritE-Chinese	0.198	0.0656	3.02	0.03
BritE-MalE	0.134	0.0782	1.72	0.522
BritE-MalE,Malay	0.047	0.0756	0.63	0.989
BritE-MalE,Chinese	0.08	0.0882	0.91	0.945
Malay-Chinese	0.046	0.0614	0.74	0.976
Malay-MalE	-0.018	0.0757	-0.24	1
Malay-MalE,Malay	-0.105	0.0732	-1.44	0.705
Malay-MalE,Chinese	-0.073	0.0858	-0.85	0.959
Chinese-MalE	-0.064	0.0731	-0.88	0.952
Chinese-MalE,Malay	-0.151	0.0707	-2.13	0.271
Chinese-MalE,Chinese	-0.118	0.0834	-1.42	0.717
MalE-MalE,Malay	-0.087	0.0832	-1.04	0.904
MalE-MalE,Chinese	-0.054	0.0945	-0.57	0.993
MalE,Malay-MalE,Chinese	0.033	0.0925	0.35	0.999

Progressive: Omission

Contrast	Est.	SE	z	p
BritE-Malay	-0.041	0.0674	-0.61	0.99
BritE-Chinese	-0.112	0.0654	-1.71	0.524
BritE-MalE	-0.039	0.0778	-0.5	0.996
BritE-MalE,Malay	0.031	0.0754	0.41	0.999
BritE-MalE,Chinese	-0.086	0.088	-0.98	0.925
Malay-Chinese	-0.071	0.0614	-1.15	0.86
Malay-MalE	0.003	0.0755	0.04	1
Malay-MalE,Malay	0.072	0.0732	0.99	0.922

Contrast	Est.	SE	z	p
Malay-MalE,Chinese	-0.045	0.0858	-0.52	0.995
Chinese-MalE	0.073	0.0729	1.01	0.916
Chinese-MalE,Malay	0.143	0.0707	2.02	0.33
Chinese-MalE,Chinese	0.026	0.0834	0.31	1
MalE-MalE,Malay	0.07	0.083	0.84	0.96
MalE-MalE,Chinese	-0.048	0.0943	-0.5	0.996
MalE,Malay-MalE,Chinese	-0.117	0.0925	-1.27	0.804

Progressive: Commission

Contrast	Est.	SE	z	p
BritE-Malay	-0.047	0.0674	-0.7	0.982
BritE-Chinese	-0.385	0.0654	-5.89	<.0001
BritE-MalE	-0.047	0.0778	-0.6	0.991
BritE-MalE,Malay	0.063	0.0754	0.83	0.962
BritE-MalE,Chinese	-0.111	0.088	-1.26	0.807
Malay-Chinese	-0.338	0.0614	-5.5	<.0001
Malay-MalE	0	0.0755	0	1
Malay-MalE,Malay	0.11	0.0732	1.5	0.664
Malay-MalE,Chinese	-0.064	0.0858	-0.74	0.977
Chinese-MalE	0.338	0.0729	4.64	<.0001
Chinese-MalE,Malay	0.448	0.0707	6.33	<.0001
Chinese-MalE,Chinese	0.274	0.0834	3.29	0.013
MalE-MalE,Malay	0.109	0.083	1.32	0.775
MalE-MalE,Chinese	-0.064	0.0943	-0.68	0.984
MalE,Malay-MalE,Chinese	-0.173	0.0925	-1.87	0.418

DO: Standard

Contrast	Est.	SE	z	p
BritE-Malay	0.49	0.0675	7.27	<.0001
BritE-Chinese	0.323	0.0655	4.93	<.0001
BritE-MalE	0.125	0.0779	1.61	0.595
BritE-MalE,Malay	0.199	0.0755	2.64	0.087
BritE-MalE,Chinese	-0.043	0.0881	-0.49	0.997
Malay-Chinese	-0.167	0.0614	-2.72	0.071

Contrast	Est.	SE	z	p
Malay-MalE	-0.365	0.0755	-4.84	<.0001
Malay-MalE,Malay	-0.291	0.0732	-3.98	0.001
Malay-MalE,Chinese	-0.533	0.0858	-6.21	<.0001
Chinese-MalE	-0.198	0.0729	-2.72	0.072
Chinese-MalE,Malay	-0.124	0.0707	-1.75	0.498
Chinese-MalE,Chinese	-0.366	0.0834	-4.39	<.0001
MalE-MalE,Malay	0.074	0.083	0.9	0.948
MalE-MalE,Chinese	-0.168	0.0943	-1.78	0.479
MalE,Malay-MalE,Chinese	-0.242	0.0925	-2.62	0.093

DO: Omission

Contrast	Est.	SE	z	p
BritE-Malay	-0.062	0.0676	-0.92	0.941
BritE-Chinese	-0.23	0.0655	-3.52	0.006
BritE-MalE	-0.241	0.0779	-3.09	0.025
BritE-MalE,Malay	-0.011	0.0755	-0.14	1
BritE-MalE,Chinese	-0.196	0.0881	-2.23	0.223
Malay-Chinese	-0.168	0.0615	-2.73	0.069
Malay-MalE	-0.178	0.0756	-2.36	0.171
Malay-MalE,Malay	0.052	0.0733	0.7	0.981
Malay-MalE,Chinese	-0.134	0.0859	-1.56	0.624
Chinese-MalE	-0.01	0.0729	-0.14	1
Chinese-MalE,Malay	0.22	0.0707	3.11	0.023
Chinese-MalE,Chinese	0.034	0.0834	0.41	0.999
MalE-MalE,Malay	0.23	0.083	2.77	0.062
MalE-MalE,Chinese	0.044	0.0943	0.47	0.997
MalE,Malay-MalE,Chinese	-0.186	0.0925	-2.01	0.337

DO: Commission

Contrast	Est.	SE	z	p
BritE-Malay	-0.767	0.0674	-11.38	<.0001
BritE-Chinese	-0.482	0.0655	-7.37	<.0001
BritE-MalE	-0.34	0.0778	-4.37	<.0001
BritE-MalE,Malay	-0.575	0.0754	-7.62	<.0001

Contrast	Est.	SE	z	p
BritE-MalE,Chinese	-0.167	0.088	-1.89	0.407
Malay-Chinese	0.284	0.0614	4.63	<.0001
Malay-MalE	0.426	0.0755	5.65	<.0001
Malay-MalE,Malay	0.192	0.0732	2.62	0.092
Malay-MalE,Chinese	0.6	0.0858	6.99	<.0001
Chinese-MalE	0.142	0.0729	1.95	0.372
Chinese-MalE,Malay	-0.092	0.0708	-1.3	0.783
Chinese-MalE,Chinese	0.316	0.0835	3.78	0.002
MalE-MalE,Malay	-0.235	0.083	-2.83	0.053
MalE-MalE,Chinese	0.174	0.0943	1.84	0.44
MalE,Malay-MalE,Chinese	0.408	0.0925	4.41	<.0001

Present: Standard

Contrast	Est.	SE	z	p
BritE-Malay	0.084	0.0674	1.24	0.817
BritE-Chinese	0.099	0.0654	1.51	0.658
BritE-MalE	0.011	0.0778	0.14	1
BritE-MalE,Malay	-0.006	0.0754	-0.08	1
BritE-MalE,Chinese	-0.051	0.088	-0.58	0.992
Malay-Chinese	0.015	0.0614	0.25	1
Malay-MalE	-0.073	0.0755	-0.97	0.929
Malay-MalE,Malay	-0.09	0.0732	-1.22	0.825
Malay-MalE,Chinese	-0.135	0.0858	-1.57	0.618
Chinese-MalE	-0.088	0.0729	-1.21	0.833
Chinese-MalE,Malay	-0.105	0.0707	-1.48	0.675
Chinese-MalE,Chinese	-0.15	0.0834	-1.8	0.467
MalE-MalE,Malay	-0.017	0.083	-0.2	1
MalE-MalE,Chinese	-0.062	0.0943	-0.66	0.987
MalE,Malay-MalE,Chinese	-0.045	0.0925	-0.49	0.997

Present: Omission

Contrast	Est.	SE	z	p
BritE-Malay	-0.224	0.0674	-3.32	0.011
BritE-Chinese	-0.037	0.0654	-0.57	0.993

Contrast	Est.	SE	z	p
BritE-MalE	-0.071	0.0778	-0.92	0.942
BritE-MalE,Malay	-0.187	0.0754	-2.48	0.131
BritE-MalE,Chinese	-0.047	0.088	-0.54	0.995
Malay-Chinese	0.187	0.0615	3.04	0.028
Malay-MalE	0.153	0.0755	2.02	0.33
Malay-MalE,Malay	0.038	0.0733	0.51	0.996
Malay-MalE,Chinese	0.177	0.0859	2.06	0.308
Chinese-MalE	-0.034	0.0729	-0.47	0.997
Chinese-MalE,Malay	-0.149	0.0707	-2.11	0.28
Chinese-MalE,Chinese	-0.01	0.0834	-0.12	1
MalE-MalE,Malay	-0.115	0.083	-1.39	0.734
MalE-MalE,Chinese	0.024	0.0943	0.26	1
MalE,Malay-MalE,Chinese	0.139	0.0925	1.51	0.66

Past: Standard

Contrast	Est.	SE	*z*	*p*
BritE-Malay	0.028	0.0674	0.41	0.999
BritE-Chinese	0.183	0.0655	2.8	0.057
BritE-MalE	-0.005	0.078	-0.06	1
BritE-MalE,Malay	0.114	0.0754	1.51	0.656
BritE-MalE,Chinese	-0.05	0.088	-0.57	0.993
Malay-Chinese	0.156	0.0614	2.53	0.114
Malay-MalE	-0.032	0.0757	-0.43	0.998
Malay-MalE,Malay	0.086	0.0732	1.18	0.847
Malay-MalE,Chinese	-0.078	0.0858	-0.9	0.946
Chinese-MalE	-0.188	0.0732	-2.57	0.105
Chinese-MalE,Malay	-0.069	0.0708	-0.98	0.925
Chinese-MalE,Chinese	-0.233	0.0835	-2.79	0.058
MalE-MalE,Malay	0.119	0.0832	1.43	0.712
MalE-MalE,Chinese	-0.045	0.0945	-0.48	0.997
MalE,Malay-MalE,Chinese	-0.164	0.0925	-1.77	0.484

Past: Omission

Contrast	Est.	SE	*z*	*p*

Contrast	Est.	SE	z	p
BritE-Malay	-0.069	0.0674	-1.03	0.909
BritE-Chinese	0.038	0.0655	0.58	0.992
BritE-MalE	0.088	0.0778	1.14	0.866
BritE-MalE,Malay	-0.034	0.0754	-0.45	0.998
BritE-MalE,Chinese	-0.051	0.088	-0.58	0.993
Malay-Chinese	0.107	0.0614	1.75	0.5
Malay-MalE	0.158	0.0755	2.09	0.293
Malay-MalE,Malay	0.035	0.0732	0.48	0.997
Malay-MalE,Chinese	0.019	0.0858	0.22	1
Chinese-MalE	0.05	0.0729	0.69	0.983
Chinese-MalE,Malay	-0.072	0.0708	-1.02	0.912
Chinese-MalE,Chinese	-0.089	0.0835	-1.06	0.896
MalE-MalE,Malay	-0.123	0.083	-1.48	0.679
MalE-MalE,Chinese	-0.139	0.0943	-1.48	0.68
MalE,Malay-MalE,Chinese	-0.017	0.0925	-0.18	1

Past: Commission

Contrast	Est.	SE	z	p
BritE-Malay	-0.1	0.0675	-1.48	0.678
BritE-Chinese	-0.03	0.0655	-0.46	0.997
BritE-MalE	0.141	0.0781	1.8	0.466
BritE-MalE,Malay	-0.002	0.0755	-0.03	1
BritE-MalE,Chinese	-0.026	0.0885	-0.29	1
Malay-Chinese	0.069	0.0614	1.13	0.868
Malay-MalE	0.24	0.0757	3.18	0.019
Malay-MalE,Malay	0.097	0.0732	1.33	0.768
Malay-MalE,Chinese	0.074	0.0862	0.86	0.956
Chinese-MalE	0.171	0.0731	2.34	0.179
Chinese-MalE,Malay	0.028	0.0707	0.4	0.999
Chinese-MalE,Chinese	0.005	0.0839	0.06	1
MalE-MalE,Malay	-0.143	0.0832	-1.72	0.52
MalE-MalE,Chinese	-0.166	0.0948	-1.75	0.498
MalE,Malay-MalE,Chinese	-0.023	0.0929	-0.25	1

Table C.3: Summary of the mixed model output with English Status as one of the predictors

	Est.	SE	df	t value	Pr(>|t|)	
(Intercept)	1.000	0.050	228.024	20.060	<0.0001	***
English_Status(L1-MalE)	-0.205	0.044	13991.020	-4.700	0.000	***
English_Status(L2-MalE)	-0.311	0.043	13991.016	-7.320	0.000	***
Feature(progressive)	-0.001	0.085	217.114	-0.010	0.993	
Feature(do)	0.052	0.085	216.285	0.610	0.543	
Feature(present)	0.039	0.085	215.465	0.450	0.651	
Feature(past)	0.127	0.085	215.465	1.490	0.138	
Grammaticality_Type(omission)	-1.753	0.070	216.080	-25.190	<0.0001	***
Grammaticality_Type(commission)	-1.705	0.098	216.080	-17.330	<0.0001	***
Proficiency	0.000	0.001	13991.041	0.070	0.947	
English_Status(L1-MalE):Feature(progressive)	0.119	0.075	13991.066	1.590	0.112	
English_Status(L2-MalE):Feature(progressive)	0.133	0.071	13991.108	1.880	0.060	.
English_Status(L1-MalE):Feature(do)	0.092	0.075	13991.064	1.230	0.219	
English_Status(L2-MalE):Feature(do)	-0.086	0.071	13991.069	-1.220	0.224	
English_Status(L1-MalE):Feature(present)	0.216	0.074	13991.007	2.900	0.004	**
English_Status(L2-MalE):Feature(present)	0.219	0.071	13991.005	3.100	0.002	**
English_Status(L1-MalE):Feature(past)	0.173	0.075	13991.024	2.330	0.020	*
English_Status(L2-MalE):Feature(past)	0.196	0.071	13991.014	2.780	0.005	**
English_Status(L1-MalE):Grammaticality_Type(omission)	0.211	0.061	13991.064	3.470	0.001	***

English_Status(L2-MalE):Grammaticality_Type(omission)	0.385	0.058	13991.069	6.670	0.000	***
English_Status(L1-MalE):Grammaticality_Type(commission)	-0.012	0.086	13991.129	-0.140	0.892	
English_Status(L2-MalE):Grammaticality_Type(commission)	0.175	0.082	13991.057	2.150	0.031	*
Feature(progressive):Grammaticality_Type(omission)	-0.028	0.121	216.494	-0.240	0.814	
Feature(do):Grammaticality_Type(omission)	-0.256	0.121	216.490	-2.130	0.035	*
Feature(present):Grammaticality_Type(omission)	0.116	0.120	215.670	0.960	0.336	
Feature(past):Grammaticality_Type(omission)	-0.059	0.120	215.670	-0.490	0.623	
Feature(progressive):Grammaticality_Type(commission)	0.233	0.139	216.390	1.670	0.095	.
Feature(do):Grammaticality_Type(commission)	0.001	0.139	216.080	0.010	0.996	
English_Status(L1-MalE):Feature(progressive):Grammaticality_Type(omission)	-0.102	0.106	13991.052	-0.970	0.333	
English_Status(L2-MalE):Feature(progressive):Grammaticality_Type(omission)	-0.126	0.100	13991.076	-1.260	0.208	
English_Status(L1-MalE):Feature(do):Grammaticality_Type(omission)	0.041	0.105	13991.080	0.390	0.700	
English_Status(L2-MalE):Feature(do):Grammaticality_Type(omission)	0.169	0.100	13991.106	1.690	0.091	.
English_Status(L1-MalE):Feature(present):Grammaticality_Type(omission)	-0.111	0.105	13991.023	-1.060	0.290	
English_Status(L2-MalE):Feature(present):Grammaticality_Type(omission)	-0.173	0.100	13991.029	-1.740	0.082	.
English_Status(L1-MalE):Feature(past):Grammaticality_Type(omission)	-0.186	0.105	13991.031	-1.760	0.078	.
English_Status(L2-MalE):Feature(past):Grammaticality_Type(omission)	-0.261	0.100	13991.034	-2.610	0.009	**

English_Status(L1-MalE):Feature(progressive):Grammaticality_Type(commission)	0.117	0.122	13991.088	0.960	0.336	
English_Status(L2-MalE):Feature(progressive):Grammaticality_Type(commission)	0.239	0.115	13991.068	2.070	0.038	*
English_Status(L1-MalE):Feature(do):Grammaticality_Type(commission)	0.515	0.122	13991.087	4.230	0.000	***
English_Status(L2-MalE):Feature(do):Grammaticality_Type(commission)	0.829	0.115	13991.057	7.190	0.000	***

Table C.4: Results of the post hoc comparisons between English Status groups by feature and grammaticality type

Copula: Standard

Contrast	Est.	SE	z	p
L1-BritE–L1-MalE	0.205	0.0436	4.7	<.0001
L1-BritE–L2-MalE	0.311	0.0425	7.32	<.0001
L1-MalE–L2-MalE	0.106	0.0338	3.14	0.005

Copula: Omission

Contrast	Est.	SE	z	p
L1-BritE–L1-MalE	-0.006	0.0437	-0.14	0.989
L1-BritE–L2-MalE	-0.073	0.0426	-1.72	0.197
L1-MalE–L2-MalE	-0.067	0.0338	-1.99	0.115

Progressive: Standard

Contrast	Est.	SE	z	p
L1-BritE–L1-MalE	0.087	0.0615	1.41	0.337
L1-BritE–L2-MalE	0.178	0.0591	3.02	0.007
L1-MalE–L2-MalE	0.092	0.0475	1.93	0.13

Progressive: Omission

Contrast	Est.	SE	z	p
L1-BritE–L1-MalE	-0.023	0.0613	-0.37	0.928
L1-BritE–L2-MalE	-0.081	0.0588	-1.37	0.357
L1-MalE–L2-MalE	-0.058	0.0475	-1.22	0.44

Progressive: Commission

Contrast	Est.	SE	z	p
L1-BritE–L1-MalE	-0.019	0.0613	-0.31	0.948
L1-BritE–L2-MalE	-0.236	0.0588	-4.01	<.0001
L1-MalE–L2-MalE	-0.217	0.0475	-4.57	<.0001

DO: Standard

Contrast	Est.	SE	z	p
L1-BritE–L1-MalE	0.114	0.0614	1.85	0.154
L1-BritE–L2-MalE	0.397	0.059	6.73	<.0001
L1-MalE–L2-MalE	0.284	0.0475	5.97	<.0001

DO: Omission

Contrast	Est.	SE	z	p
L1-BritE–L1-MalE	-0.138	0.0614	-2.26	0.062
L1-BritE–L2-MalE	-0.156	0.059	-2.65	0.022
L1-MalE–L2-MalE	-0.018	0.0475	-0.38	0.924

DO: Commission

Contrast	Est.	SE	z	p
L1-BritE–L1-MalE	-0.39	0.0613	-6.37	<.0001
L1-BritE–L2-MalE	-0.607	0.0589	-10.32	<.0001
L1-MalE–L2-MalE	-0.217	0.0475	-4.57	<.0001

Present: Standard

Contrast	Est.	SE	z	p
L1-BritE–L1-MalE	-0.011	0.0613	-0.18	0.983
L1-BritE–L2-MalE	0.092	0.0588	1.57	0.259
L1-MalE–L2-MalE	0.103	0.0475	2.18	0.075

Present: Omission

Contrast	Est.	SE	z	p
L1-BritE–L1-MalE	-0.111	0.0613	-1.81	0.166
L1-BritE–L2-MalE	-0.119	0.0589	-2.02	0.108
L1-MalE–L2-MalE	-0.008	0.0475	-0.17	0.985

Past: Standard

Contrast	Est.	SE	z	p
L1-BritE–L1-MalE	0.032	0.0613	0.52	0.864
L1-BritE–L2-MalE	0.115	0.0589	1.95	0.124

L1-MalE–L2-MalE	0.083	0.0476	1.75	0.186

Past: Omission

Contrast	Est.	SE	*z*	*p*
L1-BritE–L1-MalE	0.006	0.0613	0.1	0.995
L1-BritE–L2-MalE	-0.009	0.0589	-0.15	0.987
L1-MalE–L2-MalE	-0.015	0.0475	-0.31	0.947

Past: Commission

Contrast	Est.	SE	*z*	*p*
L1-BritE–L1-MalE	0.043	0.0614	0.71	0.76
L1-BritE–L2-MalE	-0.06	0.059	-1.03	0.561
L1-MalE–L2-MalE	-0.104	0.0476	-2.18	0.074