A CORPUS-BASED STUDY OF *GIVE*, *TAKE*, AND *MAKE* LIGHT VERB CONSTRUCTIONS IN MESOLECTAL MALAYSIAN ENGLISH

CHRISTINA ONG SOOK BENG

UNIVERSITI SAINS MALAYSIA

2023

A CORPUS-BASED STUDY OF *GIVE*, *TAKE*, AND *MAKE* LIGHT VERB CONSTRUCTIONS IN MESOLECTAL MALAYSIAN ENGLISH

by

CHRISTINA ONG SOOK BENG

**Thesis submitted in fulfillment of the requirements
for the degree of
Doctor of Philosophy**

March 2023

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

KAJIAN BERASASKAN KORPUS TERHADAP BENTUK KATA KERJA RINGAN '*GIVE*', '*TAKE*', DAN '*MAKE*' DALAM BAHASA INGGERIS MESOLEKTAL DI MALAYSIA

ABSTRAK

Penyelidikan ini bertujuan meneliti corak struktur dan fungsi aspektual bentuk kata kerja ringan (LVC) dalam bahasa Inggeris mesolektal Malaysia di samping mengenal pasti LVC yang telah dinativisasikan. LVC merujuk gabungan kata kerja yang kosong secara semantik iaitu kata kerja ringan dengan kata nama '*deverbal*'. Tiga kata kerja ringan yang dipilih adalah '*give*', '*take*', dan '*make*'. LVC yang asas terdiri daripada kata kerja ringan, dan kata nama '*deverbal*' yang didahului artikel yang tidak sahih seperti yang ditunjukkan dalam ayat berikut, '*I'll **take a look** at her feet myself*'. Kajian yang menyiasat penggunaan LVC dalam bahasa Inggeris pascakolonial mencadangkan bahawa varian yang sedang muncul seperti '***make good decision***' mungkin tidak idiomatik dan tidak diterima oleh penutur asli. Dalam proses penyelidikan ini, korpus am 100 juta perkataan bahasa Inggeris mesolektal di Malaysia yang dikenali sebagai *Corpus of Malaysian English Forum (CMEF)* telah dibina. Data korpus ini didapati daripada forum Lowyat.NET, iaitu sebuah forum internet yang popular di Malaysia dan mewakili kepelbagaian bahasa Inggeris mesolektal Malaysia yang biasa digunakan dan boleh difahami oleh masyarakat Malaysia. Korpus Nasional British (*British National Corpus, BNC*) telah digunakan sebagai korpus rujukan. Analisis struktur mendedahkan bahawa LVC bahasa Inggeris Malaysia tidak banyak menyimpang daripada struktur LVC asas kecuali penggunaan berlebihan terhadap LVC tanpa artikel. Analisis fungsi aspektual menunjukkan bahawa kebanyakkan LVC bahasa Inggeris Malaysia adalah atelik yang menggambarkan tindakan tanpa batas,

dicirikan oleh LVC tanpa artikel. Dapatan kajian kedua bercanggah dengan dapatan kajian terdahulu yang mencadangkan bahawa fungsi aspektual utama LVC adalah untuk menukar tindakan tanpa tujuan kepada tindakan yang mempunyai pencapaian. Untuk mengenalpasti LVC yang dinativisasi dalam bahasa Inggeris Malaysia, ujian '*log-likelihood*' telah digunakan untuk menilai perbezaan antara setiap corak struktur dan corak fungsi aspektual LVC yang terdapat dalam kedua-dua korpus. Struktur LVC yang dinativisasi ditakrifkan sebagai struktur yang mempunyai nilai '*log-likelihood'* yang melebihi 100. Sepuluh LVC tanpa artikel memenuhi kriteria tersebut. Kekerapan yang tinggi ini boleh dijelaskan sebagai akibat ketiadaan artikel dalam bahasa substrat di Malaysia. Berkaitan corak fungsi aspektual, nativisasi boleh dilihat dalam tiga belas LVC yang menunjukkan tindakan tanpa batas. Ini mungkin disebabkan oleh kehadiran banyak kata nama '*deverbal*' massa dan abstrak yang tidak boleh berlaku bersama artikel yang tidak sahih. Penyelidikan ini menunjukkan bahawa nativisasi struktur berlaku pada tahap leksikogrammar bahasa Inggeris mesolektal di Malaysia. Selain itu, varian-varian LVC yang digunakan secara meluas boleh dianggap sebagai tanda endonormativiti. Ini bermaksud bahasa Inggeris mesolektal di Malaysia mungkin telah melampaui tahap nativisasi berdasarkan model evolusi pascakolonial bahasa Inggeris.

A CORPUS-BASED STUDY OF *GIVE*, *TAKE*, AND *MAKE* LIGHT VERB CONSTRUCTIONS IN MESOLECTAL MALAYSIAN ENGLISH

ABSTRACT

This study investigated the structural and aspectual functional patterns of light verb constructions (LVCs) in mesolectal Malaysian English and identified nativised LVCs. LVCs refer to combinations of a semantically "empty" or light verb with a deverbal noun. The three selected light verbs are *give, take*, and *make*. A standard or prototypical LVC is made up of a light verb and a deverbal noun preceded by the indefinite article as shown in this sentence, *I'll **take a look** at her feet myself*. Studies examining the use of LVCs in postcolonial Englishes suggest emerging variants such as ***make good decision*** may not be idiomatic and may not be acceptable by native speakers. To facilitate this research, a 100-million words general corpus of mesolectal Malaysian English was created. This corpus is known as the Corpus of Malaysian English Forum (CMEF). It consists of threads from Lowyat.NET, a popular Internet forum in Malaysia, representing the mesolectal variety of Malaysian English which is commonly used and nationally intelligible. The British National Corpus (BNC) was used as the reference corpus. The structural analysis reveals that Malaysian English LVCs do not deviate much from the standard structure of LVCs except the overuse of zero-article LVCs. The aspectual functional analysis shows that most Malaysian English LVCs are atelic conveying unbounded actions, characterised by LVCs without article. The latter contradicts findings of previous related work which suggest the aspectual function of LVCs is mainly to convert aimless actions into achievements. To identify nativised LVCs in Malaysian English, the log-likelihood (LL) test was adopted to evaluate the difference between each structural and aspectual functional

LVC pattern in the two corpora. Nativised LVC structures are defined as those where the log-likelihood values are greater than 100. This is the case for ten zero article LVCs. Their high frequency could be explained by the absence of articles in the Malaysian substrate languages. As for aspectual functional patterns, nativisation is shown in thirteen LVCs indicating unbounded actions. This could be due to the presence of many abstract and mass deverbal nouns that can never co-occur with the indefinite article. The present research shows that structural nativisation is happening at the lexicogrammar level of mesolectal Malaysian English. Also, the extensively used variants of LVCs should be treated as a sign of endonormativity, which means mesolectal Malaysian English may have progressed beyond the nativisation phase according to the evolutionary model of postcolonial Englishes.

CHAPTER 1

INTRODUCTION

1.1   Introduction

A lexicogrammatical unit named light verb constructions (LVCs) is the focus of the present research. LVCs refer to combinations of a semantically "empty" verb such as *have, give, take, make*, and *do* (representing the grammar component) with a deverbal noun (representing the lexis component). The former is commonly known as light verbs (Bonial & Pollard, 2020; Brugman, 2001; Hoffmann, Hundt, & Mukherjee, 2011; Leech, Hundt, Mair, & Smith, 2009; Live, 1973; Mehl, 2019; Ronan, 2019;). Examples of prototypical or standard English LVCs include *have a look*, *give someone a kiss*, *take a shower*, *make a wish*, and *do a dance*. For discussion purposes, standard LVCs will be used throughout this study. A standard English LVC is made up of a light verb and a nominalised verb preceded by the indefinite article as shown in example (1a). Such construction is normally regarded as standard and acceptable by native speakers.

1) Purely because we'd like you to *take a look* at our Christmas Catalogue. (BNC)

2) Many children can be good in analyzing and *making good decision* in gaming too. (Lowyat.Net)

Standard English LVCs are substitutable with the corresponding verbal simplexes because the deverbal nouns carry the meaning of the constructions (Dixon, 2005; Huddleston & Pullum, 2002; Mehl, 2019; Sinclair & Fox, 1990; Stein, 1991). For instance, in example (1) *take a look* is equivalent to the verb *look*. LVC in example (2) *make good decision* is a typical Malaysian English form, it has no article and it could even take a modifier. Unlike example (1), the inclusion of modifiers may affect the

equivalency between the LVCs and simplexes which can be seen by comparing the deviated form of the LVC in example (2) with its awkward counterpart ?#*goodly decide*. LVC in example (2) is clearly different from the standard structure that native speakers are familiar with. They may even be considered as ungrammatical in the English as a native language (ENL) context. While most definitions of LVCs agree that they should be interchangeable with their simplexes and light verbs contribute little meaning to the construction (Brugman, 2001; Hoffmann, et al., 2011; Leech, et al., 2009; Shahrokny-Prehn & Hoche, 2011), there are differences in the description of the other characteristics of LVCs (Gilquin, 2019). The proportions of different LVC variants which include forms that are not standard or unidiomatic to native speakers have been attested to vary across different varieties of English. Ronan and Schneider (2015) adopted computational method to investigate frequently used LVCs in Irish English and British English. They found that Irish English favours more diverse lower frequency LVCs and has more instances of passive form, while British English prefers fewer high frequency LVCs (Ronan & Schneider, 2015). Giparaite's (2017) study examined Global Web-Based English (GloWbE) corpus to compare and contrast the modification patterns of LVCs in native and non-native varieties of English. She found the latter use more diversified modifiers than the former (Giparaite, 2017). Gilquin (2019) explored the use of LVCs among English as a Second Language (ESL – represented by Indian) and English as a Foreign Language (EFL – represented by Italian, Mexican, Spanish) speakers from different proficiency levels. The study revealed that advanced ESL speakers tend to overuse *give, take,* and *make* LVCs while EFL speakers underuse them and that both do not depict native-like use of LVCs (Gilquin, 2019). These are some of the studies comparing the frequency of LVCs in varieties of English worldwide. To explore possible grammatical innovations in non-

native varieties, particularly Malaysian English, an approach which considers all possible structural variants of LVCs (e.g. derived nouns, modifiers, and zero article) is required. This is because "a fairly productive constructional type", novel LVCs may always be introduced attributable to the unlimited possible combination of a light verb and a noun (Leech, et al., 2009, p. 166; Sundquist, 2020). Recognising its structural patterns and aspectual function is important to predict new emerging LVCs.

This chapter discusses the motivations of the current research, to investigate grammatical variation in mesolectal Malaysian English specifically the variants of LVCs. The approach to this study involves creation of a general corpus of Malaysian English made up of threads from Lowyat.Net, an Internet forum in Malaysia that represents the variety which is widely used in Malaysia. In short, this research focuses on revelation of grammatical variation using corpus-based approach. The following sections delineate the background to the study and statement of the problem. They are followed by the research objectives, research questions, scope of the study and significance of the study.

1.2   Background to the Study

Postcolonial Englishes or the English language used in countries like Malaysia, Singapore, Philippines, Hong Kong, India etc. which were once colonised by either the British or Americans have been claimed to demonstrate certain linguistic features which are different from varieties where English is spoken as a native language (Hajar & Shakila, 2014). This is mainly due to: a) deprioritising of the English language status during the post-independence period as priority is given to their national language to establish a national identity for an independent country (Tsui & Tollefson, 2004) and; b) the constant contact between English and various local languages in multicultural

societies (Tan, 2009a, 2009b). A unique characteristic many postcolonial Englishes shares is the nativisation process (Groves, 2010; Kachru, 1981; Lowenberg, 1986; Mukherjee, 2010; Richards, 1979; Schneider, 2003a). Nativisation refers to newly developed linguistic features which are systematic, extensively used and accepted by their users (Kachru, 1981). Coined by Kachru (1981), nativisation is defined as a result of constant contact between English and local languages used intranationally by non-native speakers who develop new linguistic conventions which are then accepted into their English language system. He is also known for his three concentric circles (i.e. the inner circle "where English is used as the primary language", the outer circle refers to multilingual countries once colonised by the inner circle, "where English has been institutionalised", and the expanding circle which includes the rest of the world "where English is used as the foreign language") that distinguishes English worldwide based on historical, social, and political factors (Kachru, 1985, p. 12, 1992). Twenty years later, Schneider developed an evolutionary model to explain the language ecologies of Postcolonial Englishes and he labelled the central phase as nativisation, the stage when both cultural and linguistic transformation by the settler and indigenous combine in full swing, resulting in intensification of nativised features on all linguistics levels (Schneider, 2003a, 2007). Kachruvian and Schneider's models are important for the present study, they are discussed further in Sections 2.6 and 5.3.

Research in nativised Malaysian English lexis has received much attention, but research in nativised Malaysian English grammar is under-explored. Past studies often concentrate on the lexical description of Malaysian English. Borrowed words and/or phrases deriving from local languages to fill "lexical gaps for non-existing English words" is one notable feature of Malaysian English (Lowernberg, 1986, p. 75; Tan, 2009a). Malaysia is a multi-ethnic nation made up of 69.9% of Bumiputera (which

include the Malays and indigenous people), 22.8% of Chinese, 6.6% of Indians, and 0.7% of other ethnic groups (Department of Statistics Malaysia, 2022). Their first languages namely Malay, Chinese, and Tamil are the primary local languages in Malaysia. Every ethnic group is also made up of native speakers who speak a multitude of languages (e.g. Bidayuh, Kadazan, and Mah Meri are indigenous languages while Malayalam, Telegu and Punjabi are different Indian languages) and dialects (e.g. Cantonese, Hokkien, and Hakka). Studies depicting acceptance of borrowed words from the three main local languages have proliferated since the early 21st century. The literature on Malaysian English at large has focused on lexical borrowings in the acrolectal variety (e.g. David & Dumanig, 2008; Hajar, 2013; Rita, 2014; Shakila, 2014; Tan, 2009b, and Thirusanku & Melor, 2013). These studies for the most part addressed the gap of knowledge in identifying the use of borrowed words in acrolectal Malaysian English using newspaper corpus, teaching materials and short story corpus. Findings from past lexical variation studies imply the construction of Malaysian identities and a blurring of ethnic boundaries, suggesting that Malaysians are accepting the nativised lexis even in formal contexts.

As regards nativised grammar features in the written context of Malaysian English, there have been several studies (Chai & Ong, 2019; Ho-Abdullah, 2010; Newbrook, 2006; Tan, 2013b). Most of them used newspapers as the source of data (cf. Chai & Ong, 2019; Ho-Abdullah, 2010; Newbrook, 2006; Tan, 2013b), which represents the acrolectal variety of Malaysian English. Although deviated forms are present, their frequencies do not differ much from the native speaker varieties (Ho-Abdullah, 2010; Tan, 2013b). Collins (2014) analysed the patterns of modal usage between Malaysian English and the two supervarieties, British English and American English diachronically. His findings revealed that modals have a stronger presence in

writing while quasi-modals in speech and suggest that Malaysians associate the former with conventional British English norms and the latter with American English (Collins, 2014). The other research investigated syntactic variation in the spoken form that Malaysians are familiar with which ranges from the mesolectal variety and the basilectal variety. They found distinctive forms which are colloquially used (cf. Wong, 1983) and obvious evidence of mother tongue interference (cf. Radina & Asniah, 2013; Vollmann & Wooi, 2019). Given the limited research on grammar of Malaysian English as a whole and in particular the mesolectal variety, the present study focusing on LVCs, a lexicogrammar unit is timely and necessary.

To date, patterns of nativised grammar including LVCs have received more research attention in postcolonial Asian Englishes such as Indian English (Hoffmann, et al., 2011; Moody, 2006; Mukherjee, 2010), Singaporean English (Mehl, 2019), Hong Kong English (Laporte, 2017; Mehl, 2019), and Philippines English (Borlongan & Dita, 2015). Non-standard LVCs are detected (cf. Laporte, 2017; Mukherjee, 2010) but there is little evidence of unique LVCs in these outer circle varieties of English (cf. Hoffmann, et al., 2011; Mehl, 2019). Nevertheless, the nature of light verbs (as semantically deficient verbs) allows them to co-occur with a wide range of deverbal nouns (Sundquist, 2020) which may lead to extensive, innovative use of new LVCs. In terms of frequency, ESL users are found to overuse LVCs (including both the standard form and variants of LVCs) compared to native speakers (Gilquin, 2019; Giparaite, 2017; Werner & Mukherjee, 2012).

Because of its dual senses, light and concrete, some with triple senses i.e. *have* and *do* functioning as auxiliary verbs in addition to a couple of factors which will be elaborated below, LVCs may be one of the most challenging multiword units to acquire among non-native users of English. Tendency to overuse, underuse, and

misuse LVCs has been highlighted in past ESL corpus-based studies (Gilquin, 2019). Altenberg and Granger (2001), Jukneviciene (2008), Laporte (2012), Marco (2011), Wang (2011), and Scheepers (2017) attested that even advanced learners of English have difficulties producing error-free LVCs. High frequency verbs (including light verbs) are often neglected after they are taught at an early stage causing learners to possess superficial knowledge about them (Altenberg & Granger, 2001).

As mentioned in Section 1.1, light verbs are regarded as meaningless to a certain extent because they depend on deverbal nouns to covey meaning (Huddleston & Pullum, 2002; Mehl, 2019; Sinclair & Fox, 1990). However, Brugman (2001) and Dixon (2005) argue that light verbs are not meaningless as there is a subtle functional discrepancy between LVCs and their simplexes. It is interesting to look into the polysemy of light verbs focusing on "their context-dependent contribution to the overall meaning of the constructions rather than meaning of LVCs independent of context" (Brugman, 2001, p. 552). LVCs can be interpreted to convey a casual or single occurrence of the action instead of continuous or repetitive actions (Live, 1973). Brugman (2001), Huddleston and Pullum (2002), and Wittenberg and Levy (2017) exemplify that the duration of an event is shorter when the *give* or *take* LVCs are used (see example 3a) but when the corresponding simplexes are used, the event appears to be prolonged (as shown in example 3b). LVCs may also indicate aspectual changes and specification unlike the simplex verbs (Bonial & Pollard, 2020; Ronan, 2019). The prepositional phrase following the LVC in example (4) expresses a specified time frame, i.e. duration taken to travel to Cottown.

3) (a) He *gave* a scream. (Huddleston & Pullum, 2002)

(b) He *screamed*.

4) to *take a ride* to Cottown ≠ can *ride* to the furthest (Ronan, 2019)

Despite the lack of emphasis on light verbs in formal education and the subtle aspectual function between LVCs and their corresponding simplexes, overuse of LVCs among non-native speakers of English especially among ESL speakers (Gilquin, 2019; Giparaite, 2017; Werner & Mukherjee, 2012) could be attributable to the following reasons:

a) According to Gradecak-Erdeljic (2009), the English language takes the subject verb object (SVO) sequence and that is why a basic English sentence usually requires an object to follow the verb. Quirk (1981, as cited in Algeo, 1995) and Brinton (1996, as cited in Ronan, 2019) share a similar view when they mention that "*My friend cooked*" sounds incomplete while "*My friend did the cooking*" is more acceptable. It appears that LVCs follow the SVO sequence of the English language as light verbs can be used transitively while the deverbal noun acts as the direct object; for instance, *she* (S) *is taking* (V) *a shower* (O) while the simplex is used intransitively, *she* (S) *is showering* (V). Moreover, highly polysemous verbs are usually transitive verbs which include light verbs *give*, *take*, and *make* and they remain as the most commonly used transitive verbs throughout the history of English language (Traugott, 1999 & Brinton, 2008 as cited in Sunquist, 2020).

b) The structure of LVCs allows modifiers which can enhance description of the predicate (Giparaite, 2017; Huddleston & Pullum, 2002; Live, 1973; Shahrokny-Prehn & Hoche, 2011; Stein, 1991). Elaboration of deverbal nouns can be facilitated using modifiers and determiners (Caro & Arus-Hita, 2020; Brugman, 2001; Ronan, 2019). Unlike the simplexes taking adverbs as modifier which is considered unnatural (Brinton, 1996 as cited in Ronan, 2019), LVCs are syntactically flexible, which means they can accept a range

of modifiers (Vincze, Nagy, & Berend, 2011). Sometimes, a related adverb does not exist to modify the simplex verb (*give misleading information* vs. *?inform misleadingly*). The opportunity to include an adjectival modifier which facilitates "dispersal of verbal ideas over several lexical units" (Hoppper, 1991 as cited in Brugman, 2001, p.556) may encourage the use of LVCs (Live, 1973) while reducing the use of simplexes with adverbials.

1.3   Statement of the Problem

Information on grammatical variations in Malaysia English gathered from corpus-based studies is still scarce. There are signs that progress is being made when corpus-based grammatical studies start gaining momentum but only in the English language teaching and learning context. Multiword constructions such as collocations in Malaysian student essays (cf. Ang, Hajar, Tan, & Khazriyati, 2011; Ang & Tan, 2016; Kamariah & Su'ad, 2011; Shazila & Noorzan, 2013) and phrasal verbs in Malaysian English textbooks (Ainul Azmin, Mahmoud, Rafidah, & Faizah, 2019; Rafidah, 2013; Zarifi & Mukundan, 2014) were examined. They found the use of Malaysian English multiword forms differs considerably in comparison to British English (Ang & Tan, 2016; Shazila & Noorzan, 2013; Zarifi & Mukundan, 2014). Thus far, no research on LVCs in Malaysian English has been conducted. Light verbs are worth investigating because they are highly polysemous (Altenberg & Granger, 2001; Brugman, 2001; Gradecak-Erderljic, 2009, Laporte, 2012; Mehl, 2019) and are highly frequent (Bonial & Pollard, 2020; Elenbaas, 2013; Mehl, 2019; Ronan & Schneider, 2015). High frequency verbs which include light verbs *give, take,* and *make* are preferred words in the repertoire of non-native users of English as they are regarded to be easy to use (Gilquin, 2019). That explains the tendency to overuse LVCs among

ESL users mentioned earlier. Prescriptivists found high occurrences of erroneous LVCs in non-native university student writing (Altenberg & Granger, 2001; Chi, Wong, & Wong, 1994; Eisouh, 2012; Sanguannam, 2017) while descriptivist offered evidence of forms and functions unique to different non-native varieties of English (Moody, 2006; Mukherjee, 2010; Ronan, 2019; Ronan & Schneider, 2015) as well as the native varieties (Algeo, 1995; Shahrokny-Prehn & Hoche, 2011; Smith, 2009). Given that LVCs are highly frequent and syntactically flexible, they are expected to vary across regional varieties. Investigating the use of light verbs *give, take,* and *make* is deemed sufficient for a start to provide an in-depth description of the structural and aspectual functional characteristics of Malaysian English.

The presence of several substrate languages namely Malay, Chinese, and Tamil which are the mother tongues of most Malaysians may result in a diverse array of LVCs. Also, the Malay language being the main medium of instruction of all national schools may influence the use of LVCs. There is a considerable body of work on LVCs across languages including the Malay language (Omrah Hassan & Ab Halim, 2014) and Chinese language (Lin, Xu, Jiang, & Huang, 2014). Since the primary local languages in Malaysia have LVCs, their influence on the use of English LVCs among Malaysians is inevitable. It will be interesting to discover distinct LVCs unique to Malaysian English.

The unavailability of systematic sources may have deterred researchers from looking into the grammar of Malaysian English. To date, most English corpora built in Malaysia are learner corpora, namely English of Malaysian School Students Corpus (EMAS) (Arshad, et al. 2002), Malaysian Corpus of Students' Argumentative Writing (MCSAW) (Mukundan & Kalajahi, 2013), Corpus Archive of Learner English in Sabah and Sarawak (CALES) (Botley, De Alwis, Metom, & Izza, 2005), and

Malaysian Corpus of Learner English (MACLE) (Knowles & Zuraidah, 2004) which consist of argumentative essays written by different levels of students ranging from upper secondary to college and university students. EMAS also comprises descriptive essays written by primary five, secondary one and secondary four Malaysian students. These learner corpora have assisted in the discovery of various grammatical errors in line with the linguistic demands of using native-like English in this globalised world. Another type of specialised corpus that has been commonly created to facilitate research is newspaper/print media corpus. Ho-Abdullah (2010) investigated newspaper articles published by the News Straits Times from 1990 to 1996, Tan (2013a) analysed 61 issues of the News Straits Times and 91 issues of The Star published in 2001 and 2002 respectively while Chai and Ong (2019) explored national news article from 2006 to 2012 in The Star online archive. All these corpora, as stated by Davies and Fuchs (2015), are created by individual researchers to find out certain trends. These corpora are not able to cater to researchers of World Englishes, in this case Malaysian English. Creating a web-based corpus is therefore necessary to facilitate research in World Englishes and on grammatical variation specifically, such as the current study. The reasons for the creation of Malaysia's web-based corpus are summarised below:

a) existing corpora like Malaysian learner and newspaper corpora are not able to represent Malaysian English as a whole.

b) the size of a web-based corpus will be relatively bigger compared to existing Malaysian corpora so that it can offer more examples of constructions which are non-frequent in specialised and general corpora.

c) the texts gathered online are more updated and may reflect contemporary culture (Fletcher, 2011).

As regards Malaysian English within the framework of World Englishess, its mesolectal variety requires attention at this juncture. It is the variety that asserts Malaysian identity, distinguishing it with other postcolonial Englishes and its historical input variety, British English. For some, mesolectal Malaysian English is perceived as an identity marker in this multilingual country (Baskaran, 2005; Pillai, 2012; Schneider, 2003a; Vollmann & Soon, 2019) while for others, the use of a less standard form of English indicates 'bad' English (Pillai, 2012, p.573) or "broken English" (Benson, 1990 as cited in Khaw, 2006, p. 71). The mesolectal variety is claimed to be the natural language choice to meet social needs (Schneider, 2003a). It is also the preferred medium of local communication used between Malaysians of varying ethnicity (Gill, 2002). Clearly, the mesolectal variety is frequently used in more social contexts compared to the acrolectal variety. However, there is a considerable body of work on nativised lexis (David & Dumanig, 2008; Hajar, 2013; Tan, 2009b) and some work on nativised syntax (Chai & Ong, 2019; Ho-Abdullah, 2010; Newbrook, 2006; Tan, 2013b) in the acrolectal variety, mainly represented by Malaysian English newspapers. Studies examining language used in informal contexts such as advertisements (Azirah, 2010; Nair-Venugopal, 2007), the Internet domain (Norizah & Azirah, 2009; Ong & Yuen, 2012), and a movie (Zaamah, Norazrin, & Su'ad, 2015) reveal that localised lexis which include borrowings and English lexis with local usage are quite prevalent. This shows lexical variations in mesolectal Malaysian English have been described but not as much as localised lexis in formal contexts. In an attempt to contribute to the existing knowledge of Malaysian English especially with regard to the syntactic characteristics in the mesolectal variety, detailed examination of frequency counts and statistical measures will be conducted.

1.4   Research Objectives

Having explained the gap of literature in Malaysian English and the lack of studies on lexicogrammar unit such as LVCs which may be unique to Malaysian English, this research aims to attain the following research objectives:

a) investigate the structural patterns of *give, take, make* LVCs in mesolectal Malaysian English that may deviate from the standard LVC structure

b) examine the distributions of the aspectual functions of *give, take, make* LVCs in mesolectal Malaysian English

c) discover nativised structures and nativised aspectual functions of *give, take, make* LVCs in mesolectal Malaysian English

1.5   Research Questions

The following are the questions that this research aims to address:

a) How do the structural patterns of *give, take, make* LVCs in mesolectal Malaysian English differ from the standard LVC structure?

b) How are the aspectual functions of *give, take, make* LVCs in mesolectal Malaysian English distributed?

c) What are the nativised structures and nativised aspectual functions of *give, take, make* LVCs in mesolectal Malaysian English?

1.6   Scope and Limitations of the Study

This study aims to investigate the use of LVCs in Malaysian English. To be specific, only LVCs headed by *give, take* and *make* in mesolectal Malaysian English represented by Lowyat.Net forum are analysed. Three areas listed below related to the scope of this study will be explained.

a) Reasons for investigating LVCs headed by *give, take* and *make*.

This study focuses exclusively on light verbs *give, take* and *make*. Although *have* and *do* are also common light verbs (Huddleston & Pullum, 2002; Sinclair, 2011), these two verbs have another more dominant function, which is their role as an auxiliary verb besides their heavy sense used to express concrete actions. According to Sundquist (2020, p. 374), *give, take*, and *make* in American English "get richer and more diverse" and the same light verbs as stated by Mehl (2018) are more salient compared to their heavy counterpart in the spoken form of British English. That explains the high productivity of LVCs headed by *give, take*, and *make* which may gradually increase over time. Hence, investigating them in Malaysian English may yield interesting results and contribute to the body of work on light verbs *give, take* and *make* in the outer circle varieties of English.

b) Reasons for creating and analysing a web-based corpus representing Malaysian English.

To ensure mesolectal Malaysian English is well-represented, topics discussed in Malaysia's most renowned Internet forum, Lowyat.Net were compiled. Lowyat.Net comprises eight main sections and each section contains numerous topics. The corpus is named Corpus of Malaysian English Forum (CMEF). The description of the corpus design can be seen in Chapter 3. CMEF is considered a web-based corpus because the texts are sourced from the World Wide Web, specifically Lowyat.Net, a website that allows Internet users to exchange information online. Web-based corpus is regarded as "big-data-based corpus" or ready-made web-based corpora (Loureiro-Porto, 2017, p. 450). The World Wide Web offers accessibility of countless webpages encompassing both formal and informal English texts which are relatively current from any country in the world. Moreover, Olavarria, de Ersson, and Shaw (2003 as cited in

Mukherjee & Hoffmann, 2006) state that using the World Wide Web as a corpus linguistic database for research into varieties of English is a reasonable starting point. The ease of mining data from the World Wide Web ensures a larger sized corpus compared to specialised and general corpora. Web-based corpora should be able provide adequate data and trends for in-depth studies on syntactic variations. To reiterate, the data representing Malaysian English in this study are from an Internet forum.

    c) Reasons for describing quantitative differences of LVC variants per se in Malaysian English are as follows.

The structures and aspectual functions of LVCs are rather diverse in individual varieties of English. There are many possible structures of LVCs – varying structures ranging from the basic structure to other structural variants of an LVC, such as LVCs co-occurring with zero article, multiple modifiers, and in passive voice (Bonial & Pollard, 2020, Mehl, 2019; Ronan, 2019,); likewise, the aspectual function of LVCs varies depending on their structures and complements (Bonial & Pollard, 2020; Dixon, 2005; Kearns, 2002; Ronan, 2019). Similar to other areas of lexicogrammar such as collocations, particle verbs, prepositional verbs, LVCs may also show evidence of manifestation and regional differentiation.

Identifying the proportions of different variants of LVCs in a large and representative corpus of Malaysian English and comparing them with those used in the inner circle varieties (i.e. British English in this study) is a viable method (cf. Mukherjee & Gries, 2009; Mukherjee & Hoffmann, 2006). It will: i) reveal overuse or underuse forms of Malaysian English LVCs that are acceptable in the native varieties and; ii) distinguish forms of Malaysian English LVCs that do not exist or are unacceptable by the native varieties. The first refers to forms of LVCs which diverge

from the British English and are probably manifestations of the standard form; this divergence is regarded as gradual and not categorical (Mukherjee & Hoffmann, 2006). In other words, LVC variants occurring only in the Malaysian English corpus are categorical (i.e. they may be infrequent) whereas LVC variants found in both native and non-native English corpora are probabilistic (i.e. their frequencies may be subtly different).

Researching lexicogrammatical differences quantitatively is important because "without quantitative methodology, no observer would have expected such differences to exist" (Schneider, 2007, p.87). Moreover, structural nativisation also covers quantitative differences of the investigated forms between varieties of English and those belong to the common core (Mukherjee & Gries, 2009). It is worth noting that other aspects of LVCs such as meanings reflected by the constructions independent of context (e.g. the gloss for LVC *take care* might be *to look after* or *to deal with*) will not be discussed.

Limitations of the present research which derive from the scope of study abovementioned will be reiterated in this paragraph. Firstly, LVCs headed by *have* and *do* are not considered in this research despite them being frequently used in the native varieties. They are rarely used as light verb, instead they are normally used as auxiliary and action verbs among Malaysians (as evident in the findings of the pilot study, refer to Section 3.2.1). Secondly, it may be limiting that the corpus data of this study is extracted from one specific source or website, but, Lowyat.Net is the only active Malaysian forum which sees users constantly contributing to the discussions and most threads are updated daily. The target of the self-built web-based corpus is 10% of Lowyat.Net as the forum could have billions of words as it is growing daily. Thirdly, findings of this study are not generalisable to Malaysian English as a whole

since the data mostly resemble the syntactic characteristics of the mesolectal variety

(highlighted by Baskaran, 1987 as can be seen in Section 3.4.1). In other words, the

use of LVCs in formal contexts or acrolectal Malaysian English is not retrievable

from the present data set and hence, not investigated.

1.7   Significance of the Study

Findings of this study will provide insights into the emerging variants of LVCs

and whether they share the same aspectual function with their simplexes. With regard

to the structural variations of LVCs, comparing and contrasting the frequency of the

most basic structure to the most diverse one in the same set of data can reveal variants

that Malaysian English users prefer. Also, whether the standard LVC structure should

be strictly adhered to or new variants should be considered can be decided. Conversely,

identifying the aspectual functional characteristics of LVCs may help to resolve the

issue related to interchangeability between LVCs and their simplex forms. LVCs have

been claimed to be substitutable with their corresponding simplex verbs (Elenbaas,

2013; Mehl, 2019; Ronan & Schneider, 2015) but some attested that they are not

equivalent (Bonial & Pollard, 2020; Bruening, 2016; Brugman, 2001; Sundquist,

2020). Generally, the structural patterns and aspectual functional characteristics of

LVCs identified from this study will reveal proportions of the different variants in

British English and Malaysian English especially variants that are unique only in

Malaysian English.

Research into a web-based corpus of Malaysian English will show whether

Malaysian English is systematically different from its historical input variety and

whether the characteristics are consistent within Malaysian English. Conducting a

synchronic comparison of varieties can provide sufficient evidence to find out the

degree of similarity between Malaysian English and British English with reference to lexico-grammatical markers such as LVCs. The status of Malaysian English can be assessed using tools from the corpus data. The corpus is subjected to "a test of difference", i.e. CEMF is compared to the BNC, a corpus of the native speaker input variety, in order to measure the structural differences and similarities between the varieties (Mair & Mollin, 2007, p. 345). Analysis of concordances in CMEF allows identification of idiosyncratic features (in this case variants of LVCs) that merit a categorisation as a marker of Malaysian English.

This study also enriches work in mesolectal Malaysian English which is understudied. As stated in Section 1.3, the mesolectal variety has not been given much attention despite its prevalence as the main medium of communications among multilingual and multiethnic Malaysians. Even in business settings, a professional domain, syntactical variation and other modes of speech such as code switching and mixing characterising the mesolectal variety are found in Malaysian workplace communication, for instance, business presentations, training sessions and seminar presentations (Gill, 1999b; Nair-Venugopal, 2000). Analysing the syntax of mesolectal variety is crucial because when certain localised language forms (LVCs in this case) are continuously and frequently used, they may at some point be absorbed into the acrolectal variety as suggested by Yuen (2007).

In short, this study seeks to illuminate distinctive grammatical patterns, specifically LVCs in Malaysian English using the corpus-based approach which involve synchronic comparison of Malaysian English, a postcolonial variety and British English, its historical input variety, alongside reasons leading to their emergence. Documenting nativised or idiosyncratic grammar forms may be able expedite Malaysian English to be institutionalised or in achieving endonormative

stabilisation, phase four of Schneider's Dynamic Model. It is a phase when the local community consensually accepts nativised linguistic norm and has "locally rooted linguistic self-confidence" (Schneider, 2003b, p.48).

1.8   Definitions of Key Terms

1.8.1   Light Verb Constructions (LVCs)

Light verb constructions (LVCs) refer to the pairing of a light verb and a deverbal noun (Dixon, 2005; Gipraite, 2016; Gradecak-Erdeljic, 2009; Hoffmann, et al. 2011; Kearns, 2002; Leech, et al., 2009; Mehl, 2019; Ronan, 2019). To put it simply, the standard structure of an LVC is light verb + indefinite article + deverbal noun in the base form of a verb, the meaning of the construction is contributed by the latter. LVCs in this study may not co-occur with the basic constituents per se, instead, possible variants which include deverbal noun not in plain verb base, zero article, and modifiers without corresponding adverbial form (see example 5) are also taken into account.

5)   Carol was *given misleading information*.

Terms related to structural and aspectual functional LVCs will be briefly defined in the following sub-sections.

1.8.1 (a) Structural Pattern of LVCs

Linguistically, structural pattern is also known as the syntactical form, referring specifically to the combination of a few components that make up an LVC. As stated above, an LVC is a lexicogrammatical unit that consists of a lexis and a grammar component. The former is represented by deverbal nouns, defined as a noun that happens to be identical with the verb stem (Wierzbicka, 1982; Huddleston &

Pullum, 2002) such as 'shower' in *take a shower*, this type of deverbal noun is also known as isomorphic noun. More recent studies accept deverbal nouns that are derivationally related to a verb (Sundquist, 2020) like 'assumption' in *make an assumption*, also known as non-isomorphic noun (cf. Arus-Hita & Caro, 2021; Bonial & Pollard, 2020; Mehl, 2019; Ronan, 2019; Sundquist, 2020). The grammar component is represented by light verbs, also called delexical verbs (Sinclair & Fox, 1990; Laporte, 2012). They are similar to other common action verbs but are semantically deficient (Hoffmann, et al. 2011; Leech, et al., 2009; Sundquist, 2020). The claim made by Butt (2010), Gipraite (2016), Hoffmann, et al. (2011), Kearns (2002), Leech et al. (2009), Sundquist (2020), and Wierzbicka (1982) about light verbs having semantically reduced, bleached or weakened meaning seems true based on example 5. The deverbal noun, *inform* carries the main semantic content of the construction instead of light verb *give*.

1.8.1 (b) Aspectual Function of LVCs

Aspectual function in this study refers to whether the event or action conveyed by the LVC is atelic or telic. Aspectuality is a type of meaning change related to time found in LVCs when compared to their corresponding simplexes, typically indicated by morphology (Live, 1973; Ronan, 2019). It does not involve identifying verbal meanings of the light verbs and their LVCs (e.g. LVC *take care (of)* can mean *to deal with* or *to care for*), instead, LVCs are associated to (a)telicity. Telicity refers to a sense of boundedness which suggests an event is limited in time and is usually associated with English LVCs (Bonial & Pollard, 2020). In contrast, the corresponding lexical verbs normally lend atelicity to an event because the actions are limitless (Wierzbicka, 1982). For instance, *give a scream* lends telicity, while *scream* is atelic

although both events are arguably the same. Because variants of LVCs may emerge in Malaysian English, the aspectual functions of these LVCs may be different too. The current research explores the distinction between telic and atelic LVCs based on boundedness, duration, and goal (see details in Section 2.4.2).

1.8.2 Malaysian English

Malaysian English refers to the English used by Malaysians. It is one of the varieties of postcolonial Englishes that emerged following the spread of English through British colonisation of Asia and Africa. While the Malay language is the national and official language in Malaysia, English is given the status of a second language. It is the second lingua franca (after Malay) used in both urban and rural Malaysia alongside the three main languages – Malay, Chinese, and Tamil (Asmah, 1996, as cited in Hajar, 2013). In Malaysia and countries once colonised by either the British or Americans, English is spoken as a second language (ESL), known to be developing its own norms and standards which accord with the characteristics of outer circle varieties from the classical Kachruvian three circle model (Low, 2010). This variety of English is to a certain extent different from not only the English language used in native speaking countries but also the other non-native speaking countries. Ranging from a formal standard style to an informal colloquial style (Platt & Weber, 1980; Richards, 1979; Wong, 1983), Malaysian English is one of the established ESL varieties or the postcolonial Englishes. The formal variety, used in education, business and administration, is modelled after the native speaker variety (Wong, 1983) whereas the informal variety, used in informal situations, is the deviated version of Standard British English (Crismore Ngeow, & Soo, 1996).

The model that has been consistently used to account for variation in Malaysian English is the acrolect-mesolect-basilect continuum (Platt, Weber & Ho 1984). Mesolect refers to the variety in the middle of a continuum which Platt, Weber, and Ho (1984) describe as post-creole continuum. The acrolect, almost similar to the Standard British English occupies one end of the continuum while a highly colloquial variety (basilect) used by those with little or no formal English education occupies the other end. With regard to the features of the sociolects, Baskaran's (1987) classification as shown in Table 1.1 distinguishes the characteristics of the three *lects* at three linguistics level – phonology, syntax, and lexis. This is the English used by a generation of Malaysians who are able to communicate comfortably in informal English (i.e. mesolectal variety) and most of them have little to no ability to switch to the standard varieties (Gill, 1999a).

Table 1.1 Characteristics of the three sociolects of Malaysian English

(Baskaran, 2005, p. 22)

	Acrolect Standard Malaysian English	Mesolect Dialectal Malaysian English	Basilect Patois Malaysian English
	(spoken & written) formal use and international intelligibility	(spoken & written) informal use and national intelligibility	(spoken only) colloquial use, patois intelligibility
Phonology	Slight variation tolerated so long as it is internationally intelligible.	More variation is tolerated – including prosodic features especially stress and intonation.	Severe variation – both segmental and prosodic, with intonation so stigmatised – almost unintelligible internationally.

Syntax	No deviation tolerated at all.	Some deviation is acceptable although it is not as stigmatised, as broken English (intelligibility is still there)	Substantial variation / deviation (national intelligibility)
Lexis	Variation acceptable especially for words not substitutable in an international context (or to give a more localised context).	Lexicalisations quite prevalent even for words having international English substitutes.	Major lexicalisation – heavily infused with local language items.

To summarise, the acrolectal variety is considered the standard formal and written native speaker variety of English that should be taught and learnt in Malaysian schools, the mesolectal variety is usually used for intranational communication among Malaysians of different ethnicities, while the basilectal variety is used by low proficiency English users characterised by limited vocabulary and efficiency as a medium of communication (Wong, 1983).

1.8.3 Mesolectal Malaysian English

Mesolect is regarded as a colloquial and informal variety of Malaysian English (Wong, 1983) but not to the extreme extent like the basilect which is used by the less proficient and less educated (Newbrook, 2006). Unlike the acrolect, the mesolect is nationally intelligible as it is widely used within a country by its citizens and it is also commonly used in everyday social interactions (Platt, Weber, & Ho, 1984). It also plays an important role in intranational communication (Gill, 2002). Since the mesolectal variety of Malaysian English is the focus of this study, it is important to describe its characteristics. Baskaran's (1987) work on the lectal classification was

referred to, as she is one of the proponents of this model who relates it to Malaysian English. Mesolectal Malaysian English is described as having: (i) syntactic deviation that does not affect intelligibility; (ii) lexical variation even for words with English substitutes and; (iii) significant yet tolerable phonological differences (Baskaran, 1987). Certain syntactic features of mesolectal Malaysian English such as omission of copula, the various uses of *got*, the use of *already* as perfective marker, and structural convergence with the Chinese language are highlighted in Gill (1999a), and Vollmann and Wooi (2019) studies. Indeed, syntactic variations are prevalent in this variety and there is no sign of diminishing because they are understandable by most Malaysians. In short, the syntax of mesolectal Malaysian English deviates from the Standard British English but it is acceptable by most of the locals except probably the "language purists, prescriptivists, and the English language teaching community." (Hajar & Shakila, 2014, p.21).

1.9   Conclusion

The aim of the present research is to provide insights into structural nativisation on the level of verb phrase in Malaysian English with a specific focus on LVCs. This chapter started with the background to the study which explains the need for research in mesolectal Malaysian English and the sophisticated nature of light verbs that can inform possible grammatical change in contemporary English. Then, the statement of the problem explaining the lack of emphasis on lexicogrammatical variations in mesolectal Malaysian English was established, followed by the research objectives and research questions of this study. The scope of the study presented the justifications of restricting the analysis to LVCs headed by *give, take*, and *make*, using a web-based corpus to represent mesolectal Malaysian English, and looking into probabilistic and