Investigating Lexical Variation and Change in Malaysian Twitter: A Conceptual Paper

Noraishah Gulnazir[a]
p110855@siswa.ukm.edu.my
Center for Research in Language and Linguistics,
Universiti Kebangsaan Malaysia, Malaysia

Khazriyati Salehuddin[b]
khazudin@ukm.edu.my
Center for Research in Language and Linguistics,
Universiti Kebangsaan Malaysia, Malaysia

ABSTRACT

Social media platforms such as Twitter is constantly transforming the usage of lexical items among global Internet users, including in Malaysia. Interestingly, despite adopting British English in the national education system, American English is gaining prominence among Malaysians due to the widespread dissemination of American English through the media. American English has been classified as a hyper-central language, serving as the hub for global English in Mair's theory of The World System of Englishes. Despite of the magnitude of American English as a global language, there is a dearth of research on how American English is affecting other varieties of English, especially Malaysian English. There is a need to examine the role of American English in leading global language variation and change. Thus, this conceptual paper proposes how the influence of American English on Malaysian English, in terms of lexical items can be investigated on Twitter. This paper demonstrates how two emerging American lexical items *lit* and *on fleek* can be investigated in terms of its trend of frequency and patterns of usage in Malaysian Twitter through a number of tools and methods. Results from such a study may be able to reveal the extent to which Malaysian English is influenced by American English in terms of lexical units on Twitter, shedding light on the global transformation of the English language.

Keywords: Lexical variation; lexical change; language variation; language change; Twitter

INTRODUCTION

Language variation and change manifest through phonetic, lexical, and semantic changes. The terms *lexis, lexical items,* and *lexical units* refer to words that can function alone and are oftentimes used interchangeably to signify vocabulary (Lewis et al., 1997; Caro & Mendinueta, 2017). Lexical items are basic units of meaning. They merge with a set of rules governing the language to produce larger and more complex units such as phrases and sentences.

The recent decade has witnessed a revolutionary change in the English language. For decades, English has been making its way around the world, primarily through historical outreach. The initial use and spread of English were through traditional ways such as speech, sign, and writing. The advent of digital communication in the recent years has accelerated the progress, and continues to do so, ever since English began to play its role as the dominant language on the

---

[a] *Main author*
[b] *Corresponding author*

Internet in the 1990s (Crystal, 2003). The surge in the use of digital technologies over the years with digital transformation technologies such as Cloud, Internet-of-Things, and Artificial Intelligence has brought the ubiquitous use of the Internet, especially social media to the forefront. Research has revealed that social media communication leads toward language development (Schmied, 2012) and contributes to the development and shift of language practices and linguistic repertoire (Lantz-Andersson, 2018). Sharma (2012), for example, has shown that the emergence of Facebook has expanded the features in the English language usage of college students in Nepal and has eventually influenced their identities.

One of the online platforms which has propelled language variation and change is Twitter. Founded in 2006, Twitter is a microblogging site which connects users through blogging and instant messages called *tweets.* The use of Twitter has proliferated over the years, with 206 million users worldwide towards the end of 2021 (Brian D, 2022). Twitter has changed and is still changing the social landscape through online communication. There is a wealth of data on Twitter with regard to language studies, and linguists have taken advantage of this by investigating Twitter in language learning (Rosell-Aguilar, 2018), literacy practices (Gleason, 2018), sentiment analysis (Taboada, 2016) discourse analysis (Al-Ghamdi & Albawardi, 2020), and political communication (Santoso, Utari & Kartono, 2020). Recent studies have begun using Twitter in exploring language change and variation, particularly in understanding lexis on Twitter (e.g., Huang et al., 2016; Grieve et al., 2017; Grieve et al, 2018; Grieve et al., 2019; Würschinger, 2021).

A language reaches global status when it is recognized in every country, and such is the case of English. The global spread of English has resulted in numerous varieties of English over the globe, including Malaysian English. As Baskaran (2005) puts it, "after almost two centuries of nurturing and over four decades of nursing, the English language in Malaysia has developed to become a typical progeny of New Englishes: a distinct variety in its own right" (p.18). The Malaysian variety of English has witnessed a wide expanse of language variation and change, initially due to the impact of colonialism and more recently as a result of globalisation. The historical context and the language planning policies of the country has enabled the Malaysian English to be an established variety with stylistic layers as well as distinguishing characteristics at the basilectal, mesolectal, and acrolectal levels (Bolton, Botha, & Kirkpatrick, 2020). The essence of Malaysian English is to provide a sense of identity and to build rapport through its phonology, lexical items, and syntactic structures that are rooted in the Malaysian form (Thirusanku & Yunus, 2012). Many Malaysians now use English as a primary means of communication (Kashinathan & Abdul Aziz, 2021) and traits taken from other languages have been ingrained into Malaysian English.

Generally, Malaysian English differs from other varieties of English in terms of phonology, grammar, and lexis. With regard to lexical items, there are a few aspects which characterise Malaysian English, including local and global lexical borrowings or loanwords, code-mixing and slangs (Hashim, 2020; Salehuddin, 2022). The acquisition of the English language in Malaysia is mainly through the education system, and interestingly, despite adopting British English in the national education system, American English is gaining prominence among Malaysians due to the influence of media (Shamsuddin et al., 2019). The eminence is now added with the presence of social media.

Studies have shown that English is widely used by Malaysians on social media (Ho et al., 2011; Hamat, Embi, & Hassan, 2012; Rusli et al., 2018; Yunus, Zakaria, & Suliman, 2019). New findings have revealed the emergence and innovations of jargons, memes, and slangs, especially among adolescents, who remain as the primary users of social media. With digital communication

playing such a pivotal role in language use and development as well as contributing to linguistic identities, ideally, the Malaysian variety of English should be understood by those who speak this variety of English. However, in reality, some features in the Malaysian English are unintelligible to some Malaysian English speakers, especially with the recent linguistic innovations in the variety. This is mainly because Malaysian English has emerged as a unique variety of English in its lexical usage, not only through the infusion of local traits and characteristics (Bolton, Botha & Kirkpatrick, 2020), but also through the impact of global influences and trends (Moody, 2020).

While there are numerous studies which have described lexical features in the Malaysian variety of English in terms of localizations (Baskaran, 2005; Baskaran, 2008; Hashim, 2020), there is a dearth of research which examine the emergence of lexical items in the Malaysian variety of English which are in practice due to the impact of globalisation, particularly on social media. Since there has been an influx of new global words making their way into the Malaysian variety of English, in which most of them are commonly used on social media, this conceptual paper proposes ways to investigate lexical variation and change in the Malaysian English, particularly on Twitter, to illuminate its distinctiveness as a variety of English. Based on the theory of The World System of Englishes, this study aims to understand how the global widespread and pervasiveness of American English is influencing Malaysian English on Twitter, especially in terms of lexical units.

LEXICAL VARIATION AND CHANGE

A language's stock of lexical units (lexicon) undergoes changes constantly, where there are occurrences of change in forms and meanings. Semasiological variation and onomasiological variation are the main distinctions that have been used consistently in the study of lexical items. Semasiological variation occurs when a lexical item refers to different types of referents, for instance, *pants* are synonymous with *trousers* and *underpants*. Onomasiological variation happens when a referent or category of referent is named using a variety of conceptually distinct lexical categories. For example, a particular pair of pants can be categorised as a member of *pants/trousers* category or with a different subordinate category that is *jeans*. The expansion of the semasiology-onomasiology pair has brought forward processes that affect the changes of meaning (semantic change) and word formation processes in literature. Different scholars have different classifications of word formation processes. A prominent taxonomy is by Yule (1985), which categorised the following under word formation process: coinage, borrowing, compounding, blending, clipping, backformation, conversion, acronym, derivation, prefix, suffix, and multiple processes. Primary word formation processes include inflection, derivation, compounding, clipping, borrowing, blending, truncation, ellipsis, formative extraction, and acronyms (Geeraerts, 2010 & Miller, 2014). Miller (2014) on the other hand, uses the term *lexicogenesis* to encompass word formation processes.

Past studies have shown that lexical change takes place in different ways, one of which is the replacement of certain lexical units with existing words in the language over a certain period of time. For instance, the old English rood, which used to represent the cross or crucifix symbolising the cross on which Jesus Christ died, has been replaced by the word cross over the years (Miller, 2014). Lexical change is also brought forward by the emergence of new words. This is known as neologism. Neologism is defined as "newly coined lexical units or existing lexical units that acquire a new sense (Newmark, 1988, p. 140) and are not yet included in general dictionaries (Algeo, 1991). Cook (2010) elaborated that there are two types of neologism, the first

type is the combination of words to form novel words, for instance webisode, a combination of web and episode. The second type of neologism is the existing word forms that produce new meanings, for example, using email as verb instead of a noun.

New word forms are modelled on prior knowledge, creativity, and imagination (Miller, 2014). Creativity of individuals is what yields neologisms across domains and genres. For instance, in the political realm, the withdrawal of the United Kingdom from the European Union, famously known as Brexit has been in use since 2012 and is now utilized globally (Fontaine, 2017; Jeffries & McIntyre, 2018). The blend between *British* and *exit*, in turn, has generated a plethora of new words such as *Brexitology*, *regrexit*, *breferendum*, *Brexitsphere*, *brexpats*, *Breturn*, Brexitology, and *brexiteer* through the process of word blending from Britain, British, and Brexit as the source words (Lalić-Krstin & Silaški, 2018). This demonstrates how language users are capable of exerting their creativity with regard to a socio-political context. Another example is the recent COVID-19 pandemic which has also resulted in the emergence of new word forms. Research has revealed new acronyms and abbreviations (e.g., WFH for work-from-home) as well as an increase in the use of less common expressions such as self-isolation/self-isolate, physical distancing/social distancing). Due to the adverse impact of the pandemic towards the global economy, words such as furlough and layoff also came into prominence (Asif et al., 2021).

LANGUAGE USE ONLINE AND GLOBALISATION

The emergence of world wide web and social media has allowed researchers to have a greater understanding of spontaneous, real-life language change, especially in written communication. Personal emails, chat rooms, online forums, instant text messaging applications such as WhatsApp as well as platforms such as Facebook, Twitter, Blogger, and Instagram, with their unique appeals, have enabled the global speech community to connect and express themselves via words. This is referred to as either digital networked writing (Androutsopoulos, 2011), computer-mediated communication (Romiszowski & Mason, 2013), or computer-mediated discourse (Herring & Androutsopoulos, 2015). According to Arrizabalaga (2021), past literature has delineated the following as the distinguishing features of language use on the Internet: (i) replete with acronyms, emoticons, emojis, contractions, repeated letters, capital letters with connotative meanings, inventive use of punctuation marks, unusual spellings, and self-corrections; (ii) spurred by creativity and innovation, resulting in new word forms derived from different word processes, and (iii) rife with omissions, incomplete clauses, and informal expressions.

The distinct use of language online has further contributed to language variation and change. This type of social media's linguistic transformation (Tankosić & Dovchin, 2021) is propelled by online language users. According to Graddol (2000), the Internet "has given the shift of control to ordinary users" (p. 51). The power of traditional media such as printing and broadcasting which were once the gatekeepers to promote standard language have shifted to the Internet. As a result, the Internet, especially social media, is "contributing to the fluidity and promotion of vernacular, or in-group, language" (Battarcharjee, 2009, p. 49). With such extensive and collective power belonging to social media users worldwide, the role of propelling linguistic transformation now lies with them. Language change in social media is now driven by a global network of users. The world is now witnessing the birth of global lexical items, shared by global citizens.

Giddens (1991) defines globalisation as the strengthening of global social relations that connect far-flung locations in such a way that local events are impacted by events taking place

thousands of miles away, and vice versa. Globalisation has enabled linguistics to be observed beyond traditional constructs and barriers, addressing burgeoning ideas which are defining the society. Blommaert (2010) referred to this as 'sociolinguistics of globalization'. The goal of sociolinguistics of globalisation is to connect ideas that go beyond a stratified, unidirectional perspective of the language, by understanding "trans-contextual networks, flows, and movements" (p.1). The English language is continually flowing beyond traditional geospace, causing a shift in English practices and introducing new conventions. Tseng and Hinrichs (2020) expanded on the notion of 'language mobility', in which the English language is considered to be traversing all around the world through various processes. Due to the technological advancement, the mobility of English language today is no longer constricted to geospace; English flows through digital spaces too. According to Androutsopoulos (2011), global circulations are conceptualised at two main levels. The first takes place when new genres or discourse patterns emerge on a larger scale, such as in news reporting, businesses, or popular music. The second level is when linguistic features, particularly lexical items, spread across dialects or languages.

Lexical items used globally affect collective and individual lexical repertoires. As global lexicons expand or shrink, collective and individual lexical repertoires also change - either increasing or decreasing in size. Androutsopoulos (2014) investigated language practices on Facebook and the impact it had on individual linguistic repertoires, and found a connection between the two. Tankosić and Dovchin (2021) investigated the impact of social media towards peripheral countries such as Bosnian, Serbia, and Mongolia; they found that peripheral languages adopt relocalisation. Relocalisation is one of the impacts of globalisation, in which lexical items and discourse markers from English are borrowed and re-adapted into local alphabetic, orthographic, syntactic, and grammatical systems to the extent where the original speakers are unable to understand them (Androutsopoulos, 2011).

The most important impact of global spread of English is the expansion in the varieties of English around the world, resulting into an endless list of regional, national, subnational, pidgin, and creole forms (Canagarajah, 2013). This was hypothesised earlier by Pennycook (2007), who stated that "languages will flow and change around us, new combinations of languages and cultures will be put together, texts will be sampled and mixed in ever new juxtapositions" (p. 158). This statement aligns with the concept of 'linguascapes', which Dovchin (2018) defines as "the transnational flows of linguistic resources circulating across the current world of flows, making meanings in contact with other various spatiotemporal scapes interacting with one another, and affecting the particular speakers' linguistic practices in varied ways" (p. 35). Pennycook (2007) had earlier affirmed that there is a necessity to evaluate the spread and use of English around the world while taking into account the various local contexts in which it is used, such as history and politics, the current linguascape, language ideologies, economy, and infrastructure. According to Pennycook (2007):

> At the very least, we need to understand how English is involved in global flows of culture and knowledge, how English is used and appropriated by users of English around the world, how English colludes with multiple domains of globalization, from popular culture to unpopular politics, from international capital to local transaction, from ostensible diplomacy to purported peace-keeping, from religious proselytizing to secular resistance. (p. 19)

Tseng and Hinrichs (2020) states that future research could address underexplored issues surrounding English language mobility and contact, such as English in digital media and transnational networks. New global electronic discourses have emerged from online chat, instant

messaging, weblogs, podcasts, and mobile apps as a result of the impact of digital media on language users on a daily basis, paving new research directions in global Englishes that have not yet been thoroughly explored, and such research is necessary because "it will help determine whether the media consistently and accurately reflect the "pluricentricity" of English or, on the contrary, largely misrepresent both linguistic and sociocultural reality" (Martin, 2019, p. 607).

Despite the fact that English is widely acknowledged to be prevalent in global computer-mediated communication, there is a dearth of research in studying the lexical units or grammar of English on social media (Coats, 2016). Narrowing it to the local context, generally, there is a paucity of research on lexical variation and change in Malaysia. The recent studies have mainly focused on lexical borrowing in newspapers (Kunalan, Mutty & Francis, 2021), lexical features in a movie (Nor & Zamri, 2015), and slangs and jargons (Rusli et al., 2018; Izazi & Tengku Sapora, 2020). There is a need to investigate the presence of lexical items which are widespread online at a global scale in different varieties of English, for instance, in the Malaysian English. Before investigating lexical variation and change online, it is important to first dissect the global role of the English language, which will be covered in the next section.

THE WORLD SYSTEM OF ENGLISHES

Over the past few decades, English has emerged to the fore as one of the widely used languages in the world. The study of English as a global language could be traced back to the development of the theory of World Englishes by Kachru (1985). Kachru's World Englishes have paved the way for the emergence of new theoretical frameworks, including The World System of Englishes. Proposed by Mair (2013), this theory encompasses and explains all the varieties of English around the world. There are four levels in this model, which are hyper-central variety, super-central varieties, central varieties, and peripheral varieties.
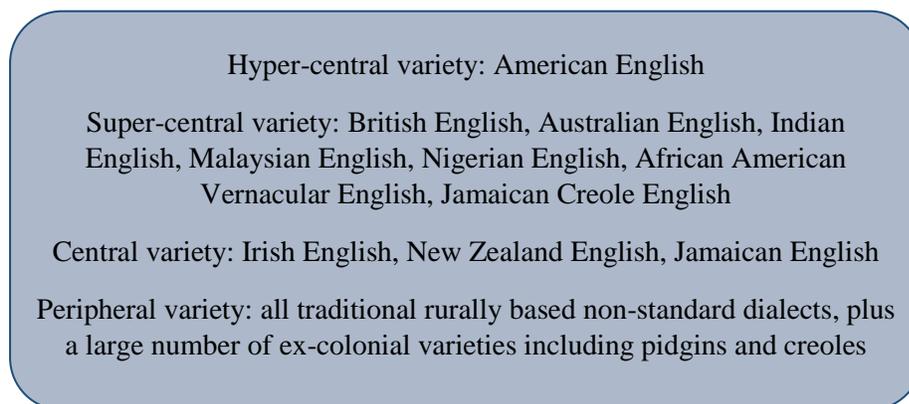
Hyper-central variety: American English

Super-central variety: British English, Australian English, Indian English, Malaysian English, Nigerian English, African American Vernacular English, Jamaican Creole English

Central variety: Irish English, New Zealand English, Jamaican English

Peripheral variety: all traditional rurally based non-standard dialects, plus a large number of ex-colonial varieties including pidgins and creoles

FIGURE 1. Mair's Theory of The World System of English (2013)

Language users of the peripheral varieties should be well-versed and at least be partially active in a number of other languages, especially those spoken by individuals at the top of the social order. For speakers from the central regions, it is common practice to adapt the pertinent super-central materials. For instance, Irish English users would adopt British English as their language references. This theory regards American English as a hyper-central language, functioning as the hub for global English.

The theory of The World System of Englishes postulates that American English has established itself as the most-central variety, having the greatest potential to impact other varieties of English. This is largely due to the global spread of American English through media - in broadcasting, films, advertisements, and newspapers (Crystal, 2003). Now, the dissemination also takes place on social media. While British English continues to be the dominant yardstick in the majority of English education systems around the world, American English is gaining traction globally, particularly in non-native countries due to the large number of native speakers and dominance over popular culture. Gilquin (2018) affirmed that the dichotomy between British and American English is propelled by three main factors: (1) education system, (2) number of native speakers, and (3) popular culture.

American popular culture is a "global juggernaut" (Crothers, 2021, p. 234), dominating movies, music, television, and now digital communication. These sources of entertainment and communication are channelled and produced in American English, which contributes to the ubiquity of American English. The population worldwide is exposed to American English on a daily basis. The presence of the global digital environment in recent years is blurring the geographic lines, causing gradual transformation towards other varieties of English. American English, being increasingly abundant on social media, especially on American-owned platforms such as Twitter, is driving the global language variation and change.

In Mair's theory of The World System of Englishes, the key indicator of the influence of American English is when lexical units from American English spread into other varieties of English. This is known as lexical Americanisms. Mair elucidates that lexical traffic or the direction of flow of lexical units will occur "downwards" instead of "upwards", which means lexical borrowings are expected to take place according to the hierarchical levels. For example, lexical units from American English (hyper-central variety) are more likely to spread into Malaysian English which is lower in hierarchy (super-central variety).

LEXICAL VARIATION AND CHANGE ON TWITTER

Social media has expanded the frequency and speed at which we could communicate; thus, language change is now more rapid than ever. Twitter as a global platform is widely utilized, and because of this the language used on Twitter is now easier to observe, disseminate, and acquire. When a public tweet is published on Twitter, it becomes available to everyone in the globe, whether through liking, replying, retweeting, or forwarding. Twitter, with its instant, real-time features, allows for the propagation of information, knowledge, communication, and ideas to take place easily which then transforms the English language. Due to the novelty of this research area, relevant studies on lexical variation and change on Twitter are elaborated in detail in this section.

With the upsurge in social media use, especially Twitter, researchers are now able to gain access to a large amount of linguistic data, which has and could fuel more research in language variation and change. One of the key features in Twitter which makes research on lexical units possible is the spatial and temporal continuity it offers (Huang et al., 2016). The development of mobile systems with Global Positioning System (GPS) has allowed tweets on Twitter to have both time annotations and spatial information. Tweets on Twitter are stored from its inception until now; therefore, it consists of a large amount of historical and real-time data. In earlier studies, tweets would be downloaded manually but the Twitter API Developer platform enables both historical and recent tweets to be mined into a corpus. This is known as corpus compilation. Corpus compilation is the act of "designing a corpus, collecting texts, encoding the corpus, assembling and storing the relevant metadata, marking up the texts where necessary and possibly adding

linguistic annotation" (McEnery & Hardie, 2012, p.241). Researchers have made use of corpus compilation using web sources in analysing lexical variation and change, because corpus analysis is the only methodology for quantitative assessments of diachronic change, and is the most popular tool for examining synchronic variation (Krug, Schlüter & Schluter, 2013).

Since tweets are geo-tagged and time-stamped, researchers have investigated lexical variation and conduct geographical analysis or the mapping of linguistic characteristics using data on Twitter. To understand research on regional patterns, we need to look back at the inception of this research area. Labov (1963) pioneered the research on regional lexical variation, particularly in phonological variation. Early research in dialect variation including Labov's (1963, 1966) has traditionally employed methods such as fieldworks by entering the community and having one-on-one interviews with informants to examine regional variations. Labov's (1966) most widely discussed study investigated sound change in New York City on the way New Yorkers use the phoneme /r/, and his subsequent research has contributed profoundly to the field of variationist sociolinguistics. However, the contribution of traditional data is not without its limitations. The main problem in collecting language data using fieldwork and interviews is the Observer's Paradox (informants modifying their speech due to the presence of an observer) and the lengthy time.

Twitter corpora serves as an alternative to combat the aforementioned issues in traditional research. According to Grieve et al. (2019), there are several advantages of using Twitter corpora, especially in conducting research for regional lexical variation. The first advantage is that corpora is easier to build using Twitter data than gathering data using fieldworks and surveys. To illustrate, before a fieldwork can take place, it is important to determine the exact linguistic variables that the researcher intends to investigate (Feagin, 2013). Twitter corpora, on the other hand, enable open-ended analysis of a much broader variety of linguistic features. In addition, Twitter corpora could eliminate Observer's Paradox and allow for investigation of language in a natural state. Another advantage of Twitter corpora is that it improves the resolution of dialect maps, enabling more informants to be sampled in more places.

More recently, Twitter data has been utilized to study lexical dialect variation (Eisenstein et al. 2014; Doyle, 2014; Jones, 2015) and to map regional dialects. Huang et al. (2016) investigated the regional variation of American English using lexical alternations i.e., variations of a particular word with the exact or similar meaning, for instance, mom/mother. By using a one-year Twitter data, they mapped lexical alternations produced by American Twitter users according to the counties in the United States through principal component analysis and regionalization methods. The research revealed unique linguistic characteristics according to the regions. Some alternations were found to be similar with certain regions while other alternations were different. People in the Northeast, for instance, preferred *bag* over *sack* and this preference is much less pronounced in the South. The word *clearly* is preferred in the East whereas *obviously* is used more frequently in the West.

To assess the generalizability of Twitter data in examining regional variations, Grieve et al. (2019) investigated lexical dialect variation, particularly lexical alternations in British Twitter. This study compared 1.8-billion-word corpus of geolocated UK Twitter data with traditional survey data from BBC Voices Project and found broad alignments between the two datasets. The findings confirmed the reliability and effectiveness of Twitter as a resource to study dialect patterns. Regional dialect mapping using Twitter data is found to be propitious compared to survey data because of the greater accuracy in the identification of regional patterns.

With each passing day, there are more new word forms making their way into the English language and these new word forms are spreading among English users. This is known as lexical

emergence. There is a dearth of research on lexical emergence because "linguists have not had access to sufficient amounts of language data with the necessary temporal resolution to track the spread of emerging word forms" (Grieve et al., 2017, p.102). Thus, to overcome this, Grieve et al. (2017) introduced a methodology to investigate emerging lexical units in American Twitter. Through calculations of relative frequency, measurement using Spearman correlation coefficient as well as concordance analysis of one-year Twitter data, 29 emerging word forms in American Twitter were revealed. Some of these new word forms include *rekt, lit, faved, on fleek, tooka, mutuals* and so on.

Following the procedures introduced in Grieve et al. (2017) to investigate lexical emergence, Grieve et al. (2019) studied lexical innovations in American Twitter by mapping the origin and diffusion of the lexical units. The research found urban regions, which are rich in culture, to be the main hotspots of lexical innovations. These regions were California, Atlanta, New York City, Washington D.C., and New Orleans. Some of these hubs of lexical innovations were also largely dominated by African American English, which is the primary source for new emerging forms in American English. This result is in line with a prominent study by Pennycook (2007) which asserts that African American culture is a powerful force in global change, especially as a tool for redefining local identities all around the world.

It is now apparent that textual data harnessed from Twitter allows for further linguistic analysis. There are several key linguistic methods which has been used in the aforementioned research to investigate lexical items from Twitter, including the distribution and trends of lexical items. Frequency distribution is frequently studied in corpus linguistics, in which the occurrence of particular lexical items in a corpus is examined. The results obtained from frequency distribution can then be inspected to observe the trends of usage of a particular lexical item in a corpus. For example, Mat Awal et al. (2021) investigated the trend of frequency of the Islamic terms *halal* and *haram* in the Malaysian Hansard Corpus and observed its patterns across 13 parliament sessions from the year 1959 until 2018. Directly relevant to this conceptual paper is the recent research by Giorgi et al. (2022) which investigated the frequency distribution of *BlackLivesMatter* - a social call against racism which has garnered attention over the years, especially on social media. The study examined the distribution of tweets on *BlackLivesMatter* across the United States for three 3-year periods: 2013 to 2015, 2016 to 2018, and 2019 to 2021.

Google Trends has also been utilized to understand the nature of data obtained online (Grieve et al., 2017; Jensen, 2017). Google Trends is used to trace specific search words or phrases either synchronically or diachronically. For example, by searching the term *selfie,* the platform reveals numerical and graphical data regarding *selfie* which can be tailored to researchers' desired region, country, categories, and duration (Figure 2). The numerical breakdown of the usage of *selfie* can be downloaded for further analysis. Google Trends also allows the generation of the geographic maps for the term *selfie* to illustrate where the query of the word is most prevalent (Figure 3).
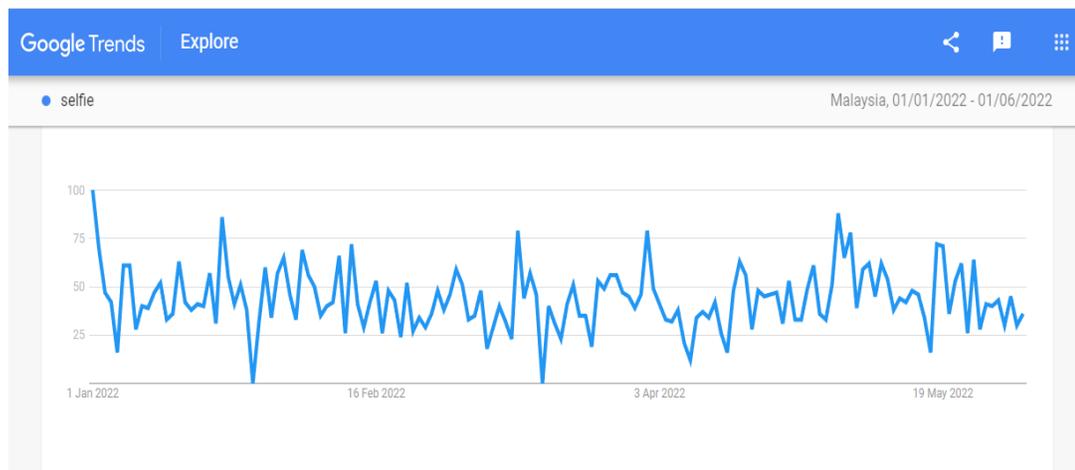
FIGURE 2. Google Trends result for the term *selfie* in Malaysia for the first six months of 2022.
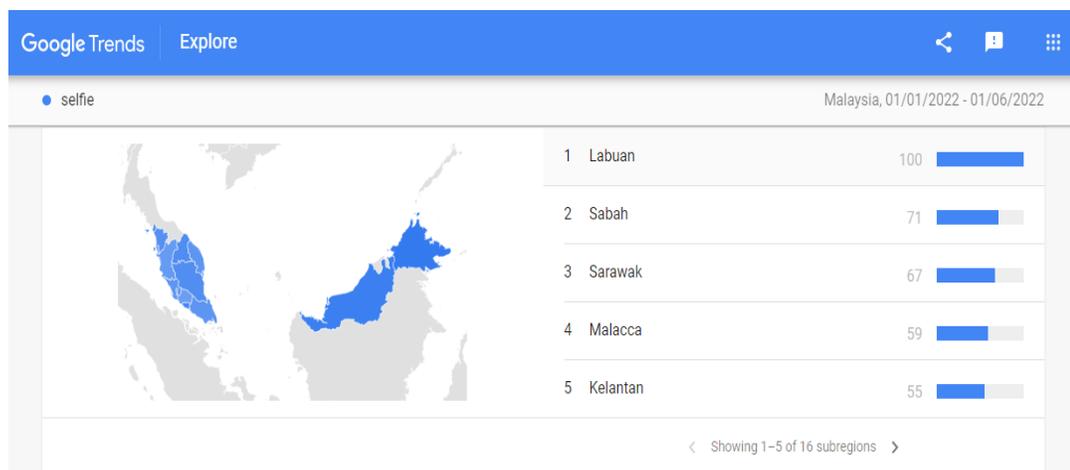Data source: Google Trends (https://www.google.com/trends).

FIGURE 3. Geographic maps generated in Google Trends for the term *selfie* in Malaysia for the first six months of 2022
according to the subregions in the country.
Data source: Google Trends (https://www.google.com/trends)

Language use on Twitter is spontaneous and recent. There are numerous foreign, new words which are now dominating the digital space. To understand the meaning and context of usage of these new words, researchers have made use of available tools, such as concordancer in corpus analysis toolkits. To understand new lexical units on Twitter, Grieve et al. (2017) used concordance analysis to understand how novel lexical items such as *famo* and *gainz* in American English were used on Twitter.

Due to the novelty of emerging lexical items, extra steps need to be taken to ensure the accuracy of meanings of the lexical items found online. Therefore, apart from concordance analysis, there are additional platforms used to further analyse novel word forms. To investigate language change on social media, webpages such as Urban Dictionary have been utilized to track the emergence of new words (Nguyen, McGillivray & Yasseri, 2018). Urban Dictionary is a

crowd-sourced online dictionary founded in 1999 and has since then been utilized to share the definitions of new word forms available online. The importance of Urban Dictionary in examining language change on Twitter is two-fold. Firstly, this user-generated site helps to track the earliest occurrence of words and phrases and secondly, Urban Dictionary enables researchers to verify the definition of terms found in Twitter corpora (Grieve et al., 2017).

Lexical units, despite its ability to stand on its own, are mostly used with other lexical units to form multi-word combinations or recurrent word sequences, with phrases consisting of at least two words or more (McEnery & Hardie, 2012). In the context of corpus studies, they are used interchangeably in literature using the following terms: *phraseological units*, *n-grams*, *multi-word lexemes*, *clusters*, *prefabricated speech or prefabs*, *fixed expressions*, *lexical bundles*, *set phrases*, *phrasemes* and *formulaic sequences* (Byrd & Coxhead, 2010; Fiedler, 2017). Two of the most-used terms are *lexical bundles* and *phraseological units*. Lexical bundles are mostly associated with academic register - in investigating research articles (Varghaei, Branch, & Khodadadi, 2022), dissertations (Narkprom & Phoocharoensil, 2022), and textbooks (Hussain, Zahra, & Abbas, 2021). According to Khayrullina and Fatkullina (2021, p. 273), phraseological units can be divided into grammatical and semantic features. Grammatical structures can be divided into communicative phrases which form sentences (*Mothers, they are like that!*), phraseological phrases (*to wither on the vine*), and unit as word forms (*So what?*). Semantic-wise, phraseological units can be classified into thematic groups according to the socio-cultural functions of different language groups (slang, socio-political, medicine etc). To illustrate, *on fleek* is commonly used in American English as a phraseological unit - *eyebrows on fleek*, and Grieve et al. (2017) investigated the generalisation of this phraseological unit in American Twitter.

THE CASE OF *LIT* AND *ON FLEEK* IN MALAYSIAN TWITTER

This paper proposes how two lexical items - *lit* and *on fleek* - can be investigated in Malaysian Twitter. These two lexical items were selected based on an earlier study by Grieve et al. (2017) which found that these lexical items emerged in American Twitter from the year 2013 onwards and has since then been circulating worldwide. Generally, *lit* refers to the past tense and past participle tense of *light*, or it means *to illuminate*. Over the years, the meaning has expanded to also describe something to be *exciting*, *good* and *intoxicating*. *On fleek*, on the other hand, is synonymous to *on point*. The study can investigate if these lexical items, which originated from American English, are influencing Malaysian English and to understand why and how these lexical items are utilised in Malaysian Twitter. The lexical items can be investigated in terms of frequency distribution, the usage of these lexical items as well as the phraseological units formed in Malaysian Twitter for a certain duration, for example, from the year 2013 (the year which these lexical items first emerged in American Twitter) until 2021 (the year prior to this paper is written). The results from such research can shed some light on the extent to which Malaysian English is influenced by American English, particularly on Twitter. The findings can be discussed according to the theory of World System of Englishes, as well as relevant insights on the impact of globalisation on lexical change and variation via digital communication.

Tweets in Malaysian Twitter containing *lit* and *on fleek* from the year 2013 until 2021 can be collected through the Twitter API Developer Platform. To comply with Twitter's Terms of Service, Academic Research access have to be obtained first before the tweets can be downloaded directly in large batches from the Tweet Downloader tool provided by Twitter. As described in Giorgi et al. (2022), since this a specialised corpus (i.e., only tweets about lit and on fleek rather

than a random sample of tweets), the following search criteria should be used: (1) keywords: lit and on fleek as the search terms; (2) country: Malaysia; and (3) tweet language: English (only). The query to be generated should be: "lit" place_country:MY lang:en. Tweets will then have to be downloaded in .csv format and will then be manually cleaned from noise (hashtags, links, and emoji).

Frequency distribution enables researchers to gain a comprehensive understanding of how lexical items are distributed in a corpus and draw conclusions from them. For example, the lexical items *lit* and *on fleek* can be investigated for its usage from the year 2013 until 2021. Frequency can be used to observe the progression of these terms in terms of its of usage over the years. To achieve this, absolute frequency (the number of times a lexical item occurs in the corpus) as well as relative frequency (the number of times a lexical item occurs in the corpus in relation to the total number of words i.e., tokens in the corpus) can be calculated, as elaborated in (Nor et al., 2019). After the calculations are performed, the data can then be visualized using a line chart. This methodology can allow the trends of usage of *lit* and *on fleek* by Malaysians to be observed across time, i.e., from the year these lexical items first appeared until present and this can help determine whether these lexical items have been consistently used by Malaysians throughout the years.

Following Grieve at al. (2017) and Jensen (2017) Google Trends can also be utilized to complement the data for trends of usage obtained on Twitter and to further understand the earliest occurrence of *lit* and *on fleek* in Malaysia. To attain this, *lit* and *on fleek* can serve as the search terms, and the following search criteria can be customised on Google Trends: (1) location: Malaysia; (2) time range: from the year 2013 until 2021; and (3) categories set to 'all' for a comprehensive result. Google Trends will then reveal the frequency of usage of these lexical items on Google during the specified time frame as well as the distribution and details of query (top and rising searches) associated to these lexical items in all the sub-regions in Malaysia (the states and federal territories in the country, i.e., the 13 states and 3 federal territories in Malaysia); all these will be visualised in geographical maps. By analysing the precise Google searches made for these lexical items, this methodology can provide support for the information gathered from Twitter and help pinpoint the regions in Malaysia that may have aided in the spread of these new word forms.

As discussed earlier, concordance helps researchers understand the ways lexical items are actually employed in a corpus. Tweets with the lexical units *lit* and *on fleek* can also be examined using the concordancer in AntConc version 4.1.1 to analyse the usage of these word forms in Malaysian Twitter. Additionally, Urban Dictionary can be used to verify and compare the usage of lexical items *lit* and *on fleek* in Malaysian Twitter with the meanings in American English. This would demonstrate whether Twitter users in Malaysia use these lexical terms similarly to American Twitter users and whether they are referring to the same meanings when they do. Similarity in the usage can suggest that Malaysian English is influenced by American English on Twitter, lending support to the Theory of World System of English. However, dissimilarity in the usage in both variety of English differ can suggest that Malaysian English has not been influenced by American English with regard to these lexical items.

The lexical items *lit* and *on fleek* are not always used individually as one lexical item. As elaborated in Grieve et al. (2017), *on fleek* is oftentimes accompanied by *eyebrows* to form the phraseological unit *eyebrows on fleek*. To compare the phraseological units between American and Malaysian English, the phraseological units for *lit* and *on fleek* in American English should first be listed based on the results on Urban Dictionary. To identify the use of these phraseological units in Malaysian Twitter data, the n-gram feature in corpus tools can be utilized. AntConc version 4.1.1 allows for the common patterns of word sequences and its frequency to be revealed through

its n-gram/cluster feature. For instance, by keying in the lexis *on fleek*, with the minimum phraseological units accepted to be two or more words as per Harris (2006) and McEnery (2011), the corpus tool would reveal the phraseological units formed using this lexical item. The data for American phraseological units and phraseological units found on Twitter can then be compared. Such an analysis would reveal the similarities and differences in terms of the phraseological units that Malaysians and Americans utilise.

It is important to note that the English language used in Malaysia is rife with code-switching, whereby language users alternate between two or more languages at once. Being a multilingual nation, there are inclusion of sequences of words adopted from other local languages in Malaysia in the English language, and this phenomenon is apparent on social media platforms such as Twitter. For example, if the primary focus of a study is the usage of *lit* and *on fleek* in Malaysian Twitter which are entirely in the English language, results should take into consideration the possibility of code switching in the corpus. This should be taken into account in the analysis and discussion. For instance, expressions such as *your kening is on fleek* by Malaysian Twitter users, with *kening* referring to *eyebrows* to replace the common phraseological unit *eyebrows on fleek* should still be taken into consideration as additional findings.

In summary, this section has propounded the methodology in analysing lexical variation and change in Malaysian Twitter. The procedures discussed here have been utilized in past research, and this conceptual paper has systematically proposed the necessary steps to investigate lexical variation and change in a particular variety of English, beginning with frequency distribution, the use of Google Trends anda concordancer, followed by using Urban Dictionary for verifications, and lastly using n-gram/cluster feature in AntConc version 4.1.1.

CONCLUSION

This conceptual paper has proposed how lexical variation and change in Malaysian English (super-central variety) can be investigated and compared to American English (hyper-central variety) to understand the extent to which Malaysian English is influenced by American English. The results from such a study can provide a general overview of the emergence of new global lexical items in Malaysian English, establish the predominant word-formational patterns as well as allow us to observe the usage of these words online. The findings can be analysed and discussed according to the theory of World System of Englishes, as well as the recent theories of globalisation and language mobility in sociolinguistics. All these can lead to theoretical contributions by illuminating the dynamics of English language variation and change at large.

The advancement of technology in propelling digital communication will most definitely result in more incoming novel lexical items, catering to either specific groups of language users, i.e., youth or English language users at large. The subject of lexical variation and change can still be explored further in the upcoming years, and it is hoped that the purported study has shed some light on this research area.

REFERENCES

Algeo, J. (1991). *Fifty Years Among New Words.* Cambridge: Cambridge University Press.

Al-Ghamdi, N. A., & Albawardi, A. H. (2020). Multivocality of Saudi COVID-19 discourse in social media posts: A socio-semiotic multimodal perspective. *GEMA Online Journal of Language Studies*, *20*(4). http://doi.org/10.17576/gema-2020-2004-13

Androutsopoulos, J. (2011). Language change and digital media: a review of conceptions and evidence. *Standard languages and language standards in a changing Europe*, *1*, 145-159.

Androutsopoulos, J. (2014). Moments of sharing: Entextualization and linguistic repertoires in social networking. *Journal of Pragmatics*, *73*, 4-18.

Arrizabalaga, B. (2021). Social Networks: A source of lexical innovation and creativity in contemporary peninsular Spanish. *Languages*, *6*(3), 138. https://doi.org/10.3390/languages6030138

Asif, M., Zhiyong, D., Iram, A., & Nisar, M. (2021). Linguistic analysis of neologism related to coronavirus (COVID-19). *Social Sciences & Humanities Open*, *4*(1), 100201.

Baskaran, L. M. (2005). *A Malaysian English primer: aspects of Malaysian English features*. University of Malaya Press.

Baskaran, L. M. (2008). Malaysian English: Morphology and syntax. In R. Mesthrie (Ed.), *Varieties of English, Volume 4: Africa, South and Southeast Asia* (pp. 610–623). Berlin/New York: Mouton de Gruyter.

Battarcharjee, D. S. (2009.) *Lexical Innovation on the Internet-Neologisms in Blogs* (Doctoral dissertation, Phd Thesis. Zurich. University of Zurich. Poland).

Blommaert, J. (2010). *The sociolinguistics of globalization*. Cambridge: Cambridge University Press.

Bolton, K., Botha, W., & Kirkpatrick, A. (Eds.). (2020). *The Handbook of Asian Englishes*. USA: John Wiley & Sons Inc.

Brian, D. (2022, January 5). *How Many People Use Twitter in 2022? [New Twitter Stats]*. https://backlinko.com/twitter-users

Byrd, P. & Coxhead, A. (2010). On the other hand: Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL, 5,* 31–64.

Canagarajah, S. (2013). *Translingual practice: Global Englishes and cosmopolitan relations.* Abingdon, UK: Routledge.

Caro, K., & Mendinueta, N. R. (2017). Lexis, lexical competence and lexical knowledge: a review. *Journal of Language Teaching & Research*, *8*(2), 205-213. https://doi.org/10.17507/jltr.0802.0

Coats, S. (2016). Grammatical feature frequencies of English on Twitter in Finland. *English in computer-mediated communication: Variation, representation, and change*, *93*, 179-209. https://doi.org/10.1515/9783110490817

Cook, C. Paul. (2010). *Exploiting Linguistic Knowledge to Infer Properties of Neologisms* (PhD. Dissertation). Toronto: University of Toronto.

Crothers, L. (2021). *Globalization and American Popular Culture*. Lanham, MD: Rowman & Littlefield.

Crystal, D. (2003). *Language and the Internet.* Cambridge: Cambridge University Press.

Dovchin, S. (2018). *Language, media and globalization in the periphery: The linguascapes of popular music in Mongolia*. New York, NY: Routledge.

Doyle, G. (2014, April). Mapping dialectal variation by querying social media. In *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics* (pp. 98-106).

Eisenstein, J., O'Connor, B., Smith, N. A., & Xing, E. P. (2014). Diffusion of lexical change in social media. *PloS one*, *9*(11), e113114.

Endarto, I. T. (2020). A corpus-based lexical analysis of Indonesian English as a new variety. *Indonesian Journal of Applied Linguistics*, *10*(1), 95-106. https://doi.org/10.17509/ijal.v10i1.24993

Feagin, C. (2013) Entering the community fieldwork. In J.K. Chambers & N. Schilling-Estes (Eds.), *The Handbook of Language Variation and Change* (2nd ed., pp. 1–13). Oxford: Blackwell.

Fiedler, S. (2017). Phraseological borrowing from English into German: Cultural and pragmatic implications. *Journal of Pragmatics*, *113*, 89-102. https://doi.org/10.1016/j.pragma.2017.03.002

Fontaine, L. (2017). The early semantics of the neologism BREXIT: A lexicogrammatical approach. *Functional Linguistics*, *4*(1), 1-15.

Geeraerts, D. (1994). Varieties of lexical variation. In W. Martin, W. Meijs, M. Moerland, E. ten Pas, P. van Sterkenburg & P. Vossen (Eds.), Proceedings of the 6th Euralex International Congress on Lexicography, Amsterdam, 78–84.

Geeraerts, D. (2010). *Theories of lexical semantics*. Oxford: Oxford University Press.

Giddens, A. (1990). The Consequences of Modernity. Stanford University Press: Stanford, CT.

Gilquin, G. (2018). American and/or British influence on L2 Englishes–Does context tip the scale (s). In S, C. Deshors (ed.), *Modeling World Englishes: Assessing the interplay of emancipation and globalization of ESL varieties*, (pp. 187-216.) Amsterdam: John Benjamins.

Giorgi, S., Guntuku, S. C., Himelein-Wachowiak, M., Kwarteng, A., Hwang, S., Rahman, M., & Curtis, B. (2022, May). Twitter Corpus of the #BlackLivesMatter Movement and Counter Protests: 2013 to 2021. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 16, pp. 1228-1235).

Gleason, B. (2018). Adolescents becoming feminist on Twitter: New literacies practices, commitments, and identity work. *Journal of Adolescent & Adult Literacy*, *62*(3), 281-289.

Graddol, D. (2000) *The Future of English. A guide to forecasting the popularity of the English language in the 21st century*. (2nd ed.). London: The British Council.

Grieve, J., Nini, A. & Guo, D. (2017). Analyzing lexical emergence in Modern American English online. *English Language and Linguistics 21*(1), 99–127. https://doi.org/10.1017/S1360674316000113

Grieve, J., Nini, A. & Guo, D. (2018). Mapping lexical innovation on American social media. *Journal of English Linguistics 46*(4): 293–319.

Grieve, J., Montgomery, C., Nini, A., Murakami, A & Guo, D. (2019). Mapping lexical dialect variation in British English using Twitter. *Frontiers in Artificial Intelligence 2,* 1-18.

Harris, A. (2006). Revisiting anaphoric islands. *Language 82*(1), 114–30.

Hamat, A., Embi, M. A., & Hassan, H. A. (2012). The use of social networking sites among Malaysian university students. *International Education Studies*, *5*(3), 56-66.

Hashim, A. (2020). Malaysian English. In Bolton, K., Botha, W., & Kirkpatrick, A. (Eds.). *The Handbook of Asian Englishes*, (p.373-397.) USA: John Wiley & Sons Inc.

Herring, Susan C., Androutsopoulos, J. (2015). Computer-mediated discourse 2.0. In Schiffrin, Deborah, Tannen, Deborah and Hamilton, Heidi, E. (Eds). *The handbook of discourse analysis* (2nd ed., 127-151), Chichester, UK: John Wiley & Sons.

Ho A. I., Hashim, R. S., Jaludin, A., & Ismail, R. (2011). Enhancing opportunities for language use through web-based social networking. *International Conference on Social Science and Humanity, 5,* 136-139.

Huang, Y., Guo, D., Kasakoff, A., & Grieve, J. (2016). Understanding US regional linguistic variation with Twitter data analysis. *Computers, environment and urban systems*, *59*, 244-255.

Hussain, G., Zahra, T., & Abbas, A. (2021). Discourse functions of lexical bundles in Pakistani Chemistry and Physics textbooks. *GEMA Online Journal of Language Studies*, *21*(1). http://doi.org/10.17576/gema-2021-2101-13

Izazi, Z. Z., & Tengku-Sepora, T. M. (2020). Slangs on social media: Variations among Malay language users on Twitter. *Pertanika Journal of Social Sciences & Humanities*, *28*(1), 17-34.

Jeffries, L., & McIntyre, D. (2019). The Devil Has All the Best Tunes: An Investigation of the Lexical Phenomenon of Brexit. In R. Page, B. Busse, & N. Nørgaard (Eds.), *Rethinking Language, Text and Context: Interdisciplinary Research in Stylistics in Honour of Michael Toolan* (1st ed., pp. 103–123). Routledge.

Jensen, M. M. (2017). What can Google Trends data tell us about dialect labels: An exploratory study. *Globe: A Journal of Language, Culture and Communication*, *5*, 48-76.

Jones, T. (2015). Toward a description of African American vernacular english dialect regions using "Black Twitter". *American Speech*, *90*(4), 403-440. https://doi.org/10.1215/00031283-3442117

Kachru, B. (1985). Standards, codification and sociolinguistic realism: The English language in the Outer Circle. In Randolph Quirk & Henry G. Widdowson (eds.), *English in the world*, (11–30). Cambridge: Cambridge University Press.

Kashinathan, S., & Aziz, A. A. (2021). ESL learners' challenges in speaking English in Malaysian classroom. *International Journal of Academic Research in Progressive Education and Development, 10*(2), 983–991.

Khayrullina, R. & Fatkullina, F. (2021). The Latest Phraseology in Internet Discourse. In Arsenteva, E. (Ed.), *The Discoursal Use of Phraseological Units* (pp. 264-283). Cambridge: Cambridge Scholars Publishing.

Krug M., Schlüter, J. & Schluter, J. (eds.). (2013). *Research Methods in Language Variation and Change.* Cambridge: Cambridge University Press.

Kunalan, R., Mutty, B., & Francis, K. (2021). Lexical borrowing in Malaysian online newspaper editorials: A corpus-based approach. *International Journal of Education, Psychology and Counseling, 6*(40), 93-98. https://doi.org/10.35631/IJEPC.640007

Labov, W. (1963) The social motivation of a sound change. *Word 19*, 273–309.

Labov, W. (1966). *The Social Stratification of English in New York City.* Cambridge: Cambridge University Press.

Lalić-Krstin, G., & Silaški, N. (2018). From Brexit to Bregret: An account of some Brexit-induced neologisms in English. *English Today*, *34*(2), 3-8.

Lantz-Andersson, A. (2018). Language play in a second language: Social media as contexts for emerging Sociopragmatic competence. *Education and Information Technologies*, *23*(2), 705-724.

Lewis, M., Gough, C., Martínez, R., Powell, M., Marks, J., Woolard, G. C., & Ribisch, K. H. (1997). *Implementing the lexical approach: Putting theory into practice* (Vol. 3, No. 1, pp. 223-232). Hove: Language Teaching Publications.

Mair, C. (2013). The World System of Englishes: Accounting for the transnational importance of mobile and mediated vernaculars. *English World-Wide, 34*(3): 253–278. https://doi.org/10.1075/eww.34.3.01mai

Martin, E. (2019). World Englishes in the Media. In Cecil L. Nelson, Zoya G. Proshina, and Daniel R. Davis (Eds.) *The Handbook of World Englishes* (2nd ed., pp. 597-615). John Wiley & Sons.

Mat Awal, N., Bakar, K. A., Abdul Rahman, A. N. C., & Abdullah, I. H. (2021). Representasi halal dan haram dalam wacana parlimen Malaysia. *GEMA Online Journal of Language Studies*, *21*(4). http://doi.org/10.17576/gema-2021-2104-10

McEnery, T., & Hardie, A. (2011). *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press.

McEnery, T., & Hardie, A. (2013). The history of corpus linguistics. In Keith Allan (ed.), *The Oxford handbook of the history of linguistics* (pp. 727–745). Oxford: Oxford University Press

Miller, D. G. (2014). *English lexicogenesis*. Oxford: Oxford University Press.

Moody A. (2020) The authority and authenticity of performative Englishes in popular culture. *World Englishes. 1–21*. https://doi.org/10.1111/weng.12486

Narkprom, N., & Phoocharoensil, S. (2022). Lexical bundles in native English speakers' and Thai writers' dissertations. *GEMA Online Journal of Language Studies, 22*(3). http://doi.org/10.17576/gema-2022-2203-03

Newmark, P. (1988). *A Textbook of Translation.* New York: Prentice Hall.

Nguyen, D., McGillivray, B., & Yasseri, T. (2018). Emo, love and god: making sense of Urban Dictionary, a crowd-sourced online dictionary. *Royal Society Open Science*, *5*(5), 172320.

Nor, N. F. M., Abdul Rahman, A. A. N. C., Jaluddin, A., Ho Abdullah, I., & Tiun, S. (2019). A corpus-driven analysis of representations around the word 'ekonomi' in Malaysian Hansard Corpus. *Gema Online Journal of Language Studies*, *19*(4). 9 http://doi.org/10.17576/gema-2019-1904-04

Pennycook, A. (2007). *Global Englishes and transcultural flows*. London, UK: Routledge.

Pillai, S. (2017). A study of the use of English among undergraduates in Malaysia and Singapore. *SARE: Southeast Asian Review of English*, *48*(1), 19-38.

Romiszowski, A., & Mason, R. (2013). Computer-mediated communication. In *Handbook of research on educational communications and technology* (pp. 402-436). Routledge.

Rosell-Aguilar, Fernando. (2018). ―Twitter as a Formal and Informal Language Learning Tool: From Potential to Evidence. *In Innovative Language Teaching and Learning at University: Integrating Informal Learning into Formal Language Education*, edited by Fernando Rosell-Aguilar, Tita Beaven, and Mara Fuertes Gutiérrez, 99–106. Retrieved from https://research-publishing.net/manuscript?10.14705/rpnet. 2018.22.780.

Rusli, M. F., Aziz, M. A., Aris, S. R. S., Jasri, N. A., & Maskat, R. (2018). Understanding Malaysian English (Manglish) jargon in social media. *Journal of Fundamental and Applied Sciences*, *10*(2S), 116-125.

Salehuddin, K. (2022). Exploring three idealised cognitive models in *The Kampung Boy* and *Budak Kampung. 3L: The Southeast Asian Journal of English Language Studies, 28*(2), 140-159. http://doi.org/10.17576/3L-2022-2802-10

Santoso, D. H., Aziz, J., Utari, P., & Kartono, D. T. (2020). Populism in new media: The online presidential campaign discourse in Indonesia. *Gema Online Journal of Language Studies*, *20*(2). http://doi.org/10.17576/gema-2020-2002-07

Schmied, J. (2012). Social digital discourse: new challenges for corpus-and sociolinguistics. *Approaches to Text and Discourse Analysis*, *10*, 43-56.

Shamsuddin, W. N. F. W., Abdullah, A. N., Tan, H., & Bolong, J. (2019). Attitudes and driving factors behind the attitudes of Malaysians university students towards British English and American English. *International Journal of Humanities, Philosophy, Language, 2* (5), 1-12.

Sharma, B. K. (2012). Beyond social networking: Performing global Englishes in Facebook by college youth in Nepal. *Journal of Sociolinguistics*, *16*(4), 483-509.

Taboada, M. (2016). Sentiment analysis: an overview from linguistics. *Annu. Rev. Linguist*, *2*, 325–347.

Tankosić, A., & Dovchin, S. (2021). The impact of social media in the sociolinguistic practices of the peripheral post-socialist contexts. *International Journal of Multilingualism*, 1-22.

Thirusanku, J., & Yunus, M. M. (2012). The many faces of Malaysian English. *International Scholarly Research Network, 2012*, 1-14. https://doi.org/10.5402/2012/138928

Tseng, A., & Hinrichs, L. (2020). Mobility and the English Language. In *The Handbook of English Linguistics*, (pp. 637-652). John Wiley & Sons Ltd.

Varghaei, E & Khodadadi, G. (2022). Comparing lexical bundles in medical research article abstracts of Iranian and foreign journals. *GEMA Online Journal of Language Studies, 22*(3). http://doi.org/10.17576/gema-2022-2203-05

Würschinger, Q. (2021). Social Networks of Lexical Innovation. Investigating the Social Dynamics of Diffusion of Neologisms on Twitter. *Front. Artif. Intell.* https://doi.org/10.3389/frai.2021.648583

Yule, G. (1985). *The study of language*: 2nd edition. Cambridge: Cambridge University Press.

Yunus, M. M., Zakaria, S., & Suliman, A. (2019). The potential use of social media on Malaysian primary students to improve writing. *International Journal of Education and Practice*, *7*(4), 450-458.

ABOUT THE AUTHORS

Noraishah Gulnazir is a MPhil candidate at the Centre for Language and Linguistics, Universiti Kebangsaan Malaysia. Her research interests include corpus linguistics, language variation and change, as well as globalisation of language.

Khazriyati Salehuddin (Ph.D) is an Associate Professor and a Psycholinguist at the Faculty of Social Sciences & Humanities, Universiti Kebangsaan Malaysia. She is currently the Editor-in-Chief of *GEMA Online® Journal of Language Studies*. For more information: goo.gl/rskBFV