Uncovering Factors Influencing Millennials' Use of Non-Standard Words in Twitter

NUR NASHATUL NASUHA NAZMAN
Pahi Secondary School, Kelantan

SU-HIE TING
KEE-MAN CHUAH
Universiti Malaysia Sarawak

ABSTRACT

Social media communication has its own language features and one aspect is modified spelling of standard words. Social media users use shortened words with full awareness of the meanings, and new non-standard words are constantly added to the repertoire of social media language. A pertinent question is whether social media users learn these non-standard words to use or whether they also contribute to the vocabulary used in social media communication. The study examined Malaysian millennials' use of non-standard words in Twitter and their reasons for shortening words. For the non-standard words, data were collected from 200 active Twitter users whereas data on reasons for shortening words were collected from 30 users. The results showed that the Malaysian millennials frequently used non-standard spelling of words. The three top words were "*ni*" (this), "*nak*" (want), and X (negation). The main reasons for the Twitter users to shorten words were the 280-character limit per tweet, user convenience, and characteristics of words. The Twitter users felt free to create new spellings of standard words at times for fun, but most of the time, they use the common non-standard words. The Malaysian millennials reported that they were inclined to shorten long and complex words, and words with many vowels. The study suggests that Twitter users balance between speed in communication and preservation of meaning when using non-standard words.

Keywords: *Social media, Twitter, non-standard words, word formation processes, shortening of words.*

INTRODUCTION

In the digital era, social media has become a powerful medium of communication globally, extending into remote areas wherever there is connectivity. As of December 2021, there are 4.95 billion internet users in the world, of which 29.03 million are in Malaysia (Statista, 2022). In other words, 88.77% of the 32.7 million Malaysian population are internet users in the year 2021. Social media platforms such as WhatsApp, Twitter, Facebook, and Instagram have made relaying of information almost instantaneous. As a result, letter writing and emails are now only relevant for official purposes of communication. Social media is gradually taking the place of face-to-face communication as well. However, social media are used not only for personal purposes, but also for increasing sales, building reputation or even for creating awareness about current issues (Bisera & Elena, 2018). For example, over the past decade, public figures, companies and organisations have turned to Twitter as a tool for rapid dissemination of information with a similar impact of word-of-mouth method (Jansen et al., 2009).

The language in social media communication warrants investigation because the creation of new words happens at a faster pace in digital communication than in face-to-face communication. In social media communication, the spelling of words has been intentionally

distorted and the words are often shortened versions of standard words (Choi et al., 2013; Han & Baldwin, 2011; Sriram et al., 2010). The shortening of words has been attributed to the character limit set for messages or the need to save time when sending messages (Ahmed, 2015). In addition to the formation of new words, existing words are also used to convey new meanings (Reed, 2014). The influence of social media is pervasive because words which have acquired new meanings in social media context eventually spill over into verbal communication (Foster, 2018). For example, the word "wall" conventionally refers to building walls, but nowadays "wall" refers to the homepage or profile of social media. In addition, social media has also introduced many non-standard words such as "unfriend", "selfie" and "emoji". The fast pace in which non-standard words emerge in social media is due to the openness of social media users to language change (Choi et al., 2013). Non-standard words can be defined as those word tokens which do not have a dictionary entry, and cannot be pronounced using the usual letter-to-phoneme conversion rules (Flint et al., 2017). The use of non-standard words and incomplete sentences give the impression of informality in communication. Despite the difficulty of analysing informal language styles, researchers such as Crystal (2011) and Schmied (2012) believe that it is an important area of research to attain a better understanding of language development in the digital era.

Thus far, researchers have concentrated on netspeak in English (e.g., Liu et al., 2011; Lusetti et al., 2018). Netspeak features include onomatopoeic spellings, phonetic replacements, shortening of words, and inanities (Izazi & Tengku-Sepora, 2020). In addition to English, Malay is also an important language for social media communication, particularly among users from Malaysia, Indonesia and Brunei. Malay is the fourth most popular language used on Twitter (Hong et al., 2011), but netspeak features of Malay is less understood compared to English. In Malay, the phenomenon of words acquiring new meanings in social media context is also happening. In Malay, the word "*membawang*" (created from the word "*bawang*" or "onions") is gaining popularity on social media and the usage is spilling into face-to-face communication. "*Membawang*" is replacing the standard word "*mulut puaka*" used to describe a person who has inclination to whine and rebuke others (Haroon, 2021). Even though some words may gain popularity over time in social media, other words have lost their popularity such as "lit" and "Gucci" (Fontein & Copp, 2018). The word "lit" refers to a person being intoxicated and hyped up on a situation while the word "Gucci" refers to something or someone who is cool or good. Although the netspeak expressions are out there for users to adopt, the preferences of social media users determine whether certain expressions gain popularity or lose traction over time.

Thus far, researchers investigating non-standard spelling of words obtained their data from the social media (e.g., Ariffin & Tiun, 2018; Kern et al., 2016, Stone & Wang, 2019) and have formulated some rules to explain the formation of non-standard words (e.g., Chekima & Rayner, 2017; Nkhata & Jimaima, 2020). However, there is a dearth of studies investigating use of non-standard words in social media communication from the users' perspective. Investigating their reasons of using non-standard words will reveal whether social media users are passive followers or active initiators of word formation processes. About 80% of Twitter users are in the 18 to 35 age group (Ortiz-Ospina, 2019). Therefore, it is vital to investigate why young Twitter users use non-standard words in tweets, whether it is to express new meanings, look trendy, save time or other yet-to-be uncovered reasons.

The study aims to describe the Malaysian millennials' use of non-standard words in Twitter and their reasons for shortening words. In this paper, words with non-standard

spelling are referred to as non-standard words, and the Twitter posts written in Malay are referred to as Malay tweets.

LITERATURE REVIEW

Communication via Twitter as Microblogging Platform

The popularity of Twitter has grown exponentially in recent years. There are about 6,000 tweets sent per second, amounting to 500 million tweets per day globally (Sayce, 2022). One key characteristic of tweets is its brevity. When Twitter was first launched, users could share their daily activities or opinion in a text-based format of not more than 140 characters, but this was later increased to 280 characters per tweet (Isaac, 2017). Tweets are mainly composed of text elements, but images and videos can be attached (Gligorić et al., 2018; Jaidka et al., 2019).

Twitter users are predominantly from the younger age groups as they prefer rapid and concise communication. Twitter was second in popularity in the year 2020 in Malaysia, accounting for 2.33% of active social media users (StatCounter Global Stats, 2020). A break-down by age group is not available for Twitter usage in Malaysia, but the global statistics give an indication. As of April 2021, about 62.5% of Twitter users were young (24% below 24 years old and 38.5% aged 25-34) (Dixon, 2022).

The Twitter platform offers live and real-time communication among users who share their thoughts on issues and events, and plays an important role to connect people (Golbeck, 2016). Twitter conversations take place seamlessly because each tweet can be followed easily, and the interactions are restricted to the user's network of friends or followers (Honeycutt & Herring, 2009). In platforms like Facebook, users have to add someone personally as friends, but in Twitter, this is not required and users therefore interact with one another more openly. Twitter was initially a personal social networking platform, but it has metamorphosised into an information sharing and product marketing powerhouse. Twitter has a new-found role as a source of content instead of merely a platform for socialisation purposes (Hughes et al., 2012). The Twitter updates became so popular that other social platforms such as Friendster and Facebook began copying the idea.

Twitter has a minimalist concept in that it only provides users with features they frequently use. It focuses on what matters the most, and despite Twitter being mostly text based, its readership is huge (Mitchell et al., 2013). The extended entities object in Twitter allows users to insert other media files such as images, videos or audio. The inserted media are auto transformed into links which do not load in the condensed view. Therefore, users still focus on the composed texts in each Tweet rather than the media files.

The text-based nature of Twitter and the 280-character limit has prompted many Twitter users to be succinct in their communication, and to be creative in coining new words to fit into the character limit. Twitter has popularised the use of hashtag, that is, using the hash symbol (#) to tag tweets to make them easily searchable (Seward, 2013). Twitter users communicate using a unique range of expressions ranging from transforming common words to the popularity of "Twittersphere's slangs and jargons" (Jadika et al., 2019). Words like hashtags, retweet (RT), tweet famous and so on made its way into the web community through Twitter (Dunlap & Lowenthal, 2009). Insights into why certain expressions gain popularity and others lose traction over time can be obtained by studying Twitter users' reasons for using words with non-standard spelling in their tweets.

Research on Twitter Communication in Malay

Studies on tweets written in Malay have found similarities in frequently used words with non-standard spelling. For example, "*nak*" ("*hendak*" for want) was found to rank second and fourth by Saloot et al. (2014) and Athirah et al. (2019) respectively. In addition, "*yg*" ("*yang*" for that is) ranked fourth and second in their Malay tweet corpus respectively. The words "*je*" (for "*sahaja*" meaning only) and "*ni*" ("*ini*" for this) are also commonly used short forms in the food tweets in Athirah's et al. (2019) study as well as in the Facebook and Twitter movie, politics and product review corpus in Chekima and Alfred's (2017) study.

Based on the preliminary research on formation of non-standard words in Malay social media communication, abbreviation, acronym or initialism are common processes and these terms are often used interchangeably. Nazman et al. (2020) found that non-standard words in Malay tweets are formed by deleting or changing vowels and consonants, and may involve combinations of shortened parts of two words. The process involving deletion of parts of a word is referred to as clipping by Nkhata and Jimaima (2020). Other researchers identified word formation processes such as conversion (Suparta et al., 2016) and derivation (Nisa, 2016). Suparta et al.'s (2016) involving a noun/verb word class change. Athirah et al. (2019) listed 16 word formation rules, but they cannot explain all the newly-formed words in Malay social media discourse. For instance, for a word ending with "a" (e.g., "*apa*" meaning what), the rule is to replace it with "e" ("*ape*"). While it can explain why social media users write "*siapa*" (who) as "*siape*", but it cannot explain why users do not write "*juga*" (also) as to "*juge*", but instead shorten it to "*gak*". Thus far, research on reasons for the usage of non-standard words in Malay social media discourse has not been studied from the Twitter users' perspective.

METHODOLOGY

The descriptive study on social media users' use of non-standard words in Twitter and their reasons for shortening words involved two groups of participants.

Texts and Participants

The data on non-standard Malay words in tweets were obtained from 200 Malaysian millennials, who were active Twitter users with a public profile. An active Twitter user is one who follows at least 30 accounts and is followed in return by at least one-third of those accounts. Another criterion used for selecting participants was their frequency of posting, that is, at least one episode of tweeting per month. Twitter users with less than 30 Malay tweets were excluded. Altogether 2,000 tweets amounting to 25,679 words were collected from the 200 Twitter users (average of 10 tweets per user).

The questionnaire data on reasons for shortening words were obtained from 30 Malaysian millennials aged 19-35 who were active users of Twitter. More of the participants were female (67%) and almost all of them were in their twenties (Table 1). Fugard and Potts (2015) state that qualitative studies require a minimum sample size of at least 12 to reach data saturation. Almost half (46.7%) of the participants accessed Twitter every day, 30% accessed Twitter three to four times per week, and 23.3% accessed Twitter only once a week.

Table 1: Demographic background of participants for the interview (n=30)

Demographic Variables		n	%
Gender	Male	10	33%
	Female	20	67%
Age (Years)	19-21	2	6.7%
	22-24	15	50%
	25-27	12	40%
	>27	1	3.3%
Frequency of accessing Twitter	Everyday	14	46.7%
	3-4 times per week	9	30%
	Once a week	7	23.3%

Instrument

To elicit the millennials' reasons for shortening words, a questionnaire was used. The first section contained questions on demographic background (gender, age, frequency of accessing Twitter). The second section presented participants with a list of 20 words with non-standard spelling frequently used in Twitter for them to indicate which word they had used when tweeting. The final section of the questionnaire had five open-ended questions:

1) When you post a tweet, do you think about how to write it?;
2) How do you decide which word should be shortened or coined?;
3) Are there any reasons for using some of these words (non-standard words)?;
4) Do you use the same non-standard words on other social media? (Instagram, Facebook, etc); and
5) Do you use different non-standard word for a same word at the different time? If yes, why?

Data Collection and Analysis Procedures

For the data on non-standard Malay words, the tweets written by the Malaysian millennials were copied and pasted into a Microsoft Word document. The data was filtered by removing extra characters such as hashtags or symbols. Then the data was inputted into the AntConc textual analysis software to generate a list of words based on frequency count of their stem forms. The top 10 words were identified for analysis of whether there is a meaning change, with reference to an established Malay dictionary, Kamus Dewan (2007).

For the data collection on reasons for shortening words in tweets, the first researcher used the direct message feature on Twitter to invite potential participants in their twenties to participate in the study. She explained the aim of the research, voluntary participation, and assured them of the anonymity of their identity. Those who consented to participate in the study were given the link to the Google form.

The participants' responses to questions on their use of non-standard words were analysed using thematic analysis. The preliminary themes were user preference, habit, and Twitter restriction on number of characters. After several rounds of recoding the tweets, the final main themes for the Malaysian millennials' reasons for using shortened words were context of tweets, character limit, habit, and fun. Concept maps are used to portray themes and sub-themes and the relationship among them.

RESULTS

This section presents the results on the top ten frequently used non-standard words in tweets written by Malaysian millennials as well as the results on their reasons for shortening words when tweeting.

Frequently Used Non-Standard Words in Tweets

Altogether, there were 11,761 non-standard words (or 45.8%) in the 25,679-word corpus. This shows that almost half of the words in the Malaysian millennials' tweets were shortened. People who are unable to use the non-standard words when tweeting would appear out of place as shortening words seem to be the usual practice.

Table 2 shows the top 10 non-standard words in tweets written by Malaysian millennials, which add up to 1471 or 12.5% of 11,761 non-standard words in the dataset. However, when computed out of the top 100 non-standard words, the top 10 words account for 49.8%, showing that these 10 words carry a great part of the meaning in the communication. Further analysis by word class revealed that most were determiners and adverbs, and there was only one verb "*nak*" (want) and one first person pronoun "*ak*" or "*aq*" (I). The millennials hardly used the second person pronoun, and when they do, they used "*ko*" for "*engkau*" (colloquial you), which is ranked 38th. The formal forms ("*anda*", "*awak*") are inappropriate in the Twitter environment. The comparison of the frequencies for the first and second person pronouns indicate that the tweets were more about self-expressions of ideas and feelings, and the frequent use of "*nak*" (want) is confirmatory.

Table 2: Top 10 non-standard words in tweets written by Malaysian millennials

Non-standard word	Standard spelling	Word class	Meaning in the Dictionary	Retention of dictionary meaning	Frequency	%
1. *Ni*	*Ini*	Determiner	*Utk menunjuk waktu, tempat atau benda yg dekat* (to show time, destination and nearby object)	Yes	905	30.6
2. *Nk*	*Hendak*	Verb	*Ingin mempunyai, mendapat atau memperoleh sesuatu* (to have, to get)	Yes	102	3.5
3. *X*	Tidak	Determiner	tiada; bukan (no)	No. X is just a letter of an alphabet	86	2.9
4. *Nie*	*Ini*	Determiner	*Utk menunjuk waktu, tempat atau benda yg dekat* (to show time, destination and nearby object)	Yes	73	2.5
5. *Je*	*Sahaja*	Adverb/ adjective	*Cuma, hanya* (only)	Yes	71	2.4
6. *Aq*	*Aku*	Pronoun	*kata ganti nama diri pertama* (first person pronoun)	Yes	51	1.7
7. *Ak*	*Aku*	Pronoun	*kata ganti nama diri pertama* (first person pronoun)	Yes	47	1.6

8. *Ja*	*Sahaja*	Adverb/ adjective	*Cuma, hanya* (only)	Yes	46	1.6
9. *Da*	*Sudah*	Adverb/ adjective	*Telah siap sedia; selesai; habis; tamat* (ready, done, the end)	Yes	46	1.6
10. *Lak*	*Pula*	Adverb	*Berulang lagi; sekali lagi* (again/well/neither)	Yes when it is a shorter version of "pula". No when it is used as a question tag. For example, "*mana lak*?" (where is it?)	44	1.5
TOTAL					1,471	49.8* 12.5**

Note: The word classes are based on the Malay dictionary, Kamus Dewan (2007).
* This percentage is computed out of the total of the top 100 non-standard words (n=1,471)
**This percentage is computed out of the grand total of non-standard words (N=11,761)

The top two words were "*ni*" (this) and "*nk*" (want), shortened from "*ini*" and "*nak*" respectively. These two words account for 30.6% and 3.5% of the total number of 100 non-standard words in Malay tweets (n=2,955). The two top ranked words, "*ini*" and "*nak*", retain the dictionary meanings. "*Ni*" is also spelt as "*nie*" by some Twitter users but this spelling is not as popular because it ranks fourth (2.5%) compared to "*ni*" which is number one. Two other words appear in the top 10 non-standard words in the tweets in two forms, that is, "*sahaja*" (only) appearing as "*je*" and "*ja*" (ranked 5th and 8th, 2.4% and 1.6% respectively) and "*aku*" (I) appearing as "*aq*" and "*ak*" (ranked 6th and 7th, 1.7% and 1.6% respectively). The last two words in the top 10 list are "*da*" for "*sudah*" (already) and "*lak*" for "*pula*" (again/well/neither).

The meaning of "*pula*" varies with context. It is shortened to "lak" (ranked 10th) and "la" (ranked 14th). When *lak*" is used for emphasis, it retains the dictionary meaning. For example, "*mana lak semua orang*?" (where is everyone?) and "*ada pokok lak sini*" (there is a tree here). However, it needs to be noted that some Twitter users use "*la*" as a shortened version of another word, "*itulah*" (that's the reason). Table 2 shows that all the top 10 shortened words retain the original meaning, except for "X". The 24th letter of the alphabet is used with the meaning of negation. The analysis of the top 10 non-standard words in Malay tweets show that meanings are generally retained, and Twitter users with a shared sociocultural background can figure out the meaning from the context.

Reasons for Using Non-Standard Words in Tweets
The results showed that the millennial Twitter users deliberately chose to use words with non-standard spelling, and the usage is not spontaneous. Out of 30 participants, 93.33% (or 28) reported they spent time to consider what to compose in their tweet based on certain reasons. Figure 1 shows that the main reasons for millennial Twitter users to shorten words in tweets were Twitter restriction, user convenience, and characteristics of words.

Main Reasons for Millennial Twitter Users to Shorten Words in Tweets
A total of 12 participants mentioned Twitter restriction: nine mentioned character limit of up to 280 characters per tweet, and three mentioned word limit. For example, Participant 4 said "No specific reason I guess. It's just that writing in full doesn't seem to fit into Twitter's character limit". Participant 21 stated that he used shortened words "to compensate tweet character limit", while Participant 16 stated that "Twitter only can type a short 97 sentence only". Because of the limit on number of characters, social media users used non-standard words more frequently in tweets than in other social media platforms like Facebook. In fact, seven participants specifically mentioned that they only use non-standard spellings on Twitter. They wanted their message to be complete in one tweet, which is why they had to figure out ways to cut down on the words to fit into one tweet as shown in the following excerpts:
- "I shortened when there are not enough characters to use" (Participant 16)
- "For some tweets, if there are words that I can cut down further, then I will use different style of the same word because of the limit" (Participant 29)

Among the social media platforms, it seems that the character limit on Twitter may spearhead the use of non-standard words which may be carried over into other platforms.

The next reason for participants to use non-standard words is user convenience, expressed in different ways by the millennial users: lazy to write (11 participants), saves time (4 participants), easier to type (3 participants), and easier to read (3 participants). Sometimes participant had more than one reason for shortening words. The tweets are written in an informal language style that reflect spoken communication. The participants felt lazy to write long words, because it slowed down the speed of their texting and, in the process, they were also inclined to make spelling mistakes. Shortened versions are definitely shorter than the standard words. An example is "*takyah*" (no need), and the full form is "*tidak payah*". Undoubtedly, using shortened words speeds up the composing of tweets. Another example is "*ini*" (this), written as "*ni*" in the tweets and also spoken in this way in face-to-face conversations. According to Participant 15, "When I'm in rush, I used shortened words" and Participant 22 stated "I rarely shorten my words except when I am in a rush". Most of these user convenience reasons were from the perspective of writing the message and only one participant talked about the ease of reading when the message contains shortened words. Participant 2 said, "Easy to understand and read". Participant 24 gave more insight on the ease of reading when she said "Easy for me and other to read a short one compared to read long words like an essay". These participants were definitely seasoned users of Twitter, which is why they have no issue with decoding of words with non-standard spelling. There is so far no study on whether it is easier and faster to read messages written in words with standard spelling or non-standard spelling. However, typing shortened words definitely saves time. Participant 14 made the time factor very clear when she wrote her reason in the questionnaire, "No particular reason, just to write faster and avoid any fatigue if it concerns on time span of long writing".
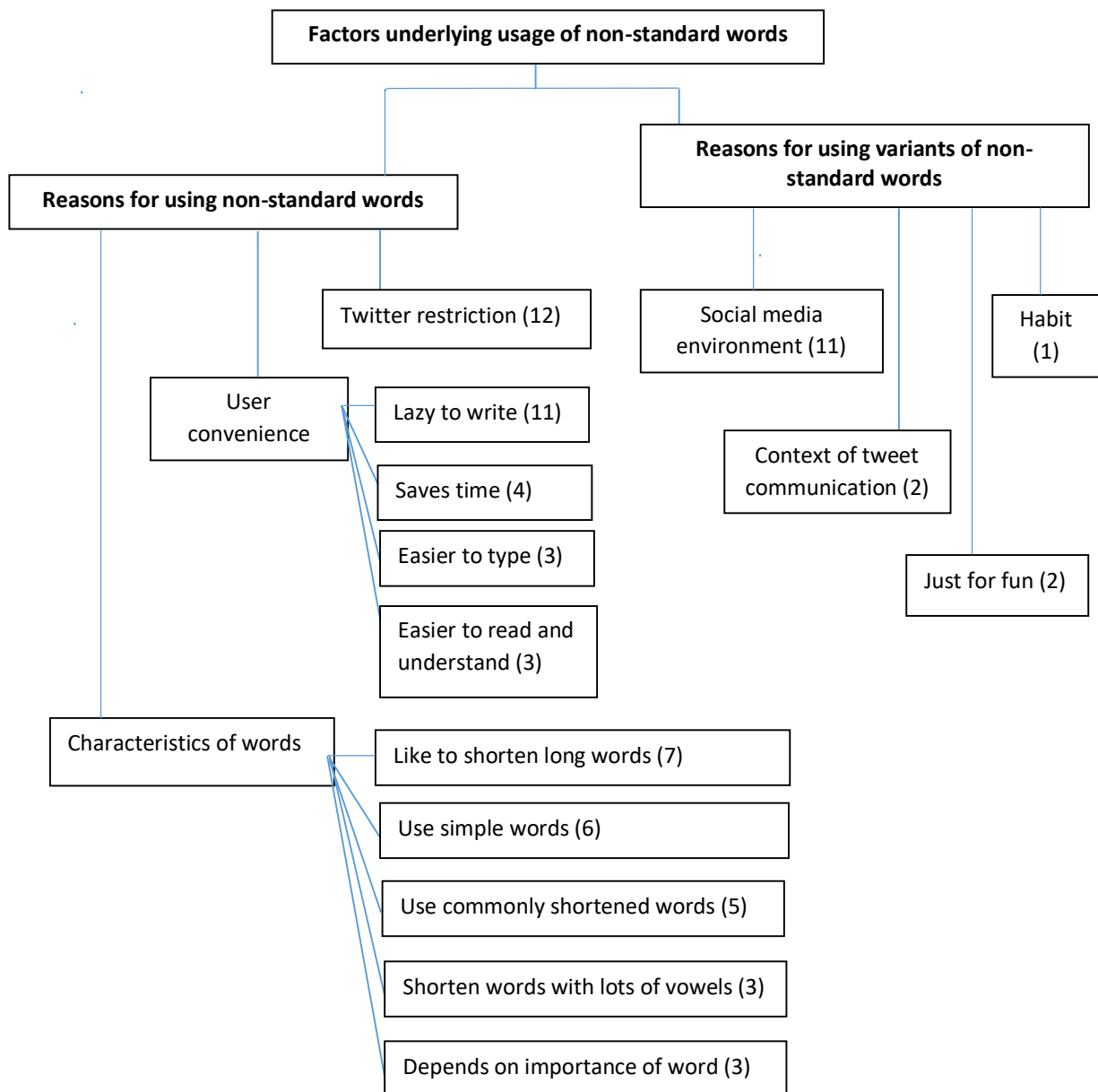
Figure 1: Malaysian millennials' reasons for using non-standard words in tweets

Interestingly, many participants gave the characteristics of words as a reason for using non-standard spelling of words. Long words are the prime target to be shortened, according to the participants. Participant 30 was explicit about the length of words that warrant shortening: "If the word is longer than five letters, I will cut and shorten it". Another six participants said that they tended to use simple words in their tweets. To ensure that their messages were understood, the participants used commonly shortened words. Participant 14 explained that he would not make up new words as others may not know what he meant, as shown in the following excerpt. This indicates that he is a passive user of shortened words rather than someone who coins shortened words to use in tweets.

[I use] The most simple and known words that people usually shortened, because if I coined new word, I afraid the message I'm trying to convey might not be perceived by the intended audience (Participant 14).

Moreover, eight participants said that they were selective about words to shorten rather than randomly shorten any word. The likely words to be shortened are those with lots of vowels and less important words (3 participants each). An example of a Malay word with many vowels is "*telefon*" (telephone) which is commonly written as "*fon*" in tweets. Another example is the word "*sekarang*" (now). Three vowels are dropped, and only the four consonants are left ("*skrg*") but the word can still be guessed and understood easily. The determiner "*ini*" (this) and the verb "*hendak*" (want) do not carry crucial meaning and are frequently shortened to "*ni*" and "*nak*" respectively.

Reasons for Using Variants of Non-Standard Words
Seven participants out of 30 participants reported that they liked to experiment with different spelling for some words. For example, the word "*dia*" has been shortened to three forms where "i" is replaced by "y" and "a" is replaced by "e": "*die*", "*dy*" and "*dye*". Apparently, the millennial Twitter users sometimes deliberately used variants of non-standard words because it is a feature of the social media environment, the context of the tweet communication, and sometimes for fun or just out of habit (Figure 1).

About one-third of the participants (n=11) said that they felt free to use variants of spelling in Twitter. For example, "*duduk*" (sit) is shortened to "*duk*" or "*dok*". The latter means a dockyard in standard usage, but the context of the shortened word can help Twitter users to figure out the meaning of "*dok*". New Twitter users may have to rely on context to decode the meaning of shortened words. However, once Twitter users have been tweeting for a while, they learn the "new vocabulary" of tweets and can understand the shortened versions without any problem.

Sometimes the specific context of the tweet communication also influences whether Twitter users can use certain variants of non-standard words. In a particular group, it may be more common to shorten "*macam*" (like) to "*mcm*" while other groups may use "*cam*" for the same word. There is probably some homogenising effect over time as Twitter users in the same group pick up certain variants of shortened words, and use them to blend into the group. Participant 18 said, "Sometimes in hassle and need to fast. It is also part of blending into the internet community". Some millennial users like Participant 21 demonstrated audience awareness when using non-standard words in their tweets. He said, "Depends on the target audience, if I write to my friends, then I will use hipster-style words like "x" (to mean "no") for my friends and for slightly more formal tweet, I would use "*tak*".

Two participants experimented with different spelling of non-standard words. For example, "*owg*" for "*orang*" when the more common version is "*org*". Participant 4 said, "sometimes, it's just for the fun of it, since Twitter is not a serious place" while Participant 21 said "I do it because I think it is fun and entertaining, since everyone on Twitter can accept". These participants saw Twitter as a platform that is open to coinage of new words.

Last but not least, there was one participant who coined words for fun. Participant 23 said, "I think it become a habit for me to use different non-standard words or slangs, plus I think most of us can accept this usage in social media". For Participant 23, it became her habit because the Twitter community accept those non-standard words and she likes to use

different types of spelling while tweeting because she believed other users can get the meaning.

DISCUSSION AND CONCLUSION

The study on factors influencing Malaysian millennials' use of non-standard words in Twitter communication shows that the Twitter users were mostly passive followers in using modified spelling of standard words although some were active initiators in coining new spellings. Two noteworthy findings will be now discussed.

Firstly, the frequently used non-standard words in Malay tweets are similar across topics and social media groups in Malaysia. In the present study, the five top words used by Malaysian millennials in their tweets were "*ni*" or "*nie*" (this), "*nak*" (want), and X (negation), "*je*" or "*ja*" (for sahaja, meaning only), and "*aq*" or "*ak*" for (aku, meaning I). The findings concur with other researchers who studied social media messages in Malay written by Malaysians. Saloot et al. (2014) found that "*nak*" ranked second in their analysis of Malay Chat-style Corpus, a sub-group of Malay tweets collected in the year 2012. Similarly, Athirah et al. (2019) found that the full word "*hendak*" ranked fourth in Malaysian tweets searched using the words "eat" and "food". The words "*je*" and "ni" were also common in both food tweets (Athirah et al., 2019) and in movie, politics and product reviews in Facebook and Twitter (Chekima & Alfred, 2017). Considering that these studies were conducted in the last decade, the frequently used non-standard words in Malay social media language have not changed much. The similarities in findings on top-ranked non-standard words also point shared norms in shortening words among Malaysians who write social media messages in Malay. In Botswana, Galebole (2020) also found that modifications in word spellings were transferred across social media platforms.

Secondly, the main reasons for millennial Twitter users to shorten words in tweets were Twitter restriction, user convenience, and characteristics of words. The result on the Twitter character limit is expected. In a sense, the result on brevity of words for user convenience is not surprising. Twitter users are pressed for time to send tweets rapidly so that there are almost instantaneous replies. Speed, informality, innovativeness, and wanting to belong to a group are factors that influence university students' use of non-standard words in WhatsApp in Botswana (Galebole, 2020). The innovativeness or creativity in coining new shortened versions of standard words in the present study is partly due to the character restriction on Twitter. The idiosyncratic versions may start out of fun, but can become habitual for certain individuals. Their social contacts can understand them, and may also use them. For example, "*macamana*" (how) is written as both "*mcm*" and "*cam*". The former cannot be pronounced whereas the latter reflects how the word is spoken. Another example is "dia" (he/she), which is written in three different forms: "*dy*", "*die*" and "*dye*". These shortened versions are among the top 100 non-standard words in the present study but the frequency is low, showing that these words have not gained traction. The university students in Galebole's (2020) study felt free to be creative in making new words because they are not expected to follow language rules. Although not widespread, the results suggest that some Malaysian millennials in the present study are active initiators of some new spellings of words, which may become popular over time. They do not have a linguistic educational background, but showed awareness of word features that make certain words the prime target for shortening, that is, long words, simple words, words with many vowels, and less important words. The Twitter environment is seen as accepting of word coinage, but based on the

comparison of the high-frequency non-standard words in Malay tweets reported in other studies, individual word coinages may stay idiosyncratic and not make it into mainstream usage. The discussion on Malaysian millennials' reasons for shortening words is limited because there are few studies on user perspectives.

Considering that the present study is on young Twitter users tweeting in Malay, an area for future research is to investigate whether there are similarities in non-standard words used by Twitter users in older age groups. It is also important to investigate non-standard usage of words across social media platforms, particularly those that do not impose character restrictions like Facebook to find out if the non-standard words used in Twitter among the millennials have gained popularity in social media discourse.

BIODATA

Nur Nashatul Nasuha Nazman graduated with a Bachelor of Arts in Linguistics and a Masters of Applied Linguistics from Universiti Malaysia Sarawak in 2022. Her research interest is in word formation and social media communication. Email: nashanazman94@gmail.com

Su-Hie Ting (correspondence author) teaches research methodology at the Faculty of Language and Communication, Universiti Malaysia Sarawak. She holds a PhD in Applied Linguistics from the University of Queensland. She has published on language choice and identity, academic writing, communication strategies, and health communication. Email: shting@unimas.my / suhieting@gmail.com

Kee-Man Chuah is a senior lecturer at the Faculty of Language and Communication, Universiti Malaysia Sarawak. He graduated with a Bachelor of Science (Hons) in Teaching of English as a Second Language and a Master of Science in Cognitive Science from Universiti Malaysia Sarawak. His areas of expertise lie in virtual reality, e-learning, and computer-assisted language learning. Email: kmchuah@unimas.my

REFERENCES

Ahmed, B. (2015). *Lexical normalisation of Twitter Data*. Paper presented 2015 Science and Information Conference (SAI), 2015, pp. 326-328. https://doi.org/jn96

Ariffin, S. N. A. N., & Tiun, S. (2018). Part-of-speech tagger for Malay social media texts. *GEMA Online® Journal of Language Studies*, *18*(4), 17-22. http://dx.doi.org/10.17576/gema-2018-1804-09

Athirah, R. R., Soon, L. K., & Haw, S. C. (2019). Exploring edit distance for normalising out-of-vocabulary Malay words on social media. In R. Černý, J. Kočí and V. Kočí (Eds.), MATEC Web of Conferences (pp. 1-7). Les Ulis, France: EDP Sciences.

Bisera, K., & Elena, S. (2018). Internet memes and their socio-linguistic features. European *Journal of Literature, Language and Linguistics Studies, 2*(4), 158-169.

Chekima, K., & Rayner, A. (2017). *Sentiment analysis of Malay social media text*. In R. Alfred, H. Iida, A. A. A. Ibrahim, & Y. Lim (Eds.), *Proceedings of the Fourth International Conference on Computational Science and Technology* (pp. 205-219). Singapore: Springer.

Choi, D., Kim, J., & Kim, P. (2013). A method for normalizing non-standard words in online social network services: A case study on Twitter. In Vinh, P., Alagar, V., Vassev, E., Khare, A. (Eds.), Context-aware systems and applications (ICCASA 2013, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Vol. 128, pp. 359-368). Cham.: Springer. https://doi.org/10.1007/978-3-319-05939-6_35

Crystal, D. (2011). *Internet linguistics: A student guide*. New York: Routledge.

Dixon, S. (2022, March 29). Twitter: Distribution of global audiences 2021, by age group. *Statista.* https://www.statista.com/statistics/283119/age-distribution-of-global-twitter-users/

Dunlap, J. C., & Lowenthal, P. R. (2009). Tweeting the night away: Using Twitter to enhance social presence. *Journal of Information Systems Education, 20*(2), 129-135.

Flint, E., Ford, E., Thomas, O., Caines, A., & Buttery, P. (2017). A text normalisation system for non-standard English words. *Proceedings of the 3rd Workshop on Noisy User Generated Text,* pp. 107-115. https://doi.org/10.18653/v1/W17-4414

Fontein, D., & Copp, E. (2018, December 19). Words and phrases to ban from your social media vocabulary*. Hootsuite*. https://blog.hootsuite.com/words-and-phrases-to-ban-from-your-social-media-vocabulary/

Foster, H. (2018). How is social media changing the English language*? Language Services Direct.* https://www.languageservicesdirect.co.uk/social-media-changing-english-language/

Fugard, A. J., & Potts, H. W. (2015). Supporting thinking on sample sizes for thematic analyses: A quantitative tool. *International Journal of Social Research Methodology, 18*(6), 669-684.

Galebole, W. (2020). Factors leading to spelling change of English language words on Whatsapp. *American Journal of Communication, 2*(1), 26-40.

Gligorić, K., Anderson, A., & West, R. (2018, April 10). *How constraints affect content: The case of Twitter's switch from 140 to 280 characters*. Association for the Advancement of Artificial Intelligence. https://www.cs.toronto.edu/~ashton/pubs/twitter-constraints-icwsm2018.pdf

Golbeck, J. (2016). Negativity and anti-social attention seeking among narcissists on Twitter: A linguistic analysis. *First Monday, 21*(3), 1-10.

Han, B., & Baldwin, T. (2011). Lexical normalisation of short text messages: Makn sens a# Twitter. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies,* pp. 368-378, Stroudsburg, PA: Association for Computational Linguistics. https://aclanthology.org/P11-1038

Haroon, R. (2021, July 18). "Mulut Puaka" wreaking havoc on social media with their whining. *The New Straits Times*. https://www.nst.com.my/opinion/columnists/2021/07/709340/mulut-puaka-wreaking-havoc-social-media-their-whining

Honeycutt, C., & Herring, S. C. (2009*). Beyond microblogging: Conversation and collaboration via Twitter. Conference Proceeding HICSS 2009 42nd Hawaii International Conference on System Sciences*, pp. 1-10.

Hong, L., Convertino, G., & Chi, E. H. (2011). Language matters in Twitter: A large scale study. *Proceedings of the International AAAI Conference on Web and Social Media, 5*(1), 518-521. https://doi.org/10.1609/icwsm.v5i1.14184

Hughes, D. J., Rowe, M., Batey, M., & Lee, A. (2012). A tale of two sites: Twitter vs. Facebook and the personality predictors of social media usage. *Computers in Human Behavior, 28*(2), 561-569.

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), 46-58.

Isaac, M. (2017, September 26). Twitter to test doubling tweet length to 280 characters*. The New York Times*. https://www.nytimes.com/2017/09/26/technology/twitter-280-characters.html?smid=url-share

Jansen, B. J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth*. Journal of the American Society for Information Science and Technology, 60*(11), 2169-2188.

Jaidka, K., Zhou, A., & Lelkes, Y. (2019). Brevity is the soul of Twitter: The constraint affordance and political discussion. *Journal of Communication*, 69(4), 345-372.

Kamus Dewan (4th ed.). (2007). Kuala Lumpur, Malaysia: Dewan Bahasa dan Pustaka.

Kern, M. L., Park, G., Eichstaedt, J. C., Schwartz, H. A., Sap, M., Smith, L. K., & Ungar, L. H. (2016). Gaining insights from social media language: Methodologies and challenges. *Psychological Methods*, 21(4), 1-17.

Lusetti, M., Ruzsics, T., Göhring, A., Samardžić, T., & Stark, E. (2018). Encoder-decoder methods for text normalization. *Proceedings of the Fifth Worksop on NLP for Similar Languages, Varieties and Dialects (VarDial 2018)*, pp. 18-28, Santa Fe, New Mexico, USA, Association for Computational Linguistics. https://aclanthology.org/W18-3902

Liu, F., Weng, F., Wang, B., & Liu, Y. (2011). Insertion, deletion, or substitution?: Normalizing text messages without pre-categorization nor supervision. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies,* pp. 71-76, Portland, Oregon, USA. Association for Computational Linguistics. https://aclanthology.org/P11-2013

Mitchell, L., Frank, M. R., Harris, K. D., Dodds, P. S., & Danforth, C. M. (2013). The geography of happiness: Connecting Twitter sentiment and expression, demographics, and objective characteristics of place. *PloS One, 8*(5), e64417.

Nazman, N., Chuah, K. M., & Ting, S. H. (2020). Tryna b Kewl: Textual analytics of distorted words among Malaysian millennials on Twitter. *International Journal of Innovative Technology and Exploring Engineering, 9*(3), 3522-3526.

Nisa, R. F. (2016). *Word formation process of non-standard vocabulary in Twitter statuses of Indonesian dangdut singers*. *Lantern, 5*(3).

Nkhata, L., & Jimaima, H. (2020). Neologisms: A morphological analysis of social media discourses on the Zambian online media. *Multidisciplinary Journal of Language and Social Sciences Education, 4*, 66-93.

Ortiz-Ospina, E. (2019, September 18). The rise of social media. *Our World in Data.* https://ourworldindata.org/rise-of-social-media

Reed, J. (2014). How social media is changing language. *OUP Blog*. https://blog.oup.com/2014/06/social-media-changing-language/

Saloot, A. M., Idris, N., Aw, A., & Thorleuchter, D. (2014). Twitter corpus creation: The case of a Malay Chat-style-text Corpus (MCC). *Digital Scholarship in the Humanities, 31*(2), 227-243. https://doi.org/10.1093/llc/fqu066

Sayce, D. (2022). The number of tweets per day. *DSayce*.

Schmied, J. (2012). Social digital discourse: New challenges for corpus-and sociolinguistics *Topics in Linguistics, 10*, 43-56.

Seward, Z. M. (2013). The first-ever hashtag, @-reply and retweet, as Twitter users invented them. *Quartz.* https://qz.com/135149/the-first-ever-hashtag-reply-and-retweet-as-twitter-users-invented-them

Sriram, B., Fuhry, D., Demir, E., Ferhatosmanoglu, H., & Demirbas, M. (2010). Short text classification in Twitter to improve information filtering. *Proceedings of 33ed International ACM SIGIR conference on Research and development in information retrieval,* pp. 841-842.

StatCounter Global Stats. (2020). Social Media Stats Malaysia, Jan-Dec 2020. https://gs.statcounter.com/social-media-stats/all/malaysia/2020

Statista. (2022). Forecast of the number of internet users in Malaysia from 2010 to 2025. https://www.statista.com/forecasts/1146231/internet-users-in-malaysia

Stone, C. B., & Wang, Q. (2019). From conversations to digital communication: The mnemonic consequences of consuming and producing information via social media. *Topics in Cognitive Science*, *11*(4), 774-793.

Suparta, I. W., Qomariana, Y., & Rahayuni, N. K. S. (2016). Word formation on Twitter posts. *Jurnal Humanis, 18*(2), 6-23.