

UNDERSTANDING MALAYSIAN ENGLISH (MANGLISH) JARGON IN SOCIAL MEDIA

M. F. Rusli^{1,*}, M. A. Aziz², S. R. S. Aris³, N. A. Jasri⁴ and R. Maskat⁵

¹Bachelor degree in Information System Engineering University Teknologi MARA, Malaysia

²Senior Lecturer at Universiti Teknologi MARA, Malaysia

³Department of Information Systems, Universiti Teknologi MARA

⁴Student at Universiti Teknologi MARA, Malaysia

⁵Doctorate in Computer Science from the University of Manchester, United Kingdom

Published online: 01 February 2018

ABSTRACT

The advent of the internet, mobile communication and media has created a new form of language such as Slang, Emoticons, Hashtag and Abbreviation as well as a combination of several languages in one word. Some go to the extent of localizing foreign language. In Malaysia, a new trend of using social media language is called Manglish, a mix language of Malay and English words that are popularized by social media users. Based on initial findings, the use of Malaysian English (Manglish) jargon can lead to confusion and miscommunication between social media users of different generations. Even though there are various translation software available, no online Manglish Jargon translator is available at present. Therefore, this work proposes the development of Manglish Jargon Translator that will reduce the miscommunication gap between social media users of all ages. Interview and survey instruments were conducted to capture user requirement and as part of the Manglish Jargon validation process.

Author Correspondence, e-mail: farhanrusli93@gmail.com

doi: <http://dx.doi.org/10.4314/jfas.v10i2s.10>



As a result, the Manglish Jargon translator was developed to further clarify Manglish Jargon terms, actual words, meanings and definitions

Keywords: Manglish, jargon, communication, social media, Malaysian.

1. INTRODUCTION

Effective communication is important to make our lives better. The way to communicate has evolves. It started off with traditional smoke signals to smart phones. The mass media technology has evolved and increasingly affected how people communicate, which frequently uses informal language. They also use unconventional structures such as dialects, abbreviations, trendy jargons, symbols and mix languages, in mostly colloquial forms over social media. In Malaysia, specifically, the code-switching usually happens between Malay and English which is typically known as Manglish [1].

For this paper, we focus on looking into new vocabulary words used in the social media. In social media, jargon and slang are widely used. The mixture of Jargon and Internet slang are acceptable and understood by most people. This was invented due to the habit of shortening words and creating new words [2]. User might misunderstand a conversation. Therefore, in order to minimize misunderstanding, we need a special translator. The main purpose of this translator is specifically to translate jargon words used in the social media for example Facebook, LinkedIn, Twitter or Instagram. At the moment, there is existing translator that interpret jargon words for Manglish words. The translator will translate Manglish into original words. Manglish word can be defined as English words that are spelt according to the way Malaysians pronounce them. The translator will also provide the original word with the right translation. By developing this translator called Malaysian English (Manglish) Jargon Translator, people across different generations are able understand and communicate better.

2. LITERATURE REVIEW

2.1 New Ways of Communicating

Web-based social networking use electronic advancements, desktop PCs and mobile technologies, such as, smartphones and tablets to make a platform for individuals or groups and associations to share or co-create on the web [3]. Social media work in a

dialogical transmission system where it originates from many sources to numerous receivers. This is absolutely the inverse with traditional media which works under a monologic transmission model which originates from one source to numerous receivers [4]. Usually websites and applications are focus to person to person communication, for example forums, and microblogging. Meanwhile, social bookmarking and wikis are for web-based social networking.

In the United States, studies showed that for social media platforms, Instagram is the most favored site while the most popular messaging applications is WhatsApp. Meanwhile, there are more than 20 million and over than 67 percent of social media users in Malaysia [5]. The number has increased tremendously over the years due to increased internet penetration rate growth in Malaysia. As reported by Suruhanjaya Komunikasi dan Multimedia (SKMM) in Sinar Harian in 2015, the average of social media users in Malaysia has at least four different social accounts and spent at least two hours a day on social media. The most popular social media platform is Facebook. As reported by Internet World Statistics website, Malaysia has about 19 million Facebook users [6].

All these online platforms revolutionize the way we communicate. From “kipidap”, “shuben” to “Maleiss”, these words that we use in daily conversation has influence our language. Language has evolved together with the evolution of technology. This is because our written language that is now on computers, tablets, and smartphones. When we communicate with each other, we tend to be less formal and more malleable. Therefore, the combination of informal, personal communication and the mass audience afforded by social media have set a new way of communication. New words are introduced with the change of spelling. Among the most famous ways to communicate include the use of acronyms such as LOL (Laughing Out Loud), emoticons such as ;-) (wink), ;-((Sad) and now we see the usage of Hashtag in Twitter in many communications. These elements sometimes add useful elements in our non-verbal communication. On the other hand, it also annoys people when they are overused.

According to Merriam-Webster dictionary, a jargon is “the technical terminology or characteristic idiom of a special activity or group or confused unintelligible language”.

Meanwhile, slang is “a colloquial variety of language used in highly informal situations”. In Malaysia, the official language is Bahasa Malaysia also known as Malay. Meanwhile, the second active language spoken in Malaysia is English. This scenario lead to the introduction of Manglish, the term used to represent the English words that are spelt according to the way Malaysians pronounce them. Manglish jargon makes it difficult for some people to understand during communication which may lead to misinterpretation and miscommunication.

2.2 Big Data Social Media Analytics

Big data processes data from external data sources and transforms it to new valuable insights. Big data as a large amount of data and illustrated big data as new term that cannot be scoped to identify its datasets by using current methodologies or data mining software tools [7]. Besides, many researchers defined big data as a remarkable opportunity for organizations to drive decisions in critical intelligent ways and it is also used to capture the challenges faced by all the researchers in order to accessing, managing, analyzing and integrating datasets of diverse data types [7-9].

As many researchers define many features of big data, basically the three “V” features are their common characteristics. They are volume, variety and velocity. Volume is concerned with the amount of data. According to data collected by the U.S. Library of Congress in April 2011, 2.7 zettabytes of data already exist nowadays in digital universe and it is increasing fast. Variety, is defined as data that is not being produced in a single category. The heterogeneity (different) of data types, representation and semantic interpretation (subjective) make this data difficult to be handled by the existing traditional analytic systems. Velocity can be defined as the rate by which data are generated and transmitted. Besides the 3 Vs, there are several features being added such as complexity, variability, value and veracity that can be used by researchers or organizations to define big data.

Generally, big data is known as structured, semi-structured and unstructured data. Structured data are generally simple, well known formatted data that are easy to manage and process with predefined relationships. The example of structured data is the data in relational database such as Main Frame, SQL Server, Oracle, DB2 etc. Unstructured data refers to data with no form of format. Some of the most common of

unstructured data are widely used in social media. Some example unstructured data are blogs, tweets, comments, likes, followers and others. Video, audio, QR code also can be categorized as unstructured data. Meanwhile, semi-structured data refers to data that does not fall into either category; structured or unstructured. Email is a good example of semi-structured data. Basically, email contains mainly unstructured data but these unstructured data are always linked to some more structured data such as To, From and Subject. Third party data such as weather, currency conversion, demographic and panel are also classified as semi-structured data.

According to Oracle in 2015, 80% of all types of data are from unstructured and semi-structured data and the fastest growing data type. Unstructured or semi-structured data are estimated to rise as they are produced and will present as an unparalleled opportunity [8]. Text Analytics is the process of mining high-quality structured data from unstructured text. There are many ways to extract and collect unstructured text, for example using API to collect tweets. For this paper, we will focus on the unstructured data that are gathered from social media networks such as Facebook, Instagram, Blog and Twitter. Data are also gathered from the popular mobile messaging applications; WhatsApp and weChat.

3. RESEARCH METHODOLOGY

This section explains about the methods of data collection and data analysis. The phases involved are namely Data Collection and Analysis and Design and Development. Detail description of the activities done will be explain in the following section.

3.1 Data Collection and Analysis Phase

During data collection and analysis phase, we conducted two interviews and survey. It started off with doing preliminary investigation, the first interview. This was done to analyze the problem associated with the use of Manglish jargon among social media users in Malaysia. We focused on the languages used in social media. The interview was done randomly with the target group aged 18-50 years old. The interview was to measure first and second-year university students and the public of a selected group of

commonly-used social media glossary. The main target is active social media users. In the initial interview, 50 people were interviewed.

Based on the findings, it was found out that the language used in daily conversation is Malay and English. Some people use a mixture of Malay and English and we also noticed the use of jargon word and slang, about 20%. We also found out that, there are a number of English words were modified and spelt differently to fit the purpose of communication among Malaysian. This is a trend among social media users. Some other way of communication is using text-based communication such as “brb”, “ttyl” etc. This scenario has shown to have an impact on the English language especially from non-English-speaking countries. The results from the investigation also revealed that a majority of the respondents have difficulties in understanding Manglish jargon. During the interview, the interviewees were asked to list out Malaysian jargons they found or read. The jargons were collected and stored in a database.

Table 1. Example of Manglish Jargon

Kipidap	Miscol	Wadehek	Mischu
Omaigad	Gais	Demn	Mekdi
Shuben	Cekidout	Ukendoit	Kepci
Uols	Okie	Dongibap	Stabak
Iols	Fwen	Lebiu	Fesbuk

Next, the second interview and data collection process were performed simultaneously. The jargons were collected from four social media platforms such as Facebook, Instagram, Blog and Twitter and two mobile messaging applications (Watsapp and weChat). These social media were selected because of their popularity among Asians especially Malaysians [5]. The process started from December 2016 until July 2017. In total, there are more than 200 Manglish jargons identified and the list adds up daily. These words need to be updated frequently since almost every day there will be new words identified. Table 1 is the example of Manglish jargons collected.

The survey was done using Google Drive application. It consists of 2 sections. Section A is demographic profile. The respondents were required to provide information such

as gender, age, time spent on social media, types of social media used, language problems encountered and other related questions. Section B is to assess the social media users' understanding of the Manglish jargon identified earlier. The survey questions were distributed to 50 users using different techniques from different sources (emails, WhatsApp messages and paper-based).

Based on the questionnaire result, 60% of the respondents were unable to figure out the jargons' meanings correctly. 30% of the respondents have guessed correctly, while 10% did not answer it. From the result, it shows that there is a need for a translator that could help to clarify the meaning of the Manglish Jargon. A high percentage of the respondents do have problems when communicating due to lack of understanding of the Manglish jargons used. Once all the required data and information are available, the design and development phase of the Manglish jargon translator took place. The feedbacks / answers received from the respondents were used and processed as part of the verification technique of the Manglish jargon.

3.2 Design Phase and Development Phase

The objective of this phase is to illustrate the desired features and operations in detail. Diagrams such as class diagram and use case were drawn using UML. The design of the diagrams is based on the data collection process and requirement analysis. Basically, this translator will translate the Manglish jargons entered by user by providing the correct spelling and English translation. The Malaysian English (Manglish) Jargon Translator prototype is developed using NetBeans 8.2 Java together with Oracle Database 11g. The prototype will be explained in the next section.

4. MALAYSIAN ENGLISH (MANGLISH) JARGON TRANSLATOR

The significance of this project is to create a system that contains a database for Malaysian English (Manglish) Jargon words that are widely used in social media. By using this system users will be able to search for the exact meaning of Manglish jargon words. The word will be searched and translated along with the details of its meaning for better understanding. Besides, the system will also keep track of the latest jargon words being used and update it for reference. This is essential where it can be used not just for the time being but also in the

long run. This prototype is at its initial stage and we have not launched this prototype to the public yet. This is because we need to improve several features before making it public.

The Manglish Jargon (MJ) translator starts when the user fill in the search form with their preferred jargon word in the search box. Then it will proceed to the next page. For example, as shown in the figure 1 below, the word “shuben” that was filled in. The system will then search the jargon and it will be translated. The jargon actual English meaning, spelling along with its explanation will then be displayed.

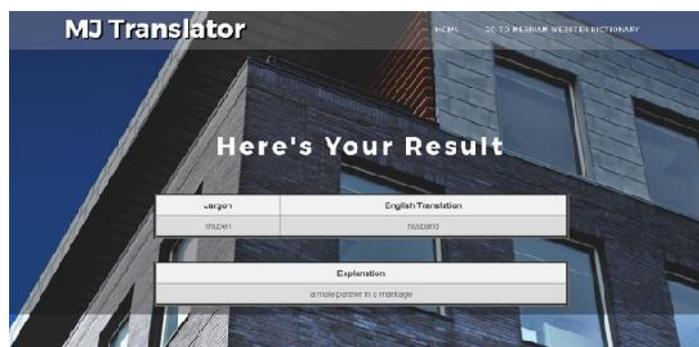


Fig.1. The translation of the jargon words

Basically, the MJ Translator is an office jargon dictionary that is useful for those who wish to interpret words or phrases that is beyond his/her understanding. This system has many features such as user-friendly interface, a translator and decoder of Malaysian English jargon and a corrector for definition of the word will be given. One of the working feature of this system is for it to share the term meanings via Twitter and email.

The MJ Translator is a straightforward interface so that users can use without hassle. The translation given for the jargon will help user to understand better. However, there are some of the limitations of this system. One of it is users are unable to add new jargon to the system. The solution for this issue is, if the new jargon is not yet in the database, the user can send it through twitter and automatically update the database. Besides that, this application is just for Malaysian English jargon translation. Therefore, the target audience is users who can understand Malay and English language. Moreover, this system does not cater for words with different dialects or slangs that cannot be recognized by this application.

5. CONCLUSION

This research is about the evolution of Malay language when communicating via popular social networks. Based on the findings, we found out that the main factor that influenced the ways language is used for communication is culture. For Malaysian, the language they use in social media is Manglish. Manglish (Malaysian English) is the modified Malay / English words that are known as jargon. The widely use of Manglish jargons may lead to misinterpreted messages when communicating. Therefore, understanding the words used in social media especially people who are new and not familiar with it is important. The significance of this project is to create a translator system that contains a database for the Malay Jargon words often used in social media. By using this translator system, users will be able to search and translate the word for better understanding. Besides that, it will also keep track of new jargon words created. The Manglish translator is only for English word that is spelt in Malay. The limitation of this research is the Manglish Jargons were identified manually. This process is time-consuming. Therefore, the next step is to create a jargon identifier that will automatically detect and add it into the database. Moreover, as continuation for this research taxonomy of the Manglish Jargon will be created.

6. REFERENCES

- [1] Kadir, Z. A., Maros, M. and Hamid, B. A. Linguistic Features in the Online Discussion Forums. *International Journal of Social Science and Humanity*, Vol. 2. No. 3, (2012b).
- [2] Thurairaj, S., Hoon, E. P., Roy, S. S., & Fong, P. W. Reflections of Students' Language Usage in Social Networking Sites: Making or Marring Academic English. *Electronic Journal of e-Learning*, 13(4), (2015), 302-316.
- [3] Kietzmann, Jan H., Kristopher Hermkens, Ian P. McCarthy, and Bruno S. Silvestre. "Social media? Get serious! Understanding the functional building blocks of social media." *Business horizons* 54, no. 3 (2011): 241-251.
- [4] Pavlik, John V. "Trends in new media research: A critical review of recent scholarship." *Sociology Compass* 7, no. 1 (2013): 1-12.
- [5] Website: <http://www.malaysiandigitalassociation.org.my/wp-content/uploads/2016/08/Malaysia-Digital-Landscape-August-2016.pdf>
- [6] <https://www.pressreader.com/malaysia/new-straits-times/20170325/281792808855455>

-
- [7] Katal, Avita, Mohammad Wazid, and R. H. Goudar. "Big data: issues, challenges, tools and good practices." In *Contemporary Computing (IC3), 2013 Sixth International Conference on*, pp. 404-409. IEEE, (2013).
- [8] Abawajy, Jemal. "Comprehensive analysis of big data variety landscape." *International journal of parallel, emergent and distributed systems* 30, no. 1 (2015): 5-14.
- [9] Fang, Hua, Zhaoyang Zhang, Chanpaul Jin Wang, Mahmoud Daneshmand, Chonggang Wang, and Honggang Wang. "A survey of big data research." *IEEE network* 29, no. 5 (2015): 6-9.

How to cite this article:

Rusli M F, Aziz M A, Aris S R S, Jasri N A, Maskat R. Understanding malaysian english (manglish) jargon in social media. *J. Fundam. Appl. Sci.*, 2018, 10(2S), 116-125.