

# Tryna b Kewl: Textual Analytics of Distorted Words among Malaysian Millennials on Twitter

Nur Nashatul Nasuha Nazman, Kee-Man Chuah, Su-Hie Ting

**Abstract:** *The presence of Malaysian millennials on social media platforms is increasingly gaining attention particularly on Twitter. Language wise, many of them are predominantly using English and Malay in their tweets but with a touch of their own “styles” in various morphological aspects. This trend eventually leads to a rampant use of distorted vocabulary, churning out many non-standard words. This study aims to address the need in classifying the types of morphological distortions of words that are widely used among the Malaysian millennials and identify the reasons behind such trend. A total of 50 active Twitter users from Malaysia aged 18 to 30 years old were randomly chosen for this study. From each user, 20 tweets of longer than 5 words were selected for lexical analysis, giving a sum of 1000 tweets (8443 words in total). Then, interviews were conducted on 30 participants to gauge the factors of using those non-standard words. The findings revealed that the words were largely distorted in terms of its inflections so as to fit some sounds. Also, most distorted words were deliberately coined so that the millennials would appear trendy, while some were merely following the usage without knowing the actual word. This study has shown that the use of distorted words among Malaysian Twitter users did not hinder effective communication.*

**Keywords:** *morphological distortion, textual analytics, Twitter*

## I. INTRODUCTION

Social media is a common tool of communication in the digital age. Among them, Twitter is one of the popular platforms that has been widely used. It is a free social networking microblogging service that allows registered members to broadcast short posts called tweets. Tweets were initially limited to 140 characters because to the constraints of Twitter’s Short Message Service (SMS) delivery system, but it was later increased to 280 characters in November 2017 or approximately 40 words. Gligoric, Anderson and West [1] studied the consequence of the switch and found out that despite the increase of characters, tweets produced are still predominantly concise. The restriction imposed by Twitter, therefore, has indirectly trained many Twitter users to be more economical and creative in their expression of thoughts, creating a plethora of newly-formed “words”, in order to ensure everything fits in one tweet.

**Revised Manuscript Received on January 05, 2020.**

**Nur Nashatul Nazman**, Department of Language and Communication, University Malaysia Sarawak, Malaysia.

**Kee-Man Chuah**, Lecturer, Department of Language and Communication, University Malaysia Sarawak, Malaysia.

**Su-Hie Ting**, Associate Professor, Department of Faculty of Language and Communication, University Malaysia Sarawak, Malaysia.

Due to Twitter’s microblogging method, users are required to make short, frequent posts to a microblog. Microblogging in Twitter may include hashtags (auto tagging of topics or words), mentions (mentioning or links to other Twitter users) or to other links from web pages, images or videos [2, 3]. From this platform, it is great if people want to quickly record their thought, opinions, ideas and create awareness. The main advantage of this platform is user-friendly and easier for users to be connected, given that the interaction is live and in real time [4]. In contrast, the character-limit set by the platform has prompted the creation of many “alien words” for different reasons [5]. This leaves ample room for further investigation to be done in relation to the content generated by the Twitter users.

Microblogging platforms such as Twitter have been the subject to many studies in the recent years. The studies in this area can be grouped into automatic sentiment analysis and opinion mining as it presents a huge source of data representing the opinions of a significant, yet totally random. Studies in the area of sentiment analysis [1, 6, 7, 8] depends upon random extraction of tweets and identify specific sentiment that researchers would like to uncover ranging from political inclination to prediction of suicidal thoughts. The second group of studies on opinion mining [5, 9] focuses on extracting keywords from tweets that match the intended target such as preference towards a product or reviews. Opinion mining is used by most marketing consultants as well as business corporations in obtaining the market trends.

The problem identified through the review of such studies is that sentiment analysis and opinion mining were done solely based on well-constructed words that convey such meaning. The thematic interpretation of extracted Twitter content has resulted in conflicting findings. This shortcoming is largely due to the informal language use, the presence of non-textual content and the use of slang words and abbreviations, that impede the accuracy of the mining process. Although the studies reviewed were involving English tweets, the growing trend of Malaysian twitter users in using slang words and abbreviations has also been studied [10]. With approximately 3.5 million users and more than 2 million tweets being generated in Malay language (Bahasa Melayu) apart from English on a daily basis [11], the potential of using Twitter’s large corpus of user-generated text-based contents for linguistic analysis is largely untapped. This study, therefore, aims to investigate the types of distortion used by