The relationship between orthographic variations and phonological and phonetic representations in Sri Lankan social media

W.A.K.V. Maheshi Wanniarachchi[1*]

[1]*Department of Linguistics, University of Kelaniya, Sri Lanka*

Abstract

Today, an increasing number of individuals are continuously occupied with social media platforms in their daily life, utilizing computers and hand-held electronic devices for sharing and viewing the content on their media walls. This research aimed to analyze the phonological and phonetic variations in social media language used by a sample of 18-30 years of age selected from different social, educational and geographical backgrounds in Sri Lanka. Sociolinguists argue that phonological variations occur due to the speed of technological innovations, and, as a result, such variations violate the traditional linguistics purism. The present study focused on identifying and analyzing the relationship between orthographic variations and the phonological and phonetic representations and their characteristics in social media language. The study enabled the identification of the frequently used phonological feature/s among youth in different situations. The data collection was limited to Facebook posts and WhatsApp conversations. Both qualitative and quantitative methods were applied to analyze the content and the frequency of occurrences. The research identified the most and the least popular phonological features in Sri Lanka. The findings highlighted not only the speed but also the freedom and other facilities as reasons which encourage the creativity of the users to form new forms of phonological features and suggested that the sociolinguists must pay more attention to the identified upcoming issues related to the language used in social media.

Introduction

Social media

Communication among humans has advanced dramatically since its inception, beginning with primitive communication via cave drawings and progressing towards the use of current advanced technology. In ancient times, people remained passive about information and its creators, as there was hardly any space for them to respond to news that they received or share their experiences and viewpoints with others. The idea of interconnectedness was implied in a short story, written by a Hungarian author called Frigyes Karinth, in 1929. However, this short story, "Chain Linked" did not carry the scientific hypothesis related to the technological awareness of online communication. Consequently, this author has influenced scientists regarding the possibility of developing such a communication system according to Seargeant and Tagg (2014, p. 1). Owing to technological developments, communication across the world has changed into new forms, particularly in relation to electronic media, starting from e-mailing and texting (SMS) to interactive modes in social communication. As a result, the world has become a global village with strong human interconnectedness by means of constant communication using devices that carry web-based applications such as computers, tablets, and smartphones. Moreover, this type of communication is increasingly expanded with the advent of new software and technological innovations. The novel terms 'netizen' or 'net citizen'   are often used to describe contemporary users who are frequently involved and meet actively via online communication.

Research on language in social media is an upcoming topic in sociolinguistics, as there is a wide range of newly implemented linguistic features, a hybrid mixture of both written and verbal forms of the language which follow unconventional spelling patterns (Sebba, 2007, p. 40). Moreover, Barton & Lee (2012) and Seargeant & Tagg (2012) pointed out that the language in social media has an interest in socially oriented linguistics because users present themselves as those who translate   local cultures and local identities with a global reach. They engage in communications with the aim of maintaining their relationships, which are mostly related to their issues and practices without having much concern about   the accuracy of the language that they use. According to Bakhtin (1984, p. 472) and Sebba (2007, p. 40), frequently used creative forms or phonological variations have become a form of 'licensed misrules' and in digital writing, this tendency has created what is labeled as 'linguistic clownery' that enables those communicators to present themselves as 'funky and cool' people.

The language behavior in social media

In any language, the spelling is considered as the consistency of phonology that supports the correct pronunciation of words, the orthography. Further, in every language, there is a set of conventions for writing which not only include the spellings but also hyphenation, capitalization, punctuation, word breaks, and emphasis. Besides, both phonology and orthography are the two sides of the same coin. They both represent the phonological properties of a word. In the Sri Lankan context, social media language is generally a mixture of both English and Sinhalese languages. Further, they use English letters for typing Sinhalese words. Additionally, many phonological variations can be identified due to adding spoken utterances (orthography) in their text messages. In the modern world, due to the advancement of web-based communication, social

media users have exponentially created a practice of using new forms of written language. Dovchin et al. (2015, p. 05) wrote that those new features in the use of language are linguistically destructive and distort the purity of the language.

This research was based on the hypothesis that social media users are the architects of new linguistic features (Jacquemet, 2005, p. 264). Based on this hypothesis, the present study aimed to analyze the relationship between orthographic variations and phonological representations in Sri Lankan social media. The objectives of this research were to investigate the types of phonological variations, their frequencies and their relationships with orthography in digital communication in the Sri Lankan context to identify the fluctuating and ever-growing language forms in digital communication, in different situations.

Methodology

The   purpose of the study was to observe and analyze empirically the language used in social media accounts, specifically on Facebook posts and WhatsApp conversations. Accordingly, the target corpus of the research consisted of samples of documents. The collection of samples was documents authored by the users of Facebook and WhatsApp, both male and female and from urban and rural areas. This survey was limited only to two social media applications, Facebook, and WhatsApp. The analysis was confined to posts on Facebook and private conversations on WhatsApp while the data was collected from a sample of users who were within the age group of 18 -30 years.

The data was collected for three months, and the corpus consisted of 50 documents. Content analysis was used to find out what were the prevailing features and language forms of social media in the Sri Lankan context, with a focus on how collations and patterns or concepts are involved in the particular mode of communication. Further, the research aimed to analyze the language used in social media, in different situations, with the use of both qualitative and quantitative types of data analysis.

Materials

Most of the selected Facebook samples (posts) were published on the walls of the selected Facebook users publicly or privately. The samples were collected individually using the "friend of a friend" method (Milroy, 1980). Since Facebook posts are always updated frequently with hundreds of comments, this research was restricted to analyze the posts shared by the selected group while the language used for giving comments or expressing their opinions was not taken into consideration. WhatsApp is a cross-platform messaging application, which provides the facility to communicate with a selected member or members, and it also provides a variety of other features like sharing photos, videos and voice messages. However, this research was limited to analyze only the text messages sent and received by the selected group of users. The sample persons engaged in WhatsApp mode of communication were arranged into conversation categories such as communication between girl-girl (G-G), girl–boy (G-B), boy-boy (B-B), group chatting (G-C), fiancé–fiancée (F-F) and child–parents (C-P).

The selected sample constituted young people of the age group18-30 years, both female and male, and belonging to different social, educational and geographical (urban and rural) backgrounds who have low or intermediate language proficiencies.

Participants / Corpus sampling

50 people were selected from each of the two social media and numbered from 1 to 50 for analyzing purposes. Both qualitative and quantitative methods were used, and those methods facilitated the identification of the phonological variations and the frequently those variations were seen in social media in different situations.

After selecting a sample for each social media, a pilot study was performed to select people and the method involved was random followed by purposive sampling. The quantitative method was a physical analysis to calculate the frequencies and to analyze the relevant features. Then, to identify each discourse, a qualitative method was used. Further, data cleaning process was conducted to recognize the information component of communication, as language conveys not only information but also many other implications such as humor, irony, evaluation, and sarcasm. Posts and conversations which violate ethical and religious boundaries or contain biased content etc. were excluded. Given below (Figure 1) is a summary of the methods applied in this research.
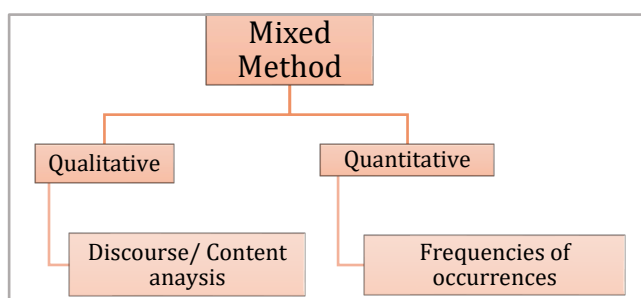

Figure 1. Recap of the Research Methods

Variables

The differences between the selected social media platforms had to be identified to define variables. For instance, when compared with Facebook, WhatsApp conversations were more individualized. The variables used in this study were as follows: the Independent variables - social media (Facebook and WhatsApp) and the dependent variables were phonological variations, different forms of spellings, prolonged vowels and consonants[1], words without vowels, different forms of capitalization, and Sinhalese words with English letters, abbreviations and pseudo language[2]

---

[1] Prolonged vowels - a deliberate extension of vowel sound/s for enhancing to maintain speaking aspects eg. nooo
Prolonged consonants - a deliberate extension of consonant sound/s for enhancing to maintain speaking aspects eg. plzzz
[2] Pseudo language – informal written or spoken resembles

Results / Discussion

Social media are internet-based sites or platforms which facilitate communication, such as sharing thoughts, views, ideas and perceptions. The first Social Network Site (SNS) was SixDegree.com which was developed in 1997. SNSs enabled their users to communicate with each other in groups and its influence was so much so that digital communication became a vital part of modern life all over the world (Baym & Boyd 2012, p. 321). Moreover, according to Boyd and Elliso (2007, p. 211) SNSs like Facebook (FB) have three main elements that can be recognized as user-generated features that allow users to construct a public or semi-public profile within a bounded system, to articulate a list of other users with whom they share a connection, and to view and traverse their list of connections and those made by others within the system. Hence, SNSs are developed into social-oriented online communication systems providing opportunities for their users to create a dynamic space for communication   allowing strong social interactions by maintaining the authenticity of speaking aspects which is a challenge in digital communication.

Phonology is one of the most discussed topics in sociolinguistic research (Baym, 2015; Sergeant & Tagg, 2014). This research aimed to analyze patterns used in internet-based communication in Sri Lankan social media. All the features mentioned below were used in the conversations to express different types of emotions which helped to maintain the meaning of speaking aspects by adding prosodic features in different situations. According to the analysis, the following phonological variations were identified in Table 1.

Table 1: Analysis of social media extracts

Linguistic feature	WhatsApp						FB		grand total	Percentage
	B-B	G-B	G-G	G-C	F-F	C-P	Total			
Different form of spellings	13	10	21	4	5	1	54	38	92	20.63
Words with prolonged vowels	**25**	**9**	**35**	**2**	**7**	**2**	**80**	3	83	18.61
Words with prolonged consonants	1	0	5	0	2	0	8	0	8	1.79
Words without vowels	32	45	84	27	22	0	210	10	220	49.33
Different forms of capitalization (English)	0	0	7	0	0	0	7	3	10	2.24
Abbreviations	7	4	10	5	0	0	26	6	32	7.17
Pseudo language				0	0	0	0	1	1	0.22
							385	61	446	

Frequency of occurrences: communication types - boy-boy (B-B), girl-girl (G-G), girl –boy (G-B), group chatting (G-C), fiancé – fiancée (F-F) and child –parents (C-P)

According to the results, in the FB posts, the most frequently used feature was different forms of spelling (38). Patterns were individualized and typed in Sinhala. The second most popular feature was the words without vowels (10) on FB. Words with prolonged vowels (3) and different forms of capitalization (3) and abbreviations (6) were equally used. In the samples collected from WhatsApp, the most frequent feature was words without vowels (210) because users had to match the speed of online communication. The next most popular feature was words with prolonged vowels (80) as users tried to maintain the aspects of speaking utterances by adding this feature to digital communication. Different forms of spelling (54), abbreviations (26) and words with prolonged consonants (8) were the other features identified in this analysis. In both WhatsApp and FB, the most frequently used feature was words without vowels (49.33%) and the least used feature was pseudo-language (0.22%) while different forms of spelling and prolonged vowels (20.63% and 18.61% respectively) were also used. Based on the results, each feature was analyzed below in detail. many occasions, English letters were added to words

Words without vowels

The most frequently used feature was words without vowels (49.33%). This feature was popular in WhatsApp communications (210) compared to FB, as a tool of writing forms in social media platforms when compared with other forms of written language (Androutsopoulos, 2011; Gimpel et al. 2011; Dresner & Herring, 2010). At present, people tend to form words without vowels as a new creative form of spellings on SNSs though it is very different to well edited written forms and their conventions or rules (Table 2 and Figure 2). The style of using only consonants when forming words was introduced through Twitter in April 2013.

However, the variations of these patterns of word forming cannot be predictable as it was subjected to change in every individualized conversation. In other words, for instance, when typing a word without vowels in online conversations, it may have different patterns for the same word. For example, "gahanawa" /gɑhanəwɑ/ is a Sinhala word which has the meaning of hitting. It can be represented in social media with different patterns in individualized communications such as "ghnaw/ ghnəw/, gahnw / gɑhnw /, ghanw / ghɑnw/, ghnwa/ ghnwɑ/ ". Some of the following patterns were identified in the samples collected for words without vowels (Figure 2).

Table 2: Different patterns of words - words without vowels

Words used in communication	Meanings
thnx,	Thanks
innwlu (innawalu /innɑwɑlu /)	To be or exist
krn (karanna/ kɑrɑnnə /)	To do something,
dnnwd (dannawda / dɑnnəwɑdə/)	Do you know?
ennda (ennada / ennədə/)	Shall I come?

Figure 2. A Sample for words without vowels

Use of prolonged vowels and consonants

Poets and novelists frequently use prolonged vowels as a linguistic tool to enhance the euphony in their literary creations, as prolonged vowels give an artistic sense to maintain the rhythm of such work. SNSs users tend to form words with prolonged vowels and consonants to give specific prosodic features in different spoken forms of the language. Thus, creative phonological forms, which we frequently see in interactive digital writing, may be interpreted as 'licensed misrule', that allow us to engage in 'linguistic clownery' (Bakhtin, 1984; Sebba, 2007). Consequently, these prolonged vowels and consonants help to maintain intonations and tones by adding a different pace and rhythm to the written forms of language in social media. The participants in these platforms aim to add prosodic features in written forms of language to maintain the means of speaking utterances in digital communications. Prolonged vowels and consonants were specifically used in the data samples collected from girls' conversation category in WhatsApp. According to sociolinguistics, females maintain their own styles in conversations because gender difference is a prevailing quality of language use. Prolonged consonants and vowels have been used as a specific characteristic to emphasize. To put it more simply, these features are used to elaborate a certain compassionate, and sentimental nature in their communications (Figure 3).

Figure 3. A sample of a prolonged vowels and consonants

Different forms of capitalization

Using different patterns of capitalization was another phonological variation which maintains authentic oral presentation of the written language on social media posts and comments. For instance, all letters in capital form are used to stress the message by adding prosodic features, or to indicate verbal authenticities of shouting, blaming or arguing (Zappavigna, 2012; Lamontagne & McCulloch, 2017). This feature was not very popular in the Sri Lankan context and only a few samples were identified in FB (3) and WhatsApp (7) under the girl-girl conversation category. Some of them were, I UPDATE MY FB PFP, JANUARYවල (Figure 4).
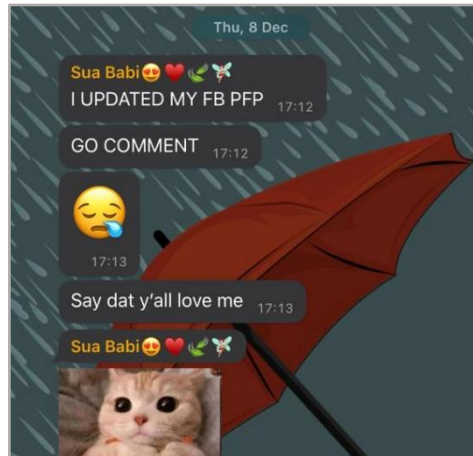

Figure 4. Different forms of capitalization

Different forms of spellings

Another feature that is deliberately used in social media was different forms of spellings for various reasons. *Robb (2014) declared that* they are used not because of the speed, but to give a creative sense and as a result, they keep on inventing new words ignoring the spelling rules of the language.

This research identified *that this feature was* used in WhatsApp (46) girl-girl conversations more frequently, compared to data samples collected from Facebook (4). For instance, dat (that), thx (thanks), qwwa/kiwwɑ/ (told), ecac/ekɑk/(one), nyt (night) are some of the different forms of spelling used. In addition, some of them had different spellings to add creativity to their conversations like බොමුත/බොමුතේ (bomutha/bomuthe for drink), laccana (beautiful), karamaganna (do it by oneself) in both WhatsApp conversations and on Facebook posts (Figure 5).

Moreover, this research observed that some of the words typed using both English and Sinhala alphabets was another feature. The patterns of using them were unpredictable. Some had English letters at the end of the Sinhala letters, and they were used to enhance creativity in the conversations. For instance, කියපන්ko (tell), පLa (go), උber (you). Some words had English or Sinhala letters in between බුRuවේ (donkey), වැtenකන් (until fallen down), ගෑNiයේ (woman), උberට (to you).

Figure 5. Different forms of spellings

Abbreviations and Pseudo code / language

Both abbreviations and pseudo-language are the shortened forms of lengthy words of word phrases which add new lexemes to the languages. In linguistics, there are different rules to form abbreviations for complex word phrases. However, in digital communication, users tend to create their own forms of abbreviations. Thus, they   have different meanings according to the context or situation (Sergeant & Tagg, 2014). Further, they frequently use them in their daily practices because using such abbreviations helps to communicate via smartphone and tablet keyboards easily and quickly. Some of them were abbreviations for word phrases, but they were typed using simple letters such as gm (good morning), gn, (good night) bs (budusaranai, blessing), gf (girlfriend) and tc (take care). Some of them were in capital letters, for instance DBN, SVT, AHS. Some abbreviations were formed to indicate single words like lv (love), u (you), hv (have) yr (you). These abbreviated forms can also be discussed as pseudo language.

Conclusion

The objective of this study was to find phonological variations in social media language used in the Sri Lankan context. Social media applications have developed sociable forms to communicate easily and quickly, owing to technology-based electronic innovations. Thus, fitting into the speed of such devices was a challenge and the users had a propensity to modify the language. However, the study revealed that not only the speed but also the freedom that users had in social communication was another reason for such modifications. As per the results, most of the phonological variations were identified in WhatsApp communications and the most frequently used feature was words typed without vowels while the least used feature was the use of pseudo language. Most of the female conversations exhibited a lot of variations used for emphasizing, which proved that the gender difference was a prevailing aspect. On the other hand, those network sites never paid attention to convention and   rules of the languages that their participants use. As a result, a lot of phonological variations could be observed in the language used in social media.

References

Androutsopoulos, J. (2013). Networked multilingualism: Some language practices on Facebook and their implications. *International Journal of Bilingualism*, 19 (2), 185-205. https://doi.org/10.1177/1367006913489198

Androutsopoulos, J. (2014). Moments of sharing: Entextualization and linguistic repertoires in social networking. *Journal of Pragmatics*, 73, 4-18. https://doi.org/10.1016/j.pragma.2014.07.013

Bakhtin, M. (1984). *The Dialogical Principle - Tzvetan Todorov*, Manchester University Press.

Bamman, D., Eisenstein, J. & Schnoebelen, T. (2014). Gender identity and lexical variation in social media, *Journal of sociolinguistic*. 18 (2) pp135-160. https://doi.org/10.1111/josl.12080

Baron, N. S. (1984). Computer mediated communication as a force in language Online. *Visible language*, (2), 118-141.

Baron, N. S. (2008). *Computer mediated communication as a force in language change.* Department of Linguistics, Brown University.

Baym, N.K. & Boyd, D. M. (2012). Socially mediated publicness: an introduction. *Journal of Broadcasting and Electronic Media*, 56 (3), 320–329.

Baym, N. K. (2015). *Personal connections in the digital age*. Second edition. Polity: Malden, MA. https://williamwolff.org/wp-content/uploads/2015/08/baym-ch1-2015.pdf

Boyd, D.M. & Ellison N.B. (2007). Social Network Sites: Definition, History and Scholarship. *Journal of computer mediated communication* 12, no.1 (2007):210-230. https://www.danah.org/papers/JMCIntro.pdf

Boyd, D.M. & Ellison, N. B. (2008). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13, 210-230. http://dx.doi.org/10.1111/j.1083-6101.2007.00393.x

Bueno, A. (2013). CALICO -Voice Computer-Mediated Communication: Effects on Pronunciation *Journal, Synchronous* (1), 1-20.

Dovchin, S. (2015). Language, multiple authenticities and social media: The online language practices of university students in Mongolia. *Journal of sociolinguistics,* 19(4), 437-459. https://doi.org/10.1111/josl.12134

Heath, M. (2018). Orthography in Social Media: Pragmatic and Prosodic Interpretations of Caps Lock, *Proceedings of the Linguistic Society of America,* 3(1):55. https://doi.org/10.3765/plsa.v3i1.4350

Heath, M. (2017). Interpretations of non-standard capitalization on Twitter. *LSO Working Papers in Linguistics*. *Proceedings from WIGL,* 14, 15-29. http://vanhise.lss.wisc.edu/ling/?q=node/167

Jacquemet, M. (2005). Transidiomatic practices: Language and power in the age of globalization. *Language and Communication* 25 (3), 257–277.

Jorgensen, J.N., Karrebaek, M. S., Madsen, L.M & Moller, J. S. (2011). Poly languaging in superdiversity, *Diversities,* Vol. 13(2)5. ©UNESCO. www.unesco.org/shs/diversities/vol13/issue2/art2

Lamontagne, J. & McCulloch, G. (2017). Wayyy longgg: Orthotactics and phonology in Lengthening on Twitter. *LSA 2017 Annual Meeting*. January 5-8.

Sebba, M. (2007). *Spelling and society: The culture and politics of orthography around the world*, Cambridge: Cambridge University Press.

Murphy, K.M. (2012). Transmodality and temporality in design interactions. *Journal of Pragmatics* 44 (14): 1966–1981.

Piotrovskaya, L. (2003). Emotional prosody and emotive intonation. *International Society of Phonetic Sciences.* https://www.semanticscholar.org/paper/Emotional-Prosody-and-Emotive-Intonation-Piotrovskaya/6847da86eb09585f2eeeb2616b7680df8efa53d5

Robb, A. (2014, April 17). How Capital Letters Became Internet Code for Yelling*. The New Republic.* https://newrepublic.com/article/117390/netiquette-capitalization-how-caps-became-code-yelling

Seargeant, P., Tagg, C. and Ngam Pramuan, W. (2012). Language choice and addressivity strategies in Thai - English social network interactions, *Journal of Sociolinguistics*, 16 (4), 510–531.

Seargeant, P., & Tagg, C. (2014). *The language of social media: Identity and community on the internet,* Palgrave Macmilllan.

Tossell, C. C, Kortum, D., Shepard C., Barg-walkow, L. H., Rahmati, A & Zhong, L. (2012). A longitudinal study of emoticon use in text messaging from smartphones *Computers in Human Behavior,* 28(2):659-663. https://doi.org/10.1016/j.chb.2011.11.012

Watt, H. J. (2010). How does the use of modern communication technology influence language and literacy development? A review. *Contemporary Issues in Communication Science and Disorders*, *37*(Fall), 141-148.

Zappavigna, M. (2011). *Ambient affiliation: A Linguistic Perspective on Twitter*. https://blog.twitter.com/en_us/a/2013/annncng-twttrAnnncng

Zappavigna, M. (2012). *Discourse of Twitter and Social Media: How We Use Language to Create Affiliation on the Web.* London: Bloomsbury.