The Language of Abbreviations and Acronyms

Sebastiaan Wegink, S5734428
https://github.com/SebastiaanWegink

Abstract

More and more people are on social media and therefore a social media language is emerging. However, the difference between the language used on social media and other platforms is becoming increasingly clear. Combining multiple Ngram Viewers and a dataset of words, would help us find the reason why this happens By doing this research, we want to make the problem clearer to people so that they can understand it better. The predicted results confirm that the use of word abbreviations in social media contributes to language difference compared to books. The predicted results are based on related papers, all of which found that word abbreviations played a major role.

1 Introduction

"Increasing numbers of people are communicating with each other through various technologies such as phone-based text messaging, Internet-based instant messaging, synchronous chat, asynchronous discussion forums, and e-mail. Many of these communications are interactive, much like a conversation but conducted at a distance (often both in time and space) and in written form" (Varnhagen et al., 2010). As more and more communication takes place online, a 'social media language' is emerging. However, according to Khalifa (2015) it is important to be aware of this when communicating online. Online communication often features language from a diverse range of individuals, which is largely unedited by proofreaders or publishers. As a result, grammar and spelling can become problematic. Therefore, we think it is important to understand how and why grammar and spelling in online communication are sometimes problematic.

Therefore, in this study we want to find out how the frequency of word acronyms and abbreviations in social media contribute to the difference in language compared to books. To answer this question, we are going to look at several points. First, we are going to look at already existing literature, related work, here we can extract useful information and apply it to our research method and results. Next, we are going to set up a research method to better understand the difference in language between books and social media. For this, using Ngram Viewers is best, we will use Google Ngram Viewer and a Twitter Ngram viewer. In this way, we are going to find out that the usage of word abbreviations in social media contributes significantly to the observed differences in language compared to books. As the frequency of word abbreviations increase within social media platforms, the divergence in language between social media and traditional written forms such as books also increases.

2 Related Work

First of all, it is important to know what abbreviation and acronyms exactly are. "Abbreviation is from the Latin word 'brevis' meaning short. It is a shortened form of a word or phrase. An abbreviation may be made by omitting certain portions from the interior or by cutting off a part. An acronym is an abbreviation formed from the initial components in a phrase or a word. Usually these components are individual letters or parts of words or names. Whereas an abbreviation maybe any type of shortened form, such as words with the middle omitted (for example Dr. for Doctor), an acronym is a word formed from the

first letter or first few letters of each word in a phrase" (Mercy and Christa, 2020).

Now we know what they are, it is also important why people use them. According to Mercy and Christa (2020) abbreviations could make a rise in the 1990s because of the 'new' widespread use of electronic communication, mainly through phones and the internet. The popularity also increased because of the popularity of text messaging. "Short message service (sms), for instance, supports message lengths of 160 characters at most. This brevity gave rise to an informal abbreviation scheme sometimes called textese, with which 10 percent or more of the words in a typical short message service (SMS) message are abbreviated. More recently, twitter, a popular social networking service began driving abbreviation use with 140 character message limits" (Mercy and Christa, 2020). This relates well to our study, the more people are driven to use abbreviations, the more the language will differ with books.

However, the a maximum character length is not the only reason why people use abbreviations and acronyms. Varnhagen et al. (2010) found that people will use abbreviations instead of the whole word because it is simply a quicker way to communicate in brief conversations. They see the use of 'the new language' as a form of linguistic and paralinguistic short cuts. This is interesting in relation to our study. We compare books to social media, on social media are mostly brief conversations and books are the opposite. Books are much more comprehensive and the authors take the time to explain everything well. Varnhagen et al. (2010) also looked at the spelling ability of people who use abbreviations and acronyms, but there was little to none correlation between these things.

Bahri et al. (2023) states that "abbreviations are also used when language shift". They make it clear in their article that there are a semantic, vocabulary, phonological, morphological and syntax changes in social media language. Among these changes, semantic and vocabulary changes are the most significant. For our study, the change in vocabulary is interesting. Bahri et al. (2023) have already indicated that abbreviations are used with a language shift. They have also indicated that the language of social media has shifted. In
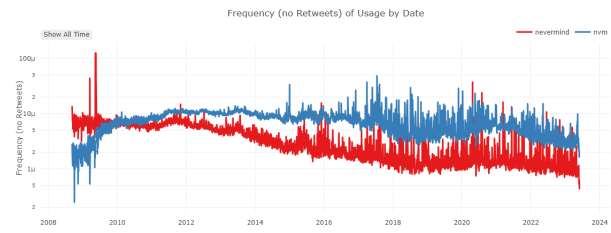
Figure 1: Example of Twitter Ngram - Blue=nvm, Red=nevermind
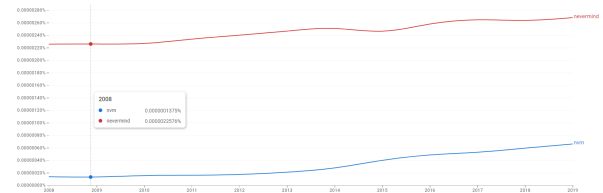
Figure 2: Example of Google Ngram - Blue=nvm, Red=nevermind

other words, part of that vocabulary change are abbreviations. The more a language shifts, the more abbreviations there will be. "The analysis of about a thousand-line corpus revealed that acronyms occupy about 40 percent of the corpus text. So, acronyms are not a small part and should not be overlooked" (Khalifa, 2015). Khalifa (2015) shows that social media language has already shifted a lot. Therefore, it already contains a lot of abbreviations and acronyms.

3 Data

For this study, we will use frequencies of word abbreviations on social media and books to analyze the difference in language usage. Therefore, we need to have a set of dependent and independent variables. In this case, the independent variable is a set of keywords and their abbreviations/acronyms and the dependent variable is the frequency of these keywords and abbreviations, see table 2.

To obtain a comprehensive dataset, we will follow multiple steps. First, we will construct a set of keywords and their corresponding abbreviations commonly used on social media platforms(see table 1). This set of keywords will capture the essence of the difference of language usage on social media compared to books. For the second step we need two tools to measure the frequencies of the words. In this case, we chose for the Google Ngram Viewer and a Twitter Ngram Viewer called storywrangler. To obtain

the dependent variable, we will input our selected keywords and their abbreviations into the Ngram Viewers. During a specified timeframe, this procedure will generate the usage frequencies on Twitter and books. Figure 1 and 2 show how we plan to use the Ngram Viewers. This way, it is easy to see which word, keyword or abbreviation, appears more frequently on social media or in a book. With the Ngram viewers that we use, it is easy to see the exact frequencies, this helps a lot with the research.

After we have collected the data, we do an in-depth analysis of the results. During the analysis we observe and compare the frequencies, the focus is on significant differences in the usage of the selected keywords and their abbreviation between books and social media.

Pre-processing

Because we are using a Twitter Ngram Viewer, we initially came across an issue, retweets. Retweets can inflate the frequency of certain words in social media. However, the Twitter Ngram Viewer from storywrangler has an option to turn off retweets. Because of that, collecting our data is a lot easier.

Table 1 provides a summary of the data that will be used in this study.

Keyword	Abbreviation/acronyms
because	BC
never mind	NVM
be right back	BRB

Table 1: Examples of word abbreviations/acronyms that we can use

4   Predicted Results

Based on the literature, we predict a higher frequency of word abbreviations in social media than books(table 2). The rise of online communication, especially through text messaging and social media platforms with character limits, such as Twitter, drives people to use word abbreviations. Also social media is being seen as a platform for brief conversations. Because of this, people are more likely to use abbreviations. In contrast, books with very extensive and 'slow' text are less likely to use abbreviations.

	Ngram Google	Ngram Twitter
Keyword	freq high	freq low
Abbreviation	freq low	freq high

Table 2: Independent and dependent variables, and their predicted results

Discussion

These finding would confirm the idea that communication platforms, driven by character limits and the need for speed, shifts social media language in a certain way.

The implications of such results extend to understanding how language evolves on social media. It would suggest that social media platforms serve as a unique linguistic environment.

5   Conclusion

This study aimed at answering the question how the frequency of word acronyms and abbreviations in social media contribute to the difference in language compared to books. We provided related work and a possible method to answer this question. The the predicted results confirms that the use of word abbreviations in social media contributes to language differences compared to books. We were also able to see that with more difference in frequency, there is more difference in language between social media and books. This is both consistent with our hypothesis.

Of course, this study also has its limitations. We would have liked to compare the language of social media with something other than books. This is because the language in books, in our view, is not really representative of language used by the average person. We would have preferred to compare it with another platform, such as 'normal' media, articles or newspapers. Unfortunately, this , at the moment, was difficult to research using an Ngram Viewer. There are currently few corpuses that contain this kind of information and also have an Ngram Viewer. Of course, further research can be done on this. Then, if there is an Ngram viewer that has information on other platforms, which is closer to 'everyday' language, you can do good research on the difference with social media language.

References

Bahri, S., E. Manullang, P. Sihombing, and K. Eleazar (2023, 07). Language change in social media. *Randwick International of Social Science Journal 4*, 713–721.

Khalifa, R. (2015, 06). Pragmatic functions of social networks' acronyms. *International Journal of English Language and Linguistics Research 3*, 37–49.

Mercy, E. and O. Christa (2020, 01). Language use in social media and natural language. *International Journal of Humanities and Social Science 10*.

Varnhagen, C. K., G. P. McFall, N. Pugh, L. Routledge, H. Sumida-MacDonald, and T. E. Kwong (2010). Lol: New language and spelling in instant messaging. *Reading and writing 23*, 719–733.