Lexical Variation and Gender in Informal Social Media Environments

Abstract

Language variation is a nuanced and complicated social phenomenon which displays intricate differences between groups of people and reveals diverse patterns of language change and use. As the use of online media as a means of informal communication in the social sphere has become more prominent, and as new forms and variants have entered the English lexicon through this increased usage, a new dimension in which language can vary has come into existence. In the following study, self-reported data from eighty-two social media-using participants is analyzed in order to search for correlations between gender and preference for use of certain abbreviations in informal social media contexts. It was hypothesized that women would noticeably prefer these nonstandard online forms more than men, but the data show few significant differences between men and women in this respect. Women seemed to slightly prefer some abbreviations, while men seemed to slightly prefer others. The means that were calculated for all of the data collected for each group were almost exactly the same, indicating that women overall do not prefer to use nonstandard forms in informal social media contexts more than men do, or vice versa. This way of quantifying preference between gender groups and the results of doing so present an interesting perspective which challenges some commonly held beliefs about typical behaviours of different genders in informal social media environments, and may shed light on sociolinguistic processes that certain lexical variables may be undergoing in these environments.

Introduction

Variation is a ubiquitous property of language that can be observed, in some form, in nearly every context in which language is being used as a communicative tool: there is almost always more

than one way to communicate the same concept. Language variation can often be observed most clearly in the lexicon. It is a phenomenon that is generally very conspicuous in the social relationships which characterize everyday life, as there are countless examples of different social groups each using varying lexical items to refer to the same core concept. Observing and investigating this social variation of language is a central aspect of the field of sociolinguistics.

As the Internet, and more specifically the popularity and influence of social media platforms as casual communicative tools, have grown and diversified, a new linguistic sphere has been born in which language variation is extremely common. The linguistic variation that we observe in social media contexts is quite different from that which is observed in spoken language, as in the overwhelming majority of social media contexts, words are being typed out as opposed to spoken. This is a very different mode of using language, in which alternative processes may be operating. Communication on social media is often very fast-paced and casual, which has led to the creation and usage of many new lexical items, most of which inherently cannot exist outside of social media, or more broadly, outside of contexts in which language is written down or typed, such as abbreviations, acronyms, and emoticons. Some have been created intentionally, while others have arisen organically (Valíková 2014:7). Additionally, many varieties of social media give us the ability to "visually construct" our social networks according to various attributes such as profession or language (Friedrich and Diniz de Figueiredo 2016), which is also a marked difference from social networks which exist organically in real life, where group membership is often not clearly visible and usually requires deliberate investigation in order to be observable. Observing and analyzing this new linguistic environment and its aspects, and the variation and change that are occurring in it, is an emerging topic in modern sociolinguistics.

Sociolinguistic interest in this area is growing, and much research has already been conducted, with interesting and multifaceted findings. Younger people tend to use abbreviations and emoticons

more often (Subramaniam and Razak 2014) and speak about different topics than their older social media-using counterparts (Schwartz et al. 2013). Findings on the relationships between gender and language variation in the social media sphere have been more varied. Some research has shown that women are significantly more inclined to use unpronounceable abbreviations and acronyms than men are (Schler et al. 2006, Bamman, Eisenstein, and Schnoebelen 2014), while other researchers have found only small differences between the kinds of language that men and women use in social media contexts (Palmer 2012). In this paper, data from social media users aged between 17 and 32 will be analyzed in an attempt to understand the relationship, if one or more does exist, between the use of abbreviations in informal social media contexts and gender.

It is worth noting that women are often considered in certain sociolinguistic contexts to be the primary users of nonstandard forms (Labov 2011; Trudgill 1972). Both Labov and Trudgill have described the tendency of men to prefer traditionally standard forms when lexical norms are unclear, and the tendency of women to prefer nonstandard variations when lexical norms are unclear and when nonstandard forms carry covert prestige. Trudgill mentions in his 1972 study that covert prestige reflects "the different sub-cultures [sic] within [this] society" (Trudgill 1972:194). Social media may be able to be treated as a subculture of modern society as a whole, and this idea of covert prestige is very applicable to the linguistic environment that exists in this particular subculture. Overt norms and prestige for lexical variants in social media, especially due to the fact that the dynamics of many social media outlets are in a constant state of flux, are most often *not* "overtly prescribed" (Labov 2011), so it follows that women may be more likely than men to tend towards nonstandard forms, such as acronyms and abbreviations, which may or may not possess covert prestige. This will be quantified, examined, and analyzed in the collected data.

Methodology

  In order to perform this analysis, an online survey was created using Google Forms, and was posted on the researcher's Facebook and Twitter accounts in order to reach the target demographic of people who actively used social media. The survey consisted of two parts. The first part consisted of five demographic questions relating to independent variables: name or pseudonym, gender, birth year, language or languages, and region in which the participant grew up. The second section of the survey asked the participant to describe how often they would use various abbreviations and acronyms in their informal social media posts. It was hypothesized that women would show a statistically significant higher preference for traditionally nonstandard forms than men.

  Fourteen different abbreviations and acronyms acted as dependent variables. These abbreviations were as follows: use of *idk* versus *I don't know*, *ik* versus *I know*, *idc* versus *I don't care*, *u* versus *you*, *ur* versus *your* or *you're*, *w/* versus *with*, *omg* versus *oh my gosh*, *ty* versus *thank you*, *yw* versus *you're welcome*, *imo* versus *in my opinion*, *ily* versus *I love you*, *nvm* versus *never mind, tbh* versus *to be honest,* and *omw* versus *on my way*. These abbreviations and their standard forms are included below in figure 1.1 for easier viewing.

  Participants were asked to describe how often they would use each form, standard or nonstandard, by selecting one of five possible responses from a multiple choice question. The form of the multiple choice question was as follows:

1. I always use [standard form].
2. I use [standard form] more often.
3. I use [standard form] and [nonstandard form] about equally often.
4. I use [nonstandard form] more often.
5. I always use [nonstandard form].

  Participants were then divided up by gender on a spreadsheet. In total, eighty-two people responded to the survey: twenty-nine men and fifty-three women. Attempts were made to even out

the gap between the numbers of male and female participants, but these unfortunately were largely

unsuccessful. In order to record and analyze this data effectively, each multiple choice response

option was assigned a number from one to five. The responses of each participant in the forms of

numbers from one to five for each dependent variable were recorded on this spreadsheet as well. (For

example, if a participant selected option 4 when responding to the multiple choice question regarding

the use of *omg* versus *oh my gosh*, their response would be recorded on their spreadsheet line as

omg/oh my gosh: 4.) The mean of the responses for each traditionally nonstandard form was

calculated for each gender, and this was used for analysis between genders, as comparison of the

means indicated whether or not one group preferred overall to use a certain form more than the other

group did. An overall mean for the use of nonstandard forms as a whole for each group was also

calculated. Thus, through this method, a mean closer to 1 indicates a stronger preference for the

standard form, while a mean closer to 5 indicates a stronger preference for the nonstandard form. All

means were rounded to one decimal place. In the results section, the means for each abbreviation for

each group, as well as the overall mean for each gender and the overall mean for each dependent

variable, will be displayed, analyzed, and discussed in an attempt to find differences between genders

in the usage of standard and nonstandard forms.

Nonstandard form	Standard form (full word or phrase)
idk	I don't know
ik	I know
idc	I don't care
u	you
ur	your, you're
w/	with
omg	oh my gosh, oh my God
ty	thank you

yw	you're welcome
imo	in my opinion
ily	I love you
nvm	never mind
tbh	to be honest
omw	on my way

Figure 1.1: dependent variables (fourteen social media abbreviations and their standard forms.)

Results

Standard form	Nonstandard form	Mean (men)	Mean (women)	Overall mean
I don't know	idk	**3.2**	**3.9**	**3.6**
I know	ik	**1.8**	**2.2**	**2.0**
I don't care	idc	**2.2**	**2.8**	**2.5**
you	u	**2.1**	**2.5**	**2.3**
your, you're	ur	**4.0**	**2.5**	**3.3**
with	w/	**3.9**	**2.9**	**3.4**
oh my gosh, oh my God	omg	**3.9**	**4.2**	**4.1**
thank you	ty	**1.9**	**1.8**	**1.9**
you're welcome	yw	**1.4**	**1.2**	**1.3**
in my opinion	imo	**3.8**	**2.6**	**3.2**
I love you	ily	**2.3**	**2.4**	**2.4**
never mind	nvm	**3.5**	**3.5**	**3.5**
to be honest	tbh	**4.1**	**4.2**	**4.2**
on my way	omw	**3.0**	**3.1**	**3.1**

Figure 1.2: table of means for each dependent variable for each group, and mean for overall preference of all participants across both groups.

Overall mean for nonstandard forms (men)	Overall mean for nonstandard forms (women)	Overall mean for nonstandard forms for both groups
2.9	**2.8**	**2.9**

Figure 1.3: table of means for total use of nonstandard forms for each group, and mean for overall preference of all participants.

As a whole, there appear to be relatively small differences between men and women in terms of preference for each abbreviation. There is a small but still noticeable ($0.5 < x < 1.0$, where x represents the difference between the mean of data from men and that of data from women) difference between gender groups on *idk* and *idc*, with women having a slightly higher preference for the nonstandard form over the standard form. Larger differences ($x \geq 1.0$) between gender groups are also visible for *ur*, *w/*, and *imo*, with men having noticeably higher preferences for each of these nonstandard forms. For *ik, u, omg, ty, yw, ily, nvm, tbh,* and *omw,* the differences in preference between genders are minimal or nonexistent ($x < 0.5$). Among the mean values which display little difference between men and women, each group prefers a somewhat similar number of nonstandard forms. Women show slightly higher preference for *ik, u, omg, ily, tbh,* and *omw* than men do, while men show slightly higher preference for *ty* and *yw* than women do. Preference between groups for *nvm* is equal. When the overall means, calculated across all dependent variables, for each group are looked at, they are almost exactly the same, with men having a very slightly higher preference for nonstandard forms than women. Men have an overall mean of 2.9 and women have an overall mean of 2.8, indicating that both groups prefer and use standard and nonstandard online social media language forms almost completely equally. (However, reducing the data to this statistic hides divisions that we see among the fourteen individual dependent variables, as described above.)

Marked differences in overall preference in both groups for different dependent variables are also noticeable. Both groups show very low preference for *ty* and *yw*, strongly preferring the standard forms *thank you* and *you're welcome,* and both groups show very high preference for *omg* and *tbh,* appearing to use the standard forms *oh my gosh* and *to be honest* significantly less often.

<u>Discussion</u>

These results clearly establish that the hypothesis was incorrect. It was speculated that women would show a noticeably higher preference for nonstandard forms than men, but men actually show a slightly higher preference for nonstandard forms based on the mean calculated for all collected data. Results did vary for each dependent variable, with men preferring some abbreviations while women preferred others. Because of the large amounts of variance between preferences between the groups for each dependent variable, and also due to the almost identical overall means for each group, it appears to be reasonable to conclude that, at least based on the data that was collected, there are no significant differences between men and women in terms of general preference for standard or nonstandard forms in informal social media contexts, but that there may be some individual preferences between genders for certain nonstandard or standard forms.

The results are quite interesting for multiple reasons. First, they defy a standard norm for what is often believed about how men and women type on social media. A commonly held societal belief is that women are much more inclined than men, and are also just quite inclined in general, to use abbreviations, acronyms, and other nonstandard forms when typing on social media, but these results show that this is not the case. When differences between the two groups for the individual dependent variables are examined, most are quite small, and both groups show noticeable preferences for certain nonstandard forms, indicating that women may only prefer certain nonstandard forms, and that the previous judgment, that they show greater preference for abbreviations in general, is likely much too

broad. Also, even when the overall mean of the data collected from men is ignored for a moment, we still see that the mean of standard versus nonstandard form use for women is only 2.8, indicating a slight preference for standard forms over nonstandard forms, which is surprising. Thus, not only does this analysis provide refutation for the notion that women use many more abbreviations and acronyms than men do, it also indicates that women in general may be less inclined to use abbreviations on social media than previously thought.

Second, the results also raise questions about the idea of social media environments as contexts in which norms are not overtly prescribed and covert prestige is operating. Women tend to prefer nonstandard variations when they have covert prestige despite being "considered incorrect" (Leith 1997) and when norms are not overtly prescribed, and to prefer standard forms when prestige is overt (Labov 2011; Trudgill 1972). It was considered in the introduction to this study that social media could be an environment in which nonstandard forms did possess covert prestige and would be preferable to women due to the fact that there were no "overtly prescribed" (Labov 2011) norms, but this is not consistent with the results, which show that women have a very slight overall preference for forms more traditionally considered standard. If all nonstandard forms had covert prestige in social media environments, it would be likely that women would show greater preference for them, and that this would be reflected in quantitative analysis, which it is not. Also, if no norms were overtly prescribed in this environment, the data would likely indicate tendencies of women to gravitate towards nonstandard forms and tendencies of men to gravitate towards standard forms, yet this is not reflected in the results either. It is very possible, due to the nuanced and constant flux of social media environments and those who exist and communicate within them, that some traditionally nonstandard forms are becoming more overtly prescribed as standard and now have overt prestige of some sort, while some traditionally standard forms (e.g. long forms of extremely common abbreviations) may be in the process of gradually becoming more overtly nonstandard. This idea of a

shift in the status of certain standard and nonstandard forms in online contexts may explain women's approximately equal preferences for standard and nonstandard forms, as well as that of men, who tend to prefer traditionally standard forms when norms are not obvious (Labov 2011). If ideas and overt prescriptions of standard and nonstandard forms in social media contexts are shifting in somewhat equal amounts, it follows that an approximately equal distribution of preferences for traditionally standard and traditionally nonstandard forms would be observed in both genders. However, it is worth nothing that, as stated previously, the nature of social media is intrinsically either often or always fluctuating and fast-paced, and is constantly undergoing subtle changes, so precisely tracking, quantifying, and analyzing these processes and changes could be very difficult.

Overall, the data and analysis indicate that there are few significant differences between men and women in terms of the use of standard and nonstandard lexical variants in informal social media environments, and that the general preferences between both groups for one over the other are nearly nonexistent, at least based on the fourteen dependent variables for which data was collected. This is likely attributable to shifting overt prescriptions of standard and nonstandard forms in constantly evolving social media environments.

References

Bamman, David, Jacob Eisenstein and Tyler Schnoebelen (2014) Gender identity and lexical variation in social media. *Journal of Sociolinguistics*, vol. 18, no. 2, pp. 135-60.

Friedrich, Patricia and Eduardo H. Diniz de Figueiredo (2016) *The Sociolinguistics of Digital Englishes.* London: Routledge.

Labov, William (2011) *Principles of Linguistic Change, Vol. 2: Social Factors.* Malden: Wiley-Blackwell.

Leith, Dick (1997) *A Social History of English.* London: Routledge.

Palmer, Jimmy (2012) The Role of Gender on Social Network Websites. University of Central Florida term paper.

Schler, Jonathan et al. (2006) Effects of Age and Gender on Blogging. *Computational Approaches to Analyzing Weblogs*, vol. 6, no. 3, pp. 191-7.

Schwartz, H. Andrew et al. (2013) Personality, Gender, and Age in the Language of Social Media: The Open-Vocabulary Approach. *PLoS ONE*, vol. 8, no. 9, art. no. e73791.

Subramaniam, Vithya and Norizan Abdul Razak (2014) Examining Language Usage and Patterns in Online Conversation: Communication Gap Among Generation Y and Baby Boomers. *Procedia - Social and Behavioral Sciences* 118: 468-74.

Trudgill, Peter (1972) Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language in Society*, vol. 1, no. 2, pp. 179-95.

Valíková, Tereza (2014) Influence of the "netspeak" on modern Internet users' language. Univerzita Palackého v Olomouci (Palacký University, Olomouc) PhD thesis.