



GÖTEBORGS UNIVERSITET
INST FÖR SPRÅK OCH LITTERATURER

ENGLISH

Gender-Related Variation in CMC Language

A Study of Three Linguistic Features on Twitter

Toni Halmetoja

BA thesis
Spring 2013

Supervisor:
Larisa Gustafsson Oldireva
Examiner:
Mats Mobärg

Abstract

Title: Gender-Related Variation in CMC Language: A Study of Three Linguistic Features on Twitter

Author: Toni Halmetoja

Supervisor: Larisa Gustafsson Oldireva

Course: EN1C03, Spring 2013, Department of Languages and Literatures, University of Gothenburg

This study examines the usage of reduced forms, first person subject ellipsis, and alternative capitalization in tweets from a gender perspective, with the data provided by a 20,000 word selection of male and female tweets. The results of the present data analysis for these features are compared to previous findings on male and female language in both spoken as well as written form in some current studies on gender-bound variation in Computer-Mediated Communication (CMC), though there are cases when a direct comparison has been found unworkable. While the present statistical observations, in most cases, are fairly equal between the genders, the following gender-related tendencies have been found manifest: female users tend to use more reduced forms, and to avoid capitalization at the beginning of a sentence. Additionally, female tweets display a wider variety of unique reduced forms. The frequency of reduced forms may be due to the formation of loose social circles on Twitter, while contrasts in capitalization may be related to gender-specific tone or auto-correction. Ultimately, findings of this study may indicate that CMC skews the traditional concept of male and female language use, based on the data examined.

Keywords: alternative capitalization, CMC, first person subject ellipsis, gender, gender-specific, microblog, netspeak, reduced forms, reductions, Twitter, variation.

Table of Contents

1. Introduction.....	1
1.1 What is Twitter?.....	1
2. Aim and Research Question.....	3
3. Theoretical Background.....	4
3.1 Gender Differences in Language Use.....	4
3.2 CMC.....	6
3.3 Reduced Forms	7
3.4 Upper and Lower Case in CMC.....	10
3.5 First Person Subject Ellipsis	11
3.6 Specific Features Investigated	11
4. Data and Methodology.....	12
5. Results and Discussion	13
5.1 Reduced Forms	13
5.2 Alternative Capitalization	21
5.3 First Person Subject Ellipsis	25
6. Conclusion	27
References.....	30
Appendix 1.....	33
The Data.....	33
Appendix 2.....	34
Reduced Forms in Male Tweets	34
Reduced Forms in Female Tweets	36

1. Introduction

Computer-mediated Communication (CMC) is a field of study that is becoming increasingly relevant as CMC becomes more and more integrated into our lives and available to more people. The fact is that at least thirty-two percent of all adults have Internet access at home, and the number increases every year (Gallup 2013, online). The Internet is a constant source of new linguistic tendencies, from nontraditional reduced forms to idioms indecipherable for all but regular users, to the extent that linguists like Crystal argue for it as a new field of study - Crystal calls it *Internet Linguistics* (2011).

There is far too much data available on the Internet for anyone to study comprehensively, so the focus of this study lies on gender-related tendencies in online language. Men and women tend to use language differently, as studied by, for example, Labov (1972), Trudgill (1983), Chambers (2009), Baron (1982), Tannen (1990), Coates (1993), and of course, Lakoff (1975), who was one of the feminist pioneers in the field. These studies have revealed that men use more nonstandard forms, along with more direct language, while women hedge their statements, use Standard English, and often hypercorrect spelling, pronunciation and grammar. While we can hardly apply these results directly to the language of CMC and the social groups of language users engaged in this form of communication, they do validate asking the question: is there a difference in the way men and women use language on the Internet? Research into this has already been made; see, for example Thomson & Murachver (2001), who find that there are gender-related stylistic differences, and Herring & Paolilo (2006), whose results indicate that there are few grammatical differences between the genders online. However, gender-related tendencies change across media; blogs do not use the same type of language as emails do, and emails do not use the same language as instant messaging does. Twitter (<https://twitter.com/>), which this study will focus on, is a relatively new phenomenon, which has to my knowledge not yet been investigated from a gender perspective.

1.1 What is Twitter?

Twitter (<http://www.twitter.com>) is a popular social media website founded in 2006, with over five hundred million active users (Lunden 2012, online). The site allows users to post short, typically blog-like paragraphs of text, viewable by the general public, up to 140 characters,

called “tweets”. The limitation is based on the length of the SMS (Sarno 2009, online) which warrants a comparison of linguistic tendencies between the media. The purpose of Twitter is to quickly share a piece of information, for example, the user’s current location, with other users who subscribe to the particular Twitter user (“followers”). Effectively, Twitter can be considered a “microblog”, similar to but not exactly the same as regular blogs.

While this study does not consider the following mechanics, they should be briefly explained for the interest of giving a better understanding of the medium. Each tweet can, if so desired, be marked with a “hashtag”, a short snippet of text preceded by a pound sign. This tag is a topic and automatically becomes a hyperlink which other users can click on in order to see all tweets featuring the same tag. There is, however, a large amount of creativity involved in how users tag their posts, from signifying a satirical intent with “#irony” to naming a news item the tweet relates to, e.g. “#election2012”. The most popular tags are shown on the main page of Twitter, which encourages users to use the existing ones in order to get more exposure; for example, many individuals will read a tweet about the 2012 elections, but if the tweet is tagged differently than the majority, it would not appear when the so-called “trending”, i.e. popular hashtag is clicked. Hashtags are typically not capitalized.

Additionally, tweets can feature tags that show to whom they reply or are directed. This is marked by the @; if one was to reply to a tweet made by the president of the United States, or simply wish him to read it, it would be tagged “@BarackObama”. The “@” sign creates a hyperlink to the account named and can also be used for other purposes when naming another Twitter user is desirable. Both types of tags count as text considering the 140 character limit.

Finally, tweets can be “retweeted” when someone else’s original post is shared with all of the retweeting user’s followers. This tag is the letters “RT” in front of the previously mentioned @Username tag, after which the original tweet follows; i.e. “RT @BarackObama Four more years.” Below is an example of hashtags, retweets, and references to who the source is with the “@” symbol.



Example 1 – Two tweets from the University of Gothenburg, featuring hashtags.

The first tweet exceeds the character limit and is thus truncated, a generally undesirable occurrence. The second tweet has been retweeted by @got_university and features hashtags, namely *#diabetes*, *#research* and *#sweden*.

2. Aim and Research Question

The purpose of this study is to investigate the occurrence of certain CMC-specific linguistic features, as well as some features of non-standard English on Twitter, and see whether or not there is variation in how these features are related to gender. Because of the limited number of characters per each tweet, as mentioned in section 1.1, tweets could be assumed to show unique characteristics in the interest of text economy, i.e. making a tweet compact enough to contain all the information the user desires it to convey, and by virtue of being a form of CMC. A study of these features would shed light on linguistic devices employed by users of social media, specifically Twitter, in order to condense the information they wish to communicate. This is interesting from the perspective of message organization and structure. Additionally, as many scholars find that males and females use English in different ways, a study of these differences in social media is worth performing. The CMC-specific features will also be analyzed taking the variable of gender into consideration.

This study focuses on a selection of CMC-specific features and one feature of non-standard English common in tweets. The question that we will attempt to answer is whether men or women use either type differently. The CMC features that are investigated are lexical and graphical, namely the use of reduced forms and the use of either exclusive lowercase or uppercase characters. The feature of non-standard English to be investigated is the linguistic characteristic of first person subject ellipsis, common in many forms of discourse. The

theoretical background is, consequently, threefold; it incorporates the research on differences between male and female language use in general, these differences in CMC, and finally, previous findings on the specific features that will be investigated; reduced forms, first person subject ellipsis, and alternate capitalization. To recap, the aims of this study are as follows:

- Investigate the linguistic features of reduced forms, the use of alternate capitalization, and first person subject ellipsis.
- Compare these features through a gender perspective analysis, with comparisons to previous research where possible.

Ultimately, this study is quantitative, with elements of a qualitative analysis. It mainly focuses on observing the frequencies of three language feature occurrences, however, hypotheses on reasons related to gender-specific trends in these frequencies will be presented.

3. Theoretical Background

This study implements findings from three fields of research. The first is that of gender differences in language use, the second is on computer-mediated communication, and the third is that on variation, in the use of reduced forms, alternate capitalization, and the use of first person subject ellipsis. This section surveys previous studies that have been looked into in order to conceptualize the present study in terms of sociolinguistic theory.

3.1 Gender Differences in Language Use

Scholars typically agree that language is used differently by the genders. These differences are not, however, absolute, but “merely probabilistic” (Chambers 2009: 119). With this in mind, we cannot expect to find any universal truths about gender-specific tendencies in language use, especially in a modest-sized study such as this one. We can, however, expect to find patterns. These patterns are described in two studies most often brought up when discussing the sociolinguistic behaviour of men and women; first in Labov’s study of 1972 (as cited in Chambers 2009) and then in Trudgill’s of 1983 (as cited in Chambers 2009). Labov concludes that “women use fewer stigmatized forms than men” and that they adhere more closely to the “prestige pattern” (1972). Trudgill’s conclusion is similar, and he states that women, “allowing for other variables such as age, education and social class” use more standard and, compared to men, more prestigious forms (1983). The same findings - that women tend to use more Standard

English - are repeated in most studies where gender is considered as a sociolinguistic variable (Chambers 2009: 115-116). Coates and Cameron (1988) say that women deviate less from Standard English than men, in virtually every social class in modern times (as cited in Chambers, 2009:116).

Lakoff, in her book *Language and Woman's Place* (1975) offers a feminist point of view of the differences. She claims that, ultimately, female language is less assertive, relying more on tag questions, hedging and politeness (47-50); also, women tend to use hypercorrect grammar and avoid colloquialisms and dialect to a greater degree than men (80, 88, 99), which is a finding relevant to the present study. Colloquial contractions are commonly found on Twitter and CMC in general; for example, shorter forms of phrases such as *wanna* and *gotta* for “want to” and “got to” respectively. These could perhaps be expected to appear less in female language use, even in CMC, but this expectation should be tested against empirical evidence.

There is hardly space to survey findings of all research that has been done on the subject, but there are many important works that should be mentioned in the field of gender-related language studies. Other important works on the subject are Baron (1982), Tannen (1990ab), Coates (1993), Cameron (1992, 2007) and finally Coates and Cameron's 1989 book *Women in their Speech Communities*, all of which examine the differences between male and female language use and the reasons behind these. Lakoff's (1975) study has, however, been chosen as the primary source due to the limited scope of this essay.

Relating these findings to the current study raises a question: what exactly constitutes a stigmatized form in online discourse? While the question is difficult to answer, it can at least be speculated on. There is no unified standard for how language should be used online, but typically, forums, message boards and Usenet, all asynchronous modes of CMC, as Twitter is, are negative on the subjects of poor grammar and excessive use of reduced forms; that is to say, many users on this type of site would criticize these types of usage. Therefore it can be assumed that, at least at first sight, Standard English remains the prestigious form. Crystal (2006: 71,84) also notes that the vast majority of websites and emails are in standard English, if fairly colloquial at times, and that prescriptivism is alive and well on the net, reinforced not only by users but also by a large number of auto-correction systems of spelling and grammar.

There is a fair amount of research on the fact that men and women do not seem to use language in the same way online, but little of it is relevant to the focus of this study; these studies

do not look at the type of features which this study does. Thus, Thomson and Murachver (2001) find that asynchronous online discourse seems to mirror real life, exhibiting a similar scale of differences between the genders; that is, women use more formal English, hedges, polite forms, and so on. Thomson and Murachver claim that “a discriminant analysis showed that it was possible to successfully classify the participants by gender with 91.4% accuracy” (Thomson and Murachver 2001: 193), based on features such as politeness, hedging, emotive and diplomatic language that would appear to lack assertiveness. When it comes to grammatical and stylistic differences, research is sparser, and the results seem to be different. Thus, Herring & Paolilo (2006) found no gender-related features in blog language use, only genre-based ones, when investigating preferences in the use of pronouns. Comparing hypothesized male and female grammatical features, they “found little evidence for systematic gender patterning of the features in a balanced corpus of weblog entries”; they conclude that “genre is a stronger predictor than author gender of the ‘gendered’ stylistic features” (Herring & Paolilo 2006). This illustrates the necessity of remaining within a certain genre or topic when studying gender-related language use.

To conclude this brief look into the research on gender and language, it can be deduced that even though men and women are traditionally considered to use language differently, it is unclear if these differences translate into an online environment, especially in the type of variation that is the focus of this study. The aforementioned research on the matter is vague and presents different conclusions; there seem to be gender-based differences in some features of CMC language use, but none in others. Taking into consideration that, to my knowledge, research into gender-based uses of reduced forms, subject ellipsis and case has not been done, it is difficult to compare our results to those of previous studies.

3.2 CMC

Computer-Mediated Communication, or CMC, is an umbrella term that effectively encompasses all forms of communication done through a computer. Such language, Crystal (2006: 52) observes, is a mixture of the spoken and written, but also has features that neither speech nor writing normally exhibit. Additionally, CMC is constantly changing, with its users adapting their language depending on their purpose, that is to signal group identity, as in the case “hackers” (Crystal 2006: 44-48, 51, 73) or to save space, which is common with media that enforces character limits, such as Twitter or SMS. Linguists have identified the unique characteristics of

computer-mediated communication. Thus, Hård af Segerstad's taxonomy of CMC features includes twenty-six different features typical of online communication (2002: 257). David Crystal, who has written books on what he calls Internet Linguistics, namely *Language and the Internet* (2006) and *Internet Linguistics* (2011), maintains that CMC characteristics are best categorized as differences in vocabulary, orthography, grammar, and pragmatics (2011: 57-77).

A problem with studies on CMC is that the Internet is not a singular, homogenous linguistic platform. Rather, it changes depending on the function in which it is currently being used; so we cannot expect a specific type of variations in all forms of CMC. This study is based on the data provided by Twitter and focuses on three of the most readily observable features, such as the use of reduced forms (both Hård af Segerstad and Crystal note that the Internet abounds in unique reductions), orthographical differences, that is, messages or single words that are either all in capital letters, or lack capitalization, and finally, the grammatical feature of subject ellipsis. While *it*-ellipsis is also common in English (Teddiman 2011: 77), it seems to appear mainly in discourse which Twitter typically does not feature; the typical tweet is not part of a discussion, but instead a short personal statement.

3.3 Reduced Forms

Crystal notes that reduced forms are one of the most remarked-upon features of CMC (2006:89). Reduced forms are used for a multitude of reasons, but as Twitter is an asynchronous medium, it can be assumed that the crucial factor is the character limit, as there is no direct pressure to reply within a given time frame. However, users often send tweets from their mobile phones, and as typing on these is, in general, slower than when using a full keyboard, it is likely that the minimization of effort also plays a role (Hård af Segerstad 2002: 188).

White's (forthcoming) study on variation in the use of reductions and their standardization shows that certain forms are reduced more often, and in various ways, such as "please" being reduced as "plz", "pls", "plse" or "pl" (11). White's study was in a very specific setting, namely during online seminars, with students who were at most novice computer users. Furthermore, students were non-native speakers, so we cannot directly compare this data to ours. Nevertheless, it is worth noting that White concludes that it is the situation, not language (in his study, Vietnamese was the native language of most students) that affects the type of reductions and their frequency. Moreover, the choice of reduction type was made by the non-native students

(“the community”) rather than their native English-speaking teachers. This finding is relevant for the present study as it does seem to show that different communities rapidly form their own set of abbreviations.

This study focuses on different types of reduced words, of which some are common in casual English and some are CMC-specific. Berglund (1999: 38-39) points out one definite area where men use common reduced forms more often than women do; in her analysis of the British National Corpus, she claims that men use the reduced form of “going to” – “gonna” more than women do.

Even though Hård af Segerstad’s taxonomy of CMC variation is useful, it does not go in-depth into different types of reduced forms; she does not classify them in further detail than conventional and unconventional abbreviations and consonant writing. Other scholars make further distinctions; thus Lee (2002: 8-10) points out that unlike in conventional writing, reductions in CMC are not restricted to acronyms and initialisms. Further, Hård af Segerstad’s classification of reduced forms found in CMC includes sentence acronyms, letter and number homophones, words combining both, reductions of individual words and combinations of the aforementioned.

Using Lee’s research as their base, Lotherington and Xu (2004: 314) conclude, from their small-scale study at York University, that reduced forms used in digital environments are acronyms, abbreviations, homophones, and “hybridized codes” i.e. combinations of letter and number homophones. It is worth noting that this study includes both English and Chinese features, some of which are not relevant for this mono-language study, such as code switching. These features have not been included in our classifications.

Yus (2011: 176-179) has a similar list of features, and specific phonetic spellings. Thus, he distinguishes different types of phonetic spellings, but this study only considers them as a group. Besides, Yus’ study includes abbreviations, acronyms, and clippings.

Crystal (2006: 90, 262-263) has also classified, although rather briefly, the variety of reduced forms. His list includes full sentence acronyms, reduced individual words (such as *p/s* for “please”), and letter or number homophones. Interestingly, he also considers the use of smileys as shorthand for emotions. Further, Crystal speculates on text economy in SMS messages, which is relevant to this study due to the similar character limit, noting that “users seem to be aware of the information value of consonants as opposed to vowels”, and he uses this

to explain what he finds to be the most common reduced form, namely words with removed vowels, such as *tXt* for “text”. Crystal also lists other characteristics, but they are not relevant to the present study.

To summarize, the previous research has classified reduced forms common in CMC as follows (with examples, if provided by authors):

Hård af Segerstad (2002)

- Conventional abbreviations
- Unconventional abbreviations
- Consonant writing

Lee (2002):

- Acronym of sentence - GTG (“Got To Go”),
- Letter homophone - U (“you’), R (“are”)
- Number homophone - 99(“Nite Nite” [’good night’])
- Combination of letter and number homophone - b4 (“before’)
- Reduction of individual word - tml (“tomorrow”), coz/cos (“because”)

Crystal (2006):

- Full sentence acronyms - GTG (“got to go”)
- Reduced individual words - pls (“please”)
- Letter/number homophones - L8R (“later”)

Yus (2011):

- Phonetic orthography
- Abbreviations
- Acronyms
- Clippings

Lotherington and Xu (2004):

- Homophonic spellings - u (“You”)
- Truncated homophonic spellings - k (“Ok”)
- Borrowed shorthand - w (“with”)
- Reduced spellings needing a gloss - wo (“without”)
- Simplified but recognizable spellings - nite (“night”)
- Alphanumeric rebus writing - sk8 (“skate”) 4 u (“for you”)

The present study will use a custom taxonomy which combines the most common features of the aforementioned classifications. This is because detailed classifications are surplus, given the limited scope of data analyzed in this essay, and characteristics mentioned in all of the previous studies are the most numerous ones in the tweets under study.

3.4 Upper and Lower Case in CMC

All-caps or no-caps text can create an impression of loudness and shouting in CMC (Driscoll & Brizee 2013, online), and could be seen as a replacement for intonation and the loudness variable, which CMC, as Crystal observes, lacks (2006: 37). While Crystal’s statement is a good definition of the feature, which should be kept in mind in the case of all-caps messages and especially those lacking capitalization, we cannot overlook the fact that most messages can be spelled in these ways due to various reasons: from mechanical issues in the form of malfunctioning keyboards or a lack of proper training in keyboard use, to laziness or hastiness. Indeed, both Crystal and Hård af Segerstad arrive at a similar conclusion. Hård af Segerstad states that “when typing on a computer keyboard, typing shift plus the letter key requires effort” (2002: 223) which can be a possible reason for the lack of capitalization, while Crystal concludes that there is a “strong tendency” to type in all lower caps in order to “save a keystroke”, i.e. to avoid having to press the shift key (2006: 90). This study will verify if these observations hold, and whether there are any gender-related differences in the use of caps.

3.5 First Person Subject Ellipsis

The Longman Grammar of Spoken and Written English states that subject ellipsis - that is, the act of leaving out the first person pronoun - is fairly common in spoken discourse (Biber et al. 1999: 1048). This typically occurs in sentences where the first word would otherwise be either simply “I”, but also when the first two words would become a contraction in spoken language, as in the case of “I’m”. Occasionally, “we” is also excluded, but this occurs less often, likely because it reduces clarity if all participants are not explicitly identified. Teddeman (2011: 71-88) observes that this is the most typical feature of ellipsis in private and public spoken dialogue as well as in correspondence, with *it*-ellipsis a close second. Androutsopoulos and Schmidt (as cited in Hård af Segerstad 2002) show that the subject pronoun is also the most common feature to remove in SMS communication. As SMS relies on limited-character messages, much like Twitter does, there is good reason to expect tweets to exhibit this grammatical reduction.

Nariyama (2004: 239) speculates that subject ellipsis only happens when the subject is obvious from the context, and this, too, would apply to both Twitter and SMS communication, given that one can see who made the tweet or sent the message, and because of this the subject may not be needed. Yus notes that the first person subject pronoun is the most frequently elided element in chat, as well, and states that this happens “since chat room messages always contain the *nick* [i.e. username] of the user at the beginning” (2011: 179). A similar conclusion was reached earlier by Lee (2002), who, in her study of one-on-one CMC, says that the subject is omitted due to both participants being explicitly identified.

3.6 Specific Features Investigated

Considering the findings of the above mentioned previous research and personal observations of commonly occurring features, this study will examine the use of the following features in relation to the variable of gender (examples provided by my own corpus of tweets):

- Reduced forms:
 - Individual reduced words: (*p/s* = “please”)
 - Colloquial informal contractions: (*wanna* = “want to”)
 - Symbols replacing words or phrases: (<3 = “love”)
 - Acronyms and initialisms: (*lol* = “laughing out loud”)
 - Number homophones: (*b4* = “before”)

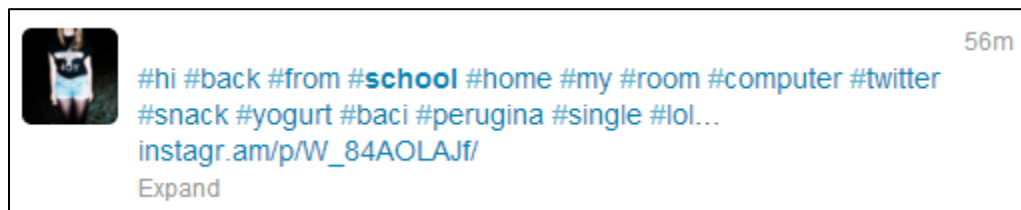
- Letter homophones: (c = “see”)
- Upper/lower case
 - All upper single words or sentences: (“LOL NO”)
 - All lower single words or sentences: (“thought i was doing worse than i actually am!!! :D”)
- First person subject ellipsis: (“learned something today”)

4. Data and Methodology

The data for this study was drawn from tweets in the months of February and March, 2013. In the interest of avoiding contrasting topic-specific vocabulary, only tweets centering on a specific topic, as per their hashtags, were chosen. This hashtag is #School, which was deemed to be a suitably neutral topic, in addition to containing a large amount of data. The #School hashtag is the tag for topics on the subject of school of any level and is typically used by a wide age range of users, from middle school to university students, which should lessen the impact of age.

A problem with this was that the tag attracted considerably more female participation, and in the interest of balanced proportion in data, a large amount of this had to be discarded; however, as the decision of what tweets to discard was done randomly, by numbering each tweet and using a random number generator to choose which to remove, it should not have an effect on the results. Gender identity was based on the user picture and the username; if both indicated male or female, the user was classified as such. Ambiguous cases were not included.

A large amount of data had to be discarded as many tweets consist exclusively of hashtags, and these were unusable for this study, as illustrated below (see example 2). Additionally, tweets consisting entirely of onomatopoeia (e.g. “hahaha”), numbers, proper nouns (e.g. “University of Alabama!”) or single words of uncertain language (e.g. “juuu”) were not included.



Example 2 – A discarded tweet consisting only of hashtags, with the username removed.

The total amount of data consists of roughly 10,000 words from male tweets and 10,000 words from female tweets. However, the actual scope of analyzed material is somewhat lower as the 20,000 words include hashtags, usernames and hyperlinks. The matter of whether or not hashtags should be considered as part of discourse is in itself difficult as oftentimes these are integrated into the message. The decision made was that if a hashtag is part of the sentence structure, adding to the sentence in a grammatical way, such as in “*#chillin* at school today”, it was considered to be part of the message rather than exclusively a tag. Hashtags can be and are sometimes capitalized; so if a message starts with a grammatically correct but lowercase hashtag, as in the example above, it was considered to be a lowercase sentence. Similarly, fully capitalized integrated hashtags were counted as fully capitalized words, typically used for emphasis. On the other hand, hashtags that are clearly meant to be topics, such as in “I wanna go home *#school*” were not included in the statistics.

The data was then analyzed manually, identifying the gender of each user by their picture and name. Finally, each tweet was checked for the features listed in 3.6.

5. Results and Discussion

The results show that there are certain gender-related tendencies in the patterns under study. These differences are not great, and would require a considerably larger corpus to clearly reveal gender-specific patterns in them. What follows is a presentation of the results provided by the selection of data compiled for this study.

5.1 Reduced Forms

A total of roughly ten thousand words per gender has provided a total of 438 reduced forms for both genders. Roughly 2.2% of words per tweet used were reduced forms, or 22 words per thousand. This is less than one might initially expect, which can be explained by the fact that Twitter is used primarily through mobile phones, which, in modern times, typically autocomplete and correct words. While this may seem like a disappointing result, it is not so, as this means that these reduced forms were deliberately used by the tweeters, rather than relying on a phone’s autocomplete.

While a significant number of these reduced forms used are very much in common use, such as *wanna* and *gotta*, especially in spoken language, many forms are CMC-specific, such as symbolic reductions as > for “is/are better than”, or <3 meaning love or strong positive connotations. Additionally, expressions of emotion such as *LOL* for “laughing out loud” were very common among both genders.

There are, as expected, topic-specific abbreviations observable, in acronyms and initialisms for university or high school names and clipped forms for subjects. These include examples such as *PE* for “Physical Education”, along with *bio* and *chem* for “biology” and “chemistry” respectively.

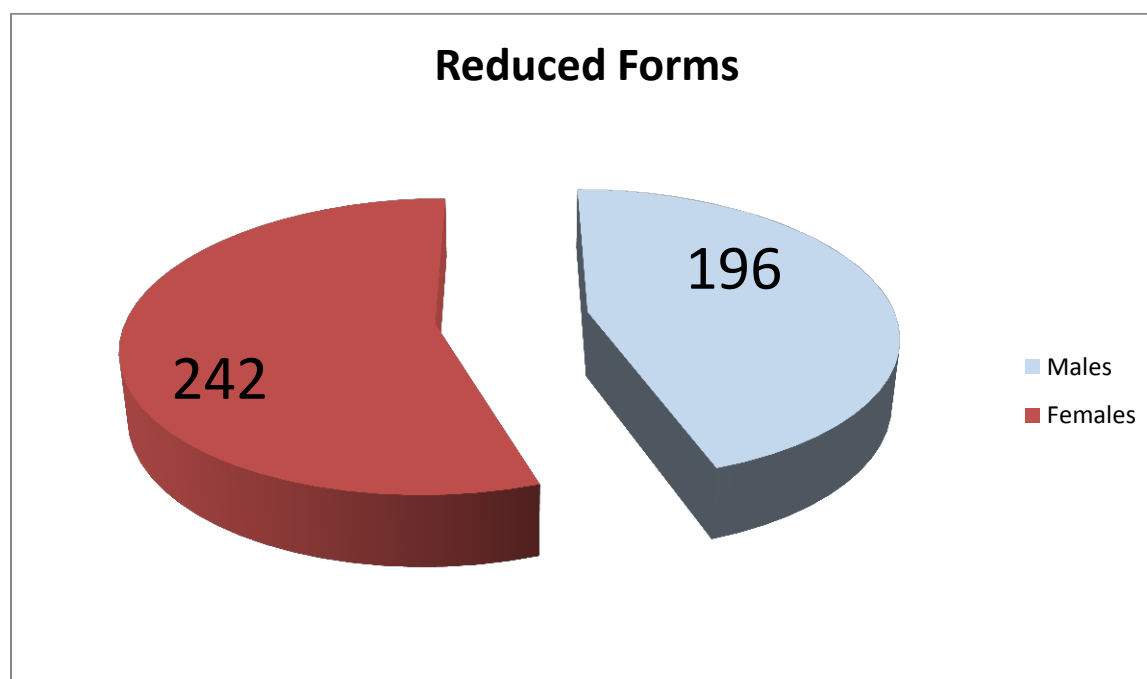


Figure 1 – Chart of reduced forms usage in male and female tweets.

As the data indicates, female individuals seem to use slightly more reduced forms. Out of ten thousand words used by males, reduced forms constitute 2% of the total, while the equivalent number for females is 2.4%, thus making the difference a fraction of a percentage point: 0.4%. While this is not enough for definite claims, there are tendencies to observe here. One difference that can be observed is that females use a wider variety of reduced forms: 137 different words in female tweets, as opposed to the male 119 separate, unique abbreviations in male tweets (see Appendix 2). A large number of these are varied spellings of the same word; the acronym */o/* can, for example, be spelled with a multiple os, as in “loool”. This is usually done for emphasis.

Additionally, words such as *because* can be reduced to *cus* or *cuZ*. All variations are, for the purposes of this study, counted as separate forms. Alternative capitalization of a reduced form is not considered a separate form for these results, as it will be discussed in section 5.2

Table 1 presents the results of the reduced forms counted and classified as proposed in section 3.6.

Table 1: Classifications of Reduced Forms

Type of Reduced Form	Males	Females
Individual reduced forms	84	99
Informal contraction	31	38
Symbols	8	15
Acronyms/initialisms	51	61
Number homophones	10	12
Letter homophones	10	17

These numbers are fairly similar. However, even though a larger corpus would be required to draw any definite conclusions, we can observe a tendency: females appear to use reduced forms more than men. Additionally, the amount of unique words is higher, displaying verbal creativity, and a more pronounced tendency to reduce words, especially individual ones. Below, we will look at some of these words in the context of tweets, with usernames and hyperlinks removed, but otherwise intact.

Tables 2-6 list the five most common reduced forms for both genders, per type, with explanations of their meanings; tables are followed by a brief discussion of each type. A full set of examples can be found in the appendix, although even in spite of the context, some acronyms could not be reliably explained.

Table 2: Reduced forms of individual words

Males	Occurrences	Females	Occurrences
to (too)	4	tho (although)	6
w/ (with)	4	chillin' (chilling)	4
chillin' (chilling)	3	jus (just)	3
dis (this)	3	pls (please)	3
ur (your)	3	to (too)	3

Table 2 shows that both males and females reduce “too” into *to*. This could also be a typo; however, considering that reduction of other words occurs frequently, there is little reason to doubt it being a genuine reduced form. This applies also to *w/* in the male tweets, which less often appears without a slash, meaning “with”. Other examples of reductions in the above examples are *your*, *tho*, and *jus*, and we can note that these appear to be the most common reductions of individual words for both genders. Finally, the male *dis* appears to imitate a spoken-like pronunciation, a tendency often observed in male tweets, which also feature words such as *da* meaning “the”, and *cos/cuz* for “because”. While it is impossible to say for sure if these are approximations of non-standard pronunciation, such a claim would support findings of previous studies of “real world” speech, namely the theory that men use dialects and colloquialisms for covert prestige. The last word in Table 2 is the female *p/s* for “please”. It is unsurprising that females use this polite form more than males, as females tend to aim for politeness in language (Lakoff 1975: 50).

Table 3: Acronyms and initialisms

Males	Occurrences	Females	Occurrences
lol (laughing out loud)	15	lol (laughing out loud)	20
hw (homework)	4	omg (oh my god)	9
Lmao (laughing my ass off)	3	AP (advanced placement)	2
omg (oh my god)	3	idk (I don't know)	2
af, a'f (as fuck)	3	smh (shaking my head)	2

In Table 3, *lol* is the most common abbreviation for both genders, used for expressing anything from the literal meaning of “laughing out loud” to mild amusement. Similar words that express emotions in these examples are *lmao* for more intensive laughter, *omg* as the typical exclamative “oh my god” and *smh*, which, as “shaking my head” to represent disapproval or disappointment. Both genders use these, as these phrases add tone and emotion to written communication, and they appear, as the table illustrates, rather often. As such, the reduction saves both time and space. Finally, we can observe the reduction of topic-specific words in the male *hw* for “homework”. The female *AP*, for “advanced placement”, may seem topic-specific, but is, in contrast, an authentic abbreviation meaning a more advanced class than normal. These forms seem to have become standardized, as there is no variation to how either is reduced, which is a result similar to

that of White’s conclusion, that certain topic-specific words become standardized within a community (2012).

Table 4: Colloquial contractions

Males	Occurrences	Females	Occurrences
wanna (want to)	10	wanna (want to)	16
gonna (going to)	5	gonna (going to)	14
hella (hell of a [lot])	5	gotta (got to)	4
imma (I’m going to)	2	ain’t (is not)	2
kinda (kind of)	2	ima (I’m going to)	1

Colloquial contractions, as Table 4 shows, are used more often by females. While the sample size is relatively small, this appears to contradict the notion of females using less reduced non-standard forms, as all of these examples are informal contractions, not used in written Standard English. However, as CMC is a hybrid of spoken and written language, it is possible that these forms simply “slip in”; alternatively, the mode of communication is so casual that females feel no need to stick to standard forms. Whatever the case, these forms are common on Twitter, more likely used to save space and time, than to mark or ignore any form of prestige.

Table 5: Letter and number homophones

Males	Occurrences	Females	Occurrences
2 (to)	6	2 (to)	8
2 (too)	1	2moro (tomorrow)	1
u (you)	6	u (you)	10
n (and)	2	n (and)	3

Judging from Table 5, Letter and number homophones are used by both genders in similar ways. Using the number 2 to approximate the “to” or “too” sound is the most common for number homophones, along with the letter homophone *u* for the pronoun “you”. These can safely be assumed to be used for the sake of economy of characters and time, but they appear far less frequently than expected. It seems that homophone-based “SMS-speak” is not used as much on Twitter.

Table 6: Symbols replacing words or phrases

Males	Occurrences	Females	Occurrences
& (and)	3	& (and)	6
<3 (love/affection)	2	<3 (love/affection)	5

As for the occurrence of symbols, shown in Table 6, the most common one is the &-symbol for “and”, which is often used to speed up typing in academic papers; hence its appearance here is no surprise. Additionally, the more CMC-specific <3, intended to look like a heart, means a strong affection or love for the message it follows. This is used more by the females, likely as part of their more emotive language, emoticons and symbols being one of the few ways to show emotion in chat. However, the sample sizes are far too small to definitely make such a conclusion.

The next section features the actual tweets, with words of each type in-context. Each selection is followed by a discussion of the features.

- 1) Male: Oh you knoww just cruisin'
- 2) Male: Way to not score in final 8 mins #badgers , now it's time to try to get some of this damn paper wrote #school
- 3) Male: @Username aite im walking tho need that lil gas 2 make 2 cdale n da morning #School
- 4) Female: jus gonna start not giving a care about a lot of stuff. #SCHOOL #BASKTBALL #FAMILY the only things on my mind perio
- 5) Female: goodday school enjoy hi! a very #goodday 2 all just came frm #school ppl #enjoy
- 6) Female: Striving to excel tho aren't I? #ryburn #school #gay #motto

In these examples, some forms are reduced in order to convey nonstandard pronunciation, such as the leaving out of the “g” in “cruising” in example (1). This can be readily observed on Twitter, understandably, as it saves a character, but in some cases the missing letter is replaced by an apostrophe, resulting in the same number of characters. In these cases, the word must be intentionally reduced to signify a seemingly relaxed (at least spelling-wise) attitude, as in the case of *cruisin'*. Both the males and the females do this about equally.

Further, in example (2) the word “minutes” is reduced to *mins*, a common reduction that saves space. This also applies to the words “though” reduced to *tho* and “little” to *lil* in example (3). In the female tweets, “just” is reduced to *jus* in example (4), and the words “from” and “people” to *frm* and *pple* respectively. The females appear to use these forms more, even though this difference is slight. These forms can be assumed to be employed simply for text economy as they hardly serve other purposes.

The second most common form of reduction is the use of acronyms and initialisms, both standard and “netspeak”-specific, as in the following examples.

- 7) Male: All this hw.. #school
- 8) Male: Yes only got #school on monday & friday this whole loving week! Lol
- 9) Female: @Username being mean trying to get stupid pictures of me! lol #school
- 10) Female: Looking rough yolo #carly #school #rough #look #horrible #stupid #vile
#like4like #likeforlike

Here, in example (7), the word “homework” becomes the initialism *hw*, and in tweets (8) and (9) we can observe the common netspeak acronym *lol* meaning, “laughing out loud”. Example (10) features another acronym common online and among young people in real life, *yolo*, meaning “you only live once”. There appears, however, to be little observable difference in the way males and females use these forms.

A common type of reduction is the reduction of “want to”, “going to” and “got to” into *wanna*, *gonna* and *gotta*. When combined, these similar reductions appear 20 times in the male tweets and 34 in the female ones. A similar phenomenon was also observed with the reduction *hella* from “hell of a lot”. These appear to be mainly pronunciation-based reductions, as in the following examples.

- 11) Male: I just wanna go outside #suchaniceday #stupid #school
- 12) Male: The shit I gotta be doing for #school. #chemistry #struggle
#whydoineedtodrawshitin3D
- 13) Female: gonna do my project now. #school

14) Female: Wen I want school to come it take 4ever but when I don't wana go 2 school its here that doesn't make

In these examples, we can observe the use of *wanna*, *gotta*, *gonna* from “going to” and a further reduced version, the female *wana* in example (14). These are features of colloquial English common in CMC, and as such one might expect women to use them less. This is, however, not the case, further supporting the fact that the genre of communication seems to decide what forms are used, rather than gender.

Tweets of both genders provide a small number of letter and number homophones. Even though males, in our examples, use more letter homophones, the numbers are too small to make a conclusion about gender-related preference. Below are two examples.

15) Male: @Username cool c u there m8! #walking #school #boring

16) Male: When u have exam yet u dun study #shady #rap #school

17) Female: Wen I want school to come it take 4ever but when I don't wana go 2 school its here that doesn't make sense cause 1day = only 24hours

18) Female: When the teachers gives u 4 questions as homework but it turns out to b 1) a. b. c. 2) a. b. c. Questions -_- like how about no #school

Here, we see that the reasons for uses are fairly similar. Examples (15) and (17) both contain number homophones which save space, as *m8* for “mate” and *4ever* for “forever”. *2* for “to” in example (17) is repeatedly observed in the writing of both genders. Examples (15), (16) and (17), in turn, use letter homophones. Of these *u* for “you” is the most common, but as in example (15), *c* for “see” appears. This form is likely a way to save characters, even though none of the tweets approaches the character limit. It also shows a certain amount of verbal creativity in the medium.

Finally, there are some examples of words being abbreviated to certain symbols, some as a form of emoticons and others simply standard symbols, such as *&* for “and”. These are illustrated in the examples below.

19) Male: In love with this photo ≤3 #school #girls #instagirls #instawe #instacool #beautiful #beauty #cute

20) Female: Just got home ≤ 3 #school #monday #fun

In examples (19) and (20), the symbol or emoticon ≤ 3 , meant to look like a sideways stylized heart, stands for “love”, or a strong affection for something. It is not entirely similar to normal smileys, which, rather than any specific word, symbolize a general emotion. Instead, ≤ 3 is often used to replace the word “love” entirely, as in example (20), where the hashtag is part of the sentence “Just got home, love school”. Example (19), however, uses both the word *love* and the emoticon, to emphasize enjoyment of the photo linked. These types of reduction were also used roughly equally by both genders, with the standard &-symbol used to save characters, while ≤ 3 is used for both text economy and to emphasize emotion.

What can be concluded about the use of reduced forms is that the females use them more in the studied material. Why this is so, is somewhat unclear; it may be a question of saving time and effort. As women are more active on this topic, that of school, they might experience having less time to write each individual tweet, and thus rely on reduced forms. Alternatively, this may simply signify a more “relaxed” attitude towards CMC, since Twitter is a very casual website, as most of social media is. It may also be that women perceive these forms as the standard in CMC, and thus apply them more than men, due to female language being closer to the standard. As for the feeling of insecurity that leads to hypercorrect grammar in female language, a phenomenon discussed in various studies (e.g. in Coates 2004), this gender-related feature is not characteristic of online communication, with its detached and less personal nature. While this does not directly explain why females use reduced forms more than males, it would explain why they appear more relaxed than in real-world face to face speech. CMC exercises little social pressure on individuals, as each user is, even if a real name is used, relatively anonymous.

5.2 Alternative Capitalization

Males and females were found to use fully capitalized words and sentences roughly equally. 97 out of ten thousand words were capitalized by male users, while the number for females was lower: 90 words in ten thousand. Only four fully capitalized tweets were found among the male samples and six were provided by the females. When it comes to sentences lacking capitalization, the difference is more significant. Here, occurrences among males amount to 75 examples, while females neglected or avoided capitalizing their sentences 120 times. Finally, individual words

lacking proper capitalization consist mainly of proper nouns, such as the name of the website, “Twitter”, typed in all lowercase. These are not as common as fully capitalized words, possibly due to auto-correction on mobile phones. Variation in capitalization is illustrated in Figure 2.

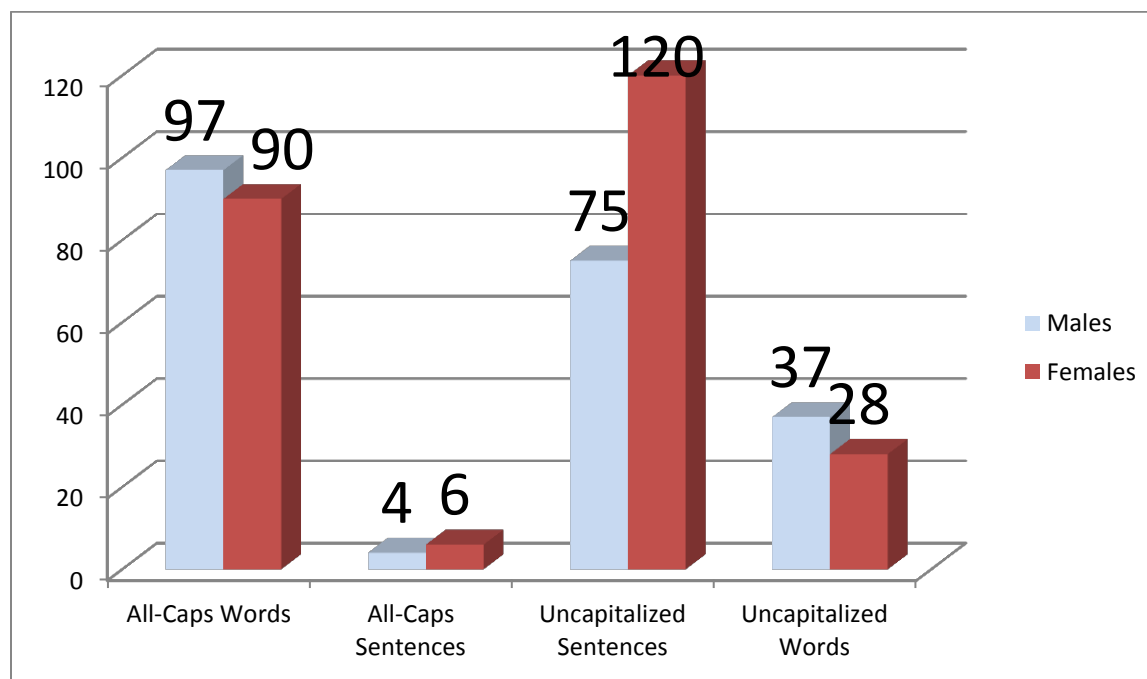


Figure 2 – Variation in capitalization by type and gender.

All-caps words were found to be used for varying purposes by both genders. In constructions such as the male example of “I am legit PISSED that spring break is over”, capitalization is clearly used for emphasis, seemingly compensating for the lack of tone and loudness in CMC. Assumedly, were this sentence spoken out loud, the word “pissed” would be emphasized and louder. Additionally, loudness is seemingly emulated in capitalizing and lengthening the acronym “LOL (laughing out loud)” in tweets such as “LOOOOL How Can I Go Roberts and See this Picture”. While LOL would normally, as an acronym, be capitalized, most occurrences were in lowercase, and this kind of capitalization underlines the loudness of laughter. Female examples of capitalization for the sake of emphasis, tone and loudness are similar, with examples such as “get resource officers back in ALL schools!” and “This made me laugh but in all seriousness school tomorrow :(WHY?” This supports Crystal’s observation of CMC language emulating the prosody and paralinguistic features of spoken language by the use of capitalization (2006: 37).

Capitalization, especially of full sentences, was also found to be used to give the impression of shouting, as mentioned by Crystal (2006: 37) and Driscoll & Brizee (2013, online). This tendency was observed in both male and female tweets. Male examples include tweets such as “HALLELUJAH. I'M PRETTY MUCH DONE MY ART HISTORY ESSAY AHHHH... now... Italian script, here I come”. Female examples are similar, as in “ahahah HAVE FUN AT SCHOOL TOMORROW. I'll be chillin, sleeping in, and watchin Netflix”. Some examples are hard to classify, however, and could either be intentional “shouting”, or simply the result of a hasty tweet made with the caps lock key active, such as the short “I GOT A STICKER #school #goodgirl #sticker”; however, as the tweet is followed by lowercase hashtags, capitalization of the main tweet seems to be intentional.

The most significant difference is manifest, as mentioned above, in the improper capitalization of sentences, i.e. not capitalizing the first letter of a sentence. Here, females outnumber men by 120 to 75. It is, however, difficult to speculate on the reasons behind this. The male and female examples appear to be fairly similar, as in the following examples:

21) Male: #school just finished, TGIF mofucker!!!!!!

22) Male: photo shop you make my life soooo much easier. #photoshop #life #easier #school #social #poster #timesaver

23) Female: cnt u till i was #food #school

24) Female: teachers: no pressure but these are the most important exams in the whole of your life ever... #PRESSURE #exams #school #scary

Examples (21) and (24) feature capitalization, but not at the beginning of the tweet. This would seem to indicate that the users understand the rules of capitalization – either on mobile phones or personal computer. Example (21) could be assumed to be uncapitalized due the hashtag present, but these can also be capitalized, as in example (25), below.

25) #School - in class -_-

It would be reasonable to assume that in other cases, the capital letters are being intentionally left out, to follow the typical hashtag format.

Finally, words that would normally be capitalized, mainly proper nouns, are lowercase, as in the examples below.

26) Male: charlie sheen una poo logetic about school smear... #CharlieSheen #dog #poo
#school #daughter #news #hot

27) Male: Im most def going to church next sunday, i need a balance #God #School #Family
& #Hoop

28) Female: The fact that about half our year did the harlem shake was just to weird #school

29) Female: Ugh it's 10:18 am and school won't end faster #dying #school #help Btw I ment
to send this friday

In example (26) the name Charlie Sheen is left uncapitalized, along with the rest of the tweet, in contrast to the hashtag with the same name. The reason for this is unclear, and whether or not it is intentional is difficult to tell. The capitalization of the hashtag may have been autocorrected. In examples (27) and (29), the names of weekdays are left uncapitalized, but in example (27) certain hashtags are capitalized. Again, this would appear to be the result of a mobile phone or computer capitalizing the beginning of each “sentence” automatically, while the words not autocorrected have been left untouched. Finally, example (28) has the name of said dance – the “Harlem Shake” – uncapitalized. This appears to be a saving effort case, as the sentence starts correctly with a capital letter.

As upper-case words indicate shouting, it is possible that female users, to a greater degree than males, prefer the tone of lowercase. Lakoff (1975: 50) observes that women use more polite forms, hedges and indirect requests; ultimately, she claims that women’s language is less assertive. Considering that using all capital letters creates an effect of shouting in text and is generally discouraged by style guides (e.g. by Driscoll & Brizee 2013, online), it may be that females want to avoid this impression more than men do. However, if this is the case, the results seem to contradict another of Lakoff’s findings, namely that women tend to hyper-correct their language. Her findings were based on spoken language and may not apply to the medium under study, especially considering that the age and social groups of the females in the present study are different from those analyzed by Lakoff. Furthermore, it is possible that the present results are not comparable with those of previous studies as they are CMC-specific; some devices used

for CMC, such as mobile phones, will automatically correct capitalization while others, such as computers, may not. It is possible that the females in our study prefer tweeting through their computers rather than mobile devices, hence utilizing less auto-correction.

Marshal (2009, online) states that lowercase can be used to create a “funky alternative to regular text” or to “communicate a demure loveliness”, which are both alternative explanations; it is possible that females wish to communicate this kind of impression more than males do.

Ultimately, it is difficult to make a definite conclusion about case-related differences as many tweets could be lowercase due to the aforementioned unfamiliarity with keyboards and phones, due to the use of older devices with less advanced auto-correction, or simply due to the economy of efforts, which is described by Crystal as the “tendency to save a key-stroke” (2006: 90). Additionally, there is the matter of auto-correction; many sentences may have originally been lowercase but capitalized by auto-correction, which would also explain why certain hashtags are capitalized while others are not, along with the inconsistent capitalization of proper nouns. So, it is hard to give a definite explanation as to why the females use less capitalization without knowing which device the tweet was made through, computer or mobile phone, what type of either, and other factors such as which browser was used if the tweet was made on computers are also worth considering. The same problem would emerge if one is to study spelling, or other aspects of CMC.

5.3 First Person Subject Ellipsis

A fair amount of messages which are reduced by removal of the first person subject were observed. Unlike in chat between two persons or multiple user chatrooms, Twitter is an asynchronous medium, which means that there is little pressure to save time on messages. Saving time is a factor in removing words such as pronouns in the case of the user already being identified by his username, but this does not directly apply to Twitter. It is possible that users still feel the need to post their tweets quickly, especially in the case of replies, to avoid seeming disinterested. Equally likely, users may be in a hurry and attempt to minimize the amount of time spent on typing out messages on their mobile phones. Of course, this tendency is common outside the Internet as well, as pointed out by grammars such as the *Longman Grammar of Spoken and Written English* (Biber et al. 1999: 1048), and it comes as no surprise to see it used online.

The results on first person subject ellipsis were fairly equally distributed between male and female users. 118 occurrences were found in male tweets, while female numbers totaled 130, as illustrated by the following chart in Figure 5.

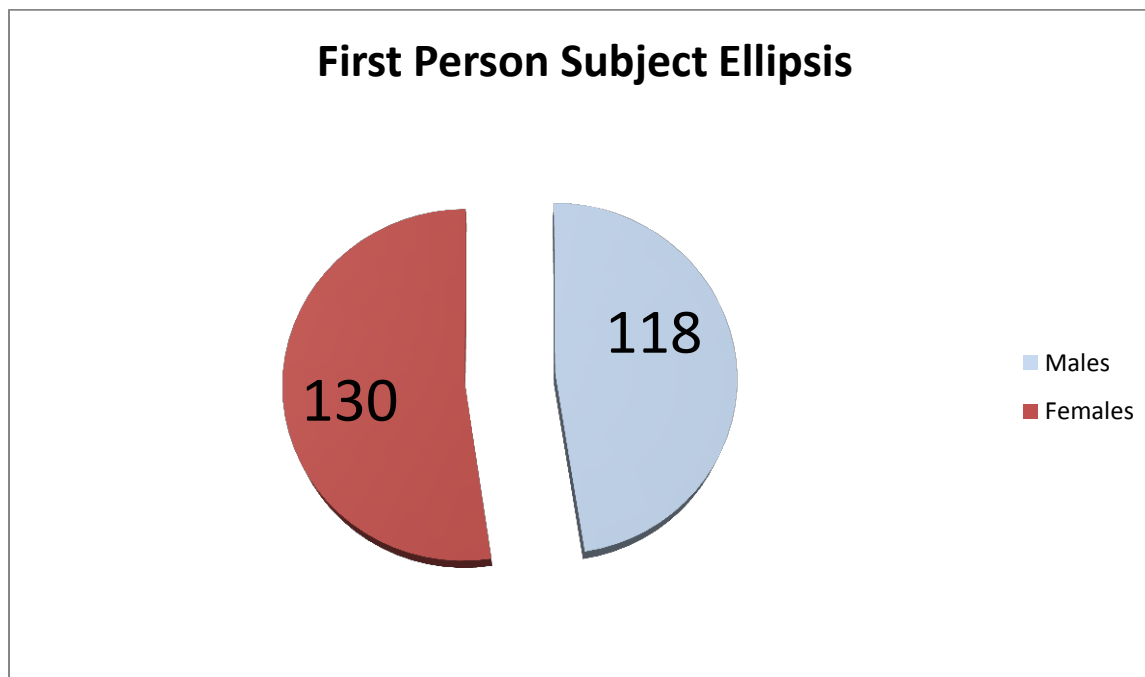


Figure 3 – Occurrences of First Person Subject Ellipsis in male and female tweets.

First person subject ellipsis appears to be a common feature on Twitter, and it typically occurs when a sentence would normally start with “I”, “I’m”. Both the pronoun and the verb are then elided. Less often, the pronoun “we” is absent. Other verbs, such as “doing”, for example, are not elided. The following are examples of this reduction.

30) Male: Got a sub oh ya [#noteacher](#) [#school](#)

31) Male: got to get off the computer. i'll be back later! [#School](#) is not fun

32) Female: Procrastinating on studying for my exams. Should probably get on that [#woops](#)
[#study](#) [#school](#)

33) Female: Really need to start doing some work.. Feeling like I'm back in [#school](#)
[#structureisgood](#) [#careerwomen](#)

It is worth noting that in examples (31) and (33), the first person pronoun is only removed in the first sentence, but not the second, while in example (32) it is present in neither. Whether or not this occurs, it seems to have no relation to gender, but rather to personal preference or other reasons on the part of the user. In example (30), the pronoun removed could either be “I”, or “we”, should the user be referring to his entire class.

While one might expect females to keep the first person pronoun more often than men, following Standard English grammar, this does not appear to be the case. Neither does it seem that males place more importance on the first person pronoun as a means of expressing identity more strongly. Indeed, it would seem that genre is more important here than gender. As Yus (2011) notes, /-ellipsis occurs most when all participants are clearly identified by another factor, which they are on Twitter, as in Yus’ study on chat. This appears to support Herring and Paolilo’s observation (2006) that at least some grammatical features – i.e. the choice of pronoun – are more dependent on genre than gender. Herring & Paolilo found that even though there were trends in the choice of the first person pronoun – such as females preferring the inclusive first person plural, and males preferring the second person – these findings would as equally support as contradict the hypothesized male and female characteristics. Instead, the study was able to conclude that genre was a stronger factor than gender in the choice of specific language features. With this in mind, considering that this study on Twitter had to discard a large amount of female data in the interest of balance, due to far more female participation, it seems likely that the “school” hashtag, specifically, is more likely to be used by females than males. Though Herring and Paolilo’s study was on blogs, its findings can be compared to those on Twitter, which is a “microblog” site.

6. Conclusion

There is hardly a way to directly compare the findings of previous studies on gender-related differences in language to similar studies on CMC. Labov (1972), Lakoff (1975), Trudgill (1983) and other linguists studied such differences in the spoken variety, typically on the basis of data provided by the working class and middle class language users, whereas the data of the present study are from an entirely different generation, most of them young individuals, who have grown up with CMC and social media. It appears that when comparing the language of the working and middle classes to that of CMC users, different contrasts can be revealed; social media may create

its own loose social group. CMC is, according to Crystal (2006), a hybrid of spoken and written language and could not be expected to be directly comparable to either form exclusively. It is a field that will require more corpus-based research, especially through the lens of gender studies, in order to obtain observations comparable to findings of previous sociolinguistic studies.

This is not to say that there are no trends, patterns, similarities—as well as differences—traceable to older studies. Even though a study of this scope is unlikely to yield any definite answers, we can observe certain tendencies in our data. Thus, we can observe that the females use reduced forms more often than the males, and their tweets display a higher variety of different forms. Females appear more prone to using them; this is surprising, as such forms have no evident gender bias. It is possible that these forms have become standardized among groups of females using Twitter, a tendency similar to that shown in White's study on the standardization of reduced forms (2012). In his study, certain forms become more common in the context of CMC within a specific social group, and this may apply to our results as well. There is hardly a singular close-knit singular speech community among females on Twitter, but it is possible that the site itself, being a social media website, forms a loose social group in which certain forms become standardized, and these standard forms are accepted by others who integrate them into their own tweets.

Additionally, females outnumber males in the avoidance of sentence capitalization. Why this is so is difficult to speculate on; it is possible, considering that the full capitalization of words is discouraged (for example, by Driscoll & Brizee 2013, online), that females wish to avoid this impression to a higher degree than men, or have adapted to a perceived lower-case standard on Twitter. It is equally possible that females simply use older devices, or different ones, with less auto-correction. Equally likely is that they simply have a different approach to CMC; thus females may think they should respond quicker, and update their tweets more frequently, thus giving less time for each individual one.

Ultimately, the language of CMC is an area which would benefit from a larger corpus-based scale study. If such studies can confirm that females tend to avoid capitalization more than males do, these tendencies could be regarded as gender-specific features of tweet language. If such studies do not confirm these tendencies, in turn, one can assume that CMC tends to skew gender-related contrasts in language usage, or that there may be a new perceived standard language on Twitter that females have adapted to more quickly than males. A larger-scale study

would also shed light on the nature of CMC in a more general gender perspective. Social media, given the popularity and activity of such websites, should also be studied to identify new linguistic features provided by the growing number of CMC users. Twitter and other websites with low character limits are also a field of study that could prove valuable in order to determine how text economy affects messages, by finding out which words, other than the first person subject, are commonly left out. This study has shown that with the exception of the two above-mentioned tendencies, there are slight differences in male and female tweets, but these observations are based on a limited scope of studied material and require further verification. Finally, other, more traditional areas in which female and male language differ should be studied, such as the use of politeness strategies, indirect requests, boosters and hedges; this will show whether CMC skews gender-specific tendencies in these spheres of language usage as well.

One word of caution to future researchers would be to make sure to account for the fact that many devices used for CMC communication use auto-correction. Therefore, nuances of capitalization and spelling may be skewed if the device which the post was made on is unknown.

7. References

- Androutopoulos, J.; Schmidt, G. 2001. "SMS-Kommunikation: Ethnografische gattungsanalyse am beispiele einer kleingruppe". *Zeitschrift für Angewandte Linguistik*.
- Baron, D. E. 1982. *Grammar and gender*. New Haven: Yale University Press.
- Berglund, Y. 1999. "Gonna and going to in the spoken component of the British National Corpus". In Mair, C; Hundt, M. (Eds.). *Corpus linguistics and linguistic theory: Papers from the twentieth international conference on English language research on computerized corpora*. Amsterdam: Rodopi, 35-49.
- Biber, D.; S. Johansson, G.; Leech, S.; Conrad, E.; Finegan, E. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.
- Cameron, D. 1992. *Feminism and linguistic theory*. 2nd ed. London: McMillan.
- Cameron, D. 2007. *The myth of Mars and Venus: Do men and women really speak different languages?* Oxford: OUP.
- Chambers, J.K. 2009. *Sociolinguistic Theory*. Revised ed. Oxford: Wiley-Blackwell
- Coates, J. 1993. *Women, men and language*. 2nd ed. London: McMillan.
- Coates, J.; Cameron, D. 1989. *Women in their speech communities*. London: Longman.
- Crystal, D. 2006. *Language and the Internet*. 2nd ed. Cambridge: Cambridge University Press
- Crystal, D. 2011. *Internet Linguistics: A student guide*. New York: Routledge.

- Driscoll, D. L; Brizee, A. 2013. "Visual-textual devices for achieving emphasis". Retrieved Feb. 25, 2013 from <http://owl.english.purdue.edu/owl/resource/609/01/>
- Gallup. 2013, Jan 14. "Home Internet access still out of reach for many worldwide". Retrieved from <http://www.gallup.com/poll/159815/home-internet-access-remains-reach-worldwide.aspx>
- Guiller, J; Durndell, A. 2007. "Students' linguistic behavior in online discussion groups: Does gender matter?". *Computers in Human Behavior*, 23(5), 2240-2255.
- Herring, S. C; Paolillo, J. C. 2006. "Gender and genre variation in weblogs". *Journal of Sociolinguistics*, 10, 439-459.
- Hård af Segerstad, Y. 2002. *Use and adaptation of written language to the conditions of computer-mediated communication*. PhD thesis, Department of Linguistics, University of Gothenburg.
- Labov, W. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Lakoff, R. 1975. *Language and woman's place*. N. Y.: Harper and Row.
- Lee, C. K. M. 2002. "Literacy practices in computer-mediated communication in Hong Kong". *Reading Matrix: An International Online Journal*, 2(2)
- Lotherington, H.; Xu, Y. 2004. "How to chat in English and Chinese: Emerging digital language conventions". *ReCALL*, 16, 308-329
- Lunden, I. 2012. "Twitter passed 500M users in June 2012...". *TechCrunch*. Retrieved 20 March, 2013 from <http://techcrunch.com/2012/07/30/analyst-twitter-passed-500m-users-in-june-2012-140m-of-them-in-us-jakarta-biggest-tweeting-city/>

Marshal, B. (2009, February 23). "A capital idea: Setting the record straight on capitalization online". *Webdesign Articles*. Retrieved 1 May from

<http://www.wpdfd.com/issues/87/a-capital-idea/>

Nariyama, S. 2004. "Subject ellipsis in English". *Journal of Pragmatics*, 36, 237-264.

Sarno, D. 2009. "Twitter creator Jack Dorsey illuminates the site's founding document". *Los Angeles Times*. February 18, 2009. Retrieved 20 March, 2013 from

<http://latimesblogs.latimes.com/technology/2009/02/twitter-creator.html>

Tannen, D. 1990a. *You just don't understand: Women and men in conversation*. N.Y.: W. Morrow.

Tannen, D. 1990b. *Gender and conversational interaction*. N.Y.: OUP.

Teddiman, L. 2011. "Subject ellipsis by text type: an investigation using ICE-GB". *Language and Computers. Corpus-based Studies in Language Use, Language Learning, and Language Documentation*, 18, 71-88.

Thomson, R; Murachver, T. 2001. "Predicting gender from electronic discourse". *British Journal of Social Psychology*, 40, 193-208.

Trudgill, P. 1983. *On Dialect: Social and Geographic Factors*. Oxford: Basil Blackwell.

White, J. R. (forthcoming). "Standardisation of reduced forms in English in an academic community of practice." Forthcoming in 2014 in *Pragmatics and Society*.

Yus, F. 2011. *Cyberpragmatics: Internet-mediated communication in context*. Amsterdam: John Benjamins.

Appendix 1

The Data

Studied material consists of roughly 10,000 words of male tweets and 10,000 words of female tweets. The actual number, disregarding hashtags that are not part of sentence structure, is somewhat lower. The total number of individual male tweets is 665, and the number of female tweets is 646. Male tweets contain on average 15 words, and the female average is marginally higher at 15.5 words per tweet. The shortest male tweet consists of two words, while the longest contains 25 words. For the females, the shortest was one acronym, and the longest contains 28 words. These tweets are reproduced in examples (34-37) below.

34) Shortest male tweet: Can't sleep. [#School](#)

35) Longest male tweet: Stupid school. My mom called them to tell them for me to go down to the office. So 20 minutes later they call me down.

36) Shortest female tweet: Fml (“fuck my life”)

37) Longest female tweet: Wen I want school to come it take 4ever but when I don't wana go 2 school its here that doesn't make sense cause 1day = only 24hours

All tweets were acquired from Twitter, using the hashtag [#school](#) as a topic, during the months of February and March, 2013.

Appendix 2

Reduced Forms in Male Tweets

Reduced Words

to (too)	4	nd (and)	1
w/ (with)	4	nothin' (nothing)	1
chillin' (chilling)	3	of (off)	1
dis (this)	3	othr (other)	1
ur (your)	3	pic (picture)	1
u (university)	3	plz (please)	1
da (the)	2	pple (people)	1
frm (from)	2	pty (party)	1
gettin' (getting)	2	std (standard)	1
jst (just)	2	sub (substitute [teacher])	1
lil (little)	2	tex (text)	1
tho (although)	2	thot (thought)	1
aite (alright)	1	thz (this)	1
ballin' (balling)	1	til (until)	1
bio (biology)	1	tims (times)	1
bout (about)	1	tmrw (tomorrow)	1
bro (brother)	1	tomar (tomorrow)	1
cos, cuz (because)	1	trucka (trucker)	1
crayz (crazy)	1	turnd (turned)	1
cruisin' (cruising)	1	w (?)	1
d (dick)	1	wass (what's)	1
def, defo (definitely)	2	wat (what)	1
diggin' (digging)	1	ya (yeah)	1
dt (detention)	1	ya (you)	1
dun (don't)	1	yer (your)	1
econ (economy)	1	yrs (years)	1
em (them)	1	TOTAL	84
fab (fabulous)	1		
feelin' (feeling)	1		
flu (influenza)	1		
jk (joke)	1		
jus (just)	1		
k (ok)	1		
mins (minutes)	1		
mofucker (motherfucker)	1		
muhfuckers (motherfuckers)	1		
mutha (mother)	1		

Acronyms/initialisms

Informal Contractions

lol, loool, loooool (laughing out loud)	15	wanna, wana (want to)	10
hw (homework)	4	gonna (going to)	5
lmao (laughing my ass off)	3	gotta (got to)	5
omg (oh my god)	3	hella (hell of a [lot])	2
af, a'f (as fuck)	3	imma (i'm going to)	2
cba (cannot be arsed)	2	kinda (kind of)	2
jgh (just got home)	2	outta (out of)	2
ny (new york)	2	ain't (is not)	1
cdale (clydesdale?)	1	c'mon (come on)	1
CLE (?)	1	guna (going to)	1
dm (?)	1	TOTAL	31
fml (fuck my life)	1		
gmornin (good morning)	1		
IEEE (?)	1		
IUP (?)	1		
lmfao (laughing my fucking ass off)	1		
P.E. (physical education)	1		
P/T (?)	1		
REAB (?)	1		
SA (South Africa)	1		
SMH (shaking my head)	1		
STE (?)	1		
TGIF (thank god it's Friday)	1		
VPN (?)	1		
scsu (university name)	1		
TOTAL	51		

Number and Letter Homophones Symbols

2 (to)	6	"&" (and)	3
2 (too)	1	"<3" (love/affection)	2
2day (today)	1	"=" (is equal to)	1
m8 (mate)	1	"@" (at)	1
no1 (no one)	1	">" (is better than)	1
u (you)	1	TOTAL	8
n (and)	6		
c (see)	2		
r (our)	1		
TOTAL	10		

Reduced Forms in Female Tweets

Reduced Words

tho (although)	6	hur (here)	1
chillin' (chilling)	4	hv (have)	1
jus (just)	3	int (intelligence)	1
pls (please)	3	j'd (jacked)	1
to (too)	3	jux, juz (just)	2
w/ (with)	3	k (ok)	1
cuz (because)	2	kik (kick)	1
dnt (don't)	2	killin' (killing)	1
gettin' (getting)	2	kno (know)	1
lil (little)	2	lookin' (looking)	1
til, till (until)	3	mcds (mcdonalds)	1
tmrw (tomorrow)	2	mon (Monday)	1
an (and)	1	tues/tue (Tuesday)	2
bio (biology)	1	ppl, pple (people)	2
boo [NAME] (?)	1	precalc (precalculus)	1
bout (about)	1	prolly (probably)	1
bro (brother)	1	rockin' (rocking)	1
bt (but)	1	sis (sister)	1
calc (calculus)	1	sub (substitute [teacher])	1
chem (chemistry)	1	takin' (taking)	1
cnt (can't)	1	temp (temperature)	1
coco (?)	1	thuurrrrr (there)	1
coop (cooperation)	1	thx (thanks)	1
crankin' (cranking)	1	tutor (tutorial)	1
dat (that)	1	v (vintage)	1
def (definitely)	1	watchin' (watching)	1
dis (this)	1	wateva (whatever)	1
em (them)	1	wen (when)	1
finna (finally)	1	whaaatevs (whatever)	1
frm (from)	1	wht (what)	1
fuckin' (fucking)	1	wippin' (whipping)	1
gals (girls)	1	wut (what)	1
goin' (going)	1	y'all (you all)	1
goodmornin'	1	yr (year)	1
gran (grandmother)	1	TOTAL	99
g-son (grandson)	1		
gud (good)	1		
hay (hate)	1		
hrs (hours)	1		
	1		

Acronyms/initialisms

Informal Contractions

lol, lolol (laughing out loud)	20	wanna/wana (want to)	16
omg, omggggg (oh my god)	9	gonna/gone/gunna (going to)	14
AP (advanced placement)	2	gotta (got to)	4
idk (I don't know)	2	ain't (is not)	2
smh (shaking my head)	2	ima (I'm going to)	1
AC (air conditioning)	1	tryna (trying to)	1
bbl (be back later)	1	TOTAL	38
bf (best friend/boyfriend)	1		
btw (by the way)	1		
cc (?)	1		
FL (?)	1		
fml (fuck my life)	1		
gm (good morning)	1		
hmu (university name?)	1		
hw (homework)	1		
idc (I don't care)	1		
IG (?)	1		
lmfao (laughing my fucking ass off)	1		
mcm (?)	1		
MSN (microsoft service network)	1		
pe (physical education?)	1		
PLL (?)	1		
PR (?)	1		
RE (?)	1		
s/o (significant other)	1		
T.G.I.F (thank god it's Friday)	1		
tb (?)	1		
tf ([what] the fuck)	1		
tt (?)	1		
TYJ (thank you Jesus)	1		
yolo (you only live once)	1		
TOTAL	61		

Number and Letter Homophones Symbols

2 (to)	8	& (and)	6
2moro (tomorrow)	1	<3 (love/affection)	5
4 (for)	1	<3 (no love/dislike)	1
4ever (forever)	1	> (is better than)	1
kn0 (know)	1	xo/xx (kiss)	2
u (you)	10	TOTAL	15
n/(n) (and)	3		
o (oh)	1		
sux (sucks)	1		
xtra (extra)	1		
y (why)	1		
TOTAL	29		



GÖTEBORGS UNIVERSITET
GÖTEBORGS UNIVERSITETSBIBLIOTEK

Dnr: A 85 592/10

Avtal för elektronisk publicering

§1. Upplåtelse för elektronisk publicering

Upphovsmannen/upphovsmännen/rättighetsinnehavaren/rättighetsinnehavarna, här kallad Författaren/Författarna, ger Göteborgs universitet rätt att publicera angivet arbete, här kallat Verket, i elektronisk form på Göteborgs universitets domän på Internet. Rätten att publicera innebär en rätt att göra Verket tillgängligt för allmänheten på Internet. Upplåtelsen medför inga inskränkningar i Författarens/Författarnas rätt att för egen del använda Verket.

§2. Giltighetstid och uppsägning

Upplåtelsen gäller till dess Författaren/Författarna skriftligen meddelar Göteborgs universitet att denna rätt skall upphöra. Vid ett upphörande har Göteborgs universitet inte längre rätt att tillgängliggöra Verket på Göteborgs universitets domän på Internet och skall med omedelbar verkan avlägsna Verket med följd att publiceringen upphör. Göteborgs universitet kan i sin tur med omedelbar verkan frånträda avtalet och avlägsna Verket med följd att publiceringen upphör. Något skäl för upphörandet behöver inte anges.

§3. Syfte med publiceringen

Den elektroniska publiceringen skall ske för universitetets forsknings-, utbildnings- och biblioteksverksamhet och ej för kommersiella ändamål.

§4. Verkets form

Författaren/författarna skall lämna över Verket i enlighet med de instruktioner som universitetet meddelar.

§5. Upphovsmannaskap

Författaren/Författarna ansvarar för att hon/han/de är upphovsman till Verket och att i förekommande fall inhämta erforderliga upphovsrättsliga medgivanden vad gäller i Verket eventuellt ingående material (bilder m.m.) som kan ha annan upphovsman, samt att Författaren/Författarna har rätt att disponera över Verket för publicering i enlighet med detta avtal.

Om det framkommer att Författaren/Författarna inte är upphovsman till del av Verket eller dess helhet eller saknar erforderliga medgivanden enligt ovan åligger det Författaren/Författarna att i förekommande fall hålla Göteborgs universitet skadeslöst.

§6. Ersättning

Inga ersättningar utgår mot bakgrund av detta avtal. Eventuellt uppkomna kostnader uppbärs av part som kostnaden hänförs till.

§7. Ikraftträdande

Detta avtal gäller omedelbart efter Författarens godkännande.

Ort & datum: Göteborg 6/17/2013

Verkets titel: Gender-related Variation in CMC Language

Namnteckning(ar) och förtydligande:

Joni Halmeoja