Spring 5-1-2014

An Examination of High School and College Students' Chatspeak Use in Twitter and Tumblr

Megan Jeanette Charters

An Examination of High School and College Students' Chatspeak Use in Twitter and Tumblr

A Capstone Project Submitted in Partial Fulfillment of the
Requirements of the Renée Crown University Honors Program at
Syracuse University

Megan Jeanette Charters
Candidate for B.A. Degree
and Renée Crown University Honors
May 2014

Honors Capstone Project in Linguistics

Capstone Project Advisor: _____
                                      Tej Bhatia, Professor

Capstone Project Reader: _____
                                      Rania Habib, Assistant Professor

Honors Director: _____
                                      Stephen Kuusisto, Director

Date: April 23, 2014

Table of Contents

Abstract

The advent of the Internet has led to the creation of new technology, and with it, a whole new language. There has also been resistance both to the technology and to the language. The best example of this would be social media, such as Twitter, Facebook, and texting. Many articles have been written on this subject and how its use is ruining the English language, saying that teenagers are more likely to use chatspeak, such as *u* for *you* and *r* for *are*, in their writing. I take the opposite position, that the English language is evolving and chatspeak is simply one consequence. This study examines the language change of teenagers using the social media platforms Twitter and Tumblr. I analyzed the tweets and Tumblr posts of 48 high school and college students in 2012 and 2013 and found that chatspeak occurred less than 3% of the time, and its usage decreased as the students got older.

Acknowledgements

I would like to thank my advisor Tej Bhatia for his encouragement and support for the last two years. I would also like to thank Rania Habib, my reader, for reading over my thesis multiple times and providing me with the tools to make this thesis a success through her class on language variation and change. I would like to thank Eric Holowartz, Hanna Richardson, Kate Hanson, and everyone else in Honors for their advice and support. Finally, I would like to thank my family and friends for supporting me throughout this process.

Executive Summary

This study examines the change in the language of high school and college students using Twitter and Tumblr in 2012 and 2013. Social media as used by teenagers is typically cited as having a detrimental effect on the English language; some sources outright insist that the English language is being ruined. However, chatspeak, one name for the language used in social media, consists of less than 5% of the total language used at any given time. Previous studies done on social media have mainly focused on two aspects: whether or not the language of social media resembles speech or writing, or on its value to advertising, marketing, and the like. Very few studies have focused solely on how social media affects language change, and those that do examine older forms of social media, such as Internet Relay Chat, an early form of instant messaging, or text messaging.

For this research, 48 students were examined; 24 students were Twitter users, and 24 were Twitter users. For each grade or class year, students were randomly selected people by doing a hashtag, or keyword, search for that grade, using tags such as #collegesenior or #classof2014. Three students were chosen from each search. Half of the 48 students were high school students, and the other half attended college. 50 posts were collected per user per year, for a total of 100 posts or tweets per student. The posts from 2012 were taken beginning in January, and the posts from 2013 were taken either starting in August or October, depending on when the data was collected.

The data was preprocessed prior to the start of the analysis. Retweets, which are Twitter posts that were originally written by another user, and non-

English posts were eliminated in order to examine the data that was written solely by the original user.

For analytical purposes, the data was then parsed, or broken into words, and then tagged with the corresponding part of speech using the Carnegie Mellon University Twitter Parts of Speech Tagger. This tagger was written specifically for parsing and tagging tweets and tags both classical parts of speech, such as nouns and verbs, and features found in many forms of social media, such as hashtags, links to other websites, and emoticons.

A total of four variables were analyzed for this study; each variable had a standard form, or a more accepted variant, and a nonstandard form, which is typically considered slang or is not accepted in more formal environments like academic settings. The first variable was the total amount of chatspeak present for each user; this allowed me to examine any overall language change taking place. The second variable was the first person singular form *I*. More specifically, the usage of lowercase *i* in place of the capital *I* was examined. The third variable was the usage of the lowercase *u* instead of the pronoun *you*. The fourth variable was apostrophe deletion; an example of this would be using *dont* instead of *don't*. These three variables are widely used in many forms of social media and are often used according to individual preference. The last variable this study examines is the usage of the abbreviations *lol* and *haha*. Both commonly denote laughter; although both *lol* and *haha* are considered chatspeak, *haha* is regarded as the standard form.

The data from Twitter was analyzed by grade, from high school freshmen to college senior, and then by gender, in order to examine possible language change in general and among males and females in particular. The data from Tumblr was only analyzed by grade, as gender data for Tumblr users was almost impossible to obtain. The data was also analyzed in real time, from 2012 to 2013. Tumblr data and Twitter data was then compared. It was expected that older students would use much less chatspeak than younger students. Additionally, it was expected that male students would use more nonstandard forms than female students.

Chatspeak made up a minute portion of the total number of tokens for both Twitter and Tumblr. 1.6% of all Twitter tokens were chatspeak tokens; in contrast, 0.67% of all Tumblr tokens consisted of chatspeak. My findings revealed that Twitter users in college, especially juniors and seniors, used more chatspeak than high school freshmen and sophomores. In contrast, high school Tumblr users use more chatspeak than their college counterparts. For both Twitter and Tumblr, the usage of nonstandard forms, such as *lol*, lowercase *u*, and lowercase *i*, decreased from 2012 to 2013, but the amounts varied from grade to grade inconsistently. The inconsistency is due to individual style, rather than language change. High school and college seniors tended to use fewer nonstandard forms than standard forms in comparison with the other years, possibly due to applications for college or jobs that require more formal language. Additionally, male Twitter users used more nonstandard forms and more chatspeak than female students, as expected from previous research. Taken together, this suggests that

the nonstandard forms are gradually being phased out in favor of the standard

forms.

Finally, limitations and areas of further research are identified. Are the

same trends described in Twitter and Tumblr present in other social media

platforms like Facebook? For instance, what societal or cultural factors, such as

race, education, or socioeconomic status, are contributing to this decline? Does

the size and density of the user's social network affect the amount and forms of

textspeak used? To what extent does the presence of a character limit affect the

amount of textspeak used?

While chatspeak may be used more in some forms of social media, it

makes up such a small percentage of the data that it is nearly non-existent. Not

only that, but chatspeak as a whole is declining over time. Claims that social

media are negatively affecting the English language are unfounded.

Chapter 1: Language Change and Social Media

1. Introduction

   The purpose of this study is to examine the language change of teenagers using social media. Youth are the agents of language change, introducing variation through their usage of slang and other nonstandard forms in their day to day interactions, and these variations can lead to language change (Labov, 2007, p. 346). I examine two social media sites that are popular with teenagers, Twitter and Tumblr, and analyze teenagers' use of nonstandard forms. I hypothesize that older teens will use more nonstandard forms than younger teens, as they have been exposed to more of that language.

   Languages always change over time as people use them. People adapt the language they speak to suit their needs, whether it is the vocabulary or the pronunciation, or even the grammar. The clearest example is a change in a language's vocabulary, or lexicon. For instance, a word or phrase is invented due to need; the words Google or Facebook have only been in use for the past ten years or so, as Google was invented in 1998 and Facebook was invented in 2006 ("Our history in depth", 2014; Carlson, 2010). Similarly, a word can go out of use; no one today would be able to rattle off the definition of *abatude* without a dictionary. Change is not limited to lexical items; all facets of a language can change over time.

   There are five types of language change: lexical change, phonetic change, morphological change, semantic change, and syntactic change. In some

languages, such as English, the spelling can change as well. For instance, *doubt* used to be spelled *dout* until the 16th century, when scholars decided to standardize spelling according to Latin and Greek (O'Conner and Kellerman, 2010). This resulted in the addition of silent letters to some words and the creation of prescriptivist Latin-based grammatical rules, such as not splitting an infinitive.

First, lexical changes affect the vocabulary; these changes are often the addition of new words or deletion of old words. For instance, whenever new technology is created, it needs a name. That name enters the language and is either adopted or cast aside in favor of a different term.

Second, phonetic changes are often subtle differences in speakers' pronunciation. English speakers from Syracuse often pronounce *bus* as *boss*; this is known as the Northern Cities Shift, as it affects cities around the Great Lakes (Labov, 2007, p. 372).

Third, morphological change is a change in a morpheme of a language. A morpheme is the smallest meaningful grammatical unit of a language. There are two types of morphemes: free and bound. *Cat* and *you* are free morphemes, meaning that they can stand alone. The *un-* in *undecided* and the past tense marker *-ed* are bound morphemes, meaning that they must be part of a larger word. One example of morphological change is in the second person pronouns in English. English originally had two forms of *you*; *thou* was the singular form, and *ye* was the plural form. *Thee* was used to address one person, and *you* was used when speaking a group. However, *thee* could be used to address someone informally; likewise, *you* could be used to address someone formally. In English,

however, it was possible to address someone formally on one occasion and informally on another. For example, someone could initially use *you* when speaking with a teacher, and then later use *thee*. This lack of distinction between the two forms meant that the system of differentiating between the singular and plural forms underwent a change. Eventually, this lead to the collapse of the system, and today, *you* is used to address both one person and a group of people (Hickey, 2012).

Fourth, syntactic change is a change in the rules or grammar of a language. English, for example, used to have a much freer word order, similar to Latin. Due to the influence of Norse languages and French in the 8th to 11th centuries, the word order shifted and became subject-verb-object (Fennell, 2001).

Finally, semantic change is the shift in a word's meaning from one meaning to another. The word *gay*, for instance, originally meant happy; it now refers to someone who is homosexual ("Gay", n.d.).

How, then, can a language change? It all starts with variations within that language. These variations are like mutations in genes. The speakers all share the same basic rules and vocabulary that define that language, just as people share the same genome. Likewise, small differences in gene expression result in different appearances for each person, and one person's speech is not the same as another's. A variation, such as using *like* to introduce a quotative phrase instead of *said*, can spread from the group it originated in to a larger group of people, which then spreads to increasing larger groups until the change has been adopted by all speakers.

Historically, languages have, for the most part, changed rather slowly, and speakers were often not aware of it. From 1350 to 1700, the Great Vowel Shift took place in England, and it affected long stressed vowels, which are found in words like *sheep* and *bought*, but not words like *only*. In Middle English, *sheep* was pronounced more like today's *shape*. The vowels raised; that is a vowel that had been pronounced in one part of the mouth was pronounced in a higher place in the mouth. The long "i" sound in Middle English, as in *sheep*, changed to the current pronunciation, as in *night* (Wheeler, 2014).

Now, with the rise of faster communication methods like the Internet, changes can spread in a matter of days or hours. The most noticeable changes are lexical changes, due to the influx of new technology. However, other linguistic changes have taken place as well. Nowhere is this change more noticeable than in social media. While social media has its roots in the Internet, the Internet itself was created to facilitate communication.

Humans need to interact with each other across long distances, and by doing so, strengthening existing relationships. When face to face communication becomes impossible, then a different method of communication is needed. This has led to many different solutions; the earliest focus on the written word, while later inventions transmit sound. The most recent solutions are capable of both.

The first is the postal system, in which letters from one person are delivered to their recipient. The earliest implementation of the postal system was in 550 B.C., at a time when hand delivery was the fastest way to communicate. Later postal systems relied on faster delivery methods; letters were carried by

horse and rider. Today, letters travel by car, airplane, or by boat to destinations all over the world. By today's standards, it is a fairly slow means of communication; the letter physically travels at the speed of the transport medium. The recipient could receive a letter days or even months after it was written, and this drove the search for a faster means of communication (Hendricks, 2013).

The second is the telegraph, which was invented in 1792 by Claude Chappe, and transmits an encoded message through a signaling medium. Initially, this medium was limited to forms like smoke signals or flags. The message was translated into semaphore, an alphabet of visual signals, and then sent from one station to the next until it reached its destination. Each station had to have a direct line of sight to the next, making this a slow and tedious process that was dependent on the weather. The advent of electricity, and later, the radio, meant that signals could be transmitted independent of the weather and over much longer distances using the dots and dashes of the Morse code (Hendricks, 2013).

The third solution is the telephone, which was invented by Alexander Graham Bell in 1876. It translates sound into electric signals and transmits them over cables to the receiving telephone, which converts the electric signals into audible sound. One of the earliest kinds of telephones was the rotary phone; these fell into disuse with the invention of touch-tone dialing in the 1960s, which uses a keypad to dial numbers instead of rotating a dial. As technology advanced, phones became smaller and more portable; these became known as cellphones or mobile phones.

The final solution, developed in the 1940s, is the computer. A computer is a device that can be programmed to carry out logical or arithmetic operations automatically. It also allows the transmission of data from one machine to another. Initially, this was done using a physical connection; now it can be done wirelessly. A few decades later, the Internet was developed, which connects computers together into networks. These networks enabled the creation of email, messages sent from one user to another over the Internet. This was the beginning of social media.

1.1 What is Social Media?

Social media is formally defined as the interaction of people as they create, share and exchange content in virtual communities and networks (Ahlqvist et al., 2008, p. 13). People mainly use it to connect with friends and family, as well as to reconnect with old acquaintances (Smith, 2011). Businesses utilize social media in order to expose and promote their products and services to large numbers of users, who in turn will pass the message on to members of their social networks. Some of the most popular social media sites include Facebook, Twitter, and Youtube.

Most, if not all, social media platforms originated on the web. One of the first social media sites was Geocities, created in 1994, where users could create their own websites. Other sites and services soon followed, such as AOL Instant Messenger in 1997 (Kaplan, 2010). The growing popularity of smartphones has led to the development of applications (apps) that allow access to these sites

without the need for a computer. Some users have never even visited the website

the app was based on, while others exclusively use the website.

1.2 Popular Social Media Platforms

Mark Zuckerberg founded Facebook, one of many online social

networking services, in 2004. The network was originally limited to universities,

but later expanded to allow users who had a registered email address. Now, it has

almost one billion users around the world, with the majority being teenagers

(Perez, 2013; Duggan, 2013). Users post status updates, which can consist of

links, photos, text, or a combination of the three. Users can expand their social

network by friending and unfriending people that they may know in real life, and

can form groups, play games, and send each other gifts (Phillips, 2007).

Twitter, nicknamed "the SMS [short message service] of the Internet", was

founded by Evan Williams, Jack Dorsey and others in 2006. It is a microblogging

service with a 140 character limit on messages. Messages, called tweets, tend to

average about 11 words per tweet (O'Connor et al., 2010, p. 122). Tweets can also

be reposted by other users; these are called retweets (RT) and used to be
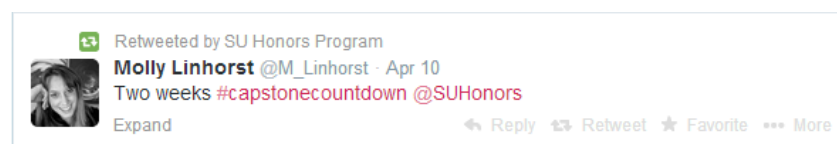
designated by "RT".

Figure 1 Picture of Retweet Icon and Tweet

Figure 1 shows a tweet by Molly Linhorst that has been retweeted by the SU

Honors Program (Linhorst, 2014). Retweets are marked by an icon that looks like

the refresh icon on a web browser, but the original tweet remains the same. Users can also send direct messages to another user by putting @ before the username, as the @ sign designates the user's handle or username on Twitter; for instance, I would send a tweet to the user Alicia Evans by starting the tweet with @AlyciaEvans1.

Tumblr was founded by David Karp in 2007 and is currently owned by Yahoo! Inc. Users can create blogs, where they can post status updates, or other content such as links to giveaways, and photos and videos. Posts can be tagged with hashtags or reblogged or liked by other users. All of these posts can be found in the user's dashboard. The dashboard displays all the posts that the user posts, along with the posts of the people the user follows. Users can also select a background theme for their blog that ranges from the default blue theme to ones that include background music and pictures; there is also the option to create a custom theme using HTML, a markup language used to create web pages.

Additional forms of social media include Pintrest, where users collect and organize pictures on glorified bulletin boards; Flickr, which is an online photo storage site; Vine, where users can create and share short seven second videos; and Youtube, where users can create and view videos.

1.3 Computer-Mediated Communication

The language used over these social media platforms, and the Internet in general, is called computer-mediated communication (CMC). Initially, CMC was restricted to emails. As a result, the language was much more formal and used in

the same way as traditional writing. Emails tend to be much more formal than other forms of social media, as they are more frequently used for business, rather than for social purposes. As the Internet has developed, the language used has become much more informal and in some places is more like a stream of consciousness than a formal business letter. It can be found not only in emails, but also in blogs and forum posts.

Chatspeak, also known as textspeak, is the use of abbreviations and slang during Internet-based communications. These often involve various acronyms and vowelless abbreviations, as well as sequences of characters intended to resemble objects. For instance, @-`-,-- represents a rose. Some of these sequences, called emoticons or smileys, represent the human face. Emoticons are used to compensate for the inability to express facial expressions and intonation in writing.

Chatspeak came into being with the introduction of text messages to mobile phones. These early messages were limited to 160 characters. Multiple key presses were needed to enter text. For instance, on a flip phone, the 1 key also has the letters *a*, *b*, and *c* associated with it. When texting, the user had to press the numeric key *1* three times to type a *c*. Typing a long message was very tedious. Chatspeak developed because of the need to save space and the need for speed.

Examples of Chatspeak and Emoticons

Text	Meaning	Emoticon
LOL	Laugh out loud	
:) or :-)	Happiness	☺
TMI	Too much information	
TL;DR	Too long; didn't read	
:O or :-O	Surprise or yelling	:O
FWIW	For what it's worth	
:(or :-(	Sadness	☹

Table 1 Some examples of chatspeak, their associated meanings and graphical representation

The rise of smartphones, cell phones that can perform many of the functions of a computer, has all but eliminated most of the concerns about speed and space, due to increase in the amount of data that can be sent or received. Blackberries and iPhones have full keyboards; messages can be as long or as short as the user desires. So why, then, does chatspeak persist? Well, one thing hasn't changed. People still want to save time; no one wants to bother spelling everything out when a simple abbreviation will do just as well. Additionally, some forms of social media have a space constraint, such as Twitter, where tweets are limited to 140 characters.

A hashtag is a symbol that is used to mark keywords or topics. It is designated by the pound symbol or number sign (#). Although the general format is to have them without spaces, on Tumblr there are hashtags with spaces in them, such as #college senior; on Twitter, it would be #collegesenior or #college #senior. However, it makes more sense if hashtags did not contain spaces, as it would be easier to separate them from other words. Because hashtags are used to categorize items, they can also be used to filter tweets and locate messages about specific subjects quickly. However, they are not standardized. For instance, one

person could mark a Syracuse football game as #SUfootball and another person as #SU.

Hashtags were originally used in tweets and later spread to other mediums. The first hashtag was used by Chris Messina in 2007, as a way to tag the locations and topics of his tweets (Biddle, 2011). As hashtags have become more popular, they have also been severely abused, leading to tags like #... or #6:18pm, which have no meaning. At the same time, people have begun to set down rules for their use, i.e. keep them short or don't use spaces.

1.4 The Media: so what's the big deal?

Chatspeak diverges from written language, and this alarms many people, including teachers and the media. Teachers argue that students frequently use chatspeak in their assignments. Most people perceive chatspeak as mixtures of characters and letters that form some sort of secret code; they see it as inferior to formal written language (Tagliamonte and Denis, 2008, p. 8). Various media articles only add to this hype, suggesting that hashtags and text speak "[erode] the English language" (Greenfield, 2011). However, there is a growing push for written language to resemble the spoken word, even going as far as to use run-on sentences. Additionally, the graphical side of chatspeak (i.e. emoticons) has nothing to do with speech.

On the other hand, linguists argue that this is simply a new and innovative use of language. Tagliamonte and Denis (2008) argue that teens don't want to use formal language; they mix different elements, combining both formal and

informal language. They wouldn't be able to do this if they didn't have command

of one language. It is clear that they have mastered the resources available to

them. We can infer that this is not the ruin of English, but rather a renovation

(Tagliamonte and Denis, 2008, p. 27).

Chapter 2: Literature Review

2. Introduction

Research on social media has been carried out in many fields including psychology, sociology, computer science, and linguistics. Topics studied extensively include sentiment analysis and second language acquisition. However, it is still difficult to find previous work on the subject. Since many papers have been written about instant messaging (IM), text messaging and Internet Relay Chat (IRC), relevant papers from all of these areas will be discussed.

The majority of research has been concentrated on its features, although there has been some research on linguistic analysis. These features include anything from pronoun usage to word frequencies. Denby (2010) examined the effect of character limitation on tweets, using messages gathered from 43 subjects. These messages were compared with the results of Ling and Baron's (2007) comparison of IM and text messages. While there were some similarities between the two studies in terms of message length and punctuation, there were also some features that were unique to Twitter, such as hashtags and pro-dropping.

The main question of these papers, regardless of what features they cover, is what is the register of social media? Is it like speech or like writing?

2.1 The Register of Social Media

The first study to answer this question was by Ferrara, Brunner and Whitmore (1991). In their experiment, 23 subjects sent e-messages, a precursor to IM, to a user who aided them in organizing their travel plans. Although these messages contained features associated with writing, they also contained ones that were associated with speech as well (Ferrara et al., 1991). Similarly, Yates (1996) compared data from corpora of both written and spoken language to data from computer conferences, finding that the language used conformed to standards that were a hybrid of the two registers. The use of pronouns and modals in the computer conferencing data resembled speech, whereas it resembled writing in other aspects, such as lexical density (Yates, 1996).

There are additional recent studies that confirm this finding as well. Palfreyman and al Khalil (2003) found that people used dialect features when they used IM in Arabic. Arabic is typically written in the standard form; the dialect is only spoken, not written. The usage of dialect features shows that the language used in social media reflects speech and not writing (Palfreyman and al Khalil, 2003). Tagliamonte and Denis (2008) compared the IM chats of 72 teenagers with speech from the same teenagers and came to a similar conclusion. Additionally, she noted that the frequencies of words such as *lol* declined as the age of the teenager increased.

2.2 Examining Sentiment Analysis

Other studies have examined sentiment analysis, which tries to determine whether a given text has a positive or a negative feel to it solely by examining the text only. O'Connor et al. (2010) compared data from various economic polls to data scraped from Twitter. They wanted to determine if analyzing the text sentiment of tweets would yield similar results to the data from polls. They found that doing so captured the broad trends that were present in the survey data. Asur and Huberman (2010) used three million tweets to make a linear regression model to predict box office revenue of movies before they are released. Sentiment analysis improved these predictions post-release.

2.3 Internet Linguistics

Internet linguistics is a relatively new branch of linguistics, proposed by David Crystal (2001) in *Language and the Internet*. Internet linguistics is the study of the language varieties and forms found on the Internet (Crystal, 2011, p. 2). This book examines the new language varieties and styles that have arisen due to influence from the Internet and other new media forms, such as texting and blogging. The term 'Internet linguistics' is less used or even accepted, as other terms, such as computer mediated communication, have been coined to focus primarily on the communicative aspect of the Internet. CMC is too broad; it covers all forms of communication, from videos to photographs and language as well. The emergence of smartphones has led to the creation of terms such as *electronically mediated communication*. However, these terms do not distinguish

between language and communication; they treat language and communication as the same thing (Crystal, 2011, pp. 1-3).

Chapter 3: Methodology

3.1 Data Collection

The ideal method would have been to collect data from as many social media platforms and people as possible. Collecting data from Facebook would have required obtaining permission from every person I wanted to collect data from due to privacy issues. While this was feasible, I wanted data that was easy to access; that meant that the data would have to be publicly available. For this reason I chose Twitter and Tumblr. Both have large user bases, and their users tend to post frequently. Additionally, many articles and papers have been written about Twitter, and very few have been written about Tumblr.

The data for both Twitter and Tumblr was collected from January 2012 and August to October 2013. In order to collect tweets older than the past week, I used the website topsy.com, which allows a user to filter tweets by user over a period of time. Tumblr posts were taken from the user's archived posts. A total of 50 posts or tweets per person per year were collected.

3.2 Selecting Subjects

Initially, I wanted to collect tweets and posts from as many people as possible. However, many people who use social media do not publicly post their ages; this makes it harder to collect data from users who are older than twenty one. The ages that people do post, if they post one at all, may not be accurate. In order to overcome this problem, a new method was devised. One potential way

would be to set up a separate account. Users would follow this account with the understanding that they were participating in a research project. A set amount of tweets or posts would be collected from each user after a certain amount of time, as Lewis Denby (2010) did. However, I chose an alternate solution: filtering users by hashtags. This meant that I would be restricted to using hashtags that were relatively common and would indicate age, such as those that tagged seniors or a particular grade or class. As a result, I chose to focus on high school students and college students because it would be too difficult to find subjects older than twenty-one or twenty-two. In the same manner, it would be difficult to find users under the age of thirteen, as users are only allowed to set up an account if they are thirteen or older.

Examples of hashtags used include #senior, #freshman, #classof2014, #college senior, and #sophomore. Care had to be taken to distinguish the rising college freshmen, who would often use both #senior and #college in their tweets, from actual college seniors, who would do the same.

Although it would be better to collect lots of data from a lot of people, in the interest of time, I decided to take a smaller number of users and collect a larger number of tweets or posts per person. I gathered 100 tweets or posts per user and decided on three subjects per grade or year, for a total of 24 high school students and 24 college students. I collected tweets from half of the students (24 Twitter users) and blog posts from the other half (24 Tumblr users). I did not care about the gender of the users, only their age.

3.3 Preprocessing the Data

In order to eliminate irrelevant data, I preprocessed the data using two different techniques. The data from Twitter and Tumblr was processed differently. In the Twitter data, I removed all retweets, since these tweets were not technically written by the user, and the results could possibly be skewed if they were included. I did not, however, remove all the reblogged posts in the Tumblr data. Doing so would leave me with very little to analyze, as the majority of Tumblr users repost other users' posts.

3.4 The Parser: A Natural Language Processing Tool

Before I could analyze the prepared data, I needed to parse it. My options were to do it by hand or by using a parser. Hand parsing the data would have been too time-consuming, so I decided to use a parser. This would allow me to feed the text into the parser, and it would automatically determine the parts of speech. I looked at three parsers: the Natural Language Toolkit (NLTK), Linguistic Inquiry and Word Count (LIWC), and the Carnegie Mellon University Parts of Speech tagger (CMU).

The first two parsers not only parse text, but perform a variety of other functions. NLTK allows users to write computer programs in Python that can process texts. It contains libraries to classify, tokenize, tag and parse text, as well as provide access to many corpora and other resources such as WordNet. LWIC, a text analysis program, looks at various language dimensions in texts. These

dimensions include positive or negative emotions, casual words, and self-references.

Either NLTK or LIWC would have suited my needs; however, additional work would have been necessary to parse the data. In the case of NLTK, I would have had to write a program to parse the collected data, as the parsers that are bundled with it are designed for novels and articles rather than tweets and text messages. I used the CMU Twitter Parts of Speech tagger, which is a parser designed to parse tweets. The parser was run on a Linux machine, as the Windows command line does not have the ability to run shell files. A shell file is a computer program designed to run by the Unix shell, a command line interpreter found on a Mac or a Linux machine. The preprocessed data was copied to the Linux machine, parsed and saved. The command I used was *./runTagger.sh --output-format conll filename.txt > filenameParsed.txt.* The preprocessed data in filename.txt was parsed and saved in filenameParsed.txt. From there, the parsed data was stored in Excel spreadsheets and then analyzed.

Chapter 4: Data Analysis

4. Introduction

I hypothesized that the age of the social media user is inversely proportional to the amount of textspeak he uses. For example, a college student would use less textspeak and more standard forms than a high school student, due to their increased exposure to the standardized language that accompanies formal academic writing. I examine several variables, including the total amount of textspeak used, the use of nonstandard forms such as lowercase *i* and *u* for the standard forms of *I* and *you*, and the use of *haha* over other textspeak traditionally used to denote laughter like *lol*.

In the analysis that follows, I first analyze the data collected from Twitter by age (grade) and then by gender. The Tumblr data is then analyzed by age; finally, the two sets of data are compared.

4.1 Analysis of Twitter Data

Examining the average total tokens of textspeak provides an overall picture of any possible language change in the data.
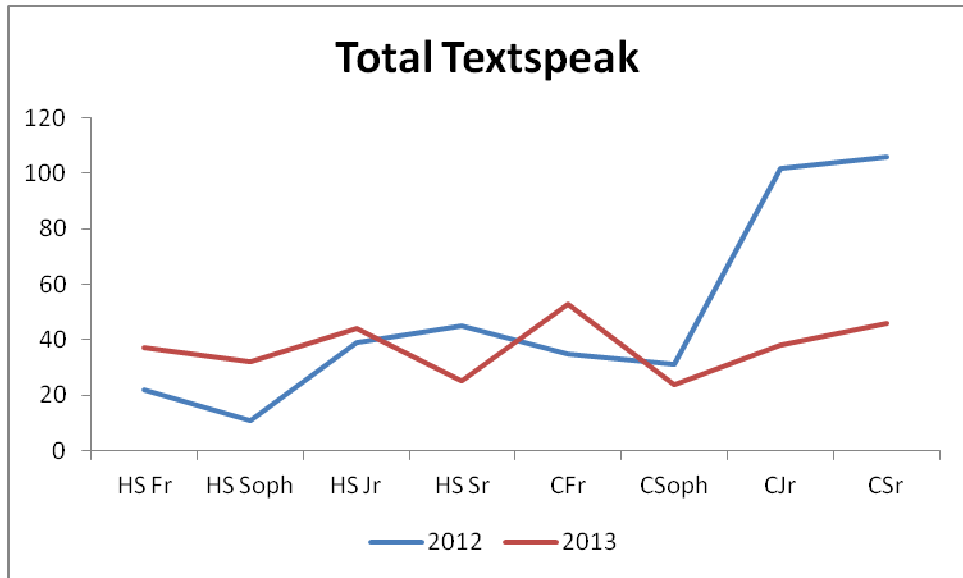
Figure 2 Table depicting total amount of textspeak among high school and college students from 2012-2013[1]

Figure 2 shows the total tokens of textspeak in 2012 and 2013 among high school and college students. In 2012, there is a gradual increase in textspeak from grade to grade. The three high school freshmen have a total of 22 tokens; the high school seniors have 45 tokens, and college seniors have a total of 106 tokens. The plotted points form an S-shaped curve. This suggests that there is indeed a change in the language in the direction of the standard form, as the younger students use less textspeak. It is true that the college students use textspeak more and high school students use it less regardless of the circumstances; the data for 2012 clearly shows this, and it is certainly one way to interpret the data. This simplistic conclusion is incorrect; examining the data for 2013 disproves this assumption.

In 2013, the total tokens increase and decrease with no apparent pattern. The college juniors, for instance, have a total of 38 tokens, whereas the college seniors have 46, but the sophomores have 24 and the freshmen have 52. The high

---

[1] The abbreviations used in the figures and tables are HS (high school) or C (college), and the grade is indicated by Fr (freshmen), Soph (sophomore), Jr (Junior), and Sr (Senior). A high school junior would be represented by HSJr.

school students decrease from 40 tokens for high school freshmen to about 22 tokens for high school seniors. The number of total textspeak tokens falls within a much narrower range of 25-52 tokens, for an average of about 38 tokens. The S-shaped curve has been replaced by what is essentially a flat line. The graph itself is much smoother and more stable, suggesting that the change has stabilized across both high school and college students.

Although the numbers vary by grade, a tentative conclusion could be drawn. It is possible that the amount of textspeak could have stabilized from 2012 to 2013. The language change that had taken place in 2012 has now become stable. If looked at from a traditional view of language change, this seems very odd, as languages typically change slowly or over a few generations, not over a matter of months. We ought to keep in mind, however, that the Internet allows for very rapid and frequent communications. Millions of tweets are posted per day. Variations are introduced and spread through retweets; the most popular ones are perpetuated in this manner until they are discarded in favor of a newer term. Taking the data from both 2012 and 2013 together, we can argue that there was a sudden change from using a lot of textspeak towards the more standard and more formal variant.

Next, I examine the usage of nonstandard forms per grade, such as apostrophe deletion and the use of lowercase *u* for *you* and *i* instead of *I*, as shown below in Figure 3.
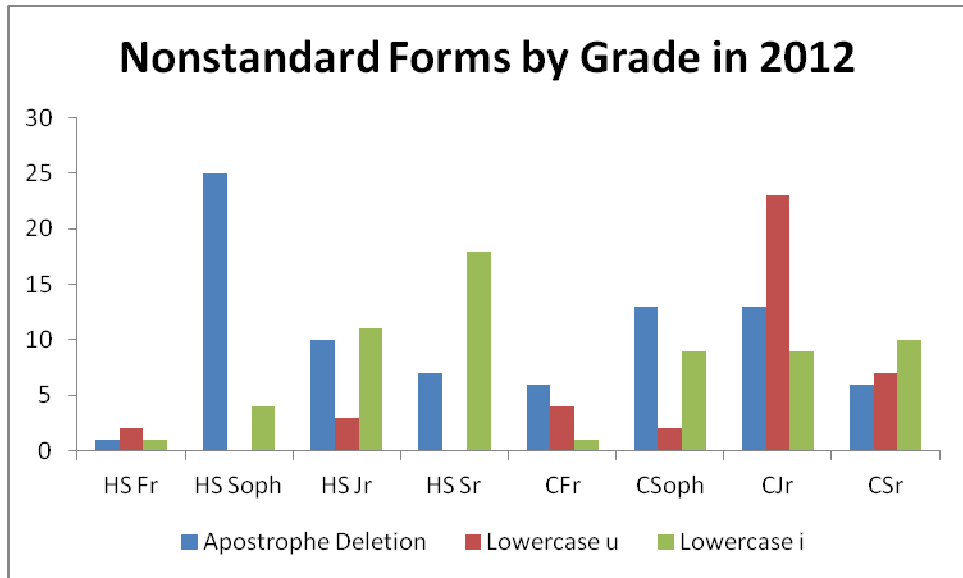
Figure 3 Total amount of nonstandard forms among high school and college students in 2012

In 2012, the number of total tokens of lowercase *i* increases from 1 for high school freshmen to 18 for high school seniors. College freshmen have a total of 1 token; the amount remains almost the same for the upper years. The number of total tokens of lowercase *u* does the opposite. Where college juniors have 23 tokens of *u*, they only have 9 of the lowercase *i*; similarly, high school juniors have 18 tokens of lowercase *i*, but only 3 lowercase *u.* Apostrophe deletion follows a similar pattern to the other two variables, increasing to 25 tokens in high school sophomores, then decreasing to 6 tokens in college freshmen, then increasing again to 13 tokens in sophomores and juniors, and finally decreasing to 6 tokens in college seniors.

There are some gaps in the data; this occurs when there are no tokens for that particular variable. For instance, the three high school sophomores and seniors I examined did not use *u* at all in 2012. However, the high school students do not use lowercase *u* as much as the college students; therefore, these gaps align with the pattern followed by other high school students.
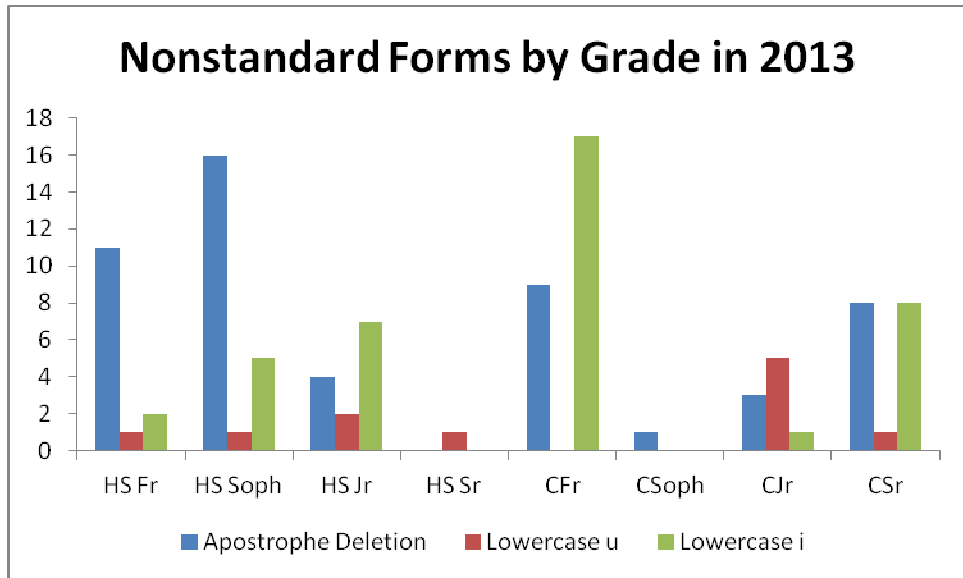
Figure 4 Total amount of nonstandard forms among high school and college students in 2013

In 2013, the number of lowercase *u* drops to under 2 tokens for almost all grades except college juniors, who have 5 tokens. Similarly, the amount of lowercase *i* has decreased as well; college sophomores did not use the lowercase *i* at all. College freshmen, however, have a total of 17 tokens. The number of apostrophe deletions has also decreased; high school seniors do not have any at all.

Across both 2012 and 2013, the number of lowercase *i* and apostrophe deletion is much higher than lowercase *u*. However, while there is a definite decrease in all three variables over time, I cannot draw any conclusions about the direction of the change with respect to grade.

If the individual data in Table 2 is examined, it is apparent that the number of tokens varies by individual. Where one high school junior may only have five lowercase *i*'s and four instances of apostrophe deletion, another has four lowercase i's and only two instances of apostrophe deletion. Additionally, the majority of users of *i* are high school students; 7 out of 12 college students do not

use the lowercase *i* at all, whereas only 4 out 12 high school students did not.

College students used half as many lowercase *i*'s as high school students. It is

impossible to determine if language change is taking place, as the number of

tokens is too low. This suggests that the usage of nonstandard forms is very much

a matter of individual style. When we take into account the usage of standard

forms, we see that the standard form is favored over the nonstandard form; in the

cases where the nonstandard form dominates, the standard form is used less often.

	Lowercase *i* in High School and College Students	
	2012	2013
HF1	0	0
HF2	0	0
HF3	0	1
HSo1	0	0
HSo2	0	6
HSo3	1	0
HJ1	1	0
HJ2	0	0
HJ3	2	4
HSr1	1	1
HSr2	3	0
HSr3	2	1
CF1	5	1
CF2	0	1
CF3	0	0
CSo1	2	1
CSo2	2	0
CSo3	0	2
CJ1	0	2
CJ2	4	0
CJ3	2	1
CSr1	0	0
CSr2	3	1
CSr3	8	1

Table 2 Total amount of lowercase *i* tokens among high school and college students in 2012-2013
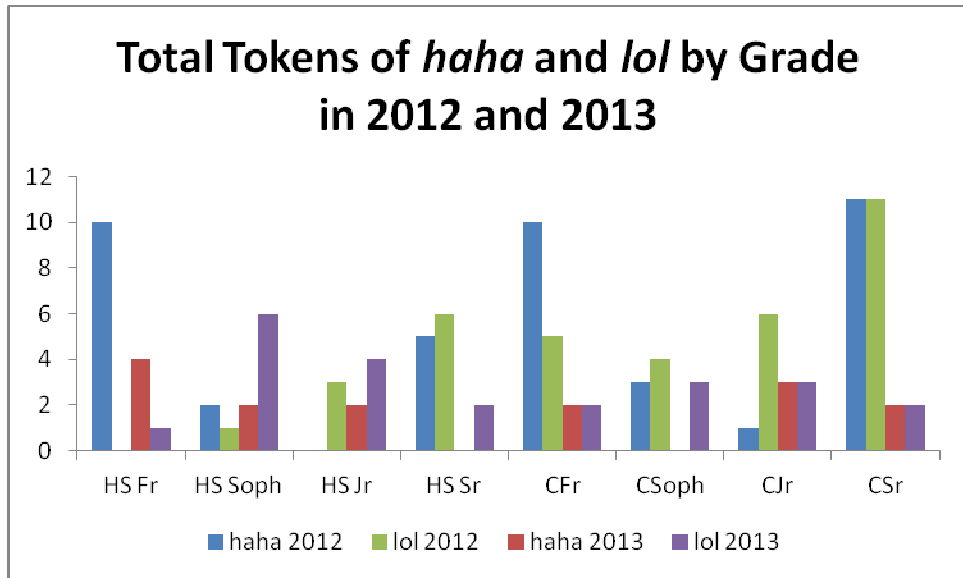
Figure 5 Number of uses of *haha* and *lol* by grade from 2012-2013

Lol and its variants are all ways of depicting laughter. In the case of *haha* and *lol*, *haha* is considered the standard form. Figure 5 shows the total tokens of *haha* and *lol* used by high school and college in 2012 and 2013. In 2012, *haha* dominates, except in a few instances where more tokens of *lol* are used. For instance, college juniors have 1 token of *haha* and 6 of *lol*; similarly, high school juniors have 0 tokens of *haha* and 3 of *lol*. This could be due either to personal preference, or it could be indicative of their social networks. More data would need to be collected in order to draw a more definite conclusion. In 2013, *lol* either dominates or has as many tokens as *haha*. The one exception is among the high school freshemen, who use *haha* more than *lol*. However, the number of tokens of both *haha* and *lol* decrease beginning with the high school seniors and continuing until the college seniors.

4.2 Discussion of Twitter Analysis:

One possible reason for the decrease in textspeak tokens in 2013 is efficiency. In the beginning, as students are less familiar with Twitter, they would use fewer characters and more textspeak in an attempt to stay under the 140 character limit. As students become more familiar with Twitter, they gradually stop relying on the textspeak and instead use the forms they would typically use in more formal settings.

A possible reason for the increase in textspeak tokens in 2012 could be societal. The increase occurs as the students graduate from middle school to high school. High school is traditionally associated with more freedom, and it is traditionally a period of rebellion and of finding one's identity. College is a continuation of that; this would explain the increase in textspeak during this time. However, as students become juniors and seniors, there is an expectation that they will find jobs or attend graduate school after college. This would explain the decrease in both textspeak and in nonstandard forms, as well as the relatively low levels of *haha* and *lol*.

4.3 Twitter and Gender

It is expected that males use more of the nonstandard forms than females due to societal and cultural expectations. Figure 6 shows the average tokens of textspeak used by males and females in 2012 and 2013.
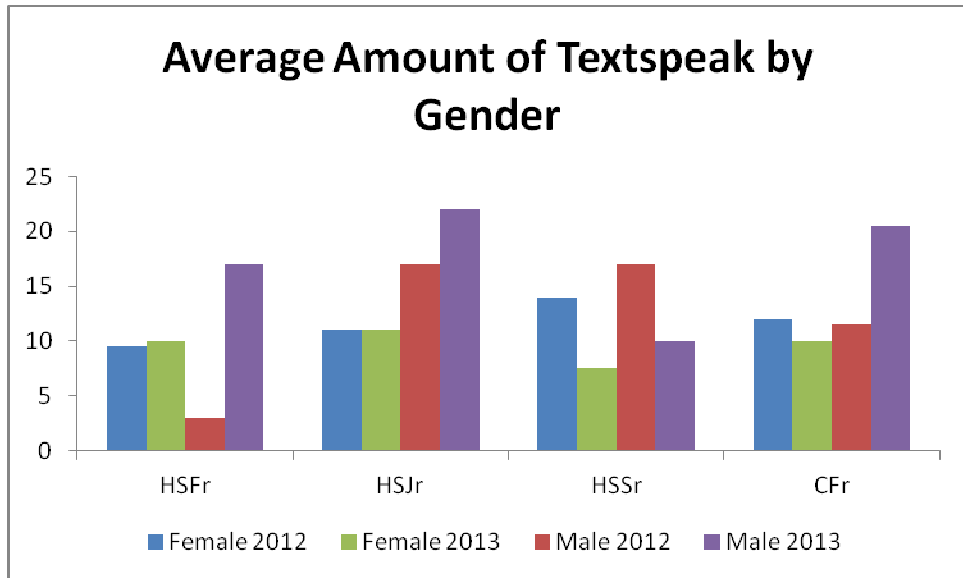
Figure 6 Average amount of textspeak by gender in 2012 and 2013

Male high school freshmen had an average of 3 total textspeak tokens in 2012; high school juniors and high school seniors had an average of 17 tokens, and college freshmen had an average of 11.5 tokens. A similar pattern is found in 2013. Among high school students, the average number of tokens increases from 17 tokens to 22 tokens from freshman to junior year. The average drops to 10 tokens in senior year and then increases to 20.5 tokens in college. The data suggest a cyclical pattern, in which high school seniors tend to use much less textspeak than any of the other grades.

The average number of tokens for females in 2012, on the other hand, exhibits a similar cyclical pattern. The amount of tokens would increase gradually until it reached 14 tokens in the senior year of high school, then decrease to 10 tokens in the sophomore year of college; the number of tokens increases during the senior year of college to 35 tokens. This pattern is also found in the 2013 data; however the increase in tokens during the college junior and senior years is much smaller, staying at about 15 tokens. This suggests that nonstandard variants cycle;

35

they fall in and out of favor according to societal or cultural factors. The most likely cause of this is college, as American high school students typically apply for college during their senior year of high school.

When the data for both genders is compared, the overall expectation is upheld. With two exceptions, males use more textspeak than females during both 2012 and 2013. In the case of the college freshmen, the difference is negligible: half a token, and as such it can be discarded. In the case of the high school freshmen, the difference is much greater: 9.5 tokens to 3.
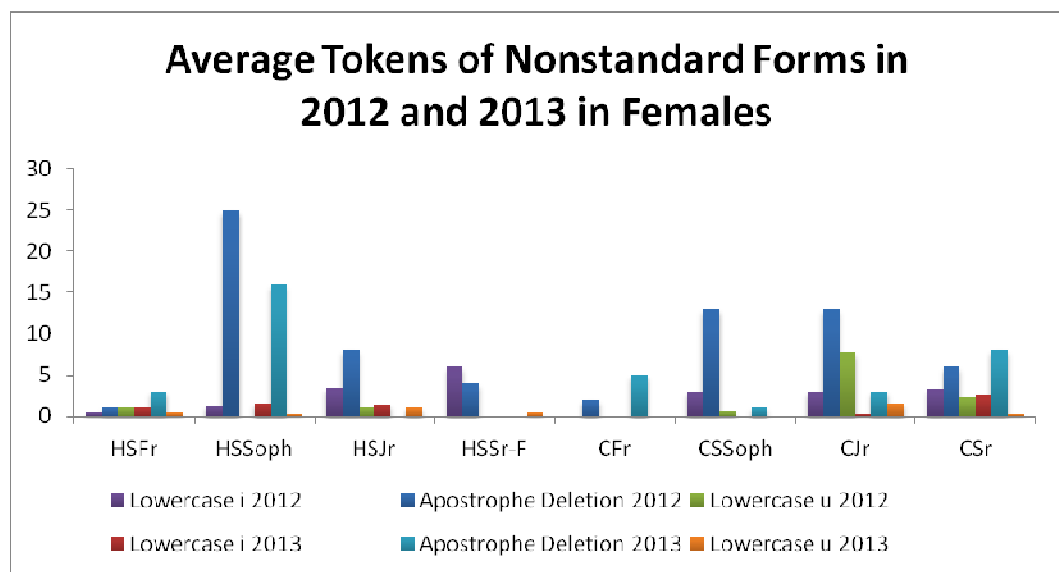

Figure 7 Average tokens of nonstandard forms by female high school and college students in 2012-2013

When the nonstandard forms, the lowercase *u* and *i* and the apostrophe deletion, are examined, a different pattern emerges. The first thing I notice is that the all three variables for both years follows a cyclical pattern. It peaks in the high school and college sophomore and junior years and drops, in some cases to zero, during the freshman and senior years. The second thing is that, in terms of frequency, apostrophe deletion has the greatest number of tokens, suggesting that

it is both easier and more common to use. The third thing is that there is an

inverse relationship between the number of lowercase *i*'s and the number of

lowercase *u*'s, suggesting that the focus of tweets is on the individual who is
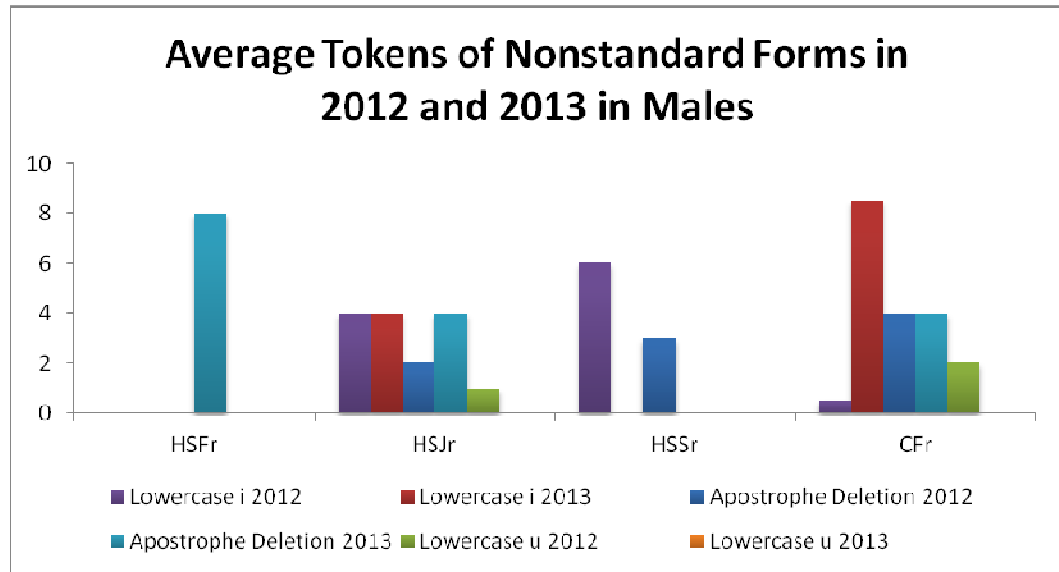
composing the message.

Figure 8 Average usage of nonstandard forms by male high school and college students in 2012-
2013

Whereas all three variables for females occurred in cycles, only two do

for the males. The only variable to not follow a cyclical pattern is the apostrophe

deletion, which increases linearly over time. The use of lowercase *i*, however,

goes from 0 to 1 to 0 to 2 over the span of 5 years. The use of lowercase *u* goes

from 0 to 4 to 6 to .33 in the same period of time. The same inverse pattern in the

use of lowercase *i*'s and *u*'s that was present in the female data is present here as

well.

Average Tokens of *lol* by Gender in 2012 and 2013				
	Female *lol* 2012	Male *lol* *2012*	Female *lol* 2013	Male *lol* *2013*
HSFr	0	0	1	0
HSJr	1	2	0	4
HSSr	4	2	1	1
CFr	0	5	1	1

Table 3 Average usage of *lol* by male and female high school and college students in 2012-2013

Average Tokens of *haha* by Gender in 2012 and 2013				
	Female *haha* 2012	Male *haha* 2012	Female *haha* 2013	Male *haha* 2013
HSFr	8	2	0	4
HSJr	0	0	0.5	1
HSSr	0	5	0	0
CFr	4	6	2	0

Table 4 Average usage of *lol* by male and female high school and college students in 2012-2013

In 2012, male high school freshmen and seniors use fewer *lol*s and more *hahas* than their female counterparts. The low numbers could be due to the fact that Twitter is mainly used for short conversations and status reports. As the numbers are so low, it is very difficult to determine if language change is taking place.

4.3.1 Discussion of Twitter Analysis by Gender:

The overall trend of males using more nonstandard forms than females could be due to cultural reasons. American females are expected to be quiet and lady-like; consequently, they tend to use more standard forms. American males, on the other hand, are given much more freedom to express themselves and to do

and say what they want, whereas females are encouraged to conform to the status quo.

4.4 Analysis of Tumblr Data:

Unlike Twitter, the data presented here can only by analyzed with respect to the users' age or grade, as many users did not post information, such as profile pictures, that would allow me to draw any conclusions about their gender. Regardless, I expect to see patterns similar to those found on Twitter.
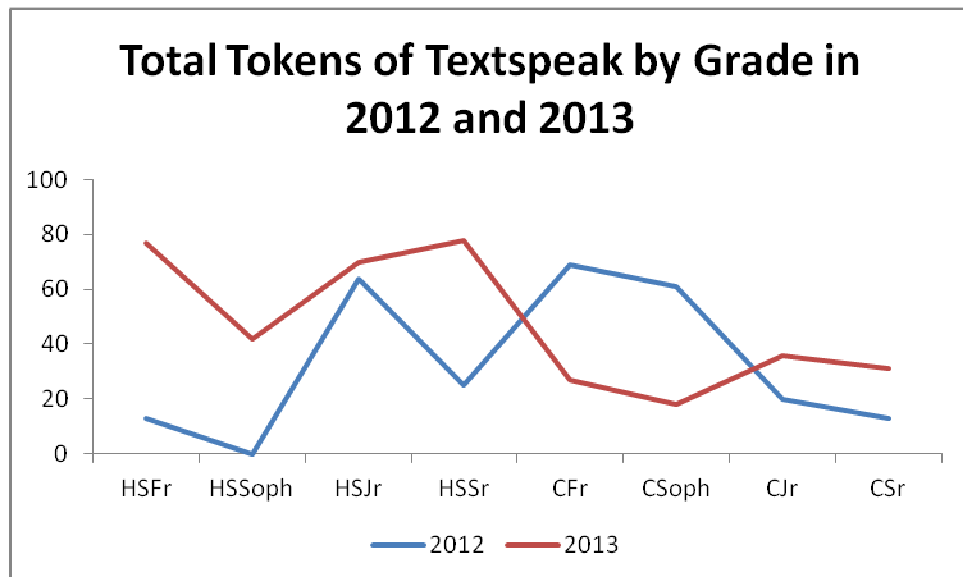

Figure 9 Average usage of textspeak by high school and college students in 2012-2013

Figure 9 shows the average amount of textspeak used by high school and college students in 2012 and 2013. In 2012, there is an increase in the total amount of textspeak among high school students from 15 to 65 tokens; this decreases by the senior year to 25 tokens. The amount increases in college freshmen to 70 tokens and then decreases to 10 tokens in college seniors. There is no clear pattern here. The amount of tokens for high school seniors and college

seniors are lower for anyone else except high school freshmen. In 2013, the high

school freshmen start at 78 tokens; this amount gradually declines to about 15

tokens by the sophomore year of college. This then increases to 30 tokens among

college seniors. As with the data from 2012, the data shows no obvious pattern.
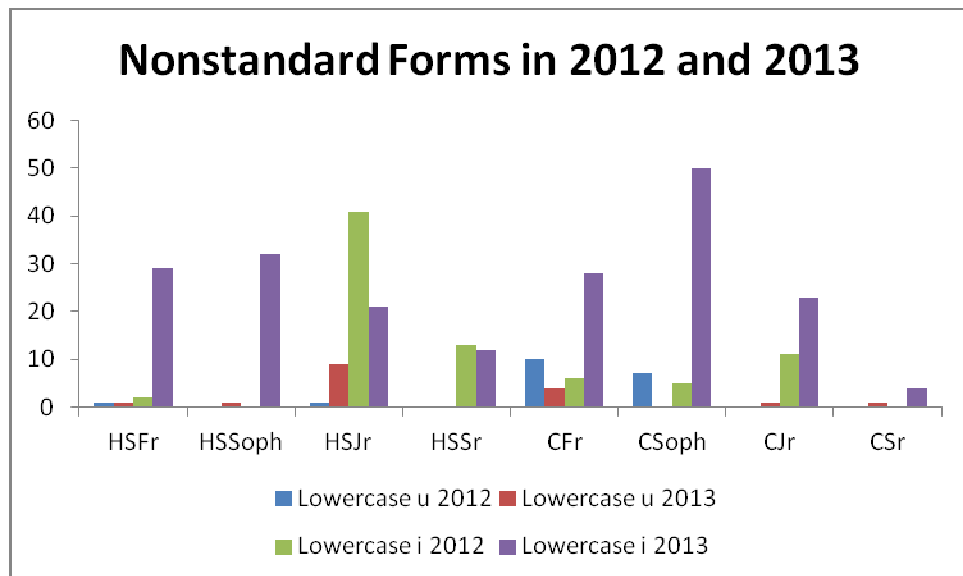

Figure 10 Average usage of nonstandard forms by high school and college students in 2012-2013

Figure 10 shows the average tokens of nonstandard forms in 2012 and

2013. In 2012, the high school students have almost no occurrences of the

lowercase *u*, as do the college juniors and seniors. The college freshmen,

combined, have 10 tokens, and the college sophomores have 7. In 2013, only the

high school juniors and college freshmen have more than 3 tokens of the

lowercase *u*. This suggests that *u* is much less prevalent than the lowercase *i*, as

the overall amount of tokens is much greater. It also suggests that the usage of *u*

could be cyclical in nature, but more data would be needed to ascertain this.

For the lowercase *i* in 2012, the high school freshmen and sophomores

have 2 tokens total; the high school juniors have 41, and the high school seniors

have 13. The amount of tokens the college students have drops from 6 to 0 over the four years. This suggests a cyclical pattern, peaking at the high school juniors, and then dropping off. In 2013, the high school freshmen and sophomores have more tokens of lowercase *i* than in 2012, but the high school seniors and juniors have less. The amount of tokens increase throughout the first two years of college, peaking at the sophomore year and then declining to 3 tokens in the senior year. This cyclical pattern is the inverse of the one seen in the previous year, where the high school students had more tokens of lowercase *i*.
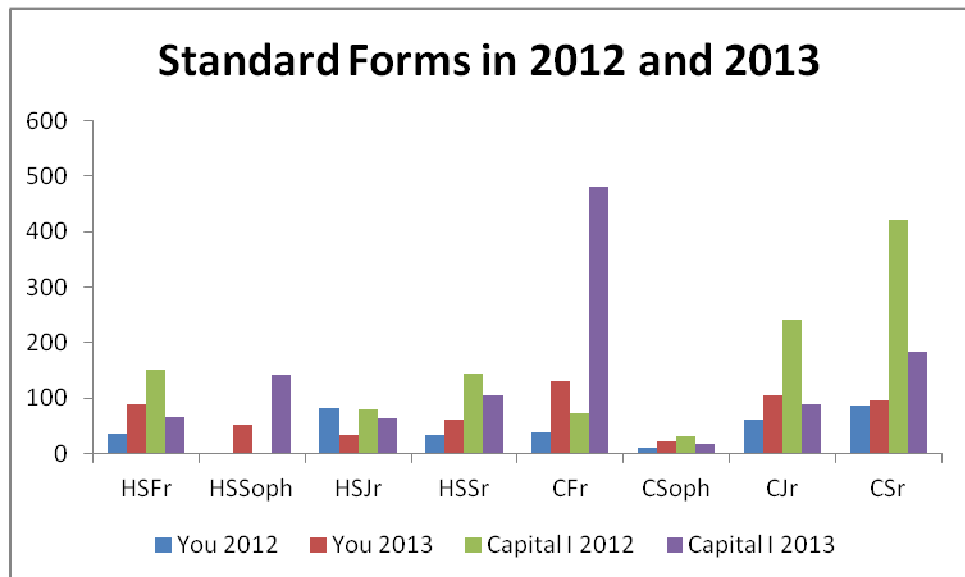

Figure 11 Average usage of standard forms by high school and college students in 2012-2013

Figure 11 shows the average usage of standard forms in 2012 and 2013. For college juniors and seniors, there is an increase in the number of standard forms used; at the same time, there is a decrease in the number of nonstandard forms used. This increase in the usage of the standard form can also be seen in the high school junior and senior years. For example, high school juniors have a total of 63 tokens of capital *i* in 2013; high school seniors have 106 tokens. Likewise,

college juniors have 90 tokens and college seniors have 184 tokens. However, high school juniors only have 21 tokens of lowercase *i*; seniors have a mere 12. College juniors have 23 tokens, whereas college seniors have a total of 4 tokens of lowercase *i*. It is also worth noting that the number of standard forms is ten times higher than the nonstandard forms. This suggests that the standard version of the variant is favored over the nonstandard form.

Total Tokens of *haha* by Grade

	2012	2013
HSF	0	1
HSSoph	0	1
HSJr	5	1
HSSr	5	7
CFr	1	6
CSoph	3	1
CJr	1	2
CSr	1	1

Table 5 Number of tokens of *haha* by high school and college students in 2012-2013

Total Tokens of *lol* by Grade

	2012	2013
HSFr	0	1
HSSoph	0	1
HSJr	1	9
HSSr	4	3
CFr	5	1
CSoph	13	5
CJr	1	5
CSr	5	7

Table 6 Number of tokens of *haha* and *lol* high school and college students in 2012-2013

In 2012, the amount of *haha* used begins at 0 for high school freshmen and sophomores, and increases to 5 for high school juniors and seniors; it then drops to 1 token for college freshmen and remains between 1 and 3 tokens for

42

college sophomores through college seniors. A similar pattern is seen for *haha* in 2013, only high school seniors and college freshmen have the highest number of tokens at 7 and 6, respectively. However, the total amount of tokens of *lol* is greatest among college students for both 2012 and 2013; they have 24 total in 2012 and 18 in 2013. High school students have 5 in 2012 and 14 in 2013. This confirms Tagliamonte and Denis' (2008: 13) observation that an increase in *haha* tokens leads to a decrease in *lol* tokens.

4.4.1 Discussion of Tumblr Analysis:

Younger Tumblr users use more textspeak than older users. This could be due to cultural reasons. Tumblr grants its users anonymity; signing up for an account requires a valid email address and nothing else, unlike Facebook, which requires the user's name, an email address and other information like the user's birthdate. Twitter requires the user's name and email address. This allows users to express themselves freely, with very little fear that authority figures can track them. This in turn leads to a greater use of textspeak in younger users, as the American education system encourages students to use the more accepted standard forms.

However, this does not explain the fact that older users use less textspeak. This could be due to the fact that older users have had more experience with the medium and decided that they do not want to conform to the stereotype of being a teen whose messages are filled with unreadable gibberish. A similar reason could be given for the patterns seen in all nonstandard forms examined, that an increase

in the standard form leads to a decrease in the tokens of the nonstandard form. To the younger users, Tumblr is this new website where they can do or say whatever they want. As time goes on, this freedom becomes less of a novelty, so they revert to what they know and what they were taught in school. As there is no post length restriction, it would be logical to assume that this would mean that the amount of textspeak would decrease.

Another reason would be the presence, or perceived presence, of an authority figure. Many teens tend to censor their language when they are around an authority figure, or when they believe that such a person may have access to their writing. Tumblr provides the outlet that teens need to express themselves freely, as most adults tend to use Twitter or Facebook as their social media platform of preference.

4.5 Comparing Twitter and Tumblr:

I would assume that the amount of nonstandard forms and textspeak in Twitter would be much greater than in Tumblr. However, it seems that the amount of nonstandard forms and textspeak is dependent on the form of social media. Below, in Table 8, I consider the total percentage of textspeak in IM, Twitter, and Tumblr, as well as the total tokens of *haha* and *lol*. The IM data was collected in 2004 by Tagliamonte and Denis (2008); it confirms my findings for Twitter, but not Tumblr, as the total tokens of *haha* outnumber the total tokens of *lol*.

Percentage of Textspeak and Total Tokens of *haha* and *lol* in IM, Twitter, and Tumblr data

Type of Social Media	haha (total)	lol (total)	% total textspeak
IM	16,183	4,506	2.40%
Twitter	57	59	1.59%
Tumblr	36	62	1.50%

Table 8 Total tokens of *haha* and *lol* in relation to total textspeak by social media

Tagliamonte and Denis (2008: 13) examined instant messages of 72 teenagers between the ages of 15 to 20; the data was collected between 2004 and 2006. IM, a primarily conversation based medium, shows roughly 16,000 tokens of *haha*, compared with the more blog based mediums of Twitter and Tumblr, which have less than 60 tokens of *haha*. Additionally, the percentage of total textspeak is much higher in IM, 2.4%, whereas Twitter has 1.59% and Tumblr has 1.5% (Tagliamonte and Denis, 2008, p. 13). Clearly, textspeak is used much more in IM than in other forms of social media. Twitter has the second highest percentage of textspeak, and Tumblr has the least. This suggests that the purpose of the form of social media should be taken into account when performing analysis, as well as the user base. If the purpose of the social media platform is intended to be chat-based, like IM or texting, it is expected that there will be more textspeak and nonstandard forms, as long gaps between replies and answers can be taken as disinterest and can lead to the end of the conversation. If, on the other hand, the purpose of the platform is not chat-based, then less textspeak is expected. Often, as in the case of a blog, the content is intended for a specific purpose and for a specific audience. This usually means that a more formal tone and style is required.

Adults tend to use social media to share news or promote brands or for professional reasons. The language used is very formal, as they are trying to build either their own or others' reputation, and they know that consistency is important. Teens, on the other hand, discuss celebrities, pop culture, popular Internet memes, and other topics (Nagy, 2014). This is true in both Twitter and Tumblr. While some users would chat about various events in their lives, such as an upcoming exam or a party that had happened over the weekend, most users, especially on Tumblr, would promote pictures, videos, and conversations that they liked. One teen commented that she used Tumblr as a way to escape the burdens of the real world; she could just go and mindlessly browse Tumblr feeds and look at things she liked to make her feel better (Shelby_Harris, 2013).

Chapter 5: Conclusions and Further Research

This study set out to show that teens' language is changing as a result of interaction with social media. My original hypothesis stated that older teens would use less of the standard form regardless of the social media platform used. With respect to gender, male teens would use more of the nonstandard forms than the females.

Based on my analysis of the data collected, I found that Twitter users in high school, especially freshmen and sophomores, used less chatspeak than college juniors and seniors, but Tumblr users do the opposite; college Tumblr users use less chatspeak than high school users. From 2012 to 2013, both Tumblr and Twitter users use fewer nonstandard forms, like *lol*, lowercase *u*, and lowercase *i*, but the amounts varied from grade to grade inconsistently. This inconsistency is a stylistic choice, rather than language change. Female Twitter users used less chatspeak and nonstandard forms than male students, as expected from previous research. Finally, high school and college seniors tended to use less of the nonstandard forms than the standard forms than other years. This suggests that the nonstandard forms are gradually being phased out in favor of the standard forms.

First, this shows that the analysis of one social media platform cannot be generalized to another. Previous analyses would analyze one platform, like Twitter or IRC, and then try to generalize these results for the whole medium. These generalizations do not always hold, as the results here show. The analysis

from Twitter and Tumblr were completely different. When they are compared with Tagliamonte and Denis' (2008) IM data, the results suggest that how the social media platform is used affects the results. A platform that is more conversation-based would most likely have more chatspeak than one that is more fact-based or directed at a certain audience, like a blog.

Second, this shows that even with the introduction of chatspeak, people still tend to use the standard form. While some of their writing might vary according to their individual styles, such as not capitalizing a proper noun or deleting apostrophes, in the long run, the standard form is favored. Even if we take into account various factors such as message length and anonymity, we see that the standard form dominates.

Third, it shows that language change is happening much faster than it ever has been, due to the wide-reaching nature of the Internet. In the past, a variation could take years to spread throughout a population; now, a variation could take less than a year to spread and become accepted within a population.

Finally, it shows that, despite the media's claims that the English language is being ruined, that the English language is being shaped and changed just as it always has been, with teens at the center of it all.

5.1 Further Research

This study raises many questions about the variables that may affect teens' usage of chatspeak in social media. For instance, it would be interesting to examine the impact of socio-economic status and education on the amount of

textspeak and the usage of nonstandard forms. As the standard form of a language is more typically associated with both high levels of education and higher socio-economic status, I would expect that individuals matching these categories would use less textspeak and more standard forms in their writing.

It would be interesting to examine the influence of an individual's social network on the language he uses. Dense networks, where everyone knows everyone, typically resist change; looser networks are more open to change (Milroy and Milroy, 1985). An individual with fewer ties to a community starts the change, as there is little social pressure for the individual to conform to the community's speech. The change is spread by the individuals who have strong ties to the community, as they have more influence (Milroy and Milroy, 1985). In a social media platform, the individual's social network would be determined by his followers and the people he is following.

It would be interesting to examine how location affects language change. We know that change often happens first in suburban areas, and the language is slower to change in less populated areas (Mesthrie et al., 2001). It would be likely that people from less urban areas would use more of the nonstandard forms and more chatspeak than people from the cities.

It would be interesting to examine not just teens, but also adults' usage of social media. Although the majority of social media users are teens, the Pew Research Center writes that 46% of adults older than 65 use social media; 65% are between the ages of 50 to 64 (Pew Research Center, 2013). Since teens are

leading the language change, I would assume that the older adults would use less chatspeak and more standard forms than the younger generations.

A final question for research is "do teens use the same language across different social media platforms?" This follows logically from this study, where I examined two social media platforms and different sets of people in both mediums. Would the same patterns hold if the same teen's language was examined across different social media platforms?

5.2 Limitations

I faced several limitations over the course of this study. One limitation of the study was the relatively small sample size for each grade. As such, the data should not be taken as representative of the community as a whole. Another limitation concerns subject selection. I had not originally intended to analyze the data with respect to gender, and as a result only a quarter of my subjects are male. I was only able to directly compare those grades that had both male and female subjects.

References

Ahlqvist, T., Back, A., Halonen, M., and Heinonen, S. (2008). Social media roadmaps: Exploring the futures triggered by social media. *Espoo 2008. VTT Tiedotteita --Research Notes 2454*, pp. 1-79

Androutsopoulos, J. (2006). Introduction: Sociolinguistics and computer-mediated communication. *Journal of Sociolinguistics*, 10(4), pp. 419-438.

Asur, S., and Huberman, B. A. (2010). Predicting the future with social media. Arxiv preprint arXiv:1003.5699.

Baron, N. (2004). See you online: gender issues in college student use of instant messaging. *Journal of Language and Social Psychology*, 23(4), pp. 397-423.

Beese, J. (2013). *Pew breaks down social media demographics by age, gender, and race*. Retrieved April 13, 2014 from http://sproutsocial.com/insights/social-media-demographics/

Boyd, D. M. (2008). Social network sites: Definition, history, and scholarship. *Journal of computer-mediated communications*, 13, pp. 210-230.

Bylund, A. (2006). *To google or not to google*. Retrieved April 13, 2014 from http://www.fool.com/investing/dividends-income/2006/07/05/to-google-or-not-to-google.aspx

Carlson, N. (2010). At last--The full story of how Facebook was founded. *Business Insider*. Retrieved April 13, 2014 from http://www.businessinsider.com/how-facebook-was-founded-2010-3

Crystal, D. (2011). *Internet linguistics: A student guide*. New York, NY: Routledge.

Crystal, D. (2001). *Language and the internet*. New York, NY: Cambridge University Press.

Denby, L. (2010). The language of Twitter: Linguistic innovation and character limitation in short messaging. (Unpublished undergraduate dissertation). University of Leeds, UK. Retrieved April 13, 2014 from http://lewisdenby.wordpress.com/2010/06/29/twitter-research-my-dissertation-online/

Fennell, B. (2001). *A history of English: A sociolinguistic approach*. Oxford, UK: Blackwell Publishers.

Gay. (n.d.). In *Online Etymology Dictionary*. Retrieved April 13, 2014 from
http://www.etymonline.com/index.php?term=gay&allowed_in_frame=0

Hendricks, D. (2013, May 8). *Complete history of social media: Then and now*.
Retrieved April 13, 2014 from http://smallbiztrends.com/2013/05/the-
complete-history-of-social-media-infographic.html

Hickey, R. (2012). *The early modern period*. Retrieved April 13, 2014 from
https://www.uni-due.de/SHE/HE_Grammar_EME.htm

Labov, W. (2007). Transmission and diffusion. *Language*, 83(2), pp. 344-387.

Ling, R., and Baron, N.S. (2007). Text messaging and IM: Linguistic comparison
of American college data. *Journal of Language and Social Psychology*,
26, pp. 291–98.

Linhorst, M. [M_Linhorst] (2014, April 10). Two weeks #capstonecountdown
@SUHonors [Tweet] Retrieved April 13, 2014 from
https://twitter.com/M_Linhorst/status/454273922200907776

Mesthrie, R., Swann, J., Deumert, A., Leap, W. L. (2001). *Introducing
sociolinguistics*. Philadelphia, PA: John Benjamins Publishing Company.

Milroy, J., and Milroy, L. (1985). Linguistic change, social network and speaker
innovation. *Journal of Linguistics*, 21, pp. 339-384.

Nagy, E. (2014, February 18). *What your teen is really doing all day on Twitter
and Instagram*. Retrieved April 13, 2014 from
http://www.fastcompany.com/3026499/most-creative-people/what-your-
teen-is-really-doing-all-day-on-twitter-and-instagram

O'Conner, P. T., and Kellerman, S. (2010, September 26). *Should English dance
to a Latin beat?*. Retrieved April 13, 2014 from
http://www.grammarphobia.com/blog/2010/09/latin.html

O'Connor, B., Balasubramanyan, R., Routledge, B. R., and Smith, N. A. (2010).
From tweets to polls: Linking text sentiment to public opinion time series.
Fourth International AAAI Conference on Weblogs and Social Media.

Owoputi, O., O'Connor, B., Dyer, C., Gimpel, K., Schneider, N., and Smith, N. A.
(2013). Improved part-of-speech tagging for online conversational text
with word clusters. *Proceedings of NAACL 2013.*

Owoputi, O., O'Connor, B., Dyer, C., Gimpel, K., Schneider, N., and Smith, N. A.
(n.d). Twitter NLP and part-of-speech tagging. Retrieved April 20, 2014
from http://www.ark.cs.cmu.edu/TweetNLP/

Paolillo, J. (1999). The virtual speech community: Social network and language variation on IRC. *Journal of Computer-Mediated Communications*, 4(4).

Palfreyman, D., and al Khalil, M. (2003). "'A funky language for teenzz to use': Representing Gulf Arabic in instant messaging." *Journal of Computer-Mediated Communication*, 9(1).

Rosen, L. D., Chang, J., Erwin, L., Carrier, M., and Cheever, N. (2010). The relationship between "textisms" and formal and informal writing among young adults. *Communication Research*, 37(3), pp. 420-440.

Scheff, S. (2013, January 15). Why are more teens turning to Tumblr?. *Huffington Post*. Retrieved April 13, 2014 from http://www.huffingtonpost.com/sue-scheff/teens-and-tumblr_b_2482848.html

Shelby_Harris. (2013, January 15). Why are more teens turning to Tumblr?. *Huffington Post*. [weblog comment] Retrieved April 13, 2014 from http://www.huffingtonpost.com/sue-scheff/teens-and-tumblr_b_2482848.html

Social networking fact sheet. (2013). Pew Research Center. Retrieved April 13, 2014 from http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/

Squires, L. M. (2007). Whats the use of apostrophes? Gender difference and linguistic variation in instant messaging.

Thurlow, C. (2003). Generation txt? The sociolinguistics of young people's text-messaging. *Discourse Analysis Online*, 1(1), 30.

Tagliamonte, S., and D'Arcy, A. (2007). Frequency and variation in the community grammar: Tracking a new change through the generations. *Language Variation and Change*, 19, pp. 199-217.

Tagliamonte, S., and Denis, D. (2008). Linguistic ruin? Lol! Instant messaging and teen language. *American Speech*, 83(1), pp. 3-34.

Our history in depth. (2014). *Google*. Retrieved April 19, 2014 from https://www.google.com/about/company/history/

Trudgill, P. (1972). Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language in Society*, 1(2), pp. 176-195.

Wheeler, L. K. (2014, March 14). *The great vowel shift*. Retrieved April 13, 2014 from http://web.cn.edu/kwheeler/great_vowel_shift.html

Yates, S. (1996). Oral and written linguistic aspects of computer conferencing: A corpus based study. In S.C. Herring (Ed.). *Computer-Mediated Communication: Linguistic, Social, and Cross-Cultural Perspectives* (pp. 29-46). Amsterdam: John Benjamins.

Zimmer, B. (2009, May 7). *The language of social media: "Unlike" any other*. Retrieved April 1, 2014 from http://www.visualthesaurus.com/cm/wordroutes/the-language-of-social-media-unlike-any-other/