Not 2 Cryptic 2 DCode:

Paralinguistic Restitution, Deletion, and Non-standard Orthography in Text Messages

Abstract:

This thesis examines the structure of text messages. In recent years, literature speculating about electronically mediated communication has proliferated. An abundance of literature on technology and language exists, but little of it explores text messaging. The literature that looks at texting tends to focus on the social aspects of text communication or on the damage people fear it will cause to language. Little literature focuses on empirical analysis of text messaging from a linguistic perspective. Text messages are a communication medium with limitations and intricacies all their own, and they deserve attention.

The informal nature of texting allows for a variety of lexical and grammatical creativity. Letter and word deletions appear, perhaps inspired by the 160 character per message length limit. Unconventional punctuation and spelling abound. Text messaging has become a significant part of language use in our culture, especially for young people. Today, phones are used more for text messaging than for voice communication in many countries. Texting is a vital piece of the technology-mediated-communication puzzle and warrants inspection; we cannot tackle the question of what digital technology means for social interaction or for language until we understand text messaging structurally. It is worth remembering, too, that as phones and phone plans advance, the 160 character limit - one of the factors unique to texting and perhaps integral in generating the new linguistic phenomena we see in text speak - will become less meaningful. Perhaps even more critically, texting patterns are also changing as phones change. In analyzing the structure of

text messaging today, we may be capturing a unique moment in the tech-language trajectory before a new type of electronically mediated communication replaces or changes texting and we lose this piece of the language history.

In this thesis I use two pilot studies – one survey and an analysis of a collection of text messages – that illuminate patterns in text messaging language. The studies involve a total of fourteen participants in three age groups. The data corpus demonstrates both similarities and differences in the way different groups use text language. Patterns in the data corpus suggest clarifications to existing literature the on non-standard orthographic features found in text messaging and on what patterns govern lexical deletion in text messages. The survey data provides evidence of a subtler use of paralinguistic restitution[1] by college-age texters and sheds insight into new complexities within the non-standard orthography found in text messages.

1. Introduction

In a recent episode of the popular television series "Bones," the psychologist is given the task of decoding a series of text messages between a murdered man and his friends. The texts are unreadable to most of the characters, consisting of mysterious strings like *M I 2 M 2 H 6y* (translated to mean "Am I too much to handle sexy?") Once the psychologist solves all of the texts, they solve the case. All experience indicates that messages constructed completely of initializations[2] and letter/number homophones[3] are not representative of how people actually text message each other. Nonetheless, this idea about text messaging being enigmatic is indicative of a real attitude

---

[1] Ways of writing that compensate for the lost prosodic and visual cues found in face-to-face interaction (e.g. *:)*)

[2] Terms constructed out the the initials of words in a phrase, (eg. *lol* for "Laugh Out Loud")

[3] Words where a number or letter is substituted for a homophonic phoneme (e.g. *2day* for 'today")

towards the technology. Texting has captured people's interest for its cryptic nature. Since the technology is utilized most heavily by teenagers and young adults and does deviate from conventional norms of written or spoken English, it is no far leap for people who do not text to suspect the young of having a secret language. In truth, most text messages are legible to most people, whether or not they have ever seen a text message before. Still, there is something to the idea that text messages represent a particular linguistic genre. Frequent users of text messaging do deviate from standard orthography in predictable ways, and some research indicates that they might acquire a type of texting fluency (Thurlow and Poff 2010). Real texts are not riddles, but there are unexplored patterns of how orthography changes to fit the medium of text over a mobile phone. Research into the area of American text messages has just begun and it is worth studying how American English bends to fit this medium.

2. Background

2.1 What is Texting?

Text messaging or Short Message Service (SMS) started in Europe in 1993 as a free service provided by the European phone service consortium Groupe Special Mobile (GSM) (Baron 2008). Texting prototypically consists of using the keypad on a mobile phone to type messages that are sent to someone else's mobile phone. The person who receives that message can then type a message in return. Although it was initially an afterthought for GSM, the popularity of text messaging has now ballooned to the point where today SMS use actually outranks mobile phone use for voice calls in Europe and much of the world (Hård af Segerstad 2002). Although its medium is the mobile phone,

SMS is often considered under the heading of Electronically Mediated Communication[4] or EMC, along with Instant Messaging (IM), e-mail, and more.

Before talking about the linguistic characteristics of SMS, we should understand the physical and economic constraints of the medium. The original mechanism for typing with phones utilized the letters already on the number keys - with three or four letters on each number, and the entire alphabet represented on the keys 2 through 9 - as they had been since the days of rotary phones. To text in the original style, you tap the keys to create the letters. To type a "j," for instance, you hit the 5 key once, since the letters on the 5 key read "JKL." To type a "k," hit it twice. To type an "l," three times, and so on for the other keys on the key pad.

Since this typing mechanism can be laborious, predictive text was created to speed things up. Predictive texting, often called T9 for "text on 9 keys," allows the user to hit each key in a word they want only once. The phone then searches its dictionary for words that match those keystrokes, presenting what its data banks think is likely to be the most frequent word made from those keystrokes first. If that word is not the one the user intended, they can scroll through the list of possible words to pick the one they want. If the word they want is not on the list (unusual proper names, for example, would never be offered by T9) the user can spell out the word in the original multi-tap texting style. Today many phones have predictive texting that keeps track of what words users text and then both learns new words and re-orders the frequency of the words offered (for example offering "home" before "good" if a user types the word home more often) further

[4] These forms of communication which I will refer to under the heading of Electronically Mediated Communication are sometimes categorized as Computer Mediated Communication, a category that is often expanded to include text messaging. Since there terms are not used consistently within the literature as to which forms of communication they encapsulate, and since this thesis focuses on text messaging, I will call all of these forms EMC.

speeding up the texting process. Many modern phones also have another predictive feature wherein when a word is begun the phone tries to guess which word the user means and offers them options for the word before the user is finished typing it. The third advance in SMS technology is QWERTY key pads on phones so that users can text using a miniature version of the same keyboard as on a computer. This is now a common feature on mobile phones in America. The particular effort involved in typing out texts is considered one of the important constraints shaping text language. Another aspect of the device's hardware that researchers think might affect texting practices is the small screen size of the mobile phone.

Another constraint of texting is length. Text messages are generally limited to 160 characters, although the number is sometimes 148. Hård af Segerstad (2002) says that this length limit, combined with the small screen size of the mobile phone, creates a form that both permits and forces people to be concise. Other people put even more weight on this space consideration. Rau (2005) says "The basic principle of texting across languages is space saving." At the same time, while the character limit is important, most texts do not wander anywhere near that limit. Ling and Baron (2007) list the average number of characters per text in their study at 35. Therefore, the shortened speech patterns we will talk about later have more complex driving factors than length limit alone.

Another factor that must be considered in looking at the limitations that shape texting is cost, which manifests in the form of texting plans. Although originally free, soon after SMS was created phone services began to charge a small amount for each message. As the use of SMS blossomed, phone companies started to charge more and also began to offer texting packages. Today, someone might have a phone plan where they have

a texting package which lets them send, for example 1000 texts a month for a set charge, after which point a per-message charge would apply. Someone else might have an unlimited texting plan where they pay a certain amount and can send as many messages as they want per month.

Although the mechanics of SMS are similar everywhere, its popularity and the roles it occupies are different, shaped by culture, regional language, and expense. In the United States, for example, texting tends to consist of informal two person exchanges between close friends and family. In Nigeria, voice calls are prohibitively expensive and texting takes on a variety of formal and informal roles, including formal greetings and offers for job interviews (Chiluwa 2008).

2.2 SMS versus IM

SMS is a form of Electronically Mediated Communication (EMC), as are e-mail, blogs, and instant messaging (IM). While all forms of EMC share some characteristics, they also differ from each other linguistically in measurable ways (Hård af Segerstad 2002). Perhaps text messaging's most similar (and better studied) cousin is the instant message. The main difference between the two is that IM is synchronous, where text messages are asynchronous (Baron 2007). In contrast with another form of EMC, though, e-mail is also considered asynchronous, but unlike e-mails texts convey short packets of information and can turn into rapid fire conversations just like IMs. Therefore, it might make sense to call text messages semi-asynchronous. Texting and IM are also both generally informal and are one-to-one forms of communication. Many of the same characteristics that differ from standard orthography have also been found in both IM and text messages. Although IM and texting share many features, they have some key

differences that make it worthwhile to consider texting as a discrete form. One difference is that there is no added cost for sending multiple IM messages to convey one sentiment. To say:

(1) Jane23: **Hey**

Jane 23: **Do you want to**

Jane 23: **Go to dinner?**

as three instant messages provides no material benefit over sending it as one transmission. With text messaging, there are many reasons to send the above as a single message including the cost, the screen size and the keystroke effort, which brings us back to the importance of spatial limitations, as discussed above.

2.3 What is not distinct about SMS

Most of us have an intuition that there is something particular about text language. The letter/number homophones leap to mind for many people, but these exist in many other types of EMC as well. At this point it is important to mention that none of the ways in which standard orthography is manipulated in text language are entirely new. Most characteristics found in SMS exist in other forms of EMC, and most of the techniques (if not the eventual forms) stem from language manipulations used before computers existed. In terms of shortening devices, for example, initialization, clipping[5], and abbreviation[6] are all familiar forms of informal language (Rau 2005.) The word "luv" on a Valentines Day card or the word "goin'" in a handwritten note would have been as familiar forty years ago as today. The degree and type of these devices in text messages is distinctive, but the devices themselves are not.

---

[5] Deletion of the end letter in a word (e.g. *singin* for "singing")

[6] Shortening of a word (e.g. *appt* for "apartment")

2.4 What makes SMS distinct: An international overview

Before we explore new questions about American text messages, it is helpful to know what previous studies of American text messaging have found. Yet, there is surprisingly little linguistic work to be found specifically on American texting. A limitation to bear in mind regarding the linguistic data on American text messaging is that text messaging was introduced relatively recently in the US: 2002, nine years after its introduction in Europe. This likely accounts for the dearth of linguistic literature on American text messaging. Therefore, we should use an international perspective to get a handle on the fundamental characteristics of texting. A growing interest in texting is apparent all around the world and in past years studies on the linguistic features of SMS have been undertaken in Norway, Italy, South Africa, Germany, Finland, and Britain, among others (Thurlow & Poff 2010). Although some of the studies in other countries began a few years earlier than the studies on American texting, all such studies are still relatively new and represent the beginnings of work into the area of text message Linguistics.

Just how distinct text messages are has been a matter of debate. In a study of French SMS, Yvon Franscois says that texts are "characterized by massive and systematic deviations from the orthographic norm" (2009). On the other hand, Borochovsky et. al. (2010) calls Hebrew texting "a mix of written and spoken language [. . .] indeed similar to the spoken and written language." One factor shaping disagreement about the distinctiveness of texting is that different languages adapt themselves differently to SMS. In a study of French, Spanish, and English text messaging, Rau (2005) found evidence of initialization, clipping, and abbreviation in all three languages, but found that the degree

and types of each device differ in each language, depending on what is best suited to the language. For example, he found that the most frequent shortening device in the English data corpus was initialization, followed by letter/number homophones, then "different" spelling[7], then abbreviation. Spanish, on the other hand, favors abbreviation over initialization, ostensibly since Spanish has longer words and a complex inflectional system. It should be noted that the data is suspect since it was taken from texting "guide books," which are presumably guides as to how to text. Nonetheless, the idea of SMS adapting to the particularities of each language holds across studies.

Although they differ as to the degree, most scholars agree that SMS is distinct. While most texting is legible to anyone, whether or not they have seen a text message before (Thurlow & Poff 2010) Thurlow and Poff characterize our intuition about the uniqueness of text messaging by means of the following observation: a person sees a certain text message and can, even out of context, tell that it is a text and not an e-mail, an IM, or any other type of speech. Thurlow (2003) characterizes the distinctiveness of British SMS in terms of the following characteristics:

- shortenings
- acronyms and initializations
- letter/number homophones
- unintentional misspellings and typos
- non-conventional spellings
- accent stylizations

[7] Similar to "accent stylization" these are ways of spelling words that differ from standard orthography and mimic speech (eg. *dunno* for "don't know")

With Grice's (1989) famous maxims of conversation in mind, Thurlow and Poff (2010) take a cross linguistic look at the research on SMS and compile a list of texting maxims. Their idea of the maxims to which SMS attempts to conform are:

- •brevity and speed
- •paralinguistic restitution
- •phonological approximation

Brevity and speed is an interesting first consideration. Scholars and lay people alike often comment on the telegraphic nature of texts. All experiments that explore length show a shorter message length than IM (Ling & Baron 2007, Hård af Segerstad 2002). Again, the average texting length is nowhere near the allowable limit, yet we still see deletions and shortenings of many kinds, many of which occur in IM but occur more frequently in texting. One idea is that this has to do with the effort and time it takes to type the messages. Thurlow and Poff suggest that it might have to with the ease of turn-taking. Either way, a variety of shortening devices are reliably demonstrated in text messaging samples and the maxim of brevity is soundly demonstrated across studies. That said, brevity and speed does not draw a clear distinction between the possibility that users want to send brief texts or want to be able to send them quickly, and shortening devices could be evidence for either. It is unclear whether the brevity and speed condition is a manifestation of the user's choice to send short messages or whether they are operating under an attempt to get their meaning across with the least effort.

Paralinguistic restitution consists of ways of writing that compensate for the lost prosodic and visual cues found in face-to-face interaction. An example of paralinguistic restitution would be capitalizing something for emphasis, instead of saying it

more loudly. Another example might be putting a smiley face emoticon (*:)*) after a message that would otherwise read neutrally where the emotional affect would be conveyed through prosodic cues and facial expression during interpersonal interaction.

By phonological approximation Thurlow and Poff mean writing a word so that it looks the way it would sound if it were spoken, like *wanna* for *want to*. That type of accent stylization has long been found in informal writing. Another common example is G-clipping *(workin* as opposed to *working*), which is frequently found in SMS but would also seem familiar in a note posted on a friend's whiteboard (Thurlow 2003). Despite the fact that phonological approximation happens in other forms of speech, Thurlow and Poff are right to propose phonological approximation as a maxim of text messaging because they use the category to include many types of this behavior that do not appear in other types of speech. This includes letter/number homophones and forms of accent stylization that would be surprising outside of EMC communication. Without understanding that in text messages words can be written in a variety of ways that convey their meaning by using sound properties, much of texting behavior would be opaque.

Building from the maxims they present, Thurlow and Poff also offer a list of distinctive features found around the world in SMS. We can think of the maxims as the guiding (but violable) principles and the distinctive features as the outputs those guiding rules generate. The features they list are:

- comparatively short length
- the relative concentration of non-standard typographic markers and
- predominantly small-talk content and solitary orientation

It should be stressed, though, that these are generalizations from the literature, not tested predictions.

Looking at a corpus of data from French SMS, Francois (2009) cites as the distinctive features of French SMS: phonetic writing, vowel deletion, acronyms, non-conventional use of letters or numbers, deviations in punctuation, misspellings, accent droppings, non-conventional abbreviations, non-conventional grammar (eg. absence of case distinction) and unintentional typos. Francois looks at the linguistic characteristics of text messaging from a different angle than most of the linguistic research, which merely looks for characteristics within a corpus of data. In her research, she tries to create a program to normalize text language. She tries using a) a technique that treats the non-standard orthography of texts like spelling errors, where the program tries to replace the deviant form with the correct one like a spell-checker, b) a technique that treats the orthography of SMS as a foreign language that requires translation, and c) a technique that treats SMS as "an alphabetic/syllabic approximation of a phonetic form," and utilizes speech recognition tools. She first tries each tactic separately but ends up combining them to the greatest effect. Although by using all three techniques simultaneously she achieved an impressive level of competency with the software, it continued to generate at least one error per text 70% of the time. Although much of her findings have more to do with computer programming than linguistic analysis, her work is helpful in that in order to work on the computer program she had to compile a list of distinctive features that were subsequently experimentally tested by the programs. In thinking about how text language is distinctive, we might also note that, when used alone, the model that treated text messages as a foreign language did the best job. She also notes that one issue in creating translation systems is the use of emoticons,

without which "the meaning of some messages cannot be disambiguated" (Yvon 2009), providing further evidence for the importance of paralinguistic restitution.

Another linguistic study that probes the distinctive features of text is Borochovsky et. al. (2010), which looked at Hebrew texts and compared them to spoken and written Hebrew. They collected 3,000 texts that took place in intimate two-person conversations for a total of 18,000 words. They then compiled a list of factors that might make texting similar to speech (i.e. informal nature and relatively fast turn-taking) and factors that might make it more similar to writing (i.e. the ability of the sender to edit) and compared their corpus to a corpus of written and of spoken Hebrew. They factors they looked at were:

- aspects of informal language
- interactivity
- abbreviations and efficiency
- manifestations of the high register

For all of those factors they analyze many sub-factors. One example they list under aspects of informal language is *lexical characteristics of informal registers* which has as a sub-section "smaller lexicon and use of hyperonyms." They hypothesize that since people have been found to use a larger vocabulary when writing than when speaking (because of the time pressure in face-to-face conversation) since texting has an intermediate amount of time pressure it should demonstrate an intermediate level of vocabulary variation. This is borne out in their data. Indeed, on most of the factors they studied they found that texts fall in between written and spoken Hebrew in frequency. Two of aberrations in their data include the occurrence of first and second person pronouns, which indicate informality but actually

occur more in texting than in spoken Hebrew, instead of somewhere on a continuum between the two. Also, the unpredicted frequency of some high-register items[8] (e.g. bound possessive pronouns) was surprising. They account for this frequency by explaining that all of the over-attested high-register items are shorter than their alternative forms, thus obeying the brevity maxim. They conclude that texting shares both features of written and of spoken Hebrew and that it should be considered a version of neither, but rather a new and distinct form of EMC.

Another study that discusses the distinctive features of text messaging is Ylva Hård af Segerstad's (2002) study which looks at how different types of Swedish EMC take on aspects of speech and writing. She captures SMS data using a combination of methods, including a web questionnaire, requesting texts from family and friends, and having research participants forward texts they sent to her, for a total corpus of 1,152 messages containing 17,024 words. She finds an average message length of 14.77 words and 64 characters. She looks at word frequency as well as punctuation, grammar, spelling and non-graphical items (i.e. symbol-word replacement). The sub-features she notes here (e.g. non-conventional punctuation, accent stylizations) are similar to the features noted in other analyses of text messages, so I won't repeat all of them here. Some features specific to Swedish include omission of subject pronouns, omission of VPs, and omission of articles, prepositions, and possessive pronouns. Although Hård af Segerstad continually returns to the idea that the different forms of EMC are shaped by their limitations (i.e. the screen size, space, limited time to edit), one observation she brings up when talking about the accent stylization - which she calls "spoken-like spelling" – is that sometimes it would

---

[8] Markers of formality or words expected to occur in more formal speech.

take fewer key strokes to type the stylized version, some accent stylizations take just as many or more key strokes than the conventionally spelled word. This insight points to a creative aspect of SMS that moves beyond the proposed limitations of the phone and into features required by a shared language of texters.

One aspect of Hård af Segerstad's study that provides useful groundwork for thinking about the syntax and structure of American text messages is that in looking at deletions she delves into an analysis of which items are deleted and talks about potential reasons. Most studies simply note that deletions occur without finishing the analysis. Hård af Segerstad's work shows that patterns of deletion are nuanced, regular, and specific. Although it is not essential to discuss at length the specific lexical deletions she analyzes since they are relevant to Swedish, the form of her analysis is useful. One example of her grammatical analysis is that of the first person singular subject pronoun *Jag* (translation: "I"). *Jag* was either the sixth most deleted token overall in SMS, or the fourth, not including punctuations. This pronoun deletion corresponds with usage conventions in American text messaging; "[I] Can't tonight. [I] Have to sleep," would be comprehensible in an American text. Such deletions occurred frequently in her corpus. Hård af Segerstad points out that one factor necessary for transmissions to be truncated in this way is that the text conversation is a deictic exchange with a great amount of assumed shared information. These contexts of shared information and closeness might be another of the shaping factors of the syntax of text. She also points out that the above exchange would be acceptable in speech, as it would in spoken English.

2.5 Structural Studies on American Text Messaging

Rich Ling and Naomi Baron conducted what they claim to be the first linguistic study of American text messaging in 2005 (Ling and Baron 2007). If we consider texting as a language, as some evidence indicates it to be, then this work was perhaps the first study of a new dialect. Baron uses the term "mapping out the territory," which is apt as their study is focused on creating a taxonomy of characteristics that occur in text messages, rather than looking for a cause of the characteristics or for underlying patterns. Much like what has been accomplished in the studies on texting in other languages, their pilot study successfully makes some generalizations about the linguistic characteristics of American SMS. Since Americans are new to SMS, but have been using the similar form of IM for many years, Ling and Baron utilize IM as a jumping off point to get a handle on what makes SMS transmissions linguistically unique.

They gather data from 23 female students from the University of Michigan who transcribed all of the texts they sent in a twenty-four hour period. Their corpus consists of 191 messages and 1,473 words. They compare this sample with a corpus of data from a study done on IM two years earlier by drawing 191 random samples from the larger data set of the previous study to match the current study. They look at:

- length
- punctuation
- shortenings
- emoticon usage

They find that texts are unsurprisingly shorter by all measures, including words per message, characters per message, and amount of messages that contain more than one sentence. IM demonstrates fewer acronyms, fewer emoticons, fewer contractions, more

punctuation, no abbreviations and many more apostrophes. Additionally, when talking about the characteristics found in text messages, they mention that they noticed, although they did not test experimentally, some unique forms of expression like the triple exclamation point, which Thurlow would call paralinguistic restitution.

Their most robust finding is the disappearance of apostrophes in text messaging. This makes sense in light of the extra keystrokes necessary on traditional phones to create an apostrophe. QWERTY phones, although they were first released in 1996, have been increasing in prevalence over time so while most of the phones in their study likely had traditional keypads, a modern study of text messaging might demonstrate different characteristics. The story of the apostrophe adds evidence to the brevity maxim discussed earlier. In thinking about the situation of the apostrophe, we can understand the comparatively fast speed of its disappearance by the fact that a missing apostrophe in unlikely to affect meaning. This is in contrast to, for example, initializations, where a more complex process of exposure and learning has to happen for them to be comprehensible. Indeed, there is an idea in popular culture that initializations are phased out as new generations take up mediums of EMC.

The limitations of this first experimental American text messaging study are significant. With a sample of only 23 people, it is,as they say, a pilot study. Any contrasts between IM and texting they found could be caused by one or two outliers. For example, the 3% difference in emoticon use between SMS and IM in their study actually represents a difference of only three emoticons. If the generalizations from this study have weight, we will see them replicated over a much larger population size in the future - optimally with males and females of different ages. Another limitation of their study from

a linguistic standpoint is the aforementioned superficiality of their study. They note characteristics, but do not address questions of the motivations behind the structure of text messaging.

Most work on SMS, linguistically focused or not, mentions some phonological anomalies. The letter/number homophones are probably the most immediately distinct feature of text language, though many studies say they are not as frequent as originally imagined (Ling & Baron 2007). While researchers often remark on these anomalies, only one study so far has focused exclusively on an examination of the phonology of American text messaging.

In 2007, Malgorzata Kuil did a study that looked mainly at the deletion of consonants, vowels and words. She predicts the deletions will follow the "figure-ground principle," which states that "figure" attributes will be retained longer than "ground" features - consonants being figure features, vowels being ground features - and the "rich-gets-richer" principle, which elaborates on the "figure-ground" principle, stating that figure and ground elements can be especially strong or weak and the weakest elements are deleted first. A word-initial or syllable-initial consonant would be stronger than a syllable-final one, and unstressed vowels are weaker than stressed vowels. Her prediction about the "figure-ground" principle was confirmed as consonants were generally retained and vowels were more often deleted in her data. On the other hand, she found that the "rich-gets-richer" principle did not hold for text messaging as, in fact, stressed vowels were more often deleted. In terms of entire words, she found that function words were more likely to be deleted and more likely to lose vowels than lexical items were. She also found many more vowel deletions in monosyllabic words than in polysyllabic words. The findings, though

noteworthy for being an exploration of phonological phenomena, are hampered to a significant degree by their data source; Kuil pulled most of the data from an online site that presents humorous examples of text language. Such examples are humorous precisely because they deviate so far from standard writing. There is no way to know if those messages are representative of the way text language is used by the population, or if they are even real examples. Other data has shown that text messages are generally not nearly as cryptic as it was originally thought (Thurlow 2003), making this a poor data source.

3. Predictions

The literature on text messaging provides a variety of suggestions for what makes texting distinct, but relatively little work has been done to understand the motivations behind the deviations from standard English that are apparent in text messages, or their complexities. This study aims to examine the differences between the understanding and use of text language between college-age texters and older texters and to provide insight into the complexities of text language. Most of the previous studies on SMS have been done with only young participants and I predicted that significant differences would be present in the ways different age groups text. I predicted that the types of non-standard orthography generally found in the literature, for example those found by Thurlow (2003) - acronyms and initializations, letter/number homophones, unintentional misspellings and typos, non-conventional spellings, accent stylizations – would all be utilized by college-age texters, but that they would be used in more restricted ways than are talked about in the literature. I predicted that if older texters made use of these devices they would do so less consistently. I will talk more about my specific predictions for the patterns in text orthography in the methods section. I also predicted that the younger texters, when presented with

hypothetical text language, would provide consistent, patterned intuitions, while the older texters would not. In combination with a more consistent use of non-standard orthography, their ability to provide these intuitions would be evidence for a type of "texting fluency."

I further predicted that Thurlow and Poff's (2010) three maxims of text messaging (brevity and speed, paralinguistic restitution, phonological approximation) would be supported by further experimentation, but be insufficient to account for all of the patterns apparent in text messages. I predicted that paralinguistic restitution, a feature mentioned but unexplored but many researchers (Thurlow 2003, Francois 2010, Hård af Segerstad 2002) would be used in a more nuanced way by college-age texters, and that word deletion, much like non-standard orthography, would be rule-governed within college-age texters and be more variable for older texters. I also predicted that consistent rules would also govern syntax within text messages for fluent texters. One specific prediction I made about the syntax was that optional information structures like topicalization and wh-cleft would disappear in text messaging.

4. Methods

I received approval from the Haverford College internal review board on October 28th 2010 to conduct a study on the grammar of text messaging consisting of text message collection and a survey requiring linguistic judgments.

4.1 Participants

I collected data from a total of fourteen participants. They were recruited by word of mouth and none of them were compensated. Two individuals participated in only the text submission and two participated in only the survey. The other ten participated in both. Participants ranged in age from twenty to fifty-six. Nine

participants had phones with QWERTY keyboards. Eight were women and six were men. Two had pay-per-text phone plans and the rest had either unlimited texts or a high number (i.e. more than they were likely to exceed) of prepaid texts per month. Seven participants were students at Haverford College, two were students at other colleges. The college students were all between age twenty and twenty-four. Two other participants were in their late twenties, and the final three were in their fifties. All who were not in college were college graduates. Participants all self-identified as users of text messaging and varied in their degree of use. Most of the participants under thirty self-reported that they sent at least two texts per day on average, while two people from the fifties age group reported sending less than two texts a day on average. The self-report measures, though, were often inaccurate and some of the older participants texted more than some of the college-aged participants.

4.2Materials

4.2.1 Data Corpus Collection: The twelve participants in the text message collection portion of this study agreed to save all of the texts they sent during one week on their mobile phones. They recorded the texts in a Microsoft Word document and sent them to me from anonymous e-mail addresses (which I created using *mail.yahoo.com)* and gave to them with the initial consent form. When they recorded the texts they were instructed to replace any words they did not feel like sharing with Xs, any texts they did not wish to share with XXX, and to replace all names with random capital letters (eg. *Mary* could become *G*). Participants also completed initial questionnaires that asked them about factors that might affect texting behavior, like phone type and texting plan. See appendix A.

I analyzed the text messages by rewriting each one in standard English and checking that translation against the text to look for lexical deletions, accent stylization/ respelling, initializations, letter deletions and letter/number homophones. An example of this correction process would be that a text reading:

(2) **a. In the library 2day**

Would be rewritten as

b. "I am//She is/They are/We are (etc.) in the library today."

and I marked it for deletion of the subject pronoun, deletion of the copula, non-standard spelling of "today" the use of the number *2* to replace "to," and an absence of final punctuation. I did not mark what I considered to be accidental spelling errors although some did occur.

I also wanted to get a sense of people's actual use of paralinguistic restitution in texts since it is mentioned but not elaborated upon in much of the literature so I marked anything I considered to be use of characters to convey lost visual or prosodic cues from speech, such as extra punctuation or emoticons.

4.2.2 Survey: I constructed a 65 item survey to test participants' judgments of theoretical text message language. See Appendix B. Part I of the survey consisted of one unit tokens which participants attempted to decode and then provided judgments for whether they had seen the token in a text before and whether they would be surprised to see the token in the future. I took a response of "yes" for the question of whether a token was surprising to mean it would be unacceptable text language. Part II consisted of strings that were meant to represent hypothetical entire text messages. Participants were asked to provide judgments of "weird" or "acceptable." I took a response

of "weird" to mean the syntax was unacceptable, since the syntax was the aspect of the messages that deviated from standard English. In Part III, they were given two pieces of a hypothetical conversation in which one speaker made a statement and the second responded in one word or item. They were then asked how the second speaker was feeling. A more detailed description of each part of the survey follows below.

Part I consisted of 40 tokens. Four of the items were nonsense strings:

--0%0—

ttcz

NTTS

krg

Two of the tokens were acronyms:

CWOT [Complete Waste of Time]

NAGI [Not a Good Idea]

presented as common in Rau (2005), which I predicted would be opaque to the participants because they were totally unfamiliar to me and acronyms require learning. Four of the items were examples of various texting behaviors (abbreviation, letter deletion, emoticon use, abbreviation again, and letter/number homophones) taken from this study's data corpus:

btw [Bye the Way]

sry [Sorry]

:[

probs [Probably]

ez [Easy]

While one might predict some interference from the fact that speakers five speakers were judging their own outputs, these items represent a small portion of the total survey, no more than one was taken from each subject, and they were distributed approximately equally across the groups. Therefore, the potential interference is minimal, and having examples of attested non-standard orthography is worthwhile in order to have items that likely to be reliably understood. If such items are not reliable understood, that finding would also be telling, since it would mean that texters are using spellings and emoticons that other participants find opaque. Three of the items presented represent use of non-prototypical emoticons:

:/

. . . :(!

:[

which I hypothesized the college texters would all accept and provide nuanced and consistent responses to, unlike the older generation, who might accept them but would translate them with less nuanced meanings. Three items were examples of what I hypothesized to be acceptable accent stylization because they represent some of the limited number of ways that English speakers are conscious that their speech does not match their orthography (e.g. g-clipping):

Gonna	[Going to]
goin	[Going]
Wut	[What]

For all of these items, the type of "accent stylization," if not the precise term, is found in pre-EMC writing. They are meant to test the hypothesis that accent stylizations are

restricted to a limited number of situations where speakers are conscious that the orthography of their language does not match their speech, which I will further elaborate on later on. Another five items were meant to be counter-examples to further support this hypothesis. The next five terms do mimic spoken pronunciation, but I predicted would not to considered acceptable text language.

chrein	[Train]
clouz	[Close]
Hawarya	[How are you]
liddul	[Little]
missya	[Miss you]

Four items were tokens that I predicted would represent acceptable letter/number homophone use because the substitutions occur in high frequency words where the homophone use saves keystrokes and the number/letter replaces a morpheme, or rhymes with terms where it replace a morpheme:

Tr8	[Trait]
2day	[Today]
missU	[Miss you]
4get	[Forget]

Eight items represent what I hypothesized would be unacceptable use of letter/number homophones:

U4ic	[Euphoric]
MerC	[Mercy]
Cnic	[Scenic]

tCup [Teacup]

CraZ [Crazy]

NtroP [Entropy]

grO [Grow]

In the above cases, the replacement took place in low frequency words which I predicted would not be considered as candidates for homophone replacement. As well, the participants could not have acquired homophone use of such words through learning since they are low frequency. They also do not rhyme with words where homophone replacement would have been learned (as with *h8* and *tr8)* and in most of the tokens, the homophone use would not save typing time. In most of the above terms, the letters or numbers do not replace morphemes. Two items were terms that I hypothesized would be rejected as unacceptable uses of stressed vowel deletion, a response to Kuil (2007) who found that stressed vowels are deleted most:

trder [trader/traitor]

hr [her/here/hour]

Three were items designed to investigate how symbol use within in or in place of a lexical term understood:

fl@n [Flatten]

$? [A question about money or costliness]

***ap** [Strap]

Part II consisted of 15 "messages." The strings included instances of determiner deletion in the subject and the predicate, determiner deletion in just the predicate, subject and verb deletion, infinitival deletion, direct object deletion (when the

subject might have been implied and when it could not have been), deletion of non-lexical items, ambiguous deletion, deletion of the subject in a sentence with a copula, a topicalized sentence with *why,* and two full (i.e. not contracted) sentences with the verb "be" in them. For the college group I predicted that only the full sentences, the subject and verb deletion, and the direct object deletion where the direct object was implied would be acceptable.

Part III consisted of ten "conversations." For this section, I did not have specific hypotheses except that the younger group would give more detailed and consistent responses. As such, my principal goal was to look for trends between groups.

5. Findings and Results

5.1 Data corpus

The data corpus was made up of texts from 12 participants and totaled 357 messages, 279 of which were longer than one word or term (e.g. *of course*). While the one- ord messages did show some features of non-standard orthography like respelling (*u* for "you*," k* for "okay") and frequent deletion of final punctuation, I exclude them from the rest of my analysis since single terms are difficult to analyze out of context and the data corpus cannot show to which question they respond.

Before entering into a discussion of trends that appear in the data, it should be stressed that any mention of item frequency may be misleading because the corpus consists of a different number of texts from each user, ranging from five for the least frequent texter to seventy-nine for the two most frequent texters. This differs from other studies on text messaging, some of which acquired the same number of texts from each participant (Thurlow 2003) and others of which gathered smaller samples per participant which therefore varied somewhat less in number (Ling & Baron 2007). This analysis, then,

should be thought of as a characterization of characteristics and patterns in text messaging and a loose suggestion of whether these characterizations and patterns are common or uncommon. The other preface required to discuss the data is that my original aim to have groups of high frequency texters (who I assumed would be college-aged) and low fequency texters (who I assumed would be middle-aged) ended up being a less helpful way to organize the data, since some of the older participants texted more than the college texters but did not pattern with them in terms of language use. Since all of the texters are also going to have more similar shared experiences with EMC with people their own ages, I regrouped the participants in the entire study into a "college group" made up of nine individuals who are currently in college, a "middle group" of two individuals in their late twenties, and an "older" group of three individuals in their fifties50's. I will generally use these groupings to talk about the data throughout this thesis. Although the "middle group" consists of one individual in the data collection section and therefore is subsumed under the "young group" for convenience, it has another participant for the survey section and is a meaningful grouping there. Also, no participants made use of the optional X (for one word) or XXX (for whole texts) censoring function.

All participants used non-standard orthography, and every type of non-standard orthography mentioned in the literature appeared in the data. Letter/number homophones, abbreviations, accent stylization and paralinguistic restitution all appeared in the older group as well as the college group, and I will talk more about the amount and types of these trends for groups later. That said, perhaps the most readily apparent generalization that can be made from the data is that particular texting "voices" are by far the strongest predictor of texting behavior. By that, I meant that for any given texting

characteristic, most participants made either frequent use of it or almost none at all. One participant used a final period only once over many texts, where others rarely omitted them. Another used the respelling *U* for "you" twelve times while all of the other participants used it a total of three times. This between-speaker variation likely has to do with both texting preferences and with phone type. Apostrophes, for example require one keystroke on some QWERTY phones and some phones with predictive texting, but many more strokes on other phones.

5.1.2 Apostrophes: One finding from this data corpus that challenges the findings in the literature is that my study does not support Ling and Baron's generalization that the apostrophe is dead. In my data corpus 51 apostrophes were deleted, but 21 were maintained. While that may seem like a trend toward deletion, six participants omitted apostrophes most or all of time but another four used them most or all of the time and two participants used them some of the time.

5.1.3 Length[9]: For the data corpus in general, I found longer messages than Ling and Baron (2007) and shorter messages than Hård af Segerstad (2002). They found an average of 35 characters and 65 characters per message, respectively, whereas I found about 39 average characters per message, including the one-word messages, and just under 50 characters per message not including one-word utterances. Hård af Segerstad's study was done in Swedish, so the findings on length are perhaps less relevant.

In looking more critically at length, I found some illuminating patterns between groups. On average, the Haverford students had the longest messages with

[9] Since I focus in these thesis on characteristics of messages longer than one word, I will also focus on the 267 messages longer than one word in my discussion of length except with factoring in the 73 single word utterances is necessary for comparison between studies.

12.64 average words per message, and an average of 1.7 sentences per message. They also had the longest sentences, with 7.18 words per sentence. The two students from other colleges sent messages with an average of 7.59 words per message and 1.17 sentences per message, with 6.88 words per sentence. In this case, the older group patterned more similarly with the non-Haverford college-age texters, and actually sent texts that were shorter than Haverford college students' but slightly longer than the other college participants. The older participants sent messages with an average of 8.47 words per message, 1.22 sentences per message, and 6.8 words per sentence. The middle group participant had an average of 11.14 words per message and 1.77 average sentences per message, which is more like the Haverford students, but 6.2 average words per sentence, which is similar to the older and non-Haverford college-age group. Again, we see significant variation between speakers:

Figure A

Group	Messages	Total Words	Total sentences	Average words/ message	Average sentences / message	Average words/ sentence
Older Group	13	159	18	12.23	1.38	8.83
Older Group	19	112	22	5.8	1.15	5.09
Older Group	14	104	16	7.4	1.14	6.5
Middle Group	35	390	62	11.14	1.77	6.2
Haverford College	12	103	20	8.58	1.66	5.15
Haverford College	39	806	88	20.6	2.25	9.15
Haverford College	10	102	13	10.2	1.3	7.8
Haverford College	18	178	32	9.88	1.77	5.56
Haverford	15	213	29	14.2	1.93	7.3

College						
Haverford College	16	171	21	10.6	1.3	8.14
Other College	61	561	75	7.48	1.22	7.48
Other College	16	171	21	10.6	1.3	8.14

Even allowing for the variation between texters, the data on length supports the possibility Haverford students' messaging style is heavily influenced by group norms that value texting in "correct" English, since "correct" English would likely lead to fewer shortening devices and longer messages.

5.2.4 Statistics of Non-Standard Orthography (Apendix C): In order to analyze the frequency of particular texting characteristics, instead of analyzing every potential feature separately, which might be fruitless and misleading in a study with few participants and different numbers of texts for each them, I looked at four aspects which I believe exemplify trends within the data. I looked at non-standard spelling, word deletion, apostrophes, and letter/number homophones. I also left the middle group participant out of the statistical analysis of non-standard orthography since, as at many points within this analysis, the middle group patterns somewhat like the older and somewhat like the younger group. Also, as only one person in this section, averages would be misleading.

Letter/number homophones are often considered by researchers to be some of the key distinctive features in English text messaging (Rau 2005, Thurlow 2003), and perhaps commonly thought of as the hallmark of American text messaging, as we saw in the episode of "Bones." Although in this sample letter/number homophones are never used to nearly the degree supposed on "Bones," they are used 12 times per week (with an average of 64.5 total messages sent a week) by the non-Haverford college-age texters. Even the older texters

use them an average of almost three times a week, which, considering they send an average of 15.3 total messages a week, is almost the same frequency per message. Not once does any Haverford participant use a letter/number homophone. Considering that letter/number homophones are perhaps the orthographic device most distinctive and specific to EMC (as opposed to, for example, abbreviations or apostrophe deletion which might happen in any type of casual writing), it follows that Haverford students would avoid them if they were adhering to a value of "correct" English in text messages.

For the other features, the data was less clear cut but telling nonetheless. For the purpose of statistical analysis, I collapsed accent stylization, abbreviation, and initializations into "non-conventional spellings." The non-Haverford college group used non-conventional spellings an average of eight times per week. The older group used it 0.4 times, or only once between all three participants. The Haverford group did use it noticeably less, only 1.3 times (over a 21.3 average total messages per participant), but controlling for the non-Haverford college-age participants sending over three times more messages, the Haverford group used respelling about half as often per message. Therefore, in terms of respelling, perhaps social values lead Haverford students to use non-conventional orthography less, but it does not disappear entirely and manifests frequently enough that they used non-standard spelling significantly more than the older group.

Apostrophes are another case where the Haverford uses text language that is closer to, but not the same as standard English. The non-Haverford college students are more likely to keep apostrophes than to delete them, but only slightly (15 apostrophes present, 12 deleted). The Haverford students are almost eight times more likely to keep the apostrophes (an average of 10.16 kept, 1.3 deleted). The older group was also more likely

to keep them (an average of 1.67 kept, 1 deleted). Averages on apostrophes are, again, somewhat misleading since most participants either never deleted them or almost never used them.

I looked at percentage of messages without deletions because word deletion is the other way of shortening messages and I assumed that the Haverford students would still make use of this device. I characterized the percentage of "messages without deletions" as messages composed of prose that forms a complete English sentence. For this measure, I looked only at word deletion, so a message such as:

(4) **ill prob be in bed in half an hour**

would count as one without deletion, despite the lack of final punctuation, omission of the apostrophe, and abbreviation of "probably", because if all of the words were corrected to standard English it would be a complete sentence.

(5) **[It is] Just tomorrow.**

would count as a message with deletion since it lacks subject and verb.

An average of only 8% of the non-Haverford college-age data and 9.67% of the older groups' messages lacked deletion. That means over 90% of their messages occurred in a reduced form. For Haverford students, 20.05% of messages occur with no deletions. That the Haverford group has twice as many messages that would read as complete English sentences further supports the idea that they attempt to use standard English in text messages. That said, that means that almost 80% of their messages do have deletions. That the vast majority of their messages do show non-standard orthography of some type demonstrates that, while the social conventions are shaping the way they text, most of the same principles that govern texting still remain in effect, if to a lesser extent.

5.2.5 Deletion: Hård af Segerstad (2002) illustrates that in Swedish texting, first and second person pronouns are often deleted. My data confirmed this finding with first and second person pronouns omitted 54 times in 279 messages. Another frequently deleted item was final punctuation, which was omitted in 78 instances and almost always in contexts where a period was called for.

Other than apostrophe omission, deletion of final punctuation and the omission of colons as in

(6) **That'll get us there around 9[:]15**

there was also a general pattern that governed word deletion within text messages. If we were guessing how word deletion might function in text messages, we might suppose that function words would be deleted and lexical items would remain intact, as in headlines or telegrams. For example, the sentence "I am going to bed now" could be condensed as "I going bed now" but instead we find

(7) **[I am] going to bed now.**

Indeed, all of the deletions in the college group were speech-like. There is frequent subject deletion:

(8) **[That] works for me.**

especially when the subject could be a first or second person pronoun:

(9) **[I] dunno havent talked to G yet**

(10) **Hey, [I] just wanted to ask you if you would be alright with me going to G's party.**

Frequent subject and copula deletions:

(11) **[I am] Doing ok.**

(12) **[There is] more shit at home. . .**

CP term deletions:

(13) **In a little, how long [are] you guys gonna be there?**

(14) **[Are] You ready now?**

Deletion of everything but the object:

(16) **Pool?**

And even utterances where only prepositional phrases remain:

(17) **In a little**

As well as deletions of words that would naturally be omitted in casual speech:

(18) **[I will] be there in 5 [minutes]**

On the other hand, instances of direct object deletion, or sentence medial deletions of function words are unattested within the young group.

The older groups' the data is more mixed. Some of their deletions are speech-like:

(19) **Ur not with D. . .[I] wish we xould(sic) treat her to dinner or something!**

(20) **[That is a] Great idea**

But two of the three participants show deletions that would not be acceptable in spoken utterances such as:

(21) **House is [a/our] job today**

(22) **Meet at [the] arrow sign later?**

(23) **We [are] here**

Demonstrating that they do not adhere as strictly to the pattern of speech-like deletion. There is also one exception to the speech-like pattern from the middle group participant in

the data corpus section where the participant says:

(24) **[The] Box [is at] at check in.**

Which would not be acceptable in speech. We might argue this is either an aberration or an example of the middle group patterning with the older group. There is also a potential counter-example in the young group where a participant begins a text with "Dear Q" and ends it with "Love, F." This participant, though, was one of the more frequent texters in the corpus and sent forty other texts within the week that would be acceptable spoken English. The salutations also bracket a long and humorous text so, context considered, I would argue that the "dear" and "love" are attempts at a humorous rhetorical device.

This type of deletion pattern makes sense in that text messages are generally understood as casual (Borohovsky-Bar et. al. 2010) and most often seen as substitutes for conversation, not for letter writing or note leaving. In thinking about efficiency, a speech-like deletion pattern also makes sense since only words which are implied can be deleted (i.e. one can never delete the subject in a text message if utterance would introduce a new subject) and subject pronouns are very likely to be implied in text messages where they are not in, say, headlines. Therefore, the type of information that can be reasonably assumed might make speech-like deletions the more efficient deletion pattern, in keeping with Thurlow and Poff's brevity maxim.

Another trend within the syntax is that it seems to generally free of optional movement constructions, with two exceptions. Most of the syntax is "simple" in the sense that propositions begin with an object, then have a verb and then a subject. Adjectives and adverbs and prepositional phrases occur, but the subject verb object order is consistent. In this case I will consider the most "simple" word order for wh-questions as beginning with

the CP term, since having the wh-word appear in its base generated position at the end of the sentence (e.g. “The man is where?”) is the less frequently attested form in English. Assuming that deleted words are implied (an assumption without which text message syntax is difficult to analyze), the only sentences where the messages do no appear in the form of subject, object, verb are

(25) **Shades of Secretariat <u>I guess</u>**

from the older group and

(26) **yeah, halloween just sneaks up <u>I feel like</u>**

From the college-age group. Without context, it is unclear whether the older group participant is actually using an optional construction or not. The participant could mean “Shades of Secretariat is what I am guessing,” which would show topicalization, but utterance could also be understood as “This place is Shades of Secretariat. I guess.” This would not be a true case of optional movement. The second message is a clear case of optional movement. I would argue that one or two transmissions out of 279 shows a considerable bias against, although not prohibition of optional movement.

5.2 Survey:

<u>5.1.2 Initial Findings:</u> For Part I where the participants judged individual text tokens, I initially predicted that the frequent texter group, who I now believe are better characterized as the college group based on actual texting frequency, would accept 16 items (40% of the total items), as enumerated above. For Part II which explored syntactic judgments, I predicted that they would accept six items (40%). I predicted that the middle group would either appear as indistinguishable from the college group or show characteristics of both groups. I predicted that the older group would be less discriminating

in both and therefore over accept in both sections. The results were that the college group accepted 45% of the items on the Part I and 43% of the items in Part II. The older group accepted significantly more items: 59% for Part I and 80% for Part II. The middle group participant accepted 41% in Part I and 70% in Part II. Although these averages support my predictions, there is significant variability within the groups.

Figure B

Participant	Age	Tokens Accepted out of 40 in Part I	Tokens Accepted out of 15 in Part II
1	56	13/ 32%	8/ 53%
2	55	38/ 95%	14/ 93%
3	52	20/ 50%	14/ 93%
4	29	12/ 30%	13/86%
5	27	21/ 52%	8/53%
6	21	13/ 32%	3/ 20%
7	21	19/ 47%	12/ 80%
8	21	21/ 52%	2/ 13%
9	21	23/ 57%	1/ 6%
10	21	17/ 42%	10/ 66%
11	20	15/ 37%	4/ 27%
12	20	20/ 50%	13/ 86%

One participant only labeled only three items total from both parts as unacceptable, while another participant labeled about half the items in Part I and all but two items in Part II unacceptable. While this variability might be illustrative of variability in attitudes as to what constitutes a proper text message, some of the variability might be due to the ambiguity of the prompts. This concern about the methodology is borne out in post-test comments that convey uncertainty about the meaning of "acceptable" and "surprising." Finding communication issues such as this one is one reason it is important to perform pilot studies before engaging in larger surveys.

The above table, though, still obscures the more subtle patterns within the data regarding types items each participant and group is accepting, which is more telling than the total amounts participants found acceptable. Therefore, to understand the nuances in how the different age groups understand texting, we need to analyze each type of item individually. In discussing the results for the different types of tokens, I will talk both in terms of amounts of participants and in percentages. It should be noted that either of these measures might be misleading; saying two participants from each group accepted something would blur over the fact that two participants comprise a minority of the college group, the entire middle group, and most of the older group. Percentages, on the other hand, represent small differences in numbers of people as large percentages, so they should be read with that understanding.

5.2.2 Universally Accepted or Rejected Items: A token that is universally accepted or rejected must therefore represent texting characteristics that are considered allowable across groups. The universally accepted items in Part I were *btw* and *missya*. In Part II, the only universally accepted item in Part II was "*Game later*?" The

findings from Part I are unsurprising since *btw* has been a common abbreviation for years and ya has been an acceptable respelling of "you." *Missya* might be notable for the fact that not a single participant was bothered by the lack of space between *miss* and *ya*, which might suggest that word boundaries are relaxed in frozen phrases in text messages. Analysis of Hawarya which I will talk about under Respelling/Accent Stylizations potentially lends some credence to this concept. There were no universally rejected items in part I or II.

5.2.3 Respelling/Accent Stylization:

Wut was one of the two least controversial respellings, being accepted by 100% of the college/middle groups and 67% of the older group. *Gunna* was the other least controversial respelling, accepted by 100% of the middle and older groups, with 14% of the college group labeling it as surprising. Since each of these represents a variation of only one participant each, I'd argue they should be called acceptable to all groups. *Goin* was rejected by 14% of the college group (eg. one person), 33% of the older group and 50% of the middle group, meaning one participant per group. Again, I would call this item acceptable and put forward the idea that people might have been wary of the lack of apostrophe sometimes represented in g-clipping. Another possibility for participants' apparent discomfort with *goin* might have to do with the fact that in most places "going" appears, so does "to," therefore if a texter is going to make use of accent stylization, it is more efficient and clearer to collapse both into *gunna*, which we have seen is considered more acceptable.

Chrein was rejected by the entire college group, half of the middle group, and 67 % of the older group. *Cloez* was rejected by 85% of the college group, the entire middle group, and 67% of the older group. *Liddle* had the same results except that it was rejected by the entire college group. The judgment that these terms are unacceptable is

supported by the fact that college participants who called the terms acceptable could not decode them. Again we should remember that one participant in the older group only called two items surprising. *Hawarya* represents a middle ground where 43% of the college group, half of the middle group, and 33% of the older group accepted it. For respelling in general the middle group participant who was the more frequent texter patterned with the college group while the less frequent texter patterned with the older group.

Figure C (Participants who accepted the token)

Group	College:	Middle	Older:
gunna	6/86%	2/100%	3/1005
goin	6/86%	1/50%	2/67%
wut	7/100%	2/100%	3/100%
clouz	1/14%	0/0%	1/33%
chrein	0/05	1/50%	1/33%
hawarya	4/57%	1/50%	2/67%
liddul	1/14%	0/0%	1/33%

Taken as whole, the data on respelling is evidence for an addendum to the previously mentioned behavior of "respelling." Only words where there is a discrepancy between orthographic and verbal presentation of which people are abundantly aware can a word be respelled. In the previous cases we see respelling that uses the g-clipping phenomenon, the mechanism that turns "going to" into "gunna," the dropping of the silent "h" in wh-words, and the fact that English orthography represents many vowel sounds many different ways. People are generally aware of these alternations, although perhaps not

using that terminology. On the other hand, although pronouncing any of the respelled tokens out loud would decode them, it takes conscious thought (and potentially linguistic study) to realize that "train" begins with a voiceless alveolar affricate, that the intervocalic "t" in "little" becomes a voiced tap, and that "close" contains a diphthong and a voiced final consonant.

5.2.4 Nonsense Strings: *CWOT* and *NAGI* were summarily rejected by all but one participant. This participant was a person in the older group who accepted 38 out of 40 total and items and who, despite not being able to decode them, said CWOT and NAGI (as well as the other nonsense strings) would be unsurprising. Again, this likely points to a flaw with the clarity of the prompt rather than to a meaningful result, so CWOT and NAGI should be called unacceptable. A similar result happened with *–0%0--,* except that a participant in the middle group who also could not decode it, accepted it. *NTTS* and *ttcz* were also for the most part rejected, but 29% of the college group said *NTTS* was unsurprising and 14 percent said *ttcz* was unsurprising. 33% of the older group said each was unsurprising, but this time that 33% represents two different participants each rating one of the two nonsense strings as unsurprising. Since people rated the capitalized one as less surprising, perhaps people are assuming that something capitalized must be an acronym they haven't learned.

Figure D (Participants who accepted the token)

Group	College	middle	Older
CWOT	0/0%	0/0%	1/33%
NAGI	0/0%	0/0%	1/33%
ttcz	1/14%	0/0%	1/33%
NTTS	2/29%	0/0%	1/33%

--0%0--	0/0%	1/50%	1/33%

5.2.5 Letter/Number Homophones: My original predictions regarding letter/number homophones and their restriction to low frequency words and words where they replace a morpheme is generally supported by the data, although some of the specifics manifested differently than I predicted. In general, as with respelling, the college group was better able to decode the tokens and less likely to accept them than the older and middle groups. Also, as with respelling/accent stylization, the participant in the middle group who is the more frequent texter patterned with the college group, and the less frequent texter patterned with the older group. The older group was far more likely to accept tokens they did not understand, although some participants in the other two groups did this occasionally as well.

Figure E

Group	Young		middle		Older	
Token	Understood	Accepted	Understood	Accepted	Understood	Accepted
Tr8	4/57%	2/20%	0/0%	1/50%	0/0%	1/33%
U4ic	3/43%	0/0%	1/50%	1/50%	2/67%	1/33%
MerC	6/86%	2/29%	1/50%	1/50%	1/33%	1/33%
2day	7/100%	6/86%	2/100%	2/100%	3/100%	3/100%
H8	7/100%	6/86%	2/100%	1/50%	2/67%	3/100%
Cnic	4/57%	2/29%	1/50%	0/0%	1/33%	1/33%
tCup	4/57%	0/0%	0/0%	0/0%	1/33%	1/33%
CraZ	7/100%	4/57%	1/50%	1/50%	2/67%	2/67%

NtroP	3/43%	0/0%	1/50%	1/50%	1/33%	1/33%
UniT	6/86%	3/43%	1/50%	1/50%	1/33%	1/33%
MissU	7/100%	7/100%	2/100%	2/100%	3/100%	3/100%
4get	7/100%	5/71%	2/100%	2/100%	3/100%	3/100%
grO	5/71%	1/14%	1/50%	0/0%	3/100%	2/67%
ez	7/100%	2/29%	1/50%	1/50%	3/100%	3/100%

Looking at some specific tokens, some patterns emerge. Comparing the two sets of tokens to my original predictions, the data for *Tr8* is clearly contradictory. While 54% of the college group was able to decode it, only 29% accepted it. None of the participants in the other groups could decode it, and only one participant in the middle and in the older group accepted it. Considering that my initial idea was that *Tr8* would be acceptable because rhyming homophones would be acceptable, this token is not a problem for the general theory of how letter/number homophones work, but rather refutes my sub-prediction involving rhymes. *MissU, H8, 2day*, and *4get* were all understood by every college texter, and missed by no more than one texter in either of the other groups (although one texter in either the middle or the older groups, again, represents a far greater portion of that group). They were all accepted by a majority of the college group and accepted entirely by the other two groups except for one middle texter rejecting *H8*.

Those four items were by far the most understood and accepted, but the other tokens were not summarily rejected. *Ez* was understood by all of the college texters and rejected by only 29%, while the entire older group both understood and accepted it. For the rest of the tokens, the college group showed a fairly high rate of comprehension, highest for *CraZ* and *UniT*, lowest for *U4ic* and *NtroP*. The mixed findings might be indicative of the fact that on many phones no keystrokes would be saved with either of those

replacements. For *U4ic* and *NtroP*, not a single college texter accepted them as text language. This finding might have to do with "euphoric" and "entropy" being very low frequency words and/or that texters disprefer two letter/number homophones within one word, which would be an interesting claim for further studies to investigate. The older group showed some acceptance of all of these terms, which further supports the idea that the older groups show a less discriminating use of text language.

That letter number homophones are understood significantly differently by the older and the college group supports the idea of a meaningful difference between the way the two groups use text language.

5.2.6 Emoticon Use: The emoticons were generally labeled as acceptable texting language: 100% across the board for *...:(*, and almost complete acceptance for both *:/* and *:[,* excepting one non-Haverford college group participant labeling *:[* as surprising and one each of the older group and middle group labeling *:/* as surprising.

Figure F

Group:	College	middle	Older
...:(	7/100%	2/100%	3/100%
:/	7/100%	1/50%	2/67%
:[	6/86%	2/100%	3/100%

While most participants accepted all emoticons, the way they decoded them was different between groups. The older group was not able to decode the emoticons most of the time,

and when they did they used two word labels like "sad face." The college group, on the other hand, gave detailed answers as to the meanings of each emoticon. One participant described *:/* as "this is problematic but ironic and/or kind of amusing." The college participants' responses also gave detailed pictures of what each emoticon meant and when it would be used. *...:(!*, for example, was described as a delayed reaction of sadness and one, therefore, of great intensity. One aspect of the findings on emoticons that runs contrary to my original hypothesis is that the responses, while detailed, are not always consistent. *:[* exemplified this effect drawing responses of both "general purpose slightly cute sad face not used to express deep sadness but instead slightly bemused sympathy," as well as "very sad," and "deeply serious sadness." Within the college group, trends for all of the emoticons were apparent but there was always at least one response that labeled the emoticon differently. Two potential options here might be that context is required for proper emoticon use, or that we are seeing a strange phenomenon where college students believe they are conveying a specific meaning, but the receiver of the message simply supplies what every meaning they think is appropriate in context. In the middle group, one participant again patterned like the older group and one patterned like the college group. The conflicting responses to some emoticons notwithstanding, this finding is evidence that college texters make a more nuanced use of paralinguistic cues in texting.

5.2.7 Letter Deletion: *Sry* was accepted and understood completely across the college and middle groups but understood by only 67% of the older group and accepted by only 33%. This could be either an example of an age specific learned term, or it could be demonstrative of some rule accepted by all college and middle texters about how high frequency words can be abbreviated. More data would be needed to discern what that

rule was. *Trder* was understood by more of the college group than the older group which is unsurprising since the data revealed them to be generally better at decoding tokens. *Trder* was accepted by only one or two participants within each group. Since this is an example of stressed vowel deletion, it its minimal acceptance goes against the finding in Kuil (2007), although Kuil did not state that stressed vowels could be deleted in all words, and since *sry* is another example of a stressed vowel deletion this is not conclusive. *Hr* is a curious case since it is an ambiguous vowel deletion. Participants offered "here," "her," and also guessed it might be the abbreviation for "hour." One participant mentioned that fact that it was ambiguous. Despite the ambiguity, about half of all groups accepted it. Therefore, whether stressed vowel deletion is a dominant or generally comprehensible type of vowel deletion remains unclear.

Group	College		Middle		Older	
Token	Accepted	Understood	Accepted	Understood	Accepted	Understood
sry	7/100%	7/100%	2/100%	2/100%	1/33%	3/100%
trder	2/29%	4/43%	1/50%	1/50%	2/67%	3/100%
hr	5/71%	5/71%	1/50%	1/50%	?	?

Figure G

5.2.8 Abbreviations:

Probs was understood and accepted as "probably" by 85% of the college group and 50% of the middle group (although, counterintuitively, it was accepted by one participant in the middle group and understood by the other). The data for the older group presents a challenge in that all of the participants accepted it, but some read it as "problems." This

finding is consistent enough for the college group and mixed enough for the other groups that for *probs* might be an example of an age-specific learned abbreviation.

5.2.9. Symbols: The symbols pattern distinctly differently from letter/number homophones and from respelling/accent stylization. *Fl@n* as "flatten" was understood by 67% of the older group, none of the middle group, and only 46% of the college group. It was accepted by the entire older group, and accepted 0% of the time in the other two groups, demonstrating that the older group is more comfortable integrating symbols into words. We might note here that on most modern phones typing *fl@n* would take more keystrokes or the same amount of keystrokes as typing *flatten*. The older and the middle group also showed high acceptance and understanding of *$?* (66%), although *$?* is another item that people offered different responses for, including "money?" and "how much does it cost?" The college group accepted and understood it 54% of the time, but this surprisingly low number might have to do with their discomfort with making a definite guesses as to meaning when meaning is context dependent. ***ap* is another initially surprising case. It was understood as "strap" by no one, but accepted by 3 people in the college group.

Figure H

Group	College		Middle		Older	
Token	Accepted	Understood	Accepted	Understood	Accepted	Understood
fl@n	0/0%	3/43%	0/0%	0/0%	3/100%	2/67%
$?	5/71%	4/57%	2/100%	2/100%	2/67%	2/67%
*ap	3/43%	0/0%	0/0%	0/0%	0/0%	0/0%

This medium level of acceptance with a complete lack of understanding is an aberration for the college-age group. One possible explanation here would be that the meaning supplied by one college participant of **ap* being a correction of some other term to "ap" was a potential meaning being considered by the other two college group participants who accepted it, although they did not offer that meaning in the survey. In the case of symbols, then, the older participants seem to be more accepting and better at decoding than the college group.

5.2.10 Syntax: Part II of the survey generally yielded less clear results. The only item in Part II that was that was universally accepted was *Game later?*. One strong tendency apparent in the data is that in the middle and older groups saw more items as grammatical. For this section the older and middle groups also patterned together; for most items all of a group or all but one member of those two groups agreed with each other. The exceptions are the first and seventh items (preposition and article deletion in the predicate and article deletion in the both predicate and the subject), where the middle participants group both rejected the items while two of the three older group participants accepted them. This shows that the middle groups is more discerning than the older, but less than the college group. Other than article deletion in the predicate, the older and middle groups rejected direct object deletion and infinitival marker deletion and accepted the other items.

The data for the college group is less clear. For them no other item was universally accepted and each item was accepted by at least one participant. That the data for the college group is much more mixed may be an artifact of the larger group size. In general they are pickier in terms of which items they will accept, especially if you exclude

the one texter who labeled all but one token as acceptable. On the other hand, this discrepancy may be due to the ambiguity of the idea of "acceptable." While some participants understood "acceptable" as "not weird" (since weird was its counter item), other people reported calling anything they would understand in a text acceptable, and even citing the reason that "sometimes people are busy and you don't want to judge them for texting quickly." This points to an issue in the wording of the question because the first idea of acceptability, not the later, was my intent.

The college group is so scattered that it makes sense to look at only at the items where all but one or two participants were in agreement, i.e. disagreement of under 30%. Less than 30% disagree in rejecting preposition and article deletion in the predicate, direction object deletion, article deletion in the subject and predicate, infinitival marker deletion, article only deletion in the predicate, and ambiguous deletion (1, 4, 7,8, 10, 11). Less than 30% disagreed in accepting complete subject and verb deletion of the infinitival marker in an interrogative, and one of the complete sentences in normal order (2, 3, 13).

These patterns of acceptance fit fairly well with the earlier conclusions about deletion patterns in text messaging. *Game later?* is clearly speech-like, and the fact that the two simple complete sentences were sometimes judged as "weird" by some of the college group might be due to *is* being spelled out in both, which might be overly formal in speech and therefore strange for a text message. On the other hand, their mixed acceptance of items that would not be acceptable in speech is somewhat confusing considering the complete dearth of non-speech-like constructions in the college-age texters messages. Therefore, while it is tempting to analyze the deeper patterns beneath what the participants

accept and reject, the data is too mixed and the understanding of the prompt too unclear to warrant deeper exploration or strong conclusions.

5.2.11 Part III: The general pattern evident from Part III is that college texters read more nuances of meaning into punctuation in one word responses. Analysis of this section is complicated by the fact that one of the three older group participants dropped out of Part III, so the control group consists of only two individuals. Perhaps tellingly, though, the participant dropped out of it because he or she found it "confusing." Despite the dropout rate, the pattern is suggestive. Although the conversations were vague enough that people gave a variety of responses making it difficult to provide statistics between groups, there are some patterns that emerge, especially regarding four of the hypothetical conversations.

One general trend is that, much like the emoticons, the college group gave more detailed and complex responses and presented their answers more confidently (i.e. "B is clearly frustrated with A" as opposed to "mad?"). In looking at specific conversations, for the first conversation where A asks about B's whereabouts and B responds with "No," the majority of the college group interpreted the response as hostile, irritated, and annoyed while the older group read total neutrality. For the conversation where A inquires about B's whereabouts and B replies with "School. . ." the majority of the college and middle group replied that the second speaker possibly thought the answer was obvious, and many read a possibility of negative emotions towards school, which none of which the older group said. For the conversation where A asks how B is and B says "Fine." One of two participants in the older group thought B actually was fine, where only three of the nine people in the college and middle group even entertained the possibility that B was

fine, and all presented alternate options. For the conversation where A asks if B is ok and B says "<3,"neither of the participants in the older group was clear as to what this would mean. All but two of the other participants called it a positive and loving response. One of the two who did not say so did not respond at all, while the other listed confusion. This might be illustrative of a case of emoticon learning where most young people know *<3* as a heart. Notably, they see as the heart versatile enough that it is understood as an appropriate response to the question "Are you ok?," although it is unlikely that the words "heart" or "love" would be.

6. Conclusion:

Most studies that have sought to characterize text messaging have analyzed only participants within one age group, usually teens or college students (Ling and Baron 2007, Thurlow 2003). Furthermore, the combination of a measure of language use and a measure of language intuition/comprehension is a novel contribution to our understanding of the language of text messaging. With 357 messages and 5514 words, my data corpus is also larger than the seminal study on American text messaging by Ling and Baron (2007), which consisted of 191 messages and 1437 words. Therefore, despite this being a pilot study, from it we can draw some meaningful suggestions about the patterns that underlie text messaging.

Texts are indeed shaped by limitations of space, as Rau (2005) argues and the constraints of typing and reading messages on a mobile phone as Hård af Segerstad (2002) finds. The physical limitations, though, are not enough to account for the complex patterns and group variations found in texts. Text messaging is a distinctive language form shaped by assumptions of shared information and informality and by social expectations regarding

the way text language will be used. Returning to my original predictions, the idea of texting fluency might be overstated since all groups responded to the survey questions consistently, although differently. Instead, we can say that college-age texters have a more strict and pattern-governed conceptions of what constitutes acceptable text language. The restrictions I predicted to would specify types of non-standard orthography are apparent for both accent stylization/respelling and for letter/number homophones. Accent stylization/respelling appears to be limited in scope to instances where people are conscious of particular phenomena where their language's orthography deviates from pronunciation. Letter/number homophones are apparently disprefered in low frequency words and where the number or letter does not replace a morpheme. They are also potentially disprefered when they occur more than once within a word.

In terms of a word deletions, I found that the college age groups' deletions are not governed by word-type, but instead all mirror the way people omit words in spoken conversation. This supports the findings of Hård af Segerstad (2002) and Borochovsky-Bar et al. (2010) that texts are casual and speech-like. As to whether the two groups had different word deletion patterns, that data seems to suggest that both groups follow the speech-like deletion pattern but, as with other texting characteristics, the older group follows the pattern less consistently. It also seems that simple syntax prevails in text messages and that optional information structure constructions occur rarely, manifesting only twice in 267 messages. This may be a manifestation of a distinct preference towards simplicity that governs texting behavior but, with no spoken or written data to compare the text messages to, it is also possible that complex syntactic constructions are simply uncommon to begin with.

My study generally supports previous characterizations of the types of non-standard orthography found in English text messages. All the characteristics listed by Ling and Baron (2007) and by Thurlow (2003), including acronyms and initializations, letter/number homophones, unintentional misspellings and typos, non-conventional spellings, accent stylizations appear in the data corpus and every participant makes use of at least one of those features. On the other hand, each participant also had a distinct texting voice where he or she made significantly greater use of certain devices and little or none of others. An unsurprising but abundantly apparent trend in the data corpus is that phone type has a large impact on those texting voices. People with predictive texting were more likely to type out entire words that those without. Those with QWERTY keypads a those with predictive texting were more likely to use apostrophes than people without. On that note, the data corpus also brought to light the that perhaps apostrophes are not being phased out as Ling and Baron suggest. It is likely this is due to changing phone technology that makes apostrophes less laborious to use. I found that overall the different age groups did not make a discernibly different use of non-standard orthography in the data corpus, but that (lack of) finding may be an effect of the study's size. More research might well find subtler use differences than my study was able to uncover.

My prediction that paralinguistic restitution would be understood in a more nuanced way by college-age texters is supported by the survey judgments on emoticons and judgments on the short scenarios. College-age texters consistently read more complex meanings into emoticons and one word responses than did older texters. One surprising aspect of this findings on emoticons is that, while all the young texters identified

complicated meanings in the non-prototypical emoticons, sometimes different participants offered opposite interpretations.

Another unexpected finding from the data corpus was the distinct difference in how Haverford College students and other college students texted. Haverford students had a complete lack of letter/number homophone use, a lower amount of overall non-standard orthographic markers, and longer average text length. While they texted differently, both groups gave similar survey responses. One possible explanation for this contrast is that, while they are adhering to different values for the way they text each other and therefore also developing different texting practices within their immediate communities, all of the college students are also highly familiar with other forms of EMC. Most or all have probably grown up using instant messaging, depend heavily on e-mail, and many probably make use of other online forums. Although I did not elicit data on how much time participants spent or had spent in the past instant messaging, e-mailing, or on blogs etc., it seems likely the entire college group spends and has spent a large amount of time with other forms of EMC. While text messaging is distinct from other forms of EMC, many features including respelling and use of punctuation and especially paralinguistic restitution are used similarly in many types of electronic communication.

While "fluency" may be an inaccurate claim to make from the data in this study, there is still noticeable gap between the way college students and middle-aged people understand text language. Being less selective in terms of what they consider acceptable text language and more comfortable with the use of words that contain symbols, even when such terms might require extra key-strokes and be opaque to their interlocutors, I propose that older texters see texting as more novel and game-like than

younger texters. The college group, I believe, sees texting as a straightforward means of communication. As such, I believe that Grician maxims provide some insight into their texting behavior.

On the assumption that the expectation of conversation is to communicate cooperatively, Grice (1989) offers four maxims to which conversation adheres. The maxims are quantity, quality, relevance, and manner. Quality means being truthful, and relevance is self-explanatory. Those two should go without saying in text messages. Quantity and manner are more helpful; quantity means communicating no more or less information than is necessary and manner means communicating in the clearest way possible. These maxims take into account the idea that people can engage with their interlocutor's mental state and from these maxims we can, for example, understand how implicatures[10] arise. Thinking somewhat about the maxim of quantity, we might understand this as being adapted to texting with the expectation that a low quantity is the expected and acceptable amount of information to be conveyed per text. This is a fairly straightforward extrapolation from Thurlow and Poff's brevity maxim. Manner is more illuminating; young texters seem to judge as inappropriate tokens that do not maximize meaning per keystroke. They are familiar enough with phones to know which words are likely to take extra keystrokes to type and find such tokens confusing or unacceptable. This is especially visible in the case of symbols where they could not decode *fl@n* - a prime example of a token that would not save typing time and might confuse their interlocutor - for perhaps precisely those reasons. In fact, almost all of the tokens that gleaned approval from college texters would save keystrokes. The manner maxim also

[10] An example of an implicature would be that in saying "some of the students passed the test" a listener would assume that some but not all of the students passed the test, although either could be true from the logical content of the statement.

seems like a solid basis for all of the instances where young texters rejected tokens even when they could decode them; despite understanding the tokens themselves they saw them as an inefficient and therefore a confusing way to convey information. Using terms that take needless keystrokes might be confusing in the same way that saying "some of the class passed the test" would be problematic if the entire class passed. A subconscious use of Grician maxims validates the idea that the college-age participants see texting as a straightforward extension of interpersonal conversation, and so more strictly bound by conversational maxims.

Caveats also exist to the findings of this thesis. The most important is its size. More research needs to be done on text messaging, and done and on a larger scale. Optimally, researchers would solicit the same number of texts per participant for statistical ease. Having no way to compensate participants, I instead sought more messages from fewer subjects. While this style has strengths of its own, including making more clear the degree to which individual texting differences have effects, the differing number of texts per participant is statistically tricky. In a small study, that one of the older participants called most items acceptable regardless of comprehension and that the post-college participants in their twenties did not always place clearly into a group creates significant data skew. Along with being able to present more robust findings on all types of measures, a larger study might also be able to uncover some trends this study could not, like clear non-standard orthography use differences between age groups. Future studies should also explore how pre-college or non-college educated people use text messaging and look at the interaction between other forms of EMC and texting practices.

Bibliography

Baron, Naomi. *Always On: Language in an Online and Mobile World*. Oxford: Oxford University Press, 2008. Print.

"The Maggots in the Meathead." Bones. Tim Southam, Emily Deschanel, David Boreanaz, T.J Thyne. Fox. Oct 14 2010.

Borochovsky-Bar, Ester, Yafit Kedmi. "The nature of SMS discourse: the case of Hebrew." *Folia Linguistica* 44.1, (2010): 1-30. Print.

Chiluwa, Innocent. "Assessing the Nigeranness of SMS test-messages in English." *English Today* 24.1 (2008): 51-56. Print.

Grice, H. Paul. "Logic and Conversation." *Studies in the Way of Words.* Cambridge: Harvard University Press, 1989. Online.

Francois, Yvon. "Rewriting the orthography of SMS messages." "The Natural Language Engineering 16.2 (2010): 133-195. Print.

Kuil, Malgorzata. "Phonology in Text Messages." *Poznan Studies in Contemporary Linguistics,* 43.2 (2007): 43-57. Print.

Ling, Rich & Baron, Naomi, Text Messaging and IM : "Linguistic Comparison of American College Data." *Journal of Language and Social Psychology* 26, (2007) 291-298. Print.

Rau, Paul. "Shortening Devices in Text Messaging: A multilingual approach." *Neuphilologische,* 106.2 (2005): 139-155. Print.

Segerstad, Yleva H. "Use and adaptation of written language to the conditions of Computer Mediated Communication." Ph.D. Dissertation, Gotenborg University. 2002. Online.

Thurlow, Crispin. "Generation Txt? The sociolinguistics of young people's text messaging." *Discourse Analysis Online*, 1.1, (2003) 1-15. Print.

Thurlow, Crispin, Michele Poff. "Language of text Messaging." [to appear (2010/11) in *Handbook of the Pragmatics of CMC*.] Print.

Other Relevant Sources include:

Androutsopoulos, Jassis. "Introduction: Sociolinguistics and computer-mediated communication. Source: Journal of sociolinguistics ." *Journal of Sociolinguistics.* 4.4 (2000): 514-533.

Androutsopoulos, Jassis. "On-Standard Spellings in Media Texts: The Case of German Fanzines." *Journal of Sociolinguistics.* 10.4 (2006): 419-438.

Carrier, Mark, Benitez, Sandra. "The effect of bilingualism on communication efficiency in text messages (SMS)." *Multilingua.* 19.2 (2010): 167 -183.

Crystal, David. *Texting: the Gr8 Db8.* (2008): Oxford: Oxford University Press. Print.

Deumert, Ana, Misinyana, Sibalawe. "Mobile language choices - The use of English and isiXhosa in text messages (SMS): Evidence from a bilingual South African sample." English World-Wide. 29.2 (2008): 117-147.

Spilioti, Tereza. "Graphemic representation of text messaging: alphabet choice and code-switches in Greek SMS." *Pragmatics.* 19.3 (2009): 393-412.

Thompson, Lee, Cupples, Julie. "Seen and not heard? Text messaging and digital society." *Social & Cultural Geography.* 9.1 (2008): 95-108.

Appendix A: Survey 1

*Please **bold** or type in your answer where appropriate.*

Age: ____ Gender: M/F

Are you in College? Y/N Are a college graduate? Y/N

Are you a Haverford Student? Y/N

Does your phone have a QWRTY (full) keypad? Y/N

Do you use predictive texting (T9*)?

*(T9 is when you press each phone key once and select from a set of possible words)

What is your texting plan? Pay per text/ ___ texts per month/ Unlimited texts

Who pays for it? Me/ Not Me

Approximately long have you been texting? ___

I use my phone more for: Texting/Calls/Unsure

Roughly, how many texts do you think you send a day? ____

Appendix B: Survey II:

Part I: Instructions: Y= yes, N=no, ? = unsure or don't know.
Imagine each of these items as part of a text message and give your honest responses. There are no right or wrong answers.

Instructions Part I	What does this mean? Put "?" if unsure.	Have you ever seen this in a text? (Y/N/?)	Is this something you would be surprised to see in a text? (Y/N/?)
1. :/			
2. Tr8			
3. CWOT			
4. --O%O--			
5. Gonna			
6. chrein			
7. NAGI			
8. btw			
Instructions Part I	What does this mean? Put "?" if unsure.	Have you ever seen this in a text? (Y/N/?)	Is this something you would be surprised to see in a text? (Y/N/?)
9. . . . :(!			
10. ttcz			
11. NTTS			
12. goin			
13. U4ic			

14. Iluvu			
15. sry			
Instructions Part I	What does this mean? Put "?" if unsure.	Have you ever seen this in a text? (Y/N/?)	Is this something you would be surprised to see in a text? (Y/N/?)
16. trder			
17. :[			
18. Wut			
19. MerC			
20. 2day			
21. H8			
22. krg			
23. clouz			
24. Cnic			
Instructions Part I	What does this mean? Put "?" if unsure.	Have you ever seen this in a text? (Y/N/?)	Is this something you would be surprised to see in a text? (Y/N/?)
25. tCup			
26. CraZ			
27. Hawarya			
28. liddul			
29. hr			

30. NtroP			
31. uniT			
32. fl@n			
33. missU			
Instructions Part I	What does this mean? Put "?" if unsure.	Have you ever seen this in a text? (Y/N/?)	Is this something you would be surprised to see in a text? (Y/N/?)
34. probs			
35. 4get			
36. grO			
37. missya			
38. $?			
39. ez			
40. *ap			

Part II: In the next part of this survey, please indicate whether the item would be acceptable text language or whether it would seem strange.

Instructions Part II	**Is this acceptable text language or does this seem weird? (A/W)**
1. I want you to go playground.	
2. Game later?	
3. Are you planning go later?	
4. Do you want to make?	

5. I liked.	
6. That is super fun.	
7. Does group think guy is good?	
8. The girl likes actor.	
9. Want go mall?	
Instructions Part II	**Is this acceptable text language or does this seem weird? (A/W/?)**
10. I want sing.	
11. The professor gave grade.	
12. Is a great game.	
13. The class is difficult.	
14. She is planning to write book.	
15. The weather is why he left.	

Part III: Read these situations and write what you would guess to be going on:

Conversation:	**How is speaker B feeling? If necessary make up a possible situation for these texts.**
A: Are you in the science building? B: No.	
A: Will you be cooking later? B:Yup!	

A: Are you hanging out with Tim? B: Yes.	
A: Where are you? B: School. . .	
A: How are you? B: Fine.	
A: You ok? B: <3	
A: It happened. B: What.	
Conversation:	**How is speaker B feeling? Make up a possible situation for these texts.**
A: She did it. B: !!!!!!	
A: Are they finished? B: Nope.	
A: He said it. B: What.	

Any other comments?:

Appendix C: Data Corpus Frequency Tables

i: College Age (Non-Haverford)

	Letter/Number Homophones*[*]	Percent of Messages without deletions	Non-Conventional Spelling*	Apostrophe Deletion*	Apostrophes*
	19	7%	12	20	15
	5	9%	4	4	15
Average	**12**	**8%**	**8**	**12**	**15**

ii: College Age (Haverford)

	Letter/Number Homophones*	Percent of Messages without deletions	Non-Conventional Spelling*	Apostrophe Deletion*	Apostrophes*
	0	26%	4	4	1
	0	21%	1	2	26
	0	7%	0	0	2
	0	4%	2	0	8
	0	4%	1	0	4
	0	13%	0	2	20
Average	**0**	**20.5%**	**1.3**	**1.3**	**10.16**

[*] Per week independent of message count.

iii: Older Group

	Letter/Number Homophones*	PErcent of Messages without deletions	Non-Conventional Spelling*	Apostrophe Deletion*	Apostrophes*
	0	9	1	2	4
	2	15	0	1	0
	6	5	0	0	1
Average	**2.67**	**9.67%**	**0.4**	**1**	**1.67**