Literacies of Bilingual Youth: A Profile of Bilingual Academic, Social, and TXT Literacies

Michelle A. McSweeney

The Graduate Center, City University of New York

Recommended Citation
McSweeney, Michelle A., "Literacies of Bilingual Youth: A Profile of Bilingual Academic, Social, and TXT Literacies" (2016). CUNY Academic Works.
https://academicworks.cuny.edu/gc_etds/1459

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
Literacies of Bilingual Youth: A profile of bilingual academic, social, and txt literacies

by

Michelle A. McSweeney

A dissertation submitted to the Graduate Faculty in Linguistics in partial fulfillment of the requirements for the degree Doctor of Philosophy, The City University of New York

2016
2016
Michelle A. McSweeney
Some rights reserved.
This work is licensed under a Creative Commons Attribution 4.0 United States License.
http://creativecommons.org/licenses/by/4.0
Literacies of Bilingual Youth: A profile of bilingual academic, social, and txt literacy

by

Michelle A. McSweeney

This manuscript has been read and accepted for the Graduate Faculty in Linguistics to satisfy the dissertation requirement for the Doctor of Philosophy

Gita Martohardjono

Date Chair of Examining Committee

Gita Martohardjono

Date Executive Officer

Committee Members:

Steve Brier

Cecelia Cutler
Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK
Abstract

This dissertation reports on a research project to identify the multiple types of language skills that urban emergent bilingual youth possess. The participants in this study are Spanish-dominant bilingual young adults enrolled in a high school completion program in New York City. They are in the process of developing both Spanish and English academic literacy skills, and it is well known that they tend to perform below the grade they are enrolled in. For this reason, they are often referred to as being “language-less” (DeCapua & Marshall, 2011; Freeman, Freeman, & Mercuri, 2002) in an academic setting. Yet, little was previously known about their linguistic skills in other language forms such as social and Txt. This research seeks to understand and document their abilities across language forms and modalities, painting a composite picture of non-traditional bilinguals students’ linguistic skills.

The aims of this dissertation are achieved through three different approaches. The first is a quantitative study into participants’ literacy skills through the use of assessments measuring academic literacy and social language awareness across written, aural, and digital modalities. The second is an in-depth analysis of the features participants use when texting. Txt is a relatively new language form, and the analysis presented in this dissertation identifies the features and patterns that illustrate its systematic and constrained nature. The third approach is a case study focused on the texting behavior between two prolific texters. The theories developed based on the texting patterns of all participants (except those two texters) are applied to this one conversation for validation. This conversation constitutes more than half of the text messages that students contributed to the project, highlighting just how important this language form is in the daily life of young adults.

A final component of this dissertation is the public availability of the text messages as an anonymized corpus along with the code and methods used to analyze the data. The text message
corpus is included here as a .csv file as well as an SQL file. The code is included as a zipfile of my GitHub Repository. These files are also available at www.byts.commons.gc.cuny.edu
Acknowledgements

It is often said that it takes a village to raise a child, I believe it is the same for a dissertation. This project is the result of an incredibly rich and diverse group of people contributing their thoughts, time, encouragement, and support; I am deeply grateful to everyone who has contributed to this project in all the different ways.

First and foremost, thank you Professor Gita Martohardjono. As an advisor, mentor, executive officer, and lab director, I have been incredibly fortunate to work with someone who is brilliant, inspired, and passionate about language and multilingualism, and flexible enough to allow this particular project to be undertaken. I am so grateful to have been a part of such a vibrant, curious, and active department.

Likewise, I am incredibly thankful to Professors Steve Brier and Cece Cutler for their contributions to this dissertation, and their perspectives on both youth culture and digital communication. At a very crucial point, Professor Brier’s passion and excitement for digital technology helped shape my ideas about the modern world. Similarly, Professor Cutler’s input at multiple stages of this project, as well as her perspectives on language and dialect have been enormously helpful in shaping this project.

A special thank you to Sarah Cacicio; this project would not have existed without your help in connecting me with the schools, teachers, and students on whom this dissertation is based. Connecting me to the schools made this project possible, but breaking down language politics throughout New York City got this project started.

Likewise, thank you to Shannon Webb for hours spent helping me form this project and your dedicated work on the LENS. Thank you to Yoshivel Chirinos as well for going with me to the schools to collect the data. Finally, thank you to everyone in the Second Language Acquisition Lab,
you have all taught me so much about language, culture, and how to create positive change from the inside of the system.

Thank you to the students and teachers who donated their time, voices, insights, and text messages. You will always remain anonymous, and I can only hope that I have captured your spirit and voice in the following pages. You are more linguistically adept than you know, and I hope your linguistic creativity and innovation is valued throughout your life.

Thank you to Matt Gold, Lisa Rhody, and the GC Digital Fellows for helping me develop the skills I needed to approach this project, the technical awareness to get creative with it, and the professionalization skills to talk about it.

Deep and heartfelt thank you to Writing Group, including Joanna Birnbaum, Ignacio Montoya, Emily Olsen, Ian Phillips, and Nadav Sabar. I am honored and humbled to be a member of such an inspiring community of scholars. Our weekly meetings have kept me sane and grounded, and made writing fun. I am incredibly grateful to have had the opportunity to share our successes and challenges, and to have grown into academics together. Out of these years came a dissertation, but more importantly, friendships built on trust, mutual support, and a curiosity about language.

A very special thank you to Ian Phillips who has spanned so many of these groups and more. I am deeply grateful to have had the opportunity to work together on so many projects, from the early visits to the Bronx to teaching at Hunter, to learning to program, Writing Group, the SLAL, and the Digital Fellows. I couldn’t have asked for a better friend in this journey.

Finally, and most importantly, thank you to my wonderful wife, Devon McSweeney, who has stood by me through this entire process both when she thought it was a good idea and when she did not. Thank you for being so patient with the crazy hours, the ambiguity, and all the challenges this project has presented, I could not have done this without you at my side. As you’ve said, “we’ve been working on this for a very long time!” 😘.
Table of Contents

Chapter 1 - Introduction ........................................................................................................... 1
  1 Background ......................................................................................................................... 2
  2 Research Statement ............................................................................................................ 8
  3 Terminology ....................................................................................................................... 8
  4 Purpose and Significance ..................................................................................................... 10
    4.1 Txt as a language form .................................................................................................. 11
    4.2 Txt and Education ......................................................................................................... 14
    4.3 The Corpus .................................................................................................................... 15
  5 Previous Work .................................................................................................................... 16
    5.1 Previous Documentation of Txt and Corpora ............................................................. 17
    5.2 Previous Research on Bilingual Computer Mediated Communication .................... 20
    5.3 Previous Research on the relationship between Education and Txt .......................... 23
  6 Research Design .................................................................................................................. 26
  7 Assumptions and Limitations ............................................................................................. 27
  8 Outline .................................................................................................................................. 31

Chapter 2 - Literacy types and modalities ............................................................................. 34
  1 Introduction ......................................................................................................................... 34
  2 Participants .......................................................................................................................... 36
    2.1 Participants - Participation ............................................................................................ 39
    2.2 Participants - Internal Review Board ........................................................................... 40
    2.3 Participants - Connecting with Schools ........................................................................ 42
  3 Measures .............................................................................................................................. 42
    3.1 Measures - Survey .......................................................................................................... 43
    3.2 Measures - Acceptability Judgment Task (AJT) ............................................................ 51
    3.3 Measures - Aural Comprehension Task ....................................................................... 59
    3.4 Measures - Chatting Game ........................................................................................... 62
    3.5 Measures - English Literacy Evaluation for Newcomer SIFE (LENS) ......................... 65
    3.6 Measures - Spanish Literacy Evaluation for Newcomer SIFE (LENS) ......................... 67
    3.7 Measures - Test of Adult Basic Education (TABE) ......................................................... 70
    3.9 Measures - Overall Academic Score ............................................................................ 74
    3.10 Relationship between Measures ................................................................................ 76
    3.11 Measures Conclusion .................................................................................................... 78
  4 Conclusion ............................................................................................................................ 78

Chapter 3 – Bilingual Txt ..................................................................................................... 80
  1 Introduction ......................................................................................................................... 80
  2 Collection and Preparation of the Text Messages ............................................................... 85
    2.1 Text Messages - De-identification ............................................................................... 87
    2.2 Text Messages – Language Tagging .............................................................................. 88
    2.3 Correlation between Academic Skills and Text Message Language Choice ................ 92
    2.4 Semantic analysis of linguistic features of select messages .......................................... 93
  3 Respellings ........................................................................................................................... 98
    3.1 Respellings - Definitions ............................................................................................... 100
    3.2 Common Respellings ................................................................................................... 104
    3.3 Uncommon or Novel Abbreviations ............................................................................. 105
    3.4 Respelling Rules ......................................................................................................... 108
  4 Abbreviations ...................................................................................................................... 109
    4.1 Reductions ..................................................................................................................... 112

viii
List of Figures

Figure 1 – Face-to-Face Language Preferences ............................................................................................48
Figure 2 – Digital Language Preferences ....................................................................................................49
Figure 3 – Frequency of Communication on a Digital Platform ....................................................................50
Figure 4 – Example of Acceptability Judgment Task ................................................................................52
Figure 5 – Paralinguistic Features ..............................................................................................................53
Figure 6 – Percent Match of Judgments with “Standard” .............................................................................57
Figure 7 – Aural Comprehension Grade Level ............................................................................................62
Figure 8 – Chatting Game Grade Level .......................................................................................................65
Figure 9 – English LENS Grade Level .........................................................................................................67
Figure 10 – Spanish LENS Grade Level .......................................................................................................70
Figure 11 = TABE – Esp Grade Level ...........................................................................................................73
Figure 12 – TABE CLAS-E Grade Level ......................................................................................................74
Figure 13 – Overall Grade Level ................................................................................................................75
Figure 14 – Intersections .............................................................................................................................81
Figure 15 – Text Message Language Choice ...............................................................................................91
Figure 16 – Emoji picture ..........................................................................................................................98
Figure 17 – Face with tears of joy ..............................................................................................................118
Figure 18 – Mobile ownership, education, and the k-cluster in the Dominican Republic .........................125
Figure 19 – “Boo” versus “boe” in hip hop lyrics, derived from all the lyrics available on RapGenius.com (Todd, 2014). ..........................................................161
Figure 20 – Respelling order of operations .................................................................................................163
Figure 21 – Iteration frequency by language frequency and sonority ............................................................228
Figure 22 – Sonority scale (Hartmann, 2014) .............................................................................................229
Figure 23 – Most common words near to iterations compared to corpus ......................................................236
Figure 24 – Frequency of most common words in corpus compare to those near to iterations ..................236
Figure 25 – Accents on the iPhone keyboard ..............................................................................................256
Figure 26 – Percent of Michael’s messages containing “lmoo” per week ....................................................273
Figure 27 – Percent of Michael and Jasmine’s messages containing “lol” per week ..................................279
Figure 28 – Percent of interactions by letter preferred by Jasmine, Michael, and all participants ...............293
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>List of all assessments</td>
<td>35</td>
</tr>
<tr>
<td>Table 2</td>
<td>Participants Per Class</td>
<td>38</td>
</tr>
<tr>
<td>Table 3</td>
<td>Number of tasks completed by number of participants</td>
<td>39</td>
</tr>
<tr>
<td>Table 4</td>
<td>Basic Demographic Information</td>
<td>46</td>
</tr>
<tr>
<td>Table 5</td>
<td>Points on Aural Comprehension Task (Craig Saffoe)</td>
<td>61</td>
</tr>
<tr>
<td>Table 6</td>
<td>Table 3.4 - Points on Aural Comprehension Task (Tyrian Purple)</td>
<td>61</td>
</tr>
<tr>
<td>Table 7</td>
<td>Average scores for all assessments</td>
<td>75</td>
</tr>
<tr>
<td>Table 8</td>
<td>Classification Scheme</td>
<td>89</td>
</tr>
<tr>
<td>Table 9</td>
<td>Commonly respelled words</td>
<td>95</td>
</tr>
<tr>
<td>Table 10</td>
<td>Three ways to interact with rules</td>
<td>101</td>
</tr>
<tr>
<td>Table 11</td>
<td>Tectism types</td>
<td>102</td>
</tr>
<tr>
<td>Table 12</td>
<td>Common and novel reductions and clusterings</td>
<td>111</td>
</tr>
<tr>
<td>Table 13</td>
<td>Common colloquial reduction abbreviations</td>
<td>113</td>
</tr>
<tr>
<td>Table 14</td>
<td>Types of vowel removal in reductions</td>
<td>114</td>
</tr>
<tr>
<td>Table 15</td>
<td>All initialisms</td>
<td>146</td>
</tr>
<tr>
<td>Table 16</td>
<td>Key features of African American English</td>
<td>167</td>
</tr>
<tr>
<td>Table 17</td>
<td>Spanish initialisms</td>
<td>171</td>
</tr>
<tr>
<td>Table 18</td>
<td>Initialisms by word function</td>
<td>173</td>
</tr>
<tr>
<td>Table 19</td>
<td>Greetings initialism summary</td>
<td>181</td>
</tr>
<tr>
<td>Table 20</td>
<td>Initialisms to comment summary</td>
<td>197</td>
</tr>
<tr>
<td>Table 21</td>
<td>Initialisms for relationships summary</td>
<td>201</td>
</tr>
<tr>
<td>Table 22</td>
<td>Timeline of “lol” research</td>
<td>207</td>
</tr>
<tr>
<td>Table 23</td>
<td>Special initialisms summary</td>
<td>222</td>
</tr>
<tr>
<td>Table 24</td>
<td>Iterated letter frequency</td>
<td>225</td>
</tr>
<tr>
<td>Table 25</td>
<td>Frequency of letters in standard written Spanish and English (averaged)</td>
<td>226</td>
</tr>
<tr>
<td>Table 26</td>
<td>Word types by category</td>
<td>230</td>
</tr>
<tr>
<td>Table 27</td>
<td>Iteration frequency by word type and type:token ratio</td>
<td>231</td>
</tr>
<tr>
<td>Table 28</td>
<td>Iterations of ‘hola’ and ‘hello’</td>
<td>232</td>
</tr>
<tr>
<td>Table 29</td>
<td>Iterations of the interjection, “mmm”</td>
<td>233</td>
</tr>
<tr>
<td>Table 30</td>
<td>Initialisms that undergo iteration</td>
<td>245</td>
</tr>
<tr>
<td>Table 31</td>
<td>Spanish Spelling Mistakes (Werbeck, 2015)</td>
<td>248</td>
</tr>
<tr>
<td>Table 32</td>
<td>English Spelling Mistakes (Wrenn, 2015)</td>
<td>249</td>
</tr>
<tr>
<td>Table 33</td>
<td>Spanish words only differentiated by accent</td>
<td>254</td>
</tr>
<tr>
<td>Table 34</td>
<td>Summary of Jasmine and Michael’s self reported language skills in Spanish and English</td>
<td>263</td>
</tr>
<tr>
<td>Table 35</td>
<td>Summary of Jasmine and Michael’s performance on language assessments</td>
<td>264</td>
</tr>
<tr>
<td>Table 36</td>
<td>Initialisms used by Jasmine and Michael</td>
<td>270</td>
</tr>
</tbody>
</table>
Chapter 1 - Introduction

The purpose of this dissertation is to develop a comprehensive picture of the literacy skills and language practices of urban bilingual youth. This topic is addressed from three perspectives; the first is a description of participants’ academic, social, and txt literacy skills. The second is an analysis of the features of bilingual Txt. The third is a case study where the analysis regarding the features of bilingual Txt are applied to one conversation. In the first (quantitative) component, the literacy skills students possess across languages, language forms and modalities is explored as well as the relationship between these types, with a critical look at the role that task type and modality play in student performance. In the second (qualitative) component, the language practices of twelve prolific bilingual texters are analyzed; hypotheses about the role of respellings in Txt and the features students use to create meaning through texted language are proposed in this section. In the third (case study) component, the hypotheses developed about the role of Txt-specific features are applied to one conversation between two participants that spans the first four months of their relationship. The result from these three components is a composite picture of the language skills and linguistic innovations students have at their disposal for communicating across registers and platforms, and how text messaging fits into bilingual students’ overall language repertoire.

One additional contribution from this project is the development of a publicly available bilingual text messaging corpus with 44,597 messages (Spanish/English BYTs Bilingual Youth Texts). These messages were downloaded directly from participants’ phones, resulting in a candid record of bilingual youth texting habits. The language in these massages is in the New York City dialect, setting the stage for future work comparing NYC Txt to other Txt dialects and creating a time capsule for future work on Txt as the language form evolves. This corpus is unique for multiple
reasons. First, there are only two text messaging corpora that are of a similar size: SMS4Science, and the NUS corpus. Both of these are multilingual corpora (SMS4Science includes Western European Languages, and NUS is in Chinese and English), though neither focuses on Spanish/English. Secondly, the BYTs corpus is freely available to researchers, and while NUS is also freely available, SMS4Science is not at the time of writing. Third, most corpora are collected from people associated with a college or university (i.e., students, professors, or college-bound adolescents), and while the BYTs corpus is collected from students, very few intend to go on to post-secondary education. Fourth, attached to this corpus is basic demographic information including age, phone type, country of origin and time stamps, which are not present in other corpora. Finally, in conjunction with this corpus is a Part-of-Speech tagger that has been optimized for this type of dataset (i.e., text messages).

1 Background

The timeliness and importance of the present work is underscored by one conversation in the BYTs corpus: this one conversation accounts for more than half of the corpus, spanning the first four months of a romantic relationship. Most of the contact these participants have with each other is via text message and FaceTime (a video chatting service). They see each other once a week, but send and receive hundreds of messages a day. In many ways, their relationship is similar to a long-distance relationship even though they live in the same borough. This is not uncommon for young adults more generally as relationships and communication becomes increasingly digitally mediated (K. D. Anderson, 2015; Boyd, 2014; Fisher & Garcia, 2015; Turkle, 2012). Yet, this means that young adults are communicating in a text-based format for more purposes and with more frequency than ever before. They are creating and interacting with print every day to establish intimacy, build

1 Some of these features are present (i.e., age range or phone type), but not all.
relationships and friendships, and maintain contact with family and friends both within their own households and across international borders. Participants in this study, and young adults more generally, are in near constant communication with significant others, friends, and family (K. D. Anderson, 2015; Boyd, 2014; Harrison & Gilmore, 2012), yet all of this communication is through print.

Linguistic research on the language used on digital platforms is still in its infancy. It has been established by previous researchers that this language form generally, and text messaging in particular, is unique, distinct from both traditional written language as well as spoken language (Bernicot, Volekaert-Legrier, Goumi, & Bert-Erboul, 2012; Crystal, 2009, 2014; Tagg, 2009). This language form goes under a variety of names, including “Netspeak” (Crystal, 2009), “Computer-Mediated Communciation (CMC)” (Janis Androutsopoulos, 2006), “Textese” (Drouin, 2011), “SMS Language,” and “Txt” (Shortis, 2007; Tagg, 2009). This dissertation uses the term, Txt, because it both captures the playfulness and flexibility of this language form as well as the fact that Txt is one form of a language, not a separate language unto itself. The decision to use “Txt” is in contrast to terms such as “Computer Mediated Communication” and “Digitally Mediated Communication” which describe communication on digital devices, including email, message boards, and other asynchronous platforms, and are therefore better suited to describe a research area rather than a language form. Therefore, this dissertation uses the term Txt to describe the language used on semi-synchronous digital interfaces such as text messaging and chatting. ‘Txt’ alone refers to features that are considered common to the language form across languages. When referring to Txt in a specific language such as Spanish or English, the language is specified.

Academic, social, and Txt are referred to as language forms here as a way to acknowledge that the language used in text messages and chat is still Spanish or still English in the same way that
both academic and social language forms are varieties of a language. The term, “form” is being used as a departure from either “register” or “dialect.” Where registers of a language are defined by the social relationship between interlocutors, and dialects are a result of a speech community, Ttxt as a language form exists across registers, dialects and languages. Furthermore, the term “form” is used to indicate that this type of language takes an identifiable and quantifiable shape that is distinct from the features found in academic or social language forms.

Much of the research related to Ttxt has been on links between texting and education rather than linguistically driven. Of course, there is plenty of anxiety about the effect that texting will have on literacy skills (Dwyer, 2012; Jury, 2010), yet the studies that examine the role that the use of text messaging has on comprehension, spelling, and literacy find a wide variety of effects (even within the same study). These studies are either inconclusive (Drouin, 2011; Drouin & Davis, 2009; Plester, Wood, & Bell, 2008), find a small positive effect (Kemp & Bushnell, 2011; Tagliamonte & Denis, 2008; Wood et al., 2011), or find a positive effect and conclude that the students who are the most proficient and creative texters have the highest overall literacy skills (Plester, Wood, & Joshi, 2009). Still others do find a negative effect of texting on overall literacy skills, but just as with the positive effects, the size is very small (Dansieh, 2011; De Jonge & Kemp, 2012; Drouin & Davis, 2009; Rosen, Chang, Erwin, Carrier, & Cheever, 2010). While these findings are inconclusive, they point to an intuition that researchers have about the relationship between different literacy types. Implicit in this line of research is the idea that the experience of texting or exposure to text messaging can have a positive or negative effect. The present study approaches this relationship from another perspective. Rather than treating text messaging as an experience or event, this study treats

---

2 There is a difference in capitalization between academic and social language forms and Ttxt because Ttxt refers to the texting language form, and is being treated as a proper noun whereas academic and social are being used in this context as adjectives to describe the language forms.
literacy as simply another form of literacy that can be measured and assessed in the same way that any literacy form (i.e., academic language) can be. This presents the opportunity to analyze the relationship between language forms without assuming a causative effect. The relationship between these forms is very complex as Txt borrows features from both academic written language as well as social face-to-face language; the result is a language form that is just as systematic and constrained as every other form. By treating Txt as a rule-governed language, it is possible to treat it as a language form in its own right, and identify the knowledge and skills required to communicate successfully in the form.

Given the relative newness of Txt, research on the role that specific-to-Txt features have in the process of constructing meaning is still in its initial stages. The relationship between Txt and other language forms (written and spoken or academic and social) is still poorly understood as well, as indicated by the numerous studies attempting to draw connections between text messaging and other forms yet largely proving inconclusive. Rather than attempt to make connections about the role of texting on a students’ ability to learn, this dissertation seeks to understand students’ linguistic skills from a holistic perspective – asking instead what literacy skills this population has in each language form independently. Traditionally, the focus of assessment is on academic literacy skills, and it is well established that the population in this study perform poorly in assessments of academic language in both their first and second languages. However, nearly nothing is known about their communication skills in other language forms.

Participants in this study are Spanish-dominant bilinguals in an alternative education program who are preparing to take the high school equivalency exam, the TASC (Test Assessing Secondary Completion). Even though they are preparing for high school equivalency, their academic reading and writing skills are at an approximately 5th grade level, indicating that they are academically
under prepared for the grade level they are in. Very few studies have been conducted assessing the literacy skills of non-traditional students in General Education Degree (GED) programs, as most studies have focused on measures of “success” such as completion, college attendance, and job outcomes (Cook et al., 2014; Fry & Center, 2010; Heckman & LaFontaine, 2006; Heckman & Rubinstein, 2001). Studies that do assess GED students’ abilities often focus on non-cognitive skills, assuming that if they pass the completion exams, they have mastered the content (Cook et al., 2014; Heckman & Rubinstein, 2001). Participants in this study are not only non-traditional, GED students, but they are also Spanish-dominant bilinguals. While this may initially seem like a very specific population, there are enough students in this program every year to justify a Spanish-language version of the TASC and Spanish-only courses to prepare. Yet this population is rarely studied, largely, in part, due to the logistical challenges associated with working with GED Students.³ The result is that little is known about the language and literacy skills of this population, and this dissertation presents the first comprehensive investigation into the literacy skills of non-traditional bilingual students.

Furthermore, the development of a large corpus of text messages provides the opportunity to analyze the features that young adults use to construct meaning on a digital platform. Much of the previous research on the role of textisms (spellings, acronyms, and punctuation uses specific to text messaging) has approached the topic from an identity and personal expression standpoint (Boyd, 2014; Burgess, 2006; del-Teso-Craviotto, 2008; Huffaker & Calvert, 2005; Mahootian, 2002; Marwick & Boyd, 2011; Montes-Alcalá, 2007; Paolillo, 2011; Shaw, 2008; Tagg, 2009). Many of these studies focus on the role that code switching plays in the establishment of a bilingual identity, and

³ As will be discussed in the limitations of this chapter and more thoroughly in Chapter 2, attrition rates are remarkably high among this population. Likewise, these students are considered a vulnerable and marginalized population, so great care was taken to ensure that they understood the research and that their privacy is maintained.
how participants navigate communication and identity expression (Burgess, 2006; Mahootian, 2002; Montes-Alcalá, 2007; Paolillo, 2011). Others have focused on identity construction in monolingual communities (del-Teso-Craviotto, 2008; Tagg, 2009), and some with special attention to adolescents (Boyd, 2014; Huffaker & Calvert, 2005; Marwick & Boyd, 2011; Shaw, 2008). While language and word choice are certainly powerful ways for individuals to perform identity, these choices are constrained by the linguistic processes at work; this situation makes the case for further linguistic studies into Txt as a language form rather than Txt as a mode of creative expression.

Studies with a specifically linguistic approach tend to analyze the pragmatic forces in text messaging. Some of these studies seek to understand how languages are mixed (Janis Androutsopoulos, 2013; Dorleijn & Nortier, 2009), how discourse is managed through code switching (Goldbarg, 2009; Hinrichs, 2006; Rowe, 2011), and how bilinguals treat different languages differently (Deumert & Masinyana, 2008). These studies, similar to the present study, seek to understand the mechanisms texters utilize to construct meaning in a completely text based platform. Additional studies have focused on specific features of text messaging such as emoticons (emotion icons) (Dresner & Herring, 2010; Provine, Spencer, & Mandell, 2007a), punctuation (Baron & Ling, 2011; Squires, 2012), abbreviations (Khalifa, 2015; McWilliam, Schepman, & Rodway, 2009; Tagliamonte & Denis, 2008; Varnhagen et al., 2009), style shifting (Iorio, 2009), and linguistic creativity (Shaw, 2008). A few studies have even focused specifically on the role that “lol” plays in text messaging, with a variety of results ranging from a marker of emotion to managing discourse (Markman, 2013; O’Neill, 2010; Uygur-Distexhe, 2014). The present study seeks to combine these approaches, developing a broad understanding of the mechanisms texters use to construct meaning.
2 Research Statement

As stated above, this dissertation seeks to develop a comprehensive picture of literacy skills among urban bilingual youth, and approaches the topic from three perspectives. The research objectives are to document the multiple literacy forms among one population by approaching the language form found in text messaging as an equally robust language form as academic or social language forms. The objects for each approach are:

1. Objectives related to literacy types
   
   To develop a comprehensive picture of participants’ proficiency in academic, social, and txt language forms; To determine if there is a relationship between language proficiency and text message language choice; To determine if there is an effect of modality on language proficiency.

2. Objectives related to the documentation and description of bilingual Ttxt
   
   To create a composite picture of the literacy skills of a representative sample of bilingual youth; To find patterns in how bilinguals incorporate both languages while texting; To document the language being used by bilingual youth in its current emergent state.

3. Objectives related to the Case Study
   
   To apply the hypotheses developed for bilingual Ttxt more broadly to one ongoing conversation to determine if the hypotheses developed regarding the features of language use in Ttxt accurately describe the discourse pattern of two bilingual texters.

3 Terminology

There are a few terms that are used in a technical sense, but also have a colloquial meaning. Though the term, Ttxt, has already been established to refer to the language form used in text messaging and chatting, the remaining terms are addressed here.
The purpose of this section is to identify the terms that will be used and why without delving too deeply into the technical differences between digital platforms (for a complete discussion of the technical differences, see Deloitte, 2014). *Texting* and *to text* will be used as verbs and refer to sending messages via Short Message Service (SMS) or iMessage. The technical differences between these formats are not relevant to this dissertation so ‘texting’ refers to both, and a *text message* refers to the message itself. A *texter* is someone who is sending a text message. A receiver is someone who is receiving the text message. Other researchers use the term, *textee* (Dyers & Davids, 2015; Widyalankara, 2015). This term will not be used here as it implies that the one receiving the message has no agency in the receiving of the message. It became very clear during discussion with participants that the receiver of a message has all of the control as to whether they actually receive the message or not. Participants control their social world through their phones, and may read or not read (or pretend not to read), or answer or not answer their messages. This power structure in text messaging is confirmed by other researchers who find that adolescents and young adults are not ‘always connected’; rather they choose when to connect and when not to (Boyd, 2014; Forgays, Hyman, & Schreiber, 2014; Vitak, 2012). Therefore, the term, *receiver*, will be used in an effort to retain the distinction between the act of sending a message (being a *texter*), and the act of receiving a message (being a *receiver*).

*Chatting* refers to messages sent via any instant message such as Google chat or AOL Instant Messenger. This is different from texting as it is more often performed on a computer rather than a mobile device, and the expectation is that it is more synchronous than a texted conversation (though this is not always the case) (Ling & Baron, 2007a). At the time of writing, it must be acknowledged that the distinction between *chatting* and *texting* is artificial. If a participant uses an iPhone, it is not possible to know if the messages participants sent were sent from their phone or written at a computer. For the purposes of this study, the messages collected from participants’ phones are
considered text messages whereas the messages they sent during the chatting activity are considered *chats*.

Throughout this dissertation, participants are referred to as emergent bilinguals. The term *emergent* is used to convey that they are Spanish dominant in all settings and are developing their English literacy skills. They are developing English academic literacy in conjunction with Spanish academic literacy. Because their language skills are actively changing and developing, their overall language skills are emerging. This term is not meant to be a judgment on their English or their Spanish, only to indicate that since they are developing linguistic skills in both languages, their bilingualism is emerging. Secondly, the term *bilingual* is being used as a proxy for “multilingual.” Approximately 25% of the students in this study speak at least one other language at home, or learned at least one other language before learning Spanish or English. These students are multilingual, and while there were no instances of students texting in Kichee, or any language distinguishable from Spanish or English, it is important to note that *bilingual* does not fully capture their language abilities even though that is the term that will be used since bilingual language skills are the focus on this dissertation.

**4 Purpose and Significance**

Digital communication technologies are becoming an increasingly prevalent and important form of communication. The participants in this study (students age 17-21, born in 1994-1998) have had access to cell phones and computers for their entire lives and have never lived in a world without the Internet. This means that Txt has always been a part of their social life and they are among the first “native speakers” of this language form. This historical context combined with many other related technological advances is often termed the “digital revolution” (Craft, 2010; Davidson, 2011; M. Davies, 2012; Giedd, 2012; Halverson, 2009). In this study, the term, “digital revolution” refers to
the rapid increase in communication technologies and access to information. The communication affordances resulting from this cultural shift are both very new and yet nearly omnipresent in American society. Because these technologies are becoming such a part of daily life, it is of increasing importance to understand the language form associated with written interactions on digital platforms.

Increases in global communication capabilities and digital communication bring up an interesting point about the relationship between language and space. Children do not learn language by watching television, videos, or talking on a phone; they learn language from the people around them (Christakis, 2009). As adolescents and adults move around the world, they carry their linguistic skills with them. Yet, in every new place, humans almost always establish a speech community with the people around them (Blommaert, Collins, & Slembrouck, 2005; Rampton, 2010). The research presented here focuses on the Spanish and English spoken among emergent bilinguals in New York City. It is commonly accepted that the dialects of language spoken in New York City are different than the dialects spoken in other places such as Los Angeles or Mexico City. So, while this dissertation seeks to find linguistic universals as well as particular dialect features, it is focused on the particular language variety used in New York City. Therefore, this research sets the stage for future investigations into comparison of Txt across the country and the world.

4.1 Txt as a language form

It would be far too easy to dismiss Txt as something young people do that is “ruining language” (K. D. Anderson, 2015; Dwyer, 2012; Ferenstein, 2012; Jury, 2010). Many of the magazine and newspaper articles that adopt this perspective are building an emotional argument based on the fear

---

4 In some parts of the world, access to digital communication technologies is greater access to electricity (Buys, Dasgupta, Thomas, & Wheeler, 2009; Corbett, 2008; Mitchell, Bull, Kiwanuka, & Ybarra, 2011; Murphy & Priebe, 2011).
that digital technology is making students less intelligent. That anxiety is a recurring theme in American discourse about the relationship between education and technology (Fleming, 1989; Morrell, 2015). While there may be some truth to a decline in American students’ academic language skills (Heller & Greenleaf, 2007; Rivkin & Schiman, 2015), that decline likely has more to do with how and whether academic language is taught in American public schools (Heller & Greenleaf, 2007; Howard, 2015) than the prevalence of text messaging. From a linguistic perspective, there is no reason to be concerned over the emergence of a new language form. Formal language will not disappear because a new language form has emerged. Likewise, it is not the case that individuals who use Txt are unable to tell the difference between an academic setting and a social setting – as some researchers seem to suggest (Dansieh, 2011; De Jonge & Kemp, 2012; Drouin, 2011). These researchers argue that because students practice Txt more than they practice academic writing, they are only able to use Txt, not academic writing. This is simply not true. While there have been news reports of teachers receiving essays full of textisms (Maltais, 2012), this behavior has yet to be confirmed (Boyd, 2014). In my own experience as an educator, students tend to write emails (rather than essays) in an overly colloquial tone that incorporates features of Txt such as abbreviations and acronyms. This is less an issue of students not knowing about academic writing and more of an issue of students not recognizing the social distance between themselves and their instructors. It is therefore not so much an issue that students are losing the ability to write academically, but rather that students are unaware of when they should be writing in an academic form. In this dissertation, I take the approach that similar to an individual who is bilingual, a student who is bi-dialectal between academic language and Txt will have more linguistic options than one who never learned Txt. If a

---

5 This not meant to be a critique of any particular part of the American public education system, though it is deeply connected to issues of inequality of access. A discussion of the reasons the education system has turned away from focusing on academic language skills and a history of how the system evolved is far beyond the scope of this dissertation.
student uses the wrong language form for a setting, that is an issue of social boundaries, not an issue of linguistic competence.

Finally, Txt is a language form being used by millions of people every day for all types of interpersonal communication (Crystal, 2009; Shortis, 2007). Most linguists approach language as a constantly changing and evolving entity, and Txt is no different. In the same way that it is important to document all languages, it is important to document this stage of Txt. David Crystal, in his introduction to the 2014 book, *SMS Communication*, identified the need for more linguistic inquiry into this language type because it presents such a unique and timely opportunity to both document and analyze a quickly evolving language form. This dissertation contributes to that growing field of research as understanding Txt among young adults in this historical moment is an important component of understanding Txt more broadly. The participants in this study are the first “native speakers” ofTxt, meaning that this language form is being “born” right now.

It is important here to account for differences between populations of texters along generational lines, not just linguistic. Many adults who use Txt (even in their native language) treat it as acquiring a second language form, often communicating in ways that seem disfluent or awkward to today’s adolescents. Such adults often have to be explicitly taught how to use the features of Txt in order to convey nuanced meaning (S. Anderson, 2015; Dobbins, 2015). The young adults in this study have always had contact with Txt in Spanish and English, they are the first generation to be using it fluently even though people a generation older are the ones who developed it. Therefore, this is the first moment when Txt is being texted by “native” speakers, making this research not only timely, but crucial in understanding the evolution of this language form. The emergence of Txt is a prime opportunity to document the widespread evolution of a language form across multiple languages and dialects and essentially do language documentation as a language is born (rather than
as it is becoming extinct (Crystal, 2014). To capture this historical moment in the evolution of Txt is to prepare to do historical linguistics on a future language, creating a time capsule for future research on Txt (among Spanish/English bilinguals) and the evolution of a language.

4.2 Txt and Education

This research is further motivated by research and writings on translanguaging. Translanguaging is a specific view of multi-lingualism and the process of moving between languages and using all of one’s language abilities for communication (Celic & Seltzer, 2011; Wei & Garcia, 2013). The focus here is on the fact that people have one language faculty, and while within society, the languages may be divided, in one person, it is one language (for example, the participants in this study do not speak Spanish and English separately, they speak Spanish/English combined) (Celic & Seltzer, 2011; Wei & Garcia, 2013). The term is most often used with respect to the classroom where students speak multiple different languages. In an academic setting, the idea of translanguaging can seem revolutionary since assessment is almost always in one language, and correspondingly, instruction is typically in one language only (Wei & Garcia, 2013). However, Txt is an ideal format to better understand how emergent bilinguals mix languages together and practice translanguaging. Even though text messaging is distinct from academic language both in language and context, the linguistic skills participants bring to Txt are the same linguistic skills they bring to the classroom (Kemp & Bushnell, 2011; Scarcella, 2003). Therefore, understanding how bilingual youth communicate on digital platforms may provide insight into the language skills they are bringing to the classroom.

In addition to documenting the language skills of bilingual youth and Txt in this historical moment, this dissertation adopts the perspective that students are still developing linguistic skills in both of their languages and they still may perform better on one type of assessment than another. By
assessing participants’ linguistic skills across language modalities (i.e., spoken, written, digital), this project seeks to understand the role that platform plays in linguistic performance among this population. By assessing participants’ linguistic skills across language forms, this project seeks to understand the relationship between academic, colloquial, and texting language forms among this population. The results from these assessments describe how different language skills interact, and the role that proficiency in different language domains has on testing outcomes.

More than 27% of the students in NYC Public Schools are Spanish/English bilinguals (NYCDOE, 2011, 2014). Spanish/English bilinguals form the largest single bilingual group in American public schools, New York City, and the United States generally (Krogstad & Gonzalez-Barrera, 2015; Lobo & Salvo, 2013). Therefore, understanding how Spanish and English are combined across language forms is crucial to understanding bilingual students.

4.3 The Corpus

The corpus that was developed out of this project is significant in multiple ways. First, it is the first publicly available corpus of Spanish/English bilingual texts, and will prove useful both for researchers of text messaging as well as researchers of the New York City variety of Spanish/English. Second, since the code used for analysis is also publicly available via GitHub⁶, it provides a template for doing this type of analysis on other text messaging corpora. Finally, the tag set developed for this bilingual texting corpus brings together previous tag sets for Spanish and English. These tag sets were standardized to each other, and a new tag for pragmatic markers was introduced. The newly added pragmatic markers tag is a response to the findings in this dissertation.

⁶ GitHub is an open source web-based repository service. It is often used for version control and project management. It was chosen for this project as it allows individuals to do “pull requests” whereby they can download and modify the code directly and build on it for their own projects.
about the role of items such as “lol”, which serve to indicate pragmatic features that are not otherwise encoded.

The Part-of-Speech tagging done for analysis of this corpus aims to serve as a first step in making bilingual texting easier. As of now, someone who uses both Spanish and English Txt has a choice to make when writing her text messages: She can choose the Spanish keyboard (with a Spanish dictionary), she can choose the English keyboard (with an English dictionary), or she can choose one language and manually turn off all dictionaries. There is a cost with each choice. Selecting just one language and one keyboard costs a language, forcing the texter to either only write in that language, or continually teach the keyboard the new words from the other language. Turning off all dictionaries has a significant cost of time, as she must type every word exactly, requiring as much as 30-40% more time (He et al., 2014; Sawyer & Hancock, 2013). All of these options penalize people who are multilingual. One solution to this problem is the development of a bilingual dictionary that would allow a texter to choose a bilingual keyboard with a bilingual dictionary attached. A bilingual dictionary for texting is not the same as having two dictionaries to look words up in. These dictionaries are based on the likelihood that one type of word would follow another type of word. They therefore require very large part-of-speech tagged corpora to develop models from. The bilingual corpus and corresponding bilingual tagger presented here are the first of their kind, and are an initial step toward solving the problem bilinguals face when texting.

5 Previous Work

This project brings together research from a variety of different fields and disciplines around the linguistics of text messaging and the broad field of Computer Mediated Communication. Because so

7 There are certainly better solutions to this that take the whole person’s language abilities into account. However, computers are very literal devices, so this is a possible solution given the technological limitations, but certainly not the best solution.
many different domains are brought together here, they are discussed individually. First, previous work on text messaging corpora on documentation of txt will be discussed then research on all forms of bilingual Computer Mediated Communication, and then research on the relationship between text messaging and education.

5.1 Previous Documentation of Txt and Corpora

Teenage cellphone ownership in the United States has increased from 63% in 2006 to over 95% since 2010 (Lenhart, 2009, 2015; Lenhart, Arafah, & Smith, 2008; Rainie, 2012). Text messaging has only been available to the public for just under twenty years (the first text message was sent on December 3rd, 1992, but not common in the U.S. until 1999) (Brenoff, 2015; Deffree, 2015; Gayomali, 2012), and has only been popular in the United States since about 2008 (Lenhart, 2009, 2012; Rainie, 2005, 2012). Much of the early research on text messaging has sought to determine if Txt is spoken language that is written (Carrier & Benitez, 2010; Crystal, 2009; Deumert & Masinyana, 2008; Soffer, 2010) or if it is a form of written communication that is not bound by the rules of prescriptive grammar (Dansieh, 2011; Drouin & Davis, 2009; Plester et al., 2008). Both of these approaches fail to explain key features of the language found on text messaging, including the unique word frequencies, prevalence of emoticons and word play, creativity, the editable and semi-ephemeral nature, and unique time constraints (Forgays et al., 2014; Gibbs, Simpson, & Bernas, 2008; Holtgraves & Paul, 2013; Tagg, 2009). A third approach recognizes the language of texting and chatting as its own form, with its own features, and subject to its own rules and norms (Janis Androuatsopoulos, 2013; Dresner & Herring, 2010; Tagg, 2009). This approach presents the unique features of Txt as characteristic of the medium and worthy of further investigation. As of yet, the research that adopts this final approach is conducted exclusively on monolingual populations. By collecting and analyzing over 11,000 text messages from monolingual English texters, and analyzing respellings, word frequencies, pattern formation, and the role of creativity, Tagg (2009) showed that
Txt is distinct from the language form commonly found in speaking or in writing. Similarly, in their research on the use of emotion icons (emoticons) in instant messaging, Dresner and Herring (2010) approach texting as a language modality distinct and independent from other modalities (i.e., speech and writing). By extension, their research on Txt seeks to understand how texters and chatters construct meaning on digital platforms. Implicitly, in their research on code switching on text messaging, both Deumert and Masinyana (2008) and Dyers and Davids (2015) treat texted/Txt as its own language form, distinct from spoken/social, and written/formal language forms. Notably, all of these studies involved looking at text messaging rather than Instant Messaging or social media sites.

What makes SMS/Texting Corpora distinct from Twitter or chat room corpora (which are publicly available), is that there are only a few ways to collect the messages, all of which have some challenges. To gather text messages, they must either be transcribed or downloaded from a texter’s personal phone. With the transcription approach, the messages participants choose are selected and there is the potential for transcription errors. With the downloading approach, participants either upload their own messages, resulting in the potential for selection bias, or the researcher has to get access to their phone or message history. So, while text messaging is arguably one of the most common ways of communication in the United States, it is still relatively difficult to conduct research on the language of text messaging. Before the present study, there have been four major collection methods:

1. Direct message donations by friends, family and acquaintances, where participants (who are known to the researcher) select a subset of messages to give to the researcher(s) (Deumert & Masinyana, 2008; Tagg, 2009);
2. Transcription by participants from their phones onto paper then collected by researcher(s), where participants look at their message history and select a few messages to give to the researcher(s) (Baron & Ling, 2011; Ling & Baron, 2007a);

3. Anonymous online donations, where participants select messages and upload them anonymously to a server (Accorsi, Patel, Lopez, Panckhurst, & Roche, 2014; Bernicot et al., 2012; Chen & Kan, 2012); and

4. Giving phones to participants and collecting either the messages or the phones after a time period, where the participants are allowed to delete any messages they do not want seen (Okuyama, 2013).

The messages for this project were collected in a novel way, by plugging students’ phones into a computer and making a backup of their entire message history. This method was beneficial because it allowed for a candid record of participants’ message history. Messages were not deleted or selected by participants, rather it was an entire history of conversations with all the people in they text with. Secondly, this method preserved conversations, so in the cases where both conversation partners were participants in the study, the conversation could be preserved in its entirety. This method has an advantage over previous collection methods since this method allows the researcher to know the demographic information of some of the respondents. Therefore this corpus allows for comparison of how a participants texts with one person versus another.

As opposed to text messaging, corpus studies on instant messaging are much more common. This may be because instant messaging data can be obtained in a laboratory or classroom setting; it is a relatively common practice for researchers to have access to a class of undergraduate students where they either participate in the study or write a paper (Gibbs et al., 2008; Ling & Baron, 2007a; Markman, 2013; Squires, 2012; Tagliamonte & Denis, 2008; Varnhagen et al., 2009). It may be easier
to collect instant messages than text messages because text messaging is seen as a more intimate platform than instant messaging (Boyd, 2014; Brake, 2013; Marwick & Boyd, 2011); researchers hypothesize that this is because text messages are more often written on a personal device such as a cellphone whereas instant messaging is most often written at a computer (Boyd, 2014; Brake, 2013; Marwick & Boyd, 2011). While instant messaging certainly shares some features with text messaging, text messaging is mobile in a way that instant messaging is not, and more people have access to mobile technologies than personal computers (Rainie, 2012, 2013), which ultimately means that text messaging is more integrated into daily life than instant messaging (Crystal, 2009; Ling & Baron, 2007a).

5.2 Previous Research on Bilingual Computer Mediated Communication

In order to situate the present study in previous research on bilingual populations and digital communications, we will turn now to a discussion of previous approaches to understanding language mixing on digital platforms. Codeswitching is the mixing of two or more languages within the same utterance. Intrasentential code switching (starting a sentence in one language and finishing in another) requires knowledge of both the structure (syntax) and form (semantics) of both languages (Auer, 2011; Gumperz, 1982; Poplack, 1980). There are various reasons people code switch; some of those reasons are social (Myers-Scotton, 1993), some are conversational (Gumperz, 1982), and others are structural (Poplack, 1980). The theories about code switching are almost exclusively developed for face-to-face conversations since the vast majority of this research has been conducted on face-to-face social interactions. The theories presented by Gumperz (1982), Poplack (1980), and Myers-Scotton (1993) have been adapted to digital and written forms, though they remain an adaptation.
There has been some research on Computer Mediated Communication among bilingual users that incorporates language choice and code switching where relevant, though it is clearly still in its infancy. Researchers have investigated the pragmatics of bilingual texting (Deumert & Masinyana, 2008; Dyers & Davids, 2015), the use of the Roman alphabet for texting in languages that use a non-Roman alphabet (Widyalankara, 2015), the efficiency of code-switching in text messages (Carrier & Benitez, 2010), language choice in chat rooms (Paolillo, 2011), and code-switching in long form blogs (Montes-Alcala, 2005). There is very active research into the role of Computer Mediated Communication and the second language classroom (Bensoussan, Avinor, Ben-Israel, & Bogdanov, 2006; Cummings, 2004; Gleason & Suvorov, 2012; Klimanova & Dembovskaya, 2013; Pasfield-Neofitou, 2011; Proctor, Uccelli, Dalton, & Snow, 2009; Yim, 2011), especially examining how intercultural communication can be facilitated on social media platforms and what the outcomes are for second language learning.

There are three studies that specifically focus on bilingual text messaging. Two of these focus on language mixing in African languages (Deumert & Masinyana, 2008; Dyers & Davids, 2015), and the third focuses on monolingual French messages that have incorporated textisms from English [i.e., “lol”] (Uygur-Distexhe, 2014). Both Deumert & Masinyana (2008) and Dyers & Davids (2015) investigated how multilingual texters treat each language separately, and both find that the pragmatics of a language are tied to the language and pragmatic features do not cross language boundaries, even in the same message. Looking at the relationship between isiXhosa and English, Deumert & Masinyana found that even when mixing languages together, bilinguals performed certain operations in one language (i.e., respelling, abbreviation, etc.) that they would not perform on the other language. More specifically, participants were willing to abbreviate and modify English words, but were not willing to do so with isiXhosa words even though there was a financial motivation to do so (2008). Dyers & Davids found that the rise of text messaging and development
of a Txt language form is resulting in the preservation of a written form of “local” languages that may otherwise rarely be written (2015). Through researching the language practices of bilingual youth, they found that participants mix Afrikaans, isiXhosa and Setswana with English for both creative and communicative purposes (2015). Additionally, they conclude that the use of both languages together is a powerful way to illustrate creativity, linguistic flexibility, and to construct identity on a texted platform (2015). Uygur-Distexhe’s research focused in on textisms of laughter, with specific attention paid to the role that “lol” plays in text messaging. She found that “lol” has a different distribution than French forms of laughter, and actually does not refer to laughter at all. Though her research is focused on words regarding laughter, she also shows how a loanword from another language can serve a discourse function in the language it is borrowed into (2014).

One final study that deserves mention here is by Okuyama on the use of text messaging by deaf students in the United Kingdom (2013). While these students are not typically considered bilingual, there is a case to be made that signed English is appreciably different from spoken or written English, in effect making these students multilingual between British Sign Language and academic British English. The study in question was a case study by Okuyama of two deaf adolescents in a boarding school (2013). They each were given a prepaid phone and could use it to text freely with each other for a period of three months. They could delete as many messages as they wished so long as they gave 90 messages per month to the researchers. Okuyama found two interesting results: the first is that the students created a speech community between them, teaching each other textisms, emoticons, and respelling techniques (Okuyama, 2013). Her second finding was that students used English words with American Sign Language grammar; she took this to indicate the strength of the relationship between spoken language forms and texted (Okuyama, 2013). This second finding also indicates a particular way that these participants combined languages in order to convey meaning.
5.3 Previous Research on the relationship between Education and Txt

The third and final area of previous research to be covered is on the relationship between education and text messaging. It seems as though with every technological advance, there are concerns about effects on human intelligence and literacy (Haas, 2013). The advent of digital technology is no different, and many researchers have sought to determine if there is a relationship between frequency or fluency of text messaging and academic literacy skills (Baron, 2005; Bushnell, Kemp, & Martin, 2011; Drouin, 2011; Drouin & Davis, 2009; Kemp, 2010a; Kemp & Bushnell, 2011; Plester et al., 2008, 2009; Wood et al., 2011). Much of this research is in response to sensationalist claims that text messaging is “ruining language” (Dwyer, 2012; Jury, 2010), or that teachers are receiving essays “full” of textisms (Crystal, 2009). While from a linguistic standpoint, these claims are theoretically impossible (no language form can ‘ruin’ another just by being used), the fears are still very strong. Therefore, the relationship between Txt and academic competencies must be understood in order to better understand the relationship between academic achievement and communication in multiple language forms.

5.3.1 Txt and academic language forms

Research on the relationship between the use of Txt and academic literacy skills has three types of results: Some researchers find a negative correlation between academic literacy skills and textism awareness and Txt use (Dansieh, 2011; De Jonge & Kemp, 2012; Drouin, 2011; Rosen et al., 2010), some find no correlation (Baron, 2005; Drouin & Davis, 2009; Kemp, 2010a; Plester et al., 2008), and still others find a positive correlation (Bushnell et al., 2011; Kemp & Bushnell, 2011; Plester et al., 2009; Powell & Dixon, 2011). The disparity in results and the weakness of the correlations indicates that there is still no conclusive evidence that communicating in Txt has any effect on

---

8 This fear is echoed with languages in contact, and the idea that there is a “pure” form of a language, and that other languages can “pollute” or degrade that.
academic literacy skills. More concretely, many of these studies confuse correlation with causation. It could be the case that students’ academic language skills are developed through a completely separate and independent process from how text messaging skills are developed. Regardless of this situation, these studies are the only studies which simultaneously investigate multiple different language forms among a single population.

Researchers who find a negative correlation between either textisms use (De Jonge & Kemp, 2012; Rosen et al., 2010) or frequency of texting (Dansieh, 2011; Drouin, 2011) argue that the negative effect is due to exposure and repeated practice of “deviant” forms. De Jonge & Kemp had high school and college students translate between prescriptive English and textisms and found that knowledge of textisms was negatively correlated with spelling awareness and reading scores (2012). Using a survey and writing task, Rosen et al. found that young adults (college students) who texted frequently and used textisms regularly performed worse than infrequent texters on a formal written assignment (2010). Dansieh surveyed 400 hundred students and found that frequent texters were more likely to adopt non-standard forms in their writing and expressed concern that second language students may not be able to differentiate the appropriate contexts for Txt versus academic writing (2011). Finally, in a study of American college students, Drouin found a negative correlation between textisms usage on social media sites and spelling and reading accuracy (2011). These researchers all cite Dixon & Kaminska (1997, 2007)’s work that found that exposure to misspelled words resulted in lowered spelling scores for both children and young adults. One assumption that these studies make is that spelling is a metric for academic literacy. While spelling is one measure of literacy and has been shown to be an important factor in academic success (Cobb, 1972; Morris & Perney, 1984; Walker, Greenwood, Hart, & Carta, 1994), it is only one component of academic skill.
On the other extreme, some researchers found a positive relationship between texting frequency, textisms use, and academic literacy skills. These researchers mostly argue that the relationship between Txt and academic literacy has to do with overall language skills rather than individual competencies. These researchers argue that the students who perform better on measures of textism awareness and Txt comprehension as well as verbal reasoning, and reading comprehension simply have an overall higher level of linguistic skill, ultimately arguing that a student who performs well on one language task will perform well on all language tasks (Bushnell et al., 2011; Kemp & Bushnell, 2011; Plester et al., 2009; Powell & Dixon, 2011).

In addition to these opposite perspectives, some researchers even find both a positive and a negative correlation between academic literacy skills and textisms usage among the same population (Kemp, 2010a; Rosen et al., 2010). These polarized findings are a testament to how divided this research is and how inconclusive many of the results are. Ultimately, this dissertation contributes to this discussion by approaching the idea from another perspective. Rather than analyzing contact withTxt as compared to academic proficiency, we will look at the relationship between academic English proficiency, social English proficiency, and English text messages. The second round of investigation focuses on overall academic proficiency and lexical density on Txt to determine if participants who have more fully developed linguistic skills in one domain transfer them to other domains. The purpose here is to better understand the relationship between language proficiencies without assuming that Txt can either diminish or enhance academic skills. The results presented in Chapter 2 suggest that the effects may be due to overall proficiency with language rather than achievement on any individual skill.
5.3.2 Written, spoken, and digital modalities

Many investigations into the relationship between Txt and academic language forms associate modality with register, drawing implicit connections between academic communication with reading and writing skills and colloquial communication with speaking and listening skills (Collier, 1989; Cummins, 1980, 1999, 2008; Scarcella, 2003). This seems to be an obvious coupling since classroom instruction focuses so much on the written word. However, in recent years, research in New Literacy Studies has taken a third approach by focusing on spoken academic communication, and how different styles of oral communication are valued in a classroom setting (Gee, 2011; Street, 2003). While this research extends the possibilities for academic skills assessment, this approach is still limited by only taking two language forms into account. By incorporating a third language form, Txt, and a third modality, digital platforms, it becomes possible to develop a more complete picture of students’ language skills.

6 Research Design

This is a mixed-methods project that involves a quantitative analysis of the relationship between academic, social, and txt literacy skills in order to determine the relationship between different language functions as well as different language modalities. The qualitative component presents hypotheses about the role that key features of Txt (such as repeated letters, abbreviations, acronyms, and special acronyms [i.e., “lol”]) play in constructing meaning on a digital platform. The final component is a case study of one conversation between two participants. Their conversation spans the first four months of their relationship. They only see each other once a week yet send hundreds of messages a day, indicating that they are using text messaging to establish intimacy in their relationship. The hypotheses developed in the qualitative chapter are applied to the case study to test their validity with this population.
Students in this study are all Spanish-speaking NYC youth, age 18-21. They are enrolled in a dual language program within Pathways to Graduation Program (a High School Equivalency Program) through the New York City public school system. They all plan to take the Test Assessing Secondary Completion (TASC) within one to three years. Students were given a series of diagnostic tests to measure their academic, social, and txt language skills. Each language form was measured in at least two modalities: the one traditionally associated with it and one other (i.e., social literacy was assessed through a spoken discussion task and a written task). This design allows language form to be separated from the modality so that a student who struggles in one modality may be able to succeed in a different modality but on the same language form. The data were collected at two schools in conjunction with a series of workshops designed to help students prepare linguistically and culturally for their goals after graduation. The results of each assessment and activity were translated into a grade-level equivalency score either based on percentage correct or with respect to a rubric. Correlations were determined based on these results.

The text messages were collected by plugging students’ phones directly into a computer and making a copy of their message history. All of the messages on their phones were downloaded. Participants had to sign a separate consent form for this and use their password to unlock their phone. Significant measures were taken to ensure privacy, detailed in Chapter 2.

7 Assumptions and Limitations

As with all research, this project makes assumptions about language and the participants and there are limitations to the research. The first major assumption has to do with treating Txt as a unique language form that is distinct from academic or social language forms. While this has been illustrated by previous researchers (Janis Androutsopoulos, 2006; Baron, 2010b; Crystal, 2009, 2014; Georgakopoulou, 2011; Tagg, 2009, 2013; Tagliamonte, 2016), it is still an assumption about the role
of Txt with respect to other language forms. The evidence that it is a unique language form is based on quantitative metrics such as word frequencies, lexical density, and spelling forms and qualitative metrics such as focus groups and reflections about texting. Collectively, these data show that Txt is a language form that is sufficiently distinct from other language forms and therefore can be analyzed as a separate entity.

The second assumption is that it is possible to measure differences in language ability in Spanish and English and that students would have different proficiency levels. While this is a widely accepted assumption, it is worth mentioning that literacy in one language is often reinforced by literacy in the other. Research on this topic often divides a student’s second language skills in a different way from their first language skills (Cummins, 1999). Researchers rarely discuss non-academic proficiency in a first language yet very often focus on both academic and social proficiency in the second language (Cummins, 1999). While we often talk about a student’s second language skills in terms of social and academic or colloquial and formal, we do not often talk about their first language skills in the same way. Rather, we say about their first language that they have not developed academic literacy skills. The assumptions about language in this dissertation and the relationship of second language proficiency types to each other builds on Scarcella’s framework for second language development (2003). In this framework, she lays out five main language components and explains that students’ linguistic development in each of the components may not necessarily be in a linear path (Scarcella, 2003). Rather, a student may be very advanced in phonological skills, but struggle immensely with pragmatic skills (Scarcella, 2003). This dissertation is built on this assumption as it considers multiple language functions and forms and seeks to isolate different language components while recognizing that a student who has the ability to perform one language function may not necessarily have acquired all of the skills in that domain yet.
The final assumption is that there is some relationship amongTxt, social, and academic language forms. This assumption is rooted in the same idea driving Scarcella’s work as well as that of researchers who have found a positive relationship between verbal reasoning and textism awareness. Namely, a student’s language skills in one language likely reinforce their language skills in their other language(s) and/or language forms; it is well established that that first language literacy predicts second language literacy in academic registers (August, Shanahan, & Escamilla, 2009). However, it is unknown if proficiency in texting in the first language predicts proficiency of texting in the second. As this dissertation is documenting the features of Txt, answering this question is beyond the scope of this dissertation.

This research had significant limitations in two primary areas in terms of cohesiveness and generalizability of the findings. The first major limitation was the rate of attrition among participants. Because of scheduling conflicts as well as the very nature of these types of academic programs (i.e., students are not mandated to attend, and are adult learners), students were frequently absent, or completely stopped coming. Additionally, students joined the study after it had already begun. In these cases, the consent forms were explained individually to the new students. These students were then allowed to participate in the activities and diagnostics. The result is that students participated in the research when they were present and did not when they were not. No students participated in all of the diagnostics, and no diagnostic was taken by all students. As will be discussed in Chapter 2, half of all of the participants only participated in one activity. Consequentially, the results and analysis are preliminary, and indicated as such where they are presented. This understandably makes it very difficult to determine if there is a significant correlation between different measures. When statistics are reported, the N is reported as well in an attempt to be clear about the numbers of participants who engaged with each task. In order to
compensate for this, I expanded the study to a second school after the first round of data collection, though there were still high attrition rates there.

A second limitation of this study is the generalizability of the findings. The participants in this study form a relatively homogenous group in that they are all within the same age range, they are all Spanish-English bilinguals in New York City, and they all know each other. While this presents a unique corpus, and allows for intact conversations, it also presents a very specific corpus and without further research, it cannot be determined if the features participants use are specific to them or if what is being observed is a much larger pattern of linguistic behaviors. The features investigated here have been reported in other corpora, and other researchers have noted them as items of importance. Yet, without further investigation of these hypotheses, they cannot be generalized to the population at large.

A final limitation is that after the analysis was conducted, there was no opportunity to go back to participants and ask more questions about their texting behavior. It would have been ideal to go back to participants after the initial analysis to ask exactly what they meant in a given context or in a given text message. Unfortunately, this was not possible both because the study ended and students moved on from their schools. As a result, the conclusions drawn from the text messages are largely speculation based on the patterns identified in the messages rather than participant insights into their own texting behavior. What this data collection method allows is an authentic record of messages without researcher involvement. In an effort to compensate for this, some of the data was set aside for the initial analysis and then used to verify the initial hypotheses. This method draws on training/test methods where a model is built using the some of the data and then validated by testing it against the remaining data. In this case, the messages between “the lovers” (the largest intact conversation in the corpus) were set aside for the initial analysis and then used to validate the hypotheses.
8 Outline

Chapter two of this dissertation outlines the methods and data collection processes for this dissertation. This chapter includes a description of the participants, the school setting and methods used in gathering the data, the design and validation of diagnostic tests, and the methods for text message collection and de-identification. The quantitative results are reported, highlighting both student performance across a variety of measures as well as the relationship between proficiency in academic language skills and text message fluency. It shows that participants who have higher academic English proficiency are more likely to text in English. However, text messages written in English tend to be less linguistically complex, suggesting that even though students who text in English have a higher academic skills level, they are still better able to communicate in Spanish Txt than in English. The second finding from this chapter is that students perform better on receptive tasks than productive tasks, illustrating that students’ linguistic skills are not always uniform across language modalities.

Chapter three introduces the respellings participants use to make meaning in Txt in both Spanish and English, investigating shape, distribution, and usage of each item to understand how respellings are used. There are three main types of respellings identified in this chapter: abbreviations, initialisms/acronyms, and iterations (repeated letters). The initialisms, “lol” (laugh out loud) and “ok” (all correct) are discussed separately as they have their own pragmatic functions. Additionally, mistakes and mistake repair is discussed. This chapter argues that abbreviations and most initialisms serve to save the positive face of the receiver by indicating that the receiver is a member of the texter’s in-group. The purpose of these is both stylization as well as adding a colloquial, intimate tone to the message. This chapter also argues that iterations are used to add an emotional overlay to the message by adding both emphasis and drawing attention to the word being iterated. The discussion of “ok” follows previous researchers in identifying it as a discourse marker.
This dissertation makes a novel contribution to the field of Computer Mediated Communication (CMC), building off Dresner and Herring’s interpretation of emoticons (2010) by identifying the use of “lol” among this population as a particle signaling a mismatch between the literal and intended meaning of an utterance. The final section in this chapter discusses the role of repair in identity presentation, showing that the primary motivating factor for mistake repair is unlikely to be disambiguation. Through analysis of the contexts in which mistake repair occurs, this section argues that participants are more likely correcting mistakes in order to present an educated version of themselves.

Chapter four is a case study of one conversation between two participants (who will be called Jasmine and Michael). This conversation chronicles the first four months of their relationship. Jasmine and Michael are both Dominican and their families knew each other in the Dominican Republic; they were reacquainted in New York City, and started dating four months before data collection began for this project. They take their relationship very seriously and are committed to each other. They both donated their entire message history, effectively meaning that the more than 30,000 messages they sent during that time are intact. Almost all of their conversation is written in English. This is taken as evidence of how seriously they are taking the relationship and the way that they present themselves to each other. The hypotheses that are developed in Chapter three are applied to Jasmine and Michael’s conversation where relevant. Notably, they use very few abbreviations, selecting instead to give a more serious, formal tone to their messages. They do, however, frequently use “lol.” In many instances, their use of “lol” likely indicates that a message should be interpreted flirtatiously. Finally, they are both very quick to correct mistakes, indicating that they are both attending very carefully to the identity that they are presenting.

The concluding chapter, Chapter 5, brings these findings together and suggests future directions for research on both text messaging and bilingual computer mediated communication. In
terms of educational outcomes, this dissertation found that regardless of English language ability, participants sent more linguistically complex text messages in Spanish. It was also found that participants performed better on measures of comprehension rather than production. That is, when students were able to write in a colloquial register, mixing languages together, they were able to illustrate that they had understood the material. These findings together indicate that assessments of comprehension that also take production into account may not capture a student’s true linguistic skills.

The overall contribution from this dissertation is the identification of the ways that young adults construct social meaning on a completely text-based register. The ability to establish intimacy, exchange empathy and build trust on a completely text-based platform is truly noteworthy, and this dissertation seeks to better understand how texters go about doing this. Ultimately we see that texters use a variety of complex cues (iterations, pragmatic particles, repair) to construct both meaning and identity in their text messages.
Chapter 2 - Literacy types and modalities

1 Introduction

This chapter presents the results of collecting multidimensional information about participants’ linguistic skills, and focuses on the quantitative dimensions of this research project, by presenting a comprehensive snapshot of the multiple literacies that students possess across multiple modalities. The results of assessments of participants’ academic, colloquial, and txt literacy skills across written, spoken, and digital platforms are presented. This information is used to determine what platforms students perform best on, and what skills they have the most proficiency in. Additionally, academic proficiency and language choice on digital platforms is identified. Through analysis of student performance on a range of linguistic measures (academic, social, and txt) across a variety of testing modalities (written, digital, and aural), this chapter addresses the relationship between features of various modes of communication.

In total, there were six assessments to measure academic and social literacy skills. Additionally, participants took a survey and donated their text messages. Each assessment in described in Table 1 along with the language form and modality it is intended to assess and how many participants took the assessment.
<table>
<thead>
<tr>
<th><strong>Test Name</strong></th>
<th><strong>Description</strong></th>
<th><strong>N</strong></th>
</tr>
</thead>
</table>
| AJT          | Acceptability Judgment Task. Written assessment to measure social language awareness in English and Spanish.  
  *Written, Social (English & Spanish)* | 48 |
| Aural Comprehension | Free-form written responses to an academic language video. Videos are in English, students can respond in a mixture of Spanish and English.  
  Assessment is only of comprehension, not formality of production.  
  *Aural, Academic (English & Spanish)* | 27 |
| Chatting Game | Free-form writing on a digital platform. Participants described academic vocabulary words to a partner via Google Chat.  
  *Digital, Academic (English & Spanish)* | 16 |
| English LENS | Literacy Evaluation for Newcomer SIFE. Written academic literacy test assessing reading comprehension in English only.  
  *Written, Academic (English)* | 28 |
| Spanish LENS | Spanish language Literacy Evaluation for Newcomer SIFE. Written academic literacy test assessing reading comprehension in Spanish only.  
  *Written, Academic (Spanish)* | 9 |
| TABE-Esp    | Test of Adult Basic Education. Written assessment of academic content awareness.  
  *Written, Academic (Spanish)* | 43 |
| TABE-CLAS E | Written assessment of academic English proficiency  
  *Written, Academic (English)* | 43 |
2 Participants

Participants in this study are all working toward their High School Equivalency (HSE) degrees in the Pathways to Graduation (P2G) program administered by the New York City Department of Education. This format is an “alternative” high school completion program. Students are enrolled in these programs for a few different reasons. First, participants stop attending traditional high schools, yet still need academic support to prepare for the HSE exam, the Test Assessing Secondary Completion (TASC). Second, students fail 12th grade or do not have enough overall credits to graduate from high school, so they go to an alternative program after age 18. Third, they enroll in public school in their late teens and do not have enough time to earn their high school diploma, so they finish in P2G. The majority of participants in this study are in P2G for the third reason: they did not have enough time in public school to earn their high school degree.

To earn an HSE degree, a student needs to pass the TASC (formerly the GED Test). They do not need to earn a certain number of classroom credits; every year approximately 750,000 people in the United States take the TASC and slightly more than half pass it. Approximately 10-15% of high school-level degrees awarded in the United States are awarded in this way (Chapman, Laird, Ifill, & KewalRamani, 2011; GED Testing Service, 2014; Heckman, Humphries, & Mader, 2010). This is a remarkably high percentage given the fact that public education is free and mandatory until age 16 in the United States. Because of this high percentage, there has been a significant amount of research to determine the effect HSE degrees have on economic success as compared to high school diplomas and non-degree holding persons. The overwhelming result is that someone with a high school diploma will likely be more financially successful than someone with an HSE degree (Heckman et al., 2010; Heckman & Rubinstein, 2001; Song & Hsu, 2008). Further research has shown that the reason for this discrepancy has far less to do with cognitive ability (cognitive ability
levels are nearly the same between the two populations), and more to do with non-intellectual skills such as persistence, focus, and motivation (Cook et al., 2014; Heckman & Rubinstein, 2001; Patterson, Song, & Zhang, 2009; Renzulli & D'Souza, 2014; Song & Hsu, 2008). Without non-academic intervention and soft skills development, the value of the HSE exam is greatly diminished for most populations. However, this research was all conducted on high-school dropouts, not late arrivals. Late arrivals tend to have higher success in HSE programs likely because they are in HSE for a very different reason (Fry & Center, 2010; Heckman & LaFontaine, 2006; Perreira, Harris, & Lee, 2006; Protopsaltis, 2005).

Participants in this study were all 18-21 years old, enrolled in Spanish-English dual language programs within the P2G model. Classes are in session Monday-Friday, 7:30am-3:30pm, and participants are allowed 10 unexcused absences before they are un-enrolled from the school. Still, completion of the program is dependent upon passing a test, and because most of the participants are older than 18 years old, they are in class voluntarily, and are free to leave should they wish. For the most part, parents are not actively involved in the school since the participants are legally adults. All of the participants in this study are preparing to take the Test Assessing Secondary Completion (TASC) in either English or Spanish within the next two years. Many indicated that they plan to go to college or the police academy once they earn the HSE. They are unable to go to these programs without a high school equivalency credential.

The study took place during school hours, in lieu of their regularly scheduled English Language class. Participants were recruited based on their participation in the Spanish-English bilingual program. Data for the study was collected in conjunction with a six-part workshop during
the stand-alone English as a Second Language (ESL) class. One hundred participants across two schools participated in the study by filling out a consent form and completing one task associated with the project.

One school is located in Manhattan (School_MN) and the other is located in the Bronx (School_BX). Three participants had attended both schools. In the first school, participants were separated into four classes based on their academic preparedness and performance on the Test of Adult Basic Education - Español (TABE-Esp). This separation is not related to their English proficiency. The four classes are High School Preparation (HSP) levels 1 and 2, and High School Equivalency (HSE) levels 1 and 2. HSE 1 and HSP 2 were combined into one class, resulting in three classes at School_MN. Table 1 shows how many participants were in each class.

Table 2 – Participants Per Class

<table>
<thead>
<tr>
<th>Class</th>
<th>Total Participants</th>
<th>Males/Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>MN HSP 1</td>
<td>44</td>
<td>25/19</td>
</tr>
<tr>
<td>MN HSP2/HSE1</td>
<td>23</td>
<td>11/12</td>
</tr>
<tr>
<td>MN HSE 2</td>
<td>7</td>
<td>7/0</td>
</tr>
<tr>
<td>BX</td>
<td>26</td>
<td>13/13</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>56/44</td>
</tr>
</tbody>
</table>

All participants learned Spanish in the home, though some participants learned a language in addition to Spanish, including Mixteco, Kichee, and Garifuna.

---

9 Stand-alone ESL means that the class is focused on developing English academic language proficiency and is often divorced from the content being delivered throughout the curriculum. Other ESL support includes push-in (an assistant enters a content-focused classroom to assist with English language support), pull-out (an assistant pulls participants out of content class to work individually or in small groups to develop English language skills), and dual-language (English and first language skills are developed in conjunction with academic content).
2.1 Participants - Participation

One hundred participants participated in the study by participating in at least one activity. However, there were eight tasks total; fifty participants only completed one task. No student completed all eight tasks. Three participants completed seven tasks and eleven participants completed six tasks. Table 2 shows how many participants completed each of the tasks. Unfortunately, participants did not complete all the same tasks. For example, of the three who completed seven tasks, one did not complete the Spanish LENS, one did not donate text messages, and one did not complete the aural assessment, resulting in very small numbers to calculate correlations with.

Table 3 – Number of tasks completed by number of participants

<table>
<thead>
<tr>
<th>Number of Tasks Completed</th>
<th>Number of Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>3</td>
<td>13</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>0</td>
</tr>
</tbody>
</table>

The primary reason for the sporadic participation was absences and scheduling issues. High school completion programs nation-wide tend to have poor attendance as participants are not required to complete a set of classes, but rather pass the TASC (Patterson et al., 2009). Because of this, participants in the upper level classes may decide to attend a math or science lesson because they feel less prepared in those subjects than in English. Since this study was conducted in English
classes, this may explain some of the attrition in the upper levels. Participants in the lower level classes drop out at an extremely high rate for reasons irrelevant to this study. Additionally, seventy-eight percent of the participants in the current study plan to take the TASC in Spanish, not English; as a result, participants are less motivated to go to English class. Finally, there was a TASC pre-test that neither the teachers nor I were aware of on one of the data-collection days, resulting in a large group of participants being absent on one day, with no option of rescheduling.

2.2 Participants - Internal Review Board

Because English is a second language for these participants and the power dynamics of being in a classroom put participants in a compromised position, special care was taken to be sure that they understood the research and the aims of the project. Consent forms were distributed in English and Spanish, and the study was explained orally in English and Spanish. A Spanish Speaking Research Assistant (SSRA), the class teacher, and I answered questions and checked for comprehension on the first day participants joined the workshop. Any participants who appeared as though they did not fully understand the project were questioned individually by myself and a SSRA until it was clear that they understood. Participants enrolled in this program are all over 18, so they were able to give consent. Parent consent was not sought. Consent forms were approved by both the CUNY Institutional Review Board and the NYC Department of Education Institutional Review Board.

The consent form has two parts. The first part is the agreement to participate in the study in general. This refers to the “low-risk” activities such as the survey, literacy tests, and other easily de-identified information. The second part specifically asked about participation in recorded interviews and donation of text messages directly from their phones. Participants had to circle ‘yes’ or ‘no’ in

---

10 For more information about the relationship between remedial education and attrition, see (Hoyt, 1999; Mamiseishvili & Deggs, 2013).
response to two questions. First, “I agree to be audio-recorded during the interview.” Second, “I agree to donate messages from my phone.” Even after participants made their choices, consent was reaffirmed before they were recorded and before they donated their messages. Finally, before donating their messages, they also had to put their phone’s passcode in to authorize the computer they were downloaded to.

Internal Review Board (IRB) approval for research on human subjects was obtained both from the City University of New York (CUNY), and the New York City Department of Education (NYCDOE). CUNY IRB Approval was granted on June 10, 2014 for project number 592008. An extension was obtained on July 1, 2015 as analysis was not yet complete on the data, so the follow-up report could not be filed. NYCDOE IRB approval was granted on August 1, 2014 for project number 735. An extension was obtained from the NYCDOE on July 15, 2015 to continue the research and analysis. The primary concern on both projects was to protect participants’ privacy. To this end, all participants have been assigned a code and that code is used in lieu of their name on all documents. The data is maintained on a secure, password protected drive in the Second Language Acquisition Lab at the CUNY Graduate Center. The text message data was de-identified computationally and then verified by two different readers to remove all proper names, location information, and phone numbers based on the procedures described in (Accorsi et al., 2014). This is described in detail in Section 4.
2.3 Participants - Connecting with Schools

A Spanish-dominant bilingual population was identified through the Program Coordinator for Immigrant Services at the NYC Department of Education's Pathways to Graduation Program. She put me in touch with dual language programs within Pathways to Graduation Program in both Manhattan and the Bronx. Both schools were looking for a way to increase student engagement and preparation for their post graduation-goals. To fulfill this need, I developed a workshop to increase participants’ communication skills and to help them reflect on the language they use in various settings. Data was collected as a part of this workshop. The workshop began on March 30, 2015 and concluded at all schools on May 15, 2015.

3 Measures

For each data collection tool, a description of the tool, the method for administering it, the rate of return, and the mean score is presented. Where appropriate, some tools have been separated into their constituent parts. Correlations related to the guiding hypotheses are presented in Section 4.

---

The original intention of this study was to select Long-Term English Learners (LTELs), but in the process of connecting with other researchers, principals, and schools, it was determined that finding and accessing a group of LTELs would be nearly impossible within a reasonable time frame. The primary concern with LTELs is that they have fallen through the cracks and are not identified within the school or grouped together into a unit. The only way to determine if a student qualifies as “long-term” is to go through their academic records to see how long they have been receiving English language support. Attrition is remarkably high, and many participants drop out before reaching high school (Kim & Garcia, 2014; Menken, Kleyn, & Chae, 2012). Information about LTELs is largely gathered by going through student records computationally to determine how many exist within the school district. The other approach is asking participants receiving English language support how many years they have been receiving it. Neither of these options was feasible for this project.
3.1 Measures - Survey

The survey consisted of twenty-nine questions across five topics (see Appendix 1). There were five questions about basic demographic background (age, gender, grade level, country of birth, and years in the U.S.), two questions about previous schooling experiences (in the U.S. and outside of the U.S.), six questions about language usage (with whom they use which language(s), and preferences for language use across situations, interlocutors, and modalities), eight questions about phone and technology use (what platforms they use in which language(s) and with what frequency, what applications they use, what kind of phone plan they have, and where they access wireless internet), and eight about texting and chatting behavior (who they communicate with on digital platforms and in what language(s), communication preferences, abbreviation and emoji use, and autocorrect settings). The survey was given in English with a Spanish translation printed on the other side of the paper. Since participants were also participants in an English class, they were instructed to complete the questions in English. Participants with the highest level of English proficiency (as determined by their teachers and enrollment in the Advanced English class) only received the English version. The survey was administered on the first day of the workshop and took participants between 45 minutes and 2 hours to complete. Participants who were unable to complete it on the first day had the opportunity to continue working on it in the following days.

3.1.1 Survey - Design and Amendment

Many of the demographic, educational, and language background questions were adapted from the language-background questionnaire used for the Second Generation Bilinguals Project also conducted in the SLA Lab at the CUNY Graduate Center. The telephone usage questions were adapted from Boyd’s research on teen mobile phone use, and addressed many of the key findings that she found in her surveys and interviews with participants (Boyd, 2014). In the text and chat
behavior section, participants were asked a variety of questions. The questions were selected for two primary reasons. First, there were questions that aimed to explore how participants use popular emergent technologies. These can be separated into two groups: social media, and person-to-person communication. Social media questions revolved around platform (i.e., Facebook, MySpace, Instagram). Person-to-person communication questions revolved around affordances (i.e., Facetime, iMessage, SMS messages, phone calls). Recent research on context collapse has explored how participants manage different social groups (i.e., family, school friends, co-workers, etc.) by differentiating which technologies they use with whom (Marwick & Boyd, 2011; Vitak, 2012). Most of this research focuses on the social media sites that participants use and how they use them. This study expands on that research by gathering information about the person-to-person tools participants use as well.

The second group of questions covered aspects of text that have been investigated for monolingual populations, including pragmatics and non-textual cues (Baron, 2004; Baron & Ling, 2011; Deumert & Masinyana, 2008), frequency of texting and preference of medium (Crosswhite, Rice, & Asay, 2014; Lenhart, 2012; Ling, Bertel, & Sundsoy, 2012), and access to the internet via mobile devices (Baron, 2010a; Mossberger, Tolbert, & McNeal, 2008; Walsh, White, Cox, & Young, 2011). Participants were very interested in answering and discussing the questions related to what platforms they used; some participants volunteered reasons for why they use one platform or another. The majority of the reasons participants offered were variations on how they prevent context collapse, and keep their social worlds separate.

The survey involved many items that had not been previously piloted. It was therefore given to fifteen bilingual undergraduate participants enrolled in the Bilingual Education Program at Hunter College and eight bilingual undergraduate participants enrolled in the Linguistics program at Lehman
College from March 1-10, 2015. The data from the piloting study is not included in the results since they come from heterogeneous backgrounds. Survey items were adapted and amended based on their feedback and the language was adapted to make it as clear and accessible as possible. The language background questions were re-worded to be certain participants would interpret the questions in a way intended by the research design. Even after these adjustments were made, participants at the first school had significant difficulty with the wording of one language background question. The form of the question was therefore altered at the second school. The primary difference is that “English” and “Spanish” were pre-filled to clarify what they were responding to and allowed a third line to write in other languages. This appeared to make the question more accessible since participants struggled with it far less at the second school.

3.1.2 Survey Results: Basic Demographic Information

All one hundred participants completed basic demographic information on the survey. Of those, fifty-five participants completed the survey. Six participants started the survey but did not complete it, and thirty-nine only completed the first page (basic demographic information) of the survey. Fifty-seven percent of the participants are male and 61% were age 18-19, and the remainder were age 20-22. More than half (57%) of the participants were born in the Dominican Republic, a percentage that is not representative of the United States as a whole, as only 3% of the Hispanic population of the United States is of Dominican origin (65% are of Mexican origin) (Motel & Patten, 2012). However, Dominicans account for 46% of the New York City Hispanic Population, whereas Mexicans only account 23% and Ecuadorians for 16% (Lobo & Salvo, 2013). This may

---

12 “Hispanic origin is based on self-described family ancestry or place of birth in response to a question on the Census Bureau’s American Community Survey. Ancestry is not necessarily the same as the place of birth of the respondent, nor is it indicative of immigrant or citizenship status. For example a U.S. citizen born in Los Angeles of Mexican immigrant parents or grandparents may (or may not) identify his or her Hispanic origin as Mexico. Likewise, some immigrants born in Mexico may identify another country as their origin depending on the place of birth of their ancestors.” (Motel & Patten, 2012)
explain why the dialect of Spanish found in New York City has noticeable influence of Dominican Spanish (García, Evangelista, Martínez, Disla, & Paulino, 1988; Otheguy, Zentella, & Livert, 2007a). Basic demographic information for the participants in this study is detailed in Table 4.

Table 4 – Basic Demographic Information

<table>
<thead>
<tr>
<th>Country of birth</th>
<th>Total</th>
<th>Male/Female</th>
<th>18-19/20-22</th>
</tr>
</thead>
<tbody>
<tr>
<td>Costa Rica</td>
<td>1</td>
<td>0/1</td>
<td>0/1</td>
</tr>
<tr>
<td>Colombia</td>
<td>4</td>
<td>3/1</td>
<td>2/2</td>
</tr>
<tr>
<td>Dominican Republic</td>
<td>57</td>
<td>31/26</td>
<td>33/24</td>
</tr>
<tr>
<td>Ecuador</td>
<td>12</td>
<td>6/6</td>
<td>8/4</td>
</tr>
<tr>
<td>El Salvador</td>
<td>5</td>
<td>3/2</td>
<td>4/1</td>
</tr>
<tr>
<td>Guatemala</td>
<td>4</td>
<td>2/2</td>
<td>2/2</td>
</tr>
<tr>
<td>Honduras</td>
<td>4</td>
<td>3/1</td>
<td>4/0</td>
</tr>
<tr>
<td>Mexico</td>
<td>12</td>
<td>8/4</td>
<td>7/5</td>
</tr>
<tr>
<td>United States</td>
<td>1</td>
<td>0/1</td>
<td>1/0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>100</td>
<td>57/43</td>
<td>61/39</td>
</tr>
</tbody>
</table>

3.1.3 Survey Results: Language

Participants report that they use Spanish more than any other language; they mostly use Spanish for both face-to-face conversations and on digital platforms for everyone except for their teachers and bosses. With their teachers, participants use a mixture of Spanish and English, but with their bosses, participants use English (both face-to-face and digitally). Overwhelmingly, the only person participants must communicate with in English is their boss. For everyone else, they use Spanish or a mixture of Spanish and English. Chart 3.1 shows the percentage of participants who usually use each language or language combination to communicate with groups of people in their lives in a face-
to-face setting. Figure 1 shows the percentage of participants who *usually* use each language or language combination to communicate with groups of people in their lives on *digital* platforms. These two tables combined show that participants almost exclusively prefer communicating in Spanish with family members (siblings, parents, and grandparents). Since Spanish is the first language of all the participants, and spoken at home, the finding that participants prefer to speak in Spanish to family members is completely mundane. However, since these participants are Spanish-dominant, it would be expected that they would prefer to speak in Spanish with everyone, but that is not the finding. With people outside of the family (friends, significant others, and teachers), participants use more Spanish and English combined, showing that their speech community outside of their house is actually a bilingual speech community, more so than an English or Spanish-speaking one.
Figure 1 – Face-to-Face Language Preferences
Figure 3 represents *how often* participants use digital tools for communication with different groups of people in their lives. This graph illustrates that participants use digital communication platforms (text messaging, chatting, FaceTime, etc) with family members (siblings, parents, grandparents) less frequently than with other groups. At first glance, this appears to be an effect of age i.e., parents and grandparents are less likely to communicate via digital platforms since they are at least 15 years older than the participants. However, that explanation does not account for siblings, who can be younger or older than participants. The second explanation is that participants use digital and non-digital divide as a way to mediate context collapse. Context collapse is a phenomenon where social worlds that may have been separate before digital communication platforms (i.e., family, friends, work) are now converging and overlapping (Boyd, 2014; Harrison &
Gilmore, 2012; Marwick & Boyd, 2011; Vitak, 2012). Participants may be controlling their social worlds by differentiating how they communicate with them. The context collapse explanation is a common finding in research on how youth manage social interactions in a digital world (Boyd, 2014; Harrison & Gilmore, 2012; Marwick & Boyd, 2011; Vitak, 2012), though most researchers focus on differentiation between different platforms (i.e., Twitter versus Facebook, Instagram versus SnapChat, etc.) rather than purely digital versus face-to-face communication as is illustrated here.

Figure 3 – Frequency of Communication on a Digital Platform

The most preferred method of communication for all groups is in person. The second most preferred for all groups except grandparents is text or chat. For grandparents, if face-to-face
communication is not an option, participants would prefer to call on the phone. The only group that participants prefer to use FaceTime with are significant others and siblings. FaceTime is a video service that can be accessed from an iPhone or Macintosh computer. It is very similar to Skype or other video messaging services. When asked why they only use FaceTime with a small group of people, they said that using it with anyone else is too intimate – that FaceTime (more than Skype) is a private space that they reserve for the people closest to them. This seems to be one way to manage context collapse – to reserve certain technologies for the people closest to them.

3.1.4 Survey Results: Years of Schooling

Fifty-eight participants responded to the questions about previous schooling experience. On average, participants had 2.3 years of school in the United States (SD = 1.59), and 9.5 years of school in their home country (SD = 2.8). Some participants had been in school in the United States for up to seven years. However, many participants are newcomers and the P2G classroom is their first academic setting in the U.S.

3.2 Measures - Acceptability Judgment Task (AJT)

The Acceptability Judgment Task (AJT) is an assessment of both social and txt language knowledge in a written modality. The assessment consisted of two parts (the complete assessment is in Appendix 2). In the first part, participants made a list of the abbreviations and acronyms that they use. This was in a competitive, group-work setting to encourage discussion and quantity. This portion was unguided, and participants were free to include any abbreviation (including those that have been established long before texting, i.e., “NYC”). For every abbreviation, participants had to provide a translation or use it in a sentence. The group who produced the most items won a bag of candy.
In the second part, participants completed the Acceptability Judgment Task; they evaluated a series of sentences containing both textisms and colloquialisms. Participants marked each sentence as either “good,” “bad,” or “I don’t know.” There are fifty-seven questions in total; thirty-eight questions evaluating textisms, fourteen evaluating colloquialisms, and five questions where participants were asked to match non-linguistic features commonly found in text messages to subtextual meaning being conveyed. The example participants were shown for training is in Figure 4.

Figure 4 – Example of Acceptability Judgment Task

<table>
<thead>
<tr>
<th></th>
<th>Good</th>
<th>Bad</th>
<th>I don't know</th>
</tr>
</thead>
<tbody>
<tr>
<td>ex what do u wanna do tonight?</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>ex who do u wanna win the election?</td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

Fourteen of the digital items are in Spanish (37%), and twenty-four are in English (63%). All of the social items are in English, and all of the items in the third type of question are in English, though they were testing non-linguistic features of the language. For example, in the textisms section, participants had to decide if “OFICIALMENTE hoy es viernes de #$muchacosabuena” (Officially, today is Friday of #$manygoodthings) is ‘good’ or ‘bad’. The item being texted here is the placement of ‘#muchacosabuena’, a very common hashtag on twitter. This hashtag is only acceptable in an utterance final position, so while that sentence is acceptable, or ‘good’, “#muchacosabuena De eso se trata esto, siempre sonreír” (#manygoodthings that's what this is about, always smiling) is unacceptable, or ‘bad.’ An example of a colloquialisms item is “haha he chirped you”, which is unacceptable or ‘bad’ whereas “haha you got chirped” is acceptable, or “good.” “Chirped” is a slang term to suggest
being insulted; it is an unaccusative\(^{13}\) verb, and therefore the act of chirping is agent-less unless a by-
phrase is inserted to increase the valency (for example, “you got chirped by a child”). Finally, the
para-linguistic questions were assessed using the format in Figure 5. Participants were presented with
a text message and a series of responses. They were then instructed to match the response to the
emotion. In the example below, both punctuation and capitalization are being tested for their non-
verbal connotations. For example, if someone responds to a message with “O.K.” they are probably
more angry or upset than if they had responded with “ok.”

Figure 5 – Paralinguistic Features

<table>
<thead>
<tr>
<th>Text message</th>
<th>Response</th>
<th>Feeling</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Broccoli for dinner.”</td>
<td>yum!</td>
<td>***happy</td>
</tr>
<tr>
<td></td>
<td>😊</td>
<td>***don’t care</td>
</tr>
<tr>
<td></td>
<td>fine.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>yuck</td>
<td>other: <em>unhappy</em></td>
</tr>
</tbody>
</table>

\(^{13}\) Unaccusative verbs are agentless, like passive verbs, there is a receiver of the action, but no one doing the action
We answered the first question together to help participants understand that the task was not “true” or “false,” but rather if this sentence “sounds okay” to write on text message or chat, it is “good,” but if it sounds awkward, “like your mom trying to be cool or someone who doesn’t know English or Spanish,” then it is “bad.” If they had never heard the term or did not have a sense of how it sounds, they were instructed to mark “I don’t know.” Participants took approximately 30 minutes to complete the task. This was their favorite task, and they wanted to explain their answers. We spent time after this task going over the lists they had made and explaining what the various acronyms and abbreviations were. During this time, participants were very excited to learn the new abbreviations they could use.

The variety of social language that these participants use is not the “standard” variety and therefore their individual language use should not be compared to more mainstream varieties. The results from the AJT will be used to determine overall patterns in the dialect that participants use. It is this composite picture that can be compared to the “standard” variety of Spanish, English, or Spanish/English bilingualism. Individual correlations for this component will not be calculated, and individual result will not be reported as the individual language that any student uses is distinct from the broader, standardized dialects, and it is not representative to compare one individual’s language against a group’s linguistic tendencies. The comparisons and analysis that follow are made between varieties, not between individuals. This distinction is crucial as it is unfair and incomplete to compare individual skills to the “standard” variety if that is not the variety of a language that an individual speaks. The following results are meant to compare how similar the variety used by this population is to the variety that is most present on platforms such as twitter.

14 Standard will be written in quotations throughout to indicate that standard is a relative term and there is no such thing as a “standard” dialect. However, there is the most common dialect and that is what is meant by “standard” in this section.
3.2.1 AJT – “Standard”

The “standard” for all of the items of the AJT is derived from usage patterns on twitter. Items were chosen based on their appearance on popular media lists of teen textisms or colloquialisms (K. D. Anderson, 2015; Benjamin, 2015; Centeno Millan, 2015; Dobbins, 2015; Erichsen, 2014; “Urban Dictionary: Textism,” 2015). For the acceptable (good) examples, the item was considered acceptable in that linguistic environment if 20 instances were found where the item was used in one linguistic context on twitter over the course of 24 hours. Twitter was chosen because it is the largest source of publically accessible and timely text that is also written in the txt form. Twenty is an arbitrary number but was considered sufficient for this study to indicate popular and widespread usage of a term. For the unacceptable (bad) examples, if 50 instances were found of the item being used on popular media sites (i.e., twitter, blogs, urban dictionary, etc.), but the item was never used in the way being tested, it was considered sufficient to indicate that the usage type was unacceptable. These items also took into account basic linguistic principles such as word category\textsuperscript{15} and syntactic environments.

3.2.1 AJT - Piloting and Amendment

The AJT was piloted on eighteen monolingual and bilingual PhD participants at the CUNY Graduate Center, four monolingual English-speaking teenagers age 15-19, and six other college educated adults age 25-27. Items were clarified and adapted based on piloting feedback. Participants in the study found the third section (matching) to be confusing. This confusion may have produced errors in the data that was reported.

\textsuperscript{15} For example, “bae” is a term of endearment for a significant other or a person someone wants to be a significant other. It stands for “before anyone else” or “beyond anyone else.” While it initially appears to be a count noun, it is currently a proper noun since it can never be combined with a possessive. For example, “my bae cooked dinner” is unacceptable. However, “bae’s cookies are the best” is acceptable.
3.2.2 AJT - Results

Fifty participants took the AJT (21 Female). For the digital colloquialisms, “do not know” was the most common answer for 21% of the items (of that 21%, 88% of the items are English, and 12% are Spanish). Overall, “do not know” was the most common answer for 9% of the Spanish items, and 27% of the English items. For the social colloquialisms (all of which are in English), “do not know” was most common for 50% of the items. Finally, for the matching component, participants did not answer 27% of the questions.

There are a variety of factors that likely contribute to participants reporting that they do not know the items. The most plausible is that the discrepancy in “do not know” responses between the digital and social items is due to language. This entire effect can be explained by the fact that the participants in this study are Spanish-dominant. Since 37% of the digital items are in Spanish, and none of the social items are in Spanish, this discrepancy could account for the entire difference in colloquial versus textisms awareness. Another possible factor contributing to the discrepancy between the social and digital could be the different types of language they are exposed to. The survey uncovered that all of the participants are Spanish dominant and most of their exposure to English is on the Internet. The vast majority of their face-to-face interactions are in Spanish, however. Participants are therefore more likely to be exposed to very few English spoken colloquialisms, but many more English textisms.

Chart 3.4 illustrates how participants performed overall as matched to the “standard” as found on twitter. For each item, if the most common response was “do not know”, the second most common response (“good” or “bad”) was used. The N’s reported in this table refer to how many questions were of each type, not to how many participants answered the question.
The first column in Figure 6 represents how participants performed overall on the textisms section of the AJT as matched with the “standard” variety; if participant judgments of textisms were exactly the same as the “standard” dialect, they would have a 100% match. Overall, they match the “standard” judgments on 76% of the textism items. The next three columns represent textisms by language. Participants match the “standard” 100% for the Spanish items, they match the standard on 75% of the bilingual items (for which the context is English and the textisms is Spanish), and 65% of the English items. This reinforces that participants are Spanish-dominant, and the variety of Spanish that they use on digital platforms is more similar to the Spanish txt “standard” than the English txt “standard”. The final column represents how closely participants matched the social items. Again, these items predominantly exist in a face-to-face setting, and are associated with colloquial speech. Participants diverged the most from the “standard” on these items; the similarity of their judgments to the English colloquial “standard” is only slightly above chance, suggesting that they may not have judged any of them the same as a speaker of “standard” colloquial English.
Overall, the results from the AJT and participant judgments as compared to monolingual and bilingual “standards” suggest that participants use a very similar to “standard” variety of Spanish and bilingual txt, but their judgments do not align with the English monolingual “standard” on either the txt language form or the colloquial language form. This population of participants is Spanish dominant, and use more Spanish than English, even when communicating on a digital platform. This is reinforced by the fact that they align more closely to the Spanish “standard” than the English one. These results do suggest that English may be more susceptible to contact influence than Spanish is in this setting. If participants are using both Spanish and English (and they report that they do use both even though they are Spanish-dominant), yet their Spanish judgments match the “standard” more closely that their English judgments, they are either using a non-standard version of the language and will adapt their language to the more mainstream variety. Or, they will continue to use the “non-standard” variety as a new variety of English. As mentioned before, the participants in this study mostly know each other. It is therefore entirely possible that the variety of English that participants speak is a new variety for this community of speakers, and it is simply a variety that does not match the monolingual English “standard.”

Additionally, the difference in closeness to “standard” of textisms versus colloquialisms may be illuminating. This discussion should be treated as preliminary as the difference in these results is not significant and could therefore be due to chance\(^\text{16}\). As reported in Section 3.1.2, participants prefer to communicate face-to-face with family members, and primarily use Spanish within this group. Language change happens most often when people have contact outside of their family, not within it, so it would be unexpected for participants to be developing fluency in new English colloquialisms when they are speaking Spanish with their family members. However, participants

\(^{16}\) These measures should be repeated with a greater number of participants.
also reported (Section 3.1.2) that they prefer digital communication for their friends, significant others, and teachers. These are also the groups from whom they will acquire the more new and modern vocabulary since they are themselves members of the larger speech community (Gumperz, 2009). Ultimately, the finding that participants use more English and use more digital communication platforms with non-family members may account for the discrepancy in how closely participants match with “standard” English colloquialisms versus textisms.

3.3 Measures - Aural Comprehension Task

Participants’ academic aural comprehension skills were assessed through a video task and open-ended questions; this task was originally designed by the SLA Lab at the CUNY Graduate Center to assess the academic listening skills of LTELs. Participants watched and listened to two videos (both in English) and then completed a written assessment of what they learned and remembered. The first video featured Craig Saffoe from the Smithsonian National Museum discussing The Big Cats Exhibit. The content of this video is academic in nature, and is in an academic spoken register. Participants watched the video three times and then answered written questions in English. The video was at a sixth grade level, and ran for approximately 3 minutes, 23 seconds. Participants required about 10 minutes to complete the answers after the third viewing. The second video was produced by the New York Times and Creattruecast.org about Tyrian Purple Dye and the role that snails had in making it before the advent of synthetic dyes. This video is much more abstract and complex than the first video. This video is at approximately an 8th grade level and ran for two minutes and fourteen second. Participants also viewed it three times and required roughly ten minutes to complete the answers.

One problem with this assessment is that participants are expected to be able to read the questions and understand them before responding. Since understanding written prose is
independent from understanding speech, the questions were discussed beforehand to check for comprehension and clarify any confusion. We then identified key vocabulary used in the question and translated it into Spanish for any participants who may not understand it in English. Finally, participants were permitted to ask for clarification for any words they did not understand since the assessment was targeting aural compression, not academic vocabulary knowledge.

This component is a part of the larger LTEL diagnostic. The LTEL diagnostic had not been piloted at the time of writing, so this component was validated with the same group of undergraduate participants from Hunter and Lehman Colleges. Unfortunately, piloting on this group did not allow for calibration of the difficulty, only to be sure that the questions were understandable and clear.

3.3.1 Aural Comprehension Task - Evaluation

The evaluation was similar to an open ended academic assessment. Each question is worth two points. Participants get one point for writing the correct answer, and one point for writing an answer that is partially correct or on topic, but not the actual answer. This schema was determined for the LTEL diagnostic, and is justified based on the goals of the tool: to determine participants’ ability to understand academic speech (not necessarily for them to recall every detail). Therefore, an answer that is somewhat related indicates that they partially understand the video.

The scores for the Craig Saffoe video were calculated. There are four questions for a total of 8 possible points, plus one open-ended question that simply asks “Do you remember anything else from the video?” Participants could earn an extra point from this question if they recalled something accurately from the video. Completeness of answers was not considered. Table 4 shows how points are distributed.
Table 5 – Points on Aural Comprehension Task (Craig Saffoe)

<table>
<thead>
<tr>
<th>Points on Craig Saffoe Video</th>
<th>Grade Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>N/A</td>
</tr>
<tr>
<td>1-2</td>
<td>3rd</td>
</tr>
<tr>
<td>3-4</td>
<td>4th</td>
</tr>
<tr>
<td>5-6</td>
<td>5th</td>
</tr>
<tr>
<td>7-8</td>
<td>6th</td>
</tr>
</tbody>
</table>

One class did not view the Tyrian Purple video because they struggled too much with the Craig Saffoe video and it did not seem fair or appropriate to present them with even more difficult content. Therefore these participants’ scores only reflect the Craig Saffoe video. The other classes had the opportunity to watch the Tyrian Purple video. The highest score from the two videos was considered to be their grade level. Table 5 shows how points are distributed for the Tyrian Purple video.

Table 6 – Table 3.4 - Points on Aural Comprehension Task (Tyrian Purple)

<table>
<thead>
<tr>
<th>Points on Tyrian Purple Video</th>
<th>Grade Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>N/A</td>
</tr>
<tr>
<td>1-2</td>
<td>6th</td>
</tr>
<tr>
<td>3-4</td>
<td>7th</td>
</tr>
<tr>
<td>5-6</td>
<td>8th</td>
</tr>
<tr>
<td>7-8</td>
<td>9th</td>
</tr>
</tbody>
</table>
3.3.2 Aural Comprehension Task – Results

Twenty-seven participants completed the aural comprehension task. The scores ranged from 3rd grade comprehension level to 9th grade comprehension level. The average aural grade level was 5.2 (SD = 1.64). Many participants could not recall anything from the Tyrian Purple video. Figure 7 shows a breakdown of student results.

Figure 7 – Aural Comprehension Grade Level

3.4 Measures - Chatting Game

The chatting game involved participants using the Google chat interface to complete a word-guessing game. Laptop computers were provided to each student by the respective schools, and Google identities were established for each classroom. There were 24 identities created since that
was the largest class. Each computer was logged into the Google account, and a chat (also called a hangout) was started before participants arrived. Participants sat down at the computers and were put into pairs, and these pairs were randomly chosen based on the computer they were sitting at. For example, Computer_1 and Computer_2 formed a pair. Each student received a word list of academic words. There were two separate word lists, List_A and List_B. These words were chosen based on sixth, seventh, and eighth grade vocabulary items in both Spanish and English. Half of the words were in Spanish and the other half in English. Participants were free to pick words from either language. Directions were given in both Spanish and English to ensure understanding.

Participants in pairs each received a different list (i.e., the student at Computer_1 got List_A, and the student at Computer_2 got List_B). These lists can be found in Appendix 3. The task was to describe the word in enough detail that the other person could guess the word. The team who guessed the most words won a bag of candy.

The chatting game was a novel activity designed specifically for this study. It was based on common word-guessing activities found in ESL classrooms, yet relied on Google chat as an interface (Ferris & Tagg, 1996). The vocabulary came from lists provided by the NYCDOE and the Common Core State Standards for grade-level. The interface and design was validated with bilinguals and monolingual adults age 20-38 using the English vocabulary items since many of the pilot study members did not speak Spanish. The results of the piloting showed that this was a particularly challenging interface to work with. Pilot participants struggled with doing academic work on a digital interface. The pilot participants said that they often used chat for social purposes, but rarely if ever tried to explain academic ideas via the interface.
3.4.1 Chatting Game - Evaluation

The Chatting Game activity was assessed based on how many words participants guessed and how complex or descriptive the language used was. The grammatical form of student response was assessed using the form aspect from the 3rd to 9th grade writing rubric from the assessment for LTELs, taking into account the unique features of the txt language register which allows for abbreviations, creativity, reduced punctuation (see Appendix 4). Taking these factors into account, each chat was treated as a very short writing sample, and a grade level assigned that corresponded to the number of words, complexity of the descriptions and complexity of the language, as described in the LTEL diagnostic rubrics.

3.4.2 Chatting Game - Results

Sixteen participants participated in the chatting game. There were two primary reasons for this low number. First, eight participants did not participate because they refused to do so. They spent that class period searching the internet for pictures and music. Secondly, there was very low turnout the whole week that the chatting game was being administered largely because it was the first week of nice weather in New York City, so many participants chose not to come to school that week. The average grade level was 5.69 (SD = 1.99). Figure 8 shows the breakdown of the results.
3.5 Measures - English Literacy Evaluation for Newcomer SIFE (LENS)

The English LENS is a diagnostic test of academic, written English designed and written in the SLA Lab at the Graduate Center at CUNY. Participants were given a paper and pencil version of this test, and were each given four passages between 3rd and 9th grade based on their teachers’ recommendations about their English proficiency. This test was originally designed for Students with Interrupted Formal Education (SIFE) in the NYC public school system. While many of the participants in this study are SIFE, they are not classified as such by the NYCDOE because of the logistics of being an HSE student. Even if they qualify as SIFE, the designation no longer applies since they are beyond the K-12 academic system. The hallmark of this test is that it is designed with
best testing practices in mind and with a focus on scaffolding idiosyncratic language and providing basic visual support. Therefore this test is grade appropriate for all bilingual participants regardless of their SIFE status. The language of this test is very important in that it maintains grade-level appropriateness yet aims to be uncomplicated, following testing and pedagogical guidelines for development of a clear assessment (McMillan, Hellsten, & Klinger, 2007; Soltero-González, Escamilla, & Hopewell, 2012; Stiggins, 1994). The English LENS was piloted and validated by the SLA Lab both with monolingual English speakers as well as with SIFE students.

3.5.1 English LENS - Evaluation

The LENS consists of a reading passage and five questions for each grade level (grades 3-9). Scores were determined by a simple percentage correct. Participants were given a test with three to four grade levels (for example, grades 3, 4, and 5, or grades 6, 7, 8, and 9). The decision of which grade to give which students was based on their teachers’ recommendation. In the event that the teacher did not know their student, the student was asked how they would self-rate their English reading skills. To pass a grade, a student had to answer 80% of the questions correctly, which means answering at least four of the five questions correctly. Participants were assigned a grade level based on the highest grade that they passed. This is similar to the computer adaptive version of this test, which assigns a grade level based on the highest passage a student passes, giving them an opportunity to take at least one grade higher. No student scored 100% on the highest grade.

3.5.2 English LENS - Results

Twenty-nine participants took the English LENS. The average grade level was 5.90 (SD = 1.99). Figure 9 below shows a breakdown of the results.
3.6 Measures - Spanish Literacy Evaluation for Newcomer SIFE (LENS)

The Spanish LENS was administered by the school officials in the traditional digital format. The test follows the same principles as the English language version in focusing on the academic language skills without any unnecessary linguistic confusion. It is not a translation: All of the passages in all of the passages are original and unique, so a student can take both the Spanish and English LENS and see novel material each time. The Spanish LENS is unique in the amount of research that went into it to align with the educational practices and academic levels in the Dominican Republic, Mexico,
The Spanish LENS has been widely tested and validated by the research team at the SLA Lab at the Graduate Center at CUNY. Piloting and item analysis was conducted before the test was made available to the NYCDOE, and the test has been actively used in a wide variety of programs, reinforcing the reliability of it.

3.6.1 Spanish LENS - Evaluation

Nine participants took the computer adaptive version of this test which allows participants to start at 3rd grade and progress in the test, receiving harder or easier passages according to their ability level. The logic behind the test allows participants the greatest opportunity to see the most passages to determine their academic written literacy level in Spanish. At the end of the test, participants receive a grade level that takes both home country curriculum and Common Core State Standards into account. The Spanish LENS allows scores that are in increments of .5. A score of 5.5 means that a participants’ reading comprehension skills are at the second half of 5th grade level.

3.6.2 Spanish LENS - Results

The Spanish LENS was only administered to nine of the study participants. The test was given at one of the two schools, but participants at the other school have not yet taken the Spanish LENS. The administration at the second school is understandably hesitant to have participants take any test that is not the Test Assessing Secondary Completion since students are tested regularly and focused on earning their high school equivalency diploma, not diagnosing their literacy level. At the second school, it was only administered to the nine participants because it was given to one class of the
three. The class it was given to has the highest turnover, and only nine participants remained from the Fall into the Spring.

The average score on the Spanish LENS was 5.22 (SD = 1.06). The highest score was grade level 7.5, meaning that that student performed as though s/he was starting the second semester of 7th grade in the U.S. or Latin America. This is representative of SIFE more broadly as their reading and writing levels tend to be at least two years below grade level (Freeman et al., 2002; Menken, Kleyn, & Chae, 2012); since these participants are in a high school equivalency program and age 18-22, the lowest grade level for participants to be at grade level is 9th grade, and no participants reached 9th grade. This shows that even though these participants are not officially identified as SIFE, their academic performance is similar to SIFE even when assessing their dominant language. Figure 10 shows a breakdown of the results.
3.7 Measures - Test of Adult Basic Education (TABE)

The Test of Adult Basic Education (TABE) is administered by the NYCDOE when participants arrive in the Pathways to Graduation (P2G) GED Program. Items on this test vary widely and are designed to determine what academic content skills participants possess. It is not only a literacy test as it is designed to measure student content knowledge and academic preparedness. Participants took two versions of the TABE: the TABE-Esp, which is the Spanish-language content test, and the TABE-CLAS E, which is the academic English proficiency test. Many participants said that they struggle with the TABE because of the time limitations and they find that the language is confusing (in both Spanish and English). Questions on the TABE range from basic science, to making lists and
reading short passages to performing math calculations. This test is used by the P2G programs to group participants into cohorts/classes and prepare them for the Test Assessing Secondary Completion (TASC). Participants are in cohorts based on their score on this test (their score on this test may diverge greatly from their English proficiency). They take the test every 500 contact hours. This means that if a student attends the program 5 days a week for 6 hours a day, they will take the TABE approximately every 4 months. This works out to about twice a year for most participants. Teachers at the school reported that often after a student has been in P2G for a year or more, their TABE scores begin to drop. I validated this finding over 2 years and found that there is a general tendency for participants to perform worse over time. The teachers assume that this is due to frustration with the test, the lack of feedback given to participants, and the perception that the TABE does not matter. The highest score for each student will be the one considered in this project, and all analyses will be run both with and without the TABE scores included because of the concerns many teachers have.

The TABE Assessment is created by CTB and McGraw-Hill, and endorsed by National Council of Teachers of Math (NTCM), National Council of Teachers of English (NTCE), and International Reading Association (IRA). The scaled score is approximately aligned to grade level (with an error of up to two grades) (“TABE Testing - CTB/McGraw-Hill,” 2015). The TABE has a wide range of questions, including Reading, Writing, Mathematics, and Spelling. The format of these questions is traditional in that they draw on knowledge gained in the test as well as knowledge about the world; this test is meant to assess academic skill in a decontextualized setting.

Participants are given a pre-test that determines if they will take the “Easy”, “Medium”, “Difficult”, and “Advanced” version of the TABE. Participants are then presented with a series of computer-adaptive questions. This test is in Spanish and is designed to measure academic skills in
Spanish, not necessarily reading comprehension, but rather knowledge of grammatical structures, ability to pick out questions from a text, and vocabulary. Participants also take the TABE-CLAS E, which is an English language test to determine English literacy skills. This test resembles an English comprehension test, though some of the language can be confusing even for a native English speaker. One issue with this test is that participants’ scores can be erratic (i.e., one time they score in the 800s and four months later, they score in the 600s on the same test).

3.7.1 TABE - Results

Forty-four participants took the TABE-Esp. The average grade level for the TABE-Esp is 5.2 (SD = 2.63). Student scores are in ranges, so the scores below are rounded to the nearest whole number. The highest possible score is 10th grade. Chart 3.9 shows a breakdown of the results.
Twenty-four participants took the TABE CLAS-E, the average score is 4.71 (SD=4.08). The highest possible score is equivalent to 8th grade. No participants scored at the 8th grade level. Scores are reported in whole numbers rather than ranges. Figure 12 shows a breakdown of the results.
3.9 Measures - Overall Academic Score

A composite academic score was calculated for each student. This score reflects participants’ academic skills more generally – not limited to language or modality, and takes all of the academic assessments into account (English LENS, Spanish LENS, Aural Comprehension Task, Chatting Game, TABE-Esp, and TABE CLAS-E). This score was calculated by taking the simple mean of all the academic scores since all of the scores are aligned to a grade level. Overall academic scores were calculated for fifty-six participants, the average grade level is 5.47 (SD = 1.69). Figure 13 shows a breakdown of the results, and Table 7 shows the average scores for all of the assessments.
Figure 13 – Overall Grade Level

Table 7 – Average scores for all assessments

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Mean, SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aural Comprehension Task</td>
<td>5.20, 1.64</td>
<td>27</td>
</tr>
<tr>
<td>Chatting Game</td>
<td>5.69, 1.99</td>
<td>16</td>
</tr>
<tr>
<td>English LENS</td>
<td>5.90, 1.99</td>
<td>29</td>
</tr>
<tr>
<td>Spanish LENS</td>
<td>5.22, 1.06</td>
<td>9</td>
</tr>
<tr>
<td>TABE-Esp</td>
<td>5.20, 2.63</td>
<td>44</td>
</tr>
<tr>
<td>TABE-CLAS E</td>
<td>4.71, 4.08</td>
<td>24</td>
</tr>
<tr>
<td>Overall</td>
<td>5.47, 1.69</td>
<td>56</td>
</tr>
</tbody>
</table>
Students performed the best on the English LENS, which is a written test of academic literacy. As mentioned before, the LENS tests were written by researchers in the Second Language Acquisition Lab, following best practices of test design and assessment. This is a diagnostic literacy test, intended to give students the opportunity to show what literacy skills they possess and highlight for teachers what skills they need to develop. The test itself has been piloted on many groups of students, and undergone extensive revisions. The fact that students performed better on this assessment than any other may be due to the fact that this test is designed for students who have had large breaks in their education or have been underschooled in other ways: namely, this population of students. This is taken to show that students perform better when assessment is contextualized and written in clear yet academic language. Secondly, the LENS is assessing academic literacy in a written format, which is the most common pairing of language form and modality, so there may be an effect of familiarity as well.

3.10 Relationship between Measures

Overall, the relationships between different types of language competencies and academic as well as linguistic experiences can potentially identify how individual skills are related. Students who performed well on the Chatting activity (average grade level is 5.2 (SD = 1.64)) also performed well on the Aural Comprehension activity (average grade level is 5.7 (SD = 1.99)); there is a positive correlation between the Chatting activity and the Aural Comprehension activity $r(13) = .56, p < .05$. These tasks (chatting game and aural comprehension) have in common that production of language was up to the student, and they were free to use any mixture of languages to express the ideas. The data from these two assessments were collected from students in an informal written register. So, if a student is able to comprehend the material and is better able to express himself when using both languages, he would have performed well on both of these tasks.
The Aural Comprehension activity and the LENS both assess receptive skills whereas the Chatting activity is exclusively productive. Therefore, if a student is stronger at receptive skills than productive, she will likely perform better on the LENS and the Aural Comprehension Tasks. However, if a student struggles with formal grammar or formal writing, but she is able to illustrate her understanding in an informal way, then she will likely perform better on the Aural Comprehension Task and the Chatting Game. There has been extensive research on the relationship between academic language skills and non-academic language skills (Bardovi-Harlig & Hartford, 1993; Cummins, 1999; Ferris & Tagg, 1996; Klein & Martohardjono, 2006; Scarcella, 2003) showing that very often students acquire the ability to communicate in the language before they acquire academic language skills.

Because home language education is one of the most significant factors in determining academic success in a second language (Collier, 1989; Dressler & Kamil, 2006; Klein & Martohardjono, 2006), it was expected that students in this study who had the most home country schooling would perform better overall on tests of academic literacy. This hypothesis was not born out in the findings. There are a variety of reasons why this might be the case. One explanation is that students who had the most years of home country (and home language) education also had the fewest years in an academic institution in the United States. This could have an effect on their academic language proficiency since they have had fewer experiences with American style testing and fewer opportunities to practice their academic English language skills in an American school system. Secondly, it is not clear from the data if students went to school consistently before entering the P2G Program. If there were significant gaps in their education, then they are likely to perform worse on all tests of academic literacy, even if overall, they attended more years of school.
3.11 Measures Conclusion

Overall, the survey, background information, and results from the academic and social diagnostics show that students’ academic skill level is similar to SIFE Students and other students who have had non-traditional schooling experiences. The demographic information shows that most students are recent immigrants from the Dominican Republic who entered the American school system after age 17. The overall academic literacy level is at grade level 5.47 (SD=1.69), putting all students more than two grade levels below their age-grade. Even though the participants in this study are not all considered SIFE (Students with Interrupted Formal Education), they are similar in the academic under-preparedness. For students who moved frequently or had breaks in their education, this finding is unsurprising. However, for students who have been in school consistently for twelve to fourteen years, it is clear that the schooling was not effective. The reasons that it was not effective are far beyond this dissertation, but it is important to note that many students are in this situation where even after spending most of their childhood in school, they are academically underprepared for age-grade level academic work. In terms of social language skills, students perform nearly identically to Latin American Spanish speakers in tests of colloquial language skills. They communicate primarily in Spanish or a mixture with nearly everyone in their life, and prefer face-to-face and texted communication over phone calls and Face Time, except for specific groups of people (i.e., FaceTime for significant others and phone calls or Skype for grandparents).

4 Conclusion

The results from the AJT showed that participants matched other Spanish-speaking Latinos on measures of colloquial awareness, but diverged from bilinguals, and severely diverged from English monolinguals, matching their judgments on only 50% of items. This shows that while participants are bilingual between Spanish and English in a dual language program, they participate in social life
as Spanish speakers, reinforced by the fact that most of their conversational partners, both face-to-face and digitally are Spanish speakers. They also prefer text messaging over other mediated forms of communication such as phone calls, Skype, emails, etc. The first preference for most participants, however, was face-to-face conversations.

Furthermore, it was discovered in this chapter that participants are at approximately a 5.5 grade level, which is far below their age-grade, indicating that they likely had non-traditional schooling experiences. This hypothesis was born out as students indicated that they had often had gaps in schooling while traveling between countries. The correlation between previous schooling experiences and academic literacy skills was not found to be significant, indicating that there may be other factors coming in to play that are affecting how students perform on diagnostics of academic literacy. It was found that participants who perform well on the Chatting Game also performed well on the Aural Comprehension task; these tasks have in common that students were free to produce their answers in any language, and in any form of the language that was most effective for them. Therefore, participants were being tested only on their ability to comprehend academic ideas and academic communication, not on their ability to communicate in it. This is significant because it suggests students may be able to understand more academic language than they are able to produce. Finally, the results of the quantitative measures give background information about students’ language and educational experiences, as well as their academic, social, and texting language skills, painting a detailed picture of the literacy skills this population of students possesses.
Chapter 3 – Bilingual Txt

1 Introduction

The language used on digital communication technologies is influenced by the cultural and technological moment it is written in. It is a language form full of adaptation and innovation, but most of the innovative features disappear as quickly as they appear. For example, in the early 2000’s, texting behavior included significant amounts of grapho-phonemic representations (‘cu@home’, ‘l8r’) (Crystal, 2009), yet there was not a single word of this type in this corpus and it is rare in other recent corpora (since 2012). What happened in that time likely has to do with the rise of smartphones and corresponding availability of QWERTY keyboards. As the technology moved away from multi-press systems, it no longer made sense to abbreviate so heavily and it was no longer fashionable to use those kinds of abbreviations. Language changes, and it changes relatively quickly in terms of human history. Add to those changes extremely rapid advances in communication technology over the past twenty years, and it is the perfect combination of factors to create a highly unstable language situation that changes rapidly — from the perspective of anyone born before 1996, that is.

The participants in this study were born after 1996, and are therefore unable to remember a time before the Internet. They are the first generation to be raised with txt as a feature of their language, and they are the first generation to have always considered it normal to communicate via

---

17 “qwerty” refers to a traditional computer keyboard because it is the first 6 letters in the upper left hand corner. Not all languages use the same keyboard, but English and Spanish use almost the same. The difference is that Spanish has an ñ whereas English does not.
18 As an example of this, ask five different native English speakers what the past tense of “writhe” is and see what they say; they are likely to give at least two different answers. “Writhe” is just one example of a word in a state of change at this historical moment.
19 1996 is generally considered the year the Internet became a reality, and the year that individuals could have access to the internet at places like libraries, schools, and even at home (Gromov, 1998; Hauben & Hauben, 1998). Some participants were alive at this point, but no one in the study was older than 3 years old.
texting. This provides a unique and time-sensitive opportunity for linguists to document the emergence of a language as it is happening. Researchers across the world are beginning to document the language of texting among monolingual populations (Accorsi et al., 2014; Akbari, 2013; Crystal, 2014; Heidari & Alibabaee, 2013; Ling, 2005; Ling & Baron, 2007b; Tagg, 2009). This chapter joins the smaller group of researchers looking at bilingual communication on digital platforms (Jannis Androutsopoulos, 2006, 2006; Bensoussan et al., 2006; Carrier & Benitez, 2010; Paolillo, 2011).

Bilingual text messages and digital communication is unique because there are two types of intersections occurring: One between the languages, and another between language and digital technologies (Figure 14).

Figure 14 – Intersections

This chapter seeks to document the end result of these two intersections as it develops while simultaneously making connections to broader linguistic principles. Intersection 1, between Spanish and English, has been occurring since people have been moving around the world. This, however, does not imply that language change as a result of this intersection has come to a halt. Much to the contrary, new dialects emerge both in response to global migration patterns (i.e., Dominicans in New York City (Otheguy & Zentella, 2012), and in response to technological innovation (mostly...
resulting in the spread of mega-languages such as English (Sonntag, 2003). These intersections are continually affecting each other, and the relationship between Spanish and English in bilinguals is being influenced by the medium that language is used on. Intersection 2, however, is where novel linguistic features are emerging the most quickly as digital communication technologies rapidly advance. This chapter, therefore, focuses on the results of Intersection 2, taking both a descriptive and theoretical approach.

There have been many significant investigations into the language of text messaging (Accorsi et al., 2014; Akbari, 2013; Bernicot et al., 2012; Carrier & Benitez, 2010; Deumert & Masinyana, 2008; Ling, 2005; Tagg, 2009). Tagg’s (2009) dissertation is among the most comprehensive. She analyzed four primary areas of text messaging innovation: respellings, creativity, word forms, and grammatical structure. Because her text messages came from speakers of a standard and well documented variety of English (London British English), she was able to make comparisons between both the spoken dialect and the txt dialect, showing conclusively through careful investigation into these measures that txt is a distinct language form from both spoken and written language and is not on a continuum between the two (Tagg, 2009). Because the participants in the present study speak a non-standard language form that is in an evolutionary state, and they are all bilingual, a comparison of this type is not appropriate for the present study. However, since there are conventions for spelling that are independent of the dialect spoken (at least in the United States), this dissertation will also investigate the types respellings that participants rely on, and what para-linguistic meaning respellings may signify.

Another group of investigations into the language of text messaging have focused on the relationship between texting and speech, with an analysis of the phonological features that arise in texted communication. Thurlow and Brown (2003), Wood et. al (2011), and Crystal (1997, 2009) all situate phonological features of text messages within a larger setting. For Thurlow and Brown,
phonological approximations were significant as markers of stylized speech and identity presentation, and are used to indicate regional dialects (2003). Wood et al. analyze phonological approximations in the context of their relationship to the acquisition of spelling and literacy, and focus on phonological awareness as an indicator of academic language skills (2011). For Crystal (1997) and Baron (2010b), the presence or absence of individual phonological features is indicative of the similarity or dissimilarity to spoken language. Following both Thurlow and Brown (2003) as well as Wood et al. (2011), phonological features are analyzed as stylized speech, especially in their relationship to African American English (AAE)\(^2\), and with relationship to participants’ academic literacy skills. In contrast to Crystal (1997) and Baron (2010b), however, phonological features will not be used to determine similarity to speech. This dissertation assumes the position that speech influences txt and txt influences speech, but that they should not be compared since txt is its own language form, not an adaptation of either speech or writing.

A third group of researchers have investigated language choice on digital platforms. Paolillo (2011) investigated language choice in Indian diaspora chat rooms, and how language was used to construct identity, mark group membership, and discuss particular topics. Barton and Lee (2012) investigated language choice on Flickr, and discovered that many people keep languages separate based on topic and intended audience, and move between languages in order to attract the widest audience possible. Finally, Carrier and Benitez (2010) documented bilingual text in terms of language efficiency, testing Thurlow and Brown’s (2003) hypothesis that efficiency is a driving force in texting, and therefore bilinguals, who have access to two languages should be more efficient than monolinguals who only have access to one. They ultimately found that bilinguals are driven to code

\(^2\) There is a lot of debate about the validity of the term, AAE since African Americans are not the only group who speak the dialect. The perspective adopted here is that AAE is the name of a dialect spoken by many people in New York City. AAE will therefore be used throughout this dissertation as a way to refer to the dialect with particular linguistic features that is spoken by many urban speech communities in the United States.
A final group of researchers have documented the pragmatic and cultural norms associated with the language of texting and chatting. Key areas of research have been to investigate the status of common emotion-icons (emoticons), abbreviations and initialisms such as “lol” (laugh out loud), and “jk” (just kidding) as well as the role that punctuation has in short messages. Through analysis of emoticon use in the context of Austin’s (1975) speech act theory, Dresner and Herring (2010) found that smiling and frowning faces are both markers of illocutionary force, not emotions. They hypothesize that texters use these features to identify the tone of a message. This research is now old enough that emoticons are being replaced with emojis, and I hypothesize that consequentially, “lol” is functioning as the pragmatic marker to indicate the illocutionary force of a message. This dissertation adopts the approach that many abbreviations and stylizations have a pragmatic role by signaling the illocutionary force and facilitating politeness.

This chapter seeks to document bilingual Txt as a language form through a thorough investigation of different types of respellings since they are one of the most salient elements of txt communication. Following Tagg (2009), respellings will be analyzed as abbreviations, initialisms, iterations, phonological influences, and mistakes. Respellings are a key component of txt as they are one of the primary spaces to develop a personal style; they are also one of the most contentiously debated areas as some people fear that the use of respellings will negatively impact the literacy skills of modern youth (Jury, 2010; Maltais, 2012; McWhorter, 2013a). The pragmatic function of each

---

21 Speech act theory takes into account the “force” of an utterance, and suggests that every utterance has three layers of interpretation, first is the locutionary (the literal meaning of the words that are used), illocutionary (the speaker’s intended meaning for the words that are used), and perlocutionary (what the listener interprets from the words that are used) (Austin, 1975).
respelling is explored, using the Gricean approach to analysis as users adhere/flout/violate prescriptive grammar rules. Once identified, respellings will be analyzed by focusing on politeness (P. Brown & Levinson, 1987) and speech act theory (Austin, 1975), with particular attention given to the status of pragmatic markers such as “lol” (P. Brown & Levinson, 1987; Horn, 2010; Levinson, 1983). Throughout this chapter, the primary goal is to document txt as urban bilingual youth have adopted it, and understand how this fits within our broader understanding of human language and human communication. Some of the specific elements analyzed in this chapter will likely disappear as this language forms evolves, making this a historical document at the time of writing.

2 Collection and Preparation of the Text Messages

Of the one hundred participants who took part in the study, forty-four participants donated some or all of the messages from their phones, thirty-four from the School_MN (2 donated WhatsApp messages in lieu of text messages), and ten from School_BX. Twenty-eight participants only donated the text messages and did not complete any other task. The text message collection day was the highest attended day of the workshop since it was the beginning of a new marking period and therefore the first day of class for many participants. Participants chose not to donate for two primary reasons. First, some participants did not use text messages, opting instead for WhatsApp, Kik, or another Over-The-Top messaging service. The second reason was for issues of privacy, many participants did not feel comfortable donating messages from their cell phones and therefore only selected a few messages to donate.
In some cases, there are intact conversations where both participants agreed to donate their messages. For example, TS001[8] and TS008[4] both donated their messages. All of their conversations are intact, but appear twice in the message record. The duplicates were removed and transferred to another file for validation, but the conversation was preserved. Messages received from WhatsApp were formatted to fit into the SMS Model by removing some of the metadata associated with them. They were included in the sample even though they are sent using a different service because students indicated that they use the two services (text messaging and WhatsApp) in the same way for communication with the same social groups. After removing duplicate messages, there are 44,597 messages in the corpus.

There are essentially three parts to this corpus. Of the 44,597 messages, 31,578 of these are one conversation between two people (one male, one female). These participants are in a relationship, and their messages span four months. These messages will be analyzed separately as a case study of an ongoing conversation and will not weigh in to the overall trends presented here. Of the remaining 13,019 messages, 2,664 were from one person. This person (male) had a series of very similar conversations) over the course of eight months. In some instances, it was not possible to obtain two-way consent, so those messages were discarded. In other instances, two-way consent was obtained because the conversation partner was in the school as well. The messages he sent are not included in the main corpus and will be analyzed separately. Another 2,644 were either completely jumbled or donated in sets of less than 80, so also will not be considered in correlations and hypothesis testing since it is not a large enough sample from any given individual. This leaves 7,711 messages in the primary corpus for analysis. Twenty-four individuals donated between 95 and 632 messages each, and these will be used to perform the general analysis.

22 This is one student’s de-identified code. The number in brackets is how many letters are in that student’s name as it may be relevant for average message length or character counts.
The text messages had to undergo three stages of processing in order to extract the relevant information for statistical analysis. The first two stages apply to all 44,597 messages. The third stage only applies to the primary corpus of messages that will be used to test the original hypotheses. First, the messages were de-identified to protect participants’ anonymity. Second, the messages were tagged as Spanish, English, Bilingual, or Other. Finally, the messages were divided into the primary messages for analysis and the secondary group (akin to creating a test set and validation set). Together, these two groups form the corpus that is available for public use, but individually, it is the primary set of messages that are used to form hypotheses in this dissertation, and the secondary group (which consists of one ongoing conversation) is used for validation.

2.1 Text Messages - De-identification

To protect student privacy, names, pseudonyms, email addresses, websites, phone numbers, school names and all other proper nouns had to be removed. Once the median group of messages were identified, a series of search and replace programs using python were created to eliminate identifying information. First the websites and embedded links were removed by removing all of the text following “www” or “http” until a space appeared. These were replaced with [WEB] the WEB indicates that it was formerly a link, thereby preserving that information for future analysis. Secondly, a list of all the words beginning with a capital letter was created and a sub-list of places and names was manually derived from that. I then used a series of compare list functions in Python to remove any identifiers on this list. The first pass on this method was to remove all numbers greater than two digits. Strings of numbers were replaced with [NUMBER, #], the NUMBER indicates that it was a number and the # indicates how many characters were in the number in order to preserve character counts. A list of all of the names of participants in the study (using the previously mentioned list), class rosters, names of anyone associated with their school and the top
100 most common first and last names in Spanish and English was made and these words were replaced with [NAME, #]. This was repeated with a list of major roads, landmarks, and neighborhoods in New York City (taken from the list of subway station names), cities and countries where participants were born, cities and towns in New York, New Jersey, and Philadelphia and replaced any instance of these with [PLACE, #]. Please see the ReadMe file affiliated with the corpus (Appendix 5).

Once the data was de-identified, over 6,000 messages were hand coded as Spanish, English, Both, or Other. No identifying information was found in these messages, so for the purposes of this dissertation (and the text messages will not be publically available until hand validation is complete), this is sufficient. The messages will not be released from the password-protected databases they are currently stored in unless a more thorough de-identification process is undertaken.

2.2 Text Messages – Language Tagging

Management of the text message data set was conducted through DiscoverText. This is a subscription-based, online program to collect, code, and otherwise manage large text-based datasets. I received a one-year subscription for free through a partnership between the creator of DiscoverText and the Oxford Internet Institute. For this project, I used Textifier, which has a data management/coding component, and Sifter, which is a machine learning package that utilizes ActiveLearning to categorize units based on a training set. This tool is available at http://discovertext.com. The first 5,495 texts were identified as English, Spanish, Bilingual, or Other. The classification scheme is in Table 8.
Many of the texts could be considered ambiguous as to what category they belong in. The ambiguity largely arose when abbreviations or other unique-to-txt elements were used. In general, if an abbreviation or initialism came from English or Spanish, that item was treated as belonging to that
language. The only exceptions are for “lol” and “ok.” In a text with only Spanish words, these two items were treated as Spanish. More accurately, in a string of text, they are treated as pragmatic markers that no longer stand for their fully spelled out versions, but rather as features that indicate saving face (Tagliamonte & Denis, 2008). Because of this pragmatic rather than semantic meaning, their use cannot be considered an instance of code switching (and therefore should not be considered using both languages). Even when used on their own (i.e., a text that is only “lol”), they are treated as neither language since in this context, “lol” is most often a response to indicate a positive sentiment rather than an actual message (Tagliamonte & Denis, 2008). Similarly, “ok” was treated as neither language when it stands alone because the item no longer indicates the English expression “okay”\textsuperscript{23}, but rather is a way to indicate a positive response to a previous sentiment.

Once 1/9 of the corpus (5,000+ messages) was manually trained, the Active Learning feature of Sifter was used to determine a preliminary coding of the remaining messages. Then, 1,000 messages were randomly validated to be sure that the machine coding was accurate enough for these purposes. The accuracy was about 89%. The machine learning had the most difficult time with texts that were completely emojis. Because emojis are not Unicode characters, they render as a code of alpha numeric characters with impossible combinations of accented letters and punctuation. The Texifiter program characterized these as 25% from each category since the characters could not be identified by the program. These should more accurately be classified as “Other.” Since the analysis focuses on relative percentages of English, Spanish, and Bilingual texts, this should not be an issue. However, this classification scheme would have to be reconsidered if this corpus is used to study emojis or other non-alpha numeric features.

\textsuperscript{23} It should be noted here that “okay” is now found in an extremely wide variety of languages, so though diachronically, it may be English, it clearly no longer is.
2.2.1 Language Tagging - The Median Group of Messages

Because the corpus of messages to be used for the general analysis is so small (7,711), it was crucial that the messages were hand-tagged to be certain that the language is correct. Since some participants donated as few as ninety-five messages, even five errors per student could significantly alter the results. Therefore this group of messages was hand-validated by a research assistant who re-tagged each message according to language (Spanish, English, Bilingual, Other). This stage was necessary because the computer-trained analysis was only about 80% correct for this group. Figure 12 shows the percentages of each language in the corpus. This table represents the averages for the corpus as a whole.

Figure 15 – Text Message Language Choice
2.3 Correlation between Academic Skills and Text Message Language Choice

One goal of this dissertation is to determine if there is a correlation between participants’ academic literacy skills and their choice of language in text messages. All the variables in this study are continuous, and significance will be reported at the p<.05 level. Pearson correlations were calculated using SPSS to determine if there is a correlation between academic language skills and percentage of English text messages or percentage of Spanish text messages and to determine if further calculations should be conducted.

Correlations were calculated for the English LENS, Spanish LENS, Aural Comprehension Task, Chatting Task, TABE-Esp, TABE CLAS-E, and composite Academic performance, along with percentage of English text messages and percentage of Spanish text messages. Correlations were chosen to identify the relationships between the variables without indicating the causality of the relationships.

As discussed, the number of participants who completed a majority of the tasks is very small. However, it was possible to obtain significance on some of the correlations. There is a very strong positive correlation between the percentage of English-only text messages participants sent and both their English LENS score \( r(11) = .84, p < .01 \), and overall academic proficiency \( r(12) = .78 p < .01 \). This supports the hypothesis that participants who have higher academic English proficiency are more likely to text in English. There was also a very strong correlation between the percentage of English-only text messages participants sent and academic English Aural Comprehension \( r(5) = .96 p < .01 \). It appears that overall communication in English is significantly correlated with the number of text messages students write in English. It is difficult to determine if this is because students have
more practice so have higher English literacy skills or is having higher English literacy skills makes students more willing and able to write text messages in English.

Similarly, there is a strong negative correlation between the percentage of Spanish-only text messages participants sent and both performance on the English LENS $r(11) = -.79 \ p < .01$, and overall academic performance $r(12) = -.72 \ p < .01$. Since there is a positive correlation between the English LENS and the number of English messages sent, and since there are very few messages written in both languages or in neither language, this finding is trivial. However, approaching the analysis from the other perspective may help to understand the relationship between academic literacy and text message language choice. Namely, students may choose to write their text messages in Spanish because they are not as confident in English and therefore less inclined to write their text messages in English.

2.4 Semantic analysis of linguistic features of select messages

The semantic component of this analysis will take lexical density and lexical variety into account. It has been well established that both the lexical density and lexical variety of speech is much lower than that of writing (Halliday, 1989; Johansson, 2009; Ure, 1971). Chafe and Tannen (1987) took a more detailed approach, and found that speech has a higher percentage of pronouns, verbs, adverbs and interjections while writing has more adjectives, passives, and modals (Chafe & Tannen, 1987).

Much of the research on texting communication focuses on the similarities and differences to speech or writing (Baron, 1998, 2010b; Holtgraves & Paul, 2013). Some researchers have argued that texting is more like speaking than writing (Baron, 2010b), and others have taken the approach that it is more similar to writing (Kemp, 2010b). This project takes the approach that texting is a register that shares some features with speech, and some with writing, but is actually a completely different form from either one (Domingo, Jewitt, & Kress, 2014; Forgays et al., 2014; Tagg, 2009).
In the interest of quantifying how bilingual texting is situated with respect to typically spoken and typically written language forms, lexical variety will be calculated. In her 2009 paper, Victoria Johansson shows that lexical variety is fundamentally related to lexical density (Johansson, 2009). The difference being that lexical variety is a measure of how many unique words there are per utterance, which lexical density is a measure of how many content words there are per utterance (omitting all function words). Building on Johansson’s work, lexical variety will be used as a proxy for lexical density. The value of lexical density is that it has been studied extensively for spoken and written language and has been used to categorize digital platforms as well (Crossley, Salsbury, & Mcnamara, 2014; Gonçalves & Sánchez, 2014; Johansson, 2009; Tagg, 2009; Ure, 1971; Vandergriff, 2013).

Speech typically has low lexical density, differentiating it from writing, which tends to be much more content-dense. However, classroom lectures tend to be more content-dense than conversations (Gee, 1998), indicating that the lexical density has more to do with academic language than the medium of delivery (Gee, 1992).

2.4.1 Lexical variety - Evaluation

Commonly respelled words were only counted as one word, so for each word in the table below, the respelled version counts as another instance of that word, not as a unique word. The lexical variety for Example 1 below is 1.08.
Table 9 – Commonly respelled words

<table>
<thead>
<tr>
<th>Common Words</th>
<th>Spelling variations</th>
</tr>
</thead>
<tbody>
<tr>
<td>And/or</td>
<td>‘n/ &amp;</td>
</tr>
<tr>
<td>Aqui</td>
<td>aki</td>
</tr>
<tr>
<td>Are</td>
<td>r</td>
</tr>
<tr>
<td>At</td>
<td>@</td>
</tr>
<tr>
<td>Because</td>
<td>b/c</td>
</tr>
<tr>
<td>Could/should/would</td>
<td>Cld/shld/wld</td>
</tr>
<tr>
<td>Is/are/was/were/be/being/been</td>
<td>Iz/r/waz/wer/b/bng/bin/bn</td>
</tr>
<tr>
<td>Of</td>
<td>O</td>
</tr>
<tr>
<td>Que</td>
<td>q/k</td>
</tr>
<tr>
<td>Quiero</td>
<td>kiero</td>
</tr>
<tr>
<td>So</td>
<td>Soooooo</td>
</tr>
<tr>
<td>To</td>
<td>2, too</td>
</tr>
<tr>
<td>Tomorrow</td>
<td>Tmw/tmrw/2mrw</td>
</tr>
<tr>
<td>Want to/got to/going to</td>
<td>Wanna/gotta/gonna/wana/gota/gona</td>
</tr>
<tr>
<td>Why</td>
<td>Y</td>
</tr>
<tr>
<td>With</td>
<td>w/</td>
</tr>
<tr>
<td>Without</td>
<td>w/o</td>
</tr>
<tr>
<td>Yes</td>
<td>y/yah/yeah/yeh</td>
</tr>
<tr>
<td>You</td>
<td>U</td>
</tr>
<tr>
<td>Your</td>
<td>Yr/ur</td>
</tr>
</tbody>
</table>
Select Interjections

<table>
<thead>
<tr>
<th>Filler Words</th>
<th>Spelling Variations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Haha</td>
<td>Ha, hahaha</td>
</tr>
<tr>
<td>Hm</td>
<td>Hmmm</td>
</tr>
<tr>
<td>Lol</td>
<td></td>
</tr>
<tr>
<td>Oh</td>
<td>Ohhh</td>
</tr>
<tr>
<td>Ok</td>
<td>k/kk</td>
</tr>
<tr>
<td>Ugh</td>
<td></td>
</tr>
<tr>
<td>Uh</td>
<td></td>
</tr>
<tr>
<td>Um</td>
<td>Ummm</td>
</tr>
</tbody>
</table>

Example 1

**Utterance**

Hahaha they're crazy 😂😂

She was laughing so hard

Hahahaha

My father is so funny. I love him hahaha he made me laugh all the time

Do something

**LV**

4/4

5/5

1/0

16/15

2/2

2.4.2 Lexical Variety – Results

Lexical variety was calculated for twenty-six participants. There was a strong negative correlation between lexical variety and percent English $r(22) = -.45, p < .05$, indicating that lexical variety

---

$^{24}$ LV is calculated for the whole language sample, so “haha” only counts one time, regardless how it is spelled.
decreased as percent of English increased. This may be due to the fact that English is a second language for all of these participants, so they may be able to converse more fluently when using Spanish or a mixture of both languages. So while the students with the highest proficiency in academic English may tend to write in English more often, English utterances on the whole were less linguistically complex than Spanish ones. Proficiency in academic English may therefore be deceptive as a measure of overall linguistic skills. This strong negative correlation could also be explained by the fact that participants give more single-word answers in English and tend to explain themselves more completely when texting in Spanish or a mixture of Spanish and English. Finally, it is important to point out that the ability to use English in a text message may not necessarily mean that a student is utilizing more complex language. In fact, the opposite appears to be true, that linguistic complexity is sacrificed for English-only communication. In a system where English is often treated as the most important language, it is important to realize that for many students, there may be linguistic and communicative sacrifices made in order to communicate in English.
3 Respellings

It is a feature of the medium itself that most of the communication is via the written (or re-written) word. Within a given language, there are a variety of ways to express a single idea — using an array of different words and syntactic combinations. Language is situationally and contextually adaptive and one language may have multiple dialects and registers associated with it. Spelling, on the other hand, is an explicitly taught device that children and adults learn after they have mastered speaking a language\textsuperscript{25}. While language is a biologically situated ability, writing is a human-developed technology. Because of this, and because of the prescriptive nature of approved and accepted orthographies\textsuperscript{26}, spelling choices are particularly constrained. It is therefore even more interesting that there are common respelling patterns in this corpus both in Spanish and in English — patterns that indicate that both visual and auditory effects have a role in txt. Finally, analysis of these messages focuses on the written features of text, but text messages are increasingly becoming multimodal by incorporating non-textual features such as pictures and emojis (and emoji pictures, see Figure 16).

Figure 16 – Emoji picture

Every single participant used at least one emoji, indicating that they have become a domesticated and expected feature of text messaging. However, a complete discussion of emojis is far beyond the

\textsuperscript{25} For first language, clearly this does not apply to order of acquisition of a second or additional language.

\textsuperscript{26} I.e., they are always top-down, and established by a centralized body.
scope of this project. Therefore, where relevant, they are included, but will not be referred to or considered in the analysis.

Careful analysis of the patterns of respelling shows that respelled words adhere to a simple underlying set of rules, and violation of these rules appears awkward or disfluent to regular texters. Participants’ reaction to nonce textisms suggests that there is a broader linguistic effect driving acceptability in respellings (Gouskova & Becker, 2013; Kawahara, 2012; Shapiro & Caramazza, 2003; Zangl & Fernald, 2007). This may be a simple case of familiarity, where texters have seen one form repeatedly and therefore expect it to only be spelled that one way. Another explanation is that there is an underlying set of rules to the formulation of respellings that operate across languages and drives language change in an organized manner (J. K. Androutsopoulos, 2000; Marsh & Desberg, 1983; C. Read & Ruyter, 1985; Weber & Henderson, 1989). These rules would be akin to linguistic universals one finds in syntactic or semantic domains.27

It is worth noting that while spelling is a forced choice of this language form, recent technological affordances and trends in communication have given individuals more options for both modality and platform. With over 300 million monthly users, and an estimated 27.6% of the U.S. population on Instagram (Smith, 2015), it is now so prevalent and popular that it demands consideration as a viable communication platform. What makes Instagram unique is the focus on images to tell a story (Hochman & Manovich, 2013; Hu, Manikonda, Kambhampati, & others, 2014). Text is kept to a minimum and users are encouraged to post photos. One participant put it best, stating “on IG (Instagram), it’s like I can let the picture talk, I don’t feel so much pressure to talk about it, I can just show what’s going on.” Many of the participants in this study indicated on their survey that they use Instagram daily, and they reported in their interviews that they have

---

27 A syntactic universal is that all languages mark tense on the verb. A semantic universal is that there is no quantifier “allnon” in any language.
conversations on Instagram. Given that it is almost completely photography based, it may come as a surprise that participants report having conversations.

The popularity of IG is taken as an indicator of a much larger cultural shift towards multi-modal communication forms that are not dependent upon the written, spoken, or signed word at all. When asked about the role that pictures play in communication, the same participant said “Of course! If my friend posts something, I can write back, ask where she was or whatever. It’s weird when you just message somebody for no reason.” This makes sense in the world of a teenager or young adult where social contact is high as compared to adults over 22 who may not have the same level of daily contact. Danah boyd pointed out in her 2014 book (It’s Complicated: The Social Lives of Networked Teens) that teenagers’ and young adults’ world is hyper-social in a way that it has never been before. With the advent of social media, there is an expectation that they will be in constant communication (boyd, 2014; Maltais, 2012). For many teenagers, the pictures on services such as Instagram and the photo attachments serve as an object to communicate about. While these tools and features are poised to influence how teens communicate via text message, they are outside the scope of this project. Since this dissertation focuses exclusively on language, affordances of non-linguistic digital communication technologies cannot be addressed here.

3.1 Respellings - Definitions

Because spelling is a forced modality of txt (in the sense that it is impossible to write a word without spelling it), texters must adhere to, violate, or flout traditional prescriptive spelling rules. The adhering, violating, and flouting vocabulary is taken from Grice’s theory politeness28 (Grice, 1970;

---

28Grice’s theory of politeness has been extremely influential in articulating how people navigate communication. He says that people adhere to, violate, or flout rules of communication. In general, people adhere to the maxims and if they do not, they are flouting them, and trying to communicate without saying something. It is rare for someone to violate a maxim, and generally results in confusion. His four maxims are

1. Quantity, where one tries to be as informative as possible without giving too much detail,
2. Quality, where one tries to be truthful.
Horn, 2010; Levinson, 1983). Flouting a spelling rule will be used to indicate that the texter has changed the spelling with a communicative or stylistic intent. The texter must be aware of the rule to flout it, and cannot be flouting the rule in error. If s/he does violate it in error but there could be communicative intent with another interpretation, it often results in a joke between the interlocutors, as is the case with uncorrected autocorrect mistakes. Violating a rule occurs when a texter does not know or is intentionally ignoring the prescriptive form. This is the fundamental difference between violating and flouting a rule since there is no communicative intent behind the violation of a rule, and the texter need not even be aware the rule exists in order to violate it. Finally, adhering to a rule means using the traditional prescriptive spelling. Most text messages adhere to spelling rules (Accorsi et al., 2014; Tagg, 2009; Wood et al., 2011), indicating that adhering to the rule is the default way texters spell. For a concise rubric, see Table 10.

Table 10 – Three ways to interact with rules

<table>
<thead>
<tr>
<th>Rule Awareness</th>
<th>Non-verbal Communicative Intent</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adhere</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Flout</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Violate</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

3. **Relation**, where one tries to be relevant and
4. **Manner**, when one tries to be clear, brief, and orderly (Grice, 1970; Horn, 2010; Levinson, 1983).

29 A good example of unintentionally flouting spelling conventions is autocorrect mistakes with double meanings. For a hilarious English-only compilation of these, see (Madison, 2011).

30 It is unlikely that texters are simply spelling phonetically because both Spanish and English orthographies are far removed from how words are pronounced yet texters write them the way they are spelled, not how they are pronounced.
While it is often impossible to analyze communicative intent in a corpus of messages, it is possible to identify regular patterns of rule defiance that occur in regular situations. Participants were explicitly asked about a selection of the rules in the acceptability judgment task in order to determine if there are patterns of writing and respelling that carry non-verbal meaning. The results from this task illustrated that participants agree on the implicit meaning behind flouting some spelling conventions.

To analyze respellings based on linguistic features and communicative motivation, five types of respellings are addressed: abbreviations (two sub-types), iterations, initialisms, phonetic influences, and mistakes. An overview of all of the types is in Table 11 and each type that will be discussed in more detail will be revisited in the respective section. Individual items may belong to more than one category, for example, “tkmmmm” for te quiero mucho (I love you a lot) (1) has a clustering (/qu/ → /k/) (2) is an initialism, and (3) is an iteration.

Table 11 – Textism types

<table>
<thead>
<tr>
<th>Type</th>
<th>Definition</th>
<th>Example</th>
<th>Percent of Respellings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iteration</td>
<td>A word or phrase spelled with one or more letters repeated where it is not repeated in the traditional spelling</td>
<td>“helloooo” for bello</td>
<td>35.4%</td>
</tr>
<tr>
<td>Abbreviation/Reduction</td>
<td>Single word with one or more letters missing with the syllable structure is still intact (almost exclusively vowels)</td>
<td>“vdd” for verdad</td>
<td>18.6%</td>
</tr>
<tr>
<td>Abbreviation/ Clustering</td>
<td>Multiple letters substituted for one letter in a word or phrase</td>
<td>/qu/ or /que/ $\rightarrow$ k “kiero” for quiero</td>
<td>12.1%</td>
</tr>
<tr>
<td>--------------------------</td>
<td>---------------------------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Initialism</td>
<td>Phrases spelled with only the first letter of each word present</td>
<td>“wya” for where you at?</td>
<td>13.3%</td>
</tr>
<tr>
<td>Mistakes</td>
<td>Mistakes are unintentional spellings of words. These are respellings that are likely based on position of letters in the keyboard or that have been repaired in a later message.</td>
<td>Message 1: … novis Message 2: *novia</td>
<td>11.7%</td>
</tr>
<tr>
<td>Phonetic Approximations</td>
<td>Words that are spelled based on the way that they sound rather than the way they are traditionally written</td>
<td>“saze” for says</td>
<td>5.7%</td>
</tr>
<tr>
<td>Syllable omission</td>
<td>When 2 or more letters are omitted from a word resulting in a reduction of the syllable structure.</td>
<td>“celu” for celular</td>
<td>2.0%</td>
</tr>
<tr>
<td>Clipping</td>
<td>When the final letter is dropped.</td>
<td>“goin” for going</td>
<td>1.2%</td>
</tr>
</tbody>
</table>

From this table, it is important to point out the difference between abbreviations and initialisms and distinguish them from acronyms since they are used slightly differently than their colloquial meanings. The term ‘abbreviation’ is used in a technical sense to describe is a single word where letters are omitted. Phrases that consist of multiple words are not abbreviations. The term ‘initialism’ is used in a technical sense to describe phrases that are represented by the first letter of each word.
These are not mutually exclusive as an initialism can also have an abbreviation embedded into it (but an abbreviation cannot have an embedded initialism). For example, “hbu” means “how about you”, incorporating two abbreviations (about \rightarrow ‘bout, you \rightarrow u) and then it becomes an initialism. The term, acronym, is not a technical term for this dissertation. Acronyms refer to initialisms that can be pronounced. Since the pronunciation of initialisms is not significant to this discussion, acronyms will be treated as any other initialism.

### 3.2 Common Respellings

Common respellings are those that are widely used and recognized by texters within a linguistic community. A respelling is considered common if any participant listed it as an “abbreviation or acronym” that they use during the brainstorming activity. For the most part, these respellings are recognized by popular media such as Urban Dictionary (“Urban Dictionary: Textism,” 2015) and Buzzfeed (S. Anderson, 2015; Dobbins, 2015).

Since this section is focused on txt respellings, abbreviations, acronyms, and initialisms that have been incorporated into academic and professional writing will be considered examples of adhering to conventional spelling norms (i.e., Mr. for ‘mister’ or Uds. for ‘ustedes’ plural you), and will not be analyzed as a part of the Txt language form. Conventional forms are those listed in the Chicago Manual of Style 15th edition for English (Staff, 2010), and the guidelines from the Dirección General de Bibliotecas de Universidad Nacional Autónoma de México for Spanish (Robles, 2015).

Alternative spellings that have been incorporated into the written tradition of language belong to the prescriptive, centralized, academic language form. They appear in formal, legal, academic, and professional documents, and are often explicitly taught in school. The focus here is on those spellings that are emerging with more casual and decentralized forms of writing (including ‘thx for

---

31 During the activity, the colloquial definitions of abbreviation and acronym were used. With participants, no distinction is made between abbreviations, acronyms, and initialisms.
‘thanks’ or ‘vdd’ for ‘verdad’ correct, the most popular and ubiquitous of which is txt. The key difference is that colloquial spellings are not explicitly taught in school, emerging instead within speech communities and spreading through conversational partners.

Common non-prescriptive respellings can give insight into how speech communities behave, and how the txt language form exists in this historical moment. They help identify Txt as a language form that is developing and changing within a speech community, and how languages evolve (Crystal, 2014). This is how we most often think of language, as the mechanism used for communication between individuals. Understanding the patterns associated with the emergence and use of common respellings will support the documentation of bilingual txt as an emergent language form, and the texting practices among bilingual texters.

3.3 Uncommon or Novel Abbreviations

Novel respellings provide different information from more common ones. Novel respellings may be made up for just one interaction or may be used by just one person, and the patterns found in novel respellings give more information about linguistic behavior of people in general. That is, if novel respellings have a systematicity to them, it can be argued that that systematicity is indicative of a broader linguistic skill at work (Brinton & Traugott, 2005; Keller, 1994). If everyone respells in the same way, then that mechanism of respelling must be the result of common factors; in this case, the universal factors are human language and the writing system. This type of analysis will be necessarily preliminary as the only languages in question are English and Spanish.

3
3

This point cannot be emphasized enough – the role of speech communities in the emergence of lexicalized abbreviations. The difference between social speech communities and academic is largely that there is a nationally-supported effort to enculturate youth into an academic speech community. There is no such organized effort to enculturate youth into a social speech community—these emerge organically resulting in a space for creativity and experimentation with language and communication that gives rise to respellings and other creative adaptations of language.

3
3

There is a distinction to be made here between ‘language’ and ‘Language.’ Language (lowercase ‘l’) refers to language situated in a speech community, this is what is thought of as “English” or “Spanish.” Language (uppercase ‘L’) refers to the innate ability of human beings to acquire and use language(s).
The first point to be made about this is that not everything can be respelled. For example, in Example 2, Conversation 1 does not make sense whereas Conversation 2 does.

Example 2

**Conversation 1** (example conversation = not attested in the data)

A:  
*Where you at*

B:  
In the park, hbu

A:  
Gh

**Conversation 2** (from the corpus)

*A and B have plans to meet after school. They discussed this via text earlier in the day.*

A:  
Wya

B:  
In the park, hbu

A:  
Omy

on my way

B:  
kk

Okay

Both of the underlined items are novel, non-standardized initialisms, but the second conversation is acceptable and the first is not. The key difference is likely recoverability and shared context. First, the first initialism (‘gh’) involves a noun (home). Any particular noun (excluding
pronouns) is highly uncommon in this corpus by the sheer fact that there are so many different nouns. They are least often used in respellings, especially when they are open class nouns such as “home.” Recoverability is postulated by others to be one contributing factor in novel items, but for these researchers, recoverability is second to other factors such as identity and creativity (Crystal, 1997; Goldbarg, 2009; Tagg, 2009). Second, in Conversation 2, B has previous knowledge that A is planning to come to the park. If this information is not shared with B, “Omy” could mean a variety of things. For example, an alternative interpretation could be that A is surprised that B is in the park and intends to say “Oh, my!” In the next message, it is clear that B has understood that A is on their way to the park. This conversation makes sense and is recoverable because B already knows what A intends to convey, and A is providing a status update that confirms the already known information.

In Conversation 1, if there is not shared information (i.e., B does not know that A is going home), this abbreviation is unacceptable because there is new information to be shared and while the “g” stands for a common verb (going), “h” could stand for literally anything, and it is unreasonable to expect that the receiver would be able to know which “h” was in question. Now consider the situation where A and B were planning to meet at home. In this case, “gh” still does not give enough context to the receiver to know what “gh” might refer to. The lack of context comes from the abbreviation of a content word (the noun, home), and open-class nouns and verbs cannot be abbreviated. Finally, consider the situation in which there was a previous arrangement for A to meet B in the park. Conversation 1 still does not make sense because A is changing the plan, and therefore has new information to share with B. The only way that novel initialisms can be used is if there are no open-class words being used, the message recipient already knows most of the information, and the message sender is giving a confirmation or update on the status. If there is any new information to share, then novel initialisms will not convey that information.
3.4 Respelling Rules

This distinction gives insight into respellings more generally. Respellings are most commonly found in message openings and closings, and least common in the body of a message (Bernicot et al., 2012; Crystal, 2009). Bernicot, Volckaert-Legrier, Goumi, and Bert-Erboul suggest that this is because greetings are the most common features of messages, so they are the most likely to be respelled for greater efficiency. They argue (along with (Thurlow & Brown, 2003) that efficiency is the driving force behind respellings. This efficiency perspective is initially very appealing. Txt is a shortened platform, and messages are often written to convey small pieces of information. However, research on the efficiency of text messaging has shown in multiple ways that efficiency is not a driving factor in txt (Carrier & Benitez, 2010; Deumert & Masinyana, 2008). Setting aside the efficiency explanation, it is possible to think of novel respellings as mechanisms of stylization and to indicate closeness, as many other researchers propose (Goldbarg, 2009; Spears, Lea, & Postmes, 2009; Tagg, 2009). Building on and incorporating this perspective, this dissertation argues that before stylization, identity construction, and creativity can be realized in a novel respelling, certain conditions of intelligibility must be met in order to license the novel usage.

The benefit of a hypothesis that requires intelligibility before stylization is that a message must be meaningful to the receiver before the stylization can be interpreted. If a message is received that uses novel respellings, but the receiver does not recognize them as respellings, they are likely to be interpreted as mistakes or otherwise nonsensical items. Therefore, in order for a texter to successfully convey identity, stylization, or creativity, the novel respelling must be the following conditions for intelligibility:

1. The message must be recoverable. That is, the information the respelled item is meant to convey must be mostly known to the receiver. The respelled form cannot involve a change or an update to the previously known information.
2. The message **cannot involve content words**. For example, open-class nouns and verbs cannot be respelled unless the abbreviation is lexicalized into the language of txt.

3. There must be a **social relationship** between the message receiver and message sender such that there is a motivation to save positive face. The polite form, must be a casual form rather than a formal form.

When these three conditions can be met, then respellings are allowed. This does not mean that respellings must be used. Rather, they are a tool for texters to encode style, identity, social closeness, and non-literal meaning into their messages. Texters use both novel and familiar respellings to achieve this goal. Some respellings may appear novel to a user yet are actually lexicalized into the dialect of the person texting. For example, when texting with someone from the Bronx, someone from Brooklyn may have to explain what “zoe” means (it means ‘to steal’). Within the speech community of Brooklyn youth, that is a common slang term. However, among other groups, it may be completely novel. Therefore, novel respellings are considered from the perspective of the texter, not the receiver.

The list of common respellings was compiled by asking participants what abbreviations they use. If an abbreviation is on the list, it is considered common. If it is not on the list, it is considered novel. The vast majority of respelled forms in this corpus are common, and appeared on the participant-generated lists. However, the novel abbreviations are particularly useful for study as they give insight into how language changes in a microcosm, between two interlocutors. The discussion of novel respellings is incorporated into their relevant sections where appropriate.

**4 Abbreviations**

Abbreviations are any word that is shortened from its prescriptive form. This is limited to single words rather than phrases or sayings. Abbreviations have a long history in many non-formal writing environments, including note-taking, list-making, and letter writing (Crowley, 1976; Daniels &
Most research and discussion of abbreviations treats all abbreviations the same (Erichsen, 2014; Kemp & Bushnell, 2011; McWilliam et al., 2009; Plester et al., 2008, 2009). A distinction will be made here between two types of reductions and clusterings as they are derived in two different ways. Reductions are words with one or more internal letters missing, but the syllable structure intact. For example, “bby” and “bb” are both representations of baby, and both representations retain the two-syllable contour of the traditional spelling and spoken forms. This is the most productive form of respelling in the data and correspondingly will receive the most attention. Clusterings are the combination of two or more letters into a single letter. This is also a common and productive form of respelling and stylization. For example, in Spanish, /qu/ and /que/ become /k/, and in English, /ks/ becomes /x/. These two categories (reductions and clusterings) account for 30.7% of the respellings (18.6%, and 12.1%, respectively).

Abbreviations also fall into the common and novel categories, with those that are widespread and recoverable by the vast majority of a speech community; these are referred to as “common abbreviations”. These are abbreviations that are recognizable and are readily identifiable to members of a txt speech community. The second type are those that are novel to a user, a speech community, or the particular communicative situation; these are referred to as “novel abbreviations.” It is difficult to determine which of the novel abbreviations will be retained, and which will be lost. Therefore, in terms of overall features of the dialect, common and novel are distinct categories, in terms of the linguistic processes that create the abbreviations, they are similar. Example 3 illustrates both reduction and cluster type abbreviations.

---

34 Each item was counted as many times at it appeared, so “verdd” and “vdd” were counted as two instances of reduction even though they are both reductions of ‘verdad’ (true).
Example 3

A: wyd vieja

What you doing vieja

What you doing lady (term of endearment)

B: Aki cn la mujeres

Aquí con la mujeres

Here with the ladies

Participant B frequently uses abbreviations and initialisms in her text messages. In this instance, she is texting a male friend who she has a sexual relationship with, though they are not dating. She uses both types of abbreviations in the message she sends. First, she employs the k-cluster in “aki”, prescriptively, this is aquí (here), and illustrates a complex usage of the k-cluster since /qu/ alone is not pronounced as /kay/, so it is playing off the substitution of “que” for “k” which further becomes substituting “qu” for “k”. Secondly, she uses a reduction, representing con (with) as “cn,” which illustrates the flexibility and creativity that can be employed in one message (setting aside for now the txt features A employs in this brief exchange). Examples of common and novel abbreviations from the corpus are in Table 12.

Table 12 – Common and novel reductions and clusterings

<table>
<thead>
<tr>
<th>Reduction</th>
<th>Clustering</th>
</tr>
</thead>
<tbody>
<tr>
<td>Common</td>
<td>‘flz’ for feliz (happy)</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Novel</td>
<td>‘rinow’ for right now</td>
</tr>
</tbody>
</table>

The information about the status of their relationship is recovered from the text message transcript.
4.1 Reductions

The first group of abbreviations addressed here are reductions. Example 4 illustrates how reductions are used in a text message

Example 4

A: Bye d vdd toi cansaicima y m siento mal
   Bye de verdad estoy cansadisima y me siento mal
   Bye for real I am very tired and feel awful

In this example, A uses three different types of reductions. The first is when she omits the /e/ from both “de” and “me” while these are different parts of speech (“de” is a preposition whereas “me” is a reflexive pronoun), they have a similar prescriptive shape, similar pronunciation, and a similar respelling. In the second reduction, “vdd” for verdad (true/real), she omits all of the internal vowels, similar to an abjad. In the third reduction, “toi” for estoy (I am), she both drops the first vowel and changes the last vowel into the stylized form. This process of dropping the leading vowel occurs frequently in both Spanish and English. Overall, Participant A illustrates the three types of reduction abbreviations in just this one message, suggesting that texters who use abbreviations use them frequently when appropriate. Some common colloquial reduction abbreviations with their alternative forms are listed in Table 13.

36 These could also be considered a clipping since technically the last letter is the only one dropped. However, given the role that phonological processes have in clipped abbreviations (i.e., dropping the final “g” in gerunds and present progressives reflects how they are spoken), it is more likely that these are the result of a reduction since in speech, “me” and “de” are pronounced with their final vowels.

37 Another interpretation of this form is that it comes from Caribbean varieties of Spanish and is not considered a marked or stylized form by the speaker. This is unlikely since almost all students had been in academic Spanish classrooms, so they are aware that they are making a choice here. Even if this form is the most common in the dialect that they speak, they are still close enough to the interlocutor to be using a casual, familiar form.

38 Alternatively, this could be interpreted as the influence of Caribbean Spanish, where it is very common to drop both initial and final vowel sounds.
Table 13 – Common colloquial reduction abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Alternative Forms</th>
<th>Full Form</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tlf</td>
<td>Tlfno</td>
<td>Telefono</td>
<td>Telephone</td>
</tr>
<tr>
<td>Vdd</td>
<td>**vdad</td>
<td>Verdad</td>
<td>True/correct</td>
</tr>
<tr>
<td>Msj</td>
<td>Mnsjs</td>
<td>Mensaje(s)</td>
<td>Message(s)</td>
</tr>
<tr>
<td>Msg</td>
<td>Msgs</td>
<td>Message(s)</td>
<td></td>
</tr>
<tr>
<td>Mnsjndo</td>
<td>**msgando</td>
<td>Mensajando</td>
<td>Messaging</td>
</tr>
<tr>
<td>Aora</td>
<td></td>
<td>Ahora</td>
<td>Now</td>
</tr>
<tr>
<td>Flz</td>
<td></td>
<td>Feliz</td>
<td>Happy</td>
</tr>
<tr>
<td>Bb</td>
<td></td>
<td>Bebe or baby</td>
<td></td>
</tr>
<tr>
<td>Bn</td>
<td></td>
<td>Bien</td>
<td>Good</td>
</tr>
<tr>
<td>Trd</td>
<td></td>
<td>Tarde</td>
<td>Late</td>
</tr>
<tr>
<td>Nmro</td>
<td></td>
<td>Numero</td>
<td>Number</td>
</tr>
</tbody>
</table>

### 4.1.1 Shape of Reductions

Almost all of the reductions in this corpus are the result of removing vowels, and there are only two examples of non-vowels being removed. First, ‘h’ is occasionally removed, as in ‘aora’ for *ahora* (now), and other silent combinations (i.e., ‘ght’ in ‘rinow’ for certain pronunciations of “right now”). This might give some insight into what factors indicate that a reduction will persist. There is a very robust pattern of vowel removal, and it appears that in English, all vowels are candidates for removal, whereas in Spanish, only ‘a’, ‘e’, and ‘o’ are candidates for removal, as indicated in Table 14.
Table 14 – Types of vowel removal in reductions

Each of these items is attested in the corpus.

<table>
<thead>
<tr>
<th></th>
<th>Spanish example</th>
<th>English example</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>‘vdd’ verdad</td>
<td>‘tht’ that</td>
</tr>
<tr>
<td>e</td>
<td>‘d’ de</td>
<td>‘rsp’ respond</td>
</tr>
<tr>
<td></td>
<td></td>
<td>‘txt’ text</td>
</tr>
<tr>
<td>i</td>
<td>*</td>
<td>‘stpd’ stupid</td>
</tr>
<tr>
<td></td>
<td></td>
<td>‘ddnt’ didn’t</td>
</tr>
<tr>
<td>o</td>
<td>‘cn’ con</td>
<td>‘wuld’ would</td>
</tr>
<tr>
<td>u</td>
<td>**</td>
<td>‘stpd’ stupid</td>
</tr>
</tbody>
</table>

* Notably, ‘i’ is very rarely omitted in Spanish. In the entire corpus, it is only omitted once. In English and Spanish, the letter ‘i’ usually represents one of three sounds: [i] as in /key/ or /quiero/ and [i] as in /in/ or /stupid/ and [ai] as in /ice cream/. In Spanish, the [i] and [ai] sounds do not appear, so the only sound ‘i’ represents is [i], which is a very high front vowel, and therefore postulated as being very salient (Carlson, Granstrom, & Klatt, 1979; Fridland, Bartlett, & Kreuz, 2004). This sound is also never omitted in English words in this corpus. However, /i/ is often omitted in Japanese text message abbreviations (Hamano, 2016), indicating that vowel-omission preference is language specific.39

** ‘u’ is also never omitted except in the clustering of ‘qu’ and ‘que’. It is the least common of the Spanish vowels, but not so uncommon that one would predict it is never omitted. In Spanish, ‘u’ makes the [u] sound. Similar to [i], this is a very high vowel. In English, ‘u’ can combine with ‘o’ to make a [ɒ].

39 The vowel omission patterns in txt across languages is beyond the scope of this project, but would be fascinating and have potentially significant consequences.
Example 5

A:  Ola ñame 😴😴

Hello yam

B:  Hola stpd

Hello stupid

In Example 5, Participant B (who is Spanish-dominant) drops the /u/ and /i/ from stupid, spelling it as “stpd”. While is the only word where /i/ is dropped and she is the only participant who does it, both “stpd” and “stupid” appear to be a nickname she regularly uses for him. When participants were asked why they rarely ‘drop’ /i/ or /u/, they said it is because these are ‘loud’ vowels. This intuition is shared in educational resources, referring to these vowels as “strong” (Libros Media, 2015) or “loud and short” (Erichsen, 2015) as compared to other vowels. While there is not a restriction on omitting /i/ and /u/ since Participant B omits them from stupid, there is a general tendency not to omit them from words in Spanish or English.41

40 Sleepy face emoji

41 This is a diagram of all the vowels in all languages. Note that [i] and [u] are both high vowels, indicating that the tongue is very close to the roof of the mouth. In this way, [i] and [u] obstruct the airflow more than other vowels do (vowels that are lower on this chart). This is a possible explanation for the tendency to not omit these vowels, as they are more similar to consonants, and therefore interpreted as more important to the audio-visual representation of the word.
Vowel removal is allowed in every position within the word (word initial, word-internal, and word-final), so long as the word can reasonably be recovered using the remaining letters (as illustrated in Example 3 in this section). In both languages, consonants and high vowels ([i] and [u]) are disproportionately preserved, suggesting that they are more important in identifying the target word. So, while low vowels may be the most sonorous (i.e., they are the “loudest”) (Parker, 2012), they give the least amount of information about what word is being conveyed because they are the least unique to any given word. Likewise, ‘e’, ‘a’, and ‘o’ are the most common letters in Spanish and in the top four in English (‘t’ is the second most common English letter). This may also contribute to their ease of omission since they appear the most often and are again the least distinctive of any given word (Grade, Chart, Assessment, & Prompt, 1965). This explanation refers back to the intelligibility rules for abbreviations identified in Section 2.3. Intelligibility is an essential feature of the words that can be abbreviated and those that cannot since the one major goal of a text message is exchange information. Additionally, there are three already existent orthographic traditions that frequently omit the vowels from words, including abjads, early childhood approximations, and telegraphic communication.

### 4.1.2 Vowel Omission Writing Systems

Abjads, early childhood approximations, and telegraphic communication are three recognizable systems that exist across authors in different languages and speech communities. The similarity between these three writing systems suggests that vowels are less critical to conveying meaning when writing than consonants.

Abjad writing systems are orthographies where each character refers to a consonant and few or no vowels are written. For example, in Arabic, the root, “k-l-m,” can be used to express anything related to speaking or conversing, depending on the context. The context indicates the meaning, which indicates the vowels inserted between each root-word letter. So, “kalima” means* to speak or*
articulate whereas “kalam” means to debate. There are many languages that use this type of writing system, including Arabic, Malay, and the Tuareg languages (Daniels & Bright, 1996). These languages use alphabetic orthographies but generally do not write vowels, though some include a word-final vowel or diacritics for children and second language learners. Most of the languages with an Abjad writing system are Semitic languages (Daniels & Bright, 1996; C. L. Meyers, 1983; Schiff & Ravid, 2004). To non-Semitic readers, this may appear like writing in abbreviations, but it is actually a fully formed way of writing where the vowels are not recorded. If a reader were to pronounce the writing, he or she would fill in the vowels as appropriate for the part of speech of the word (Daniels & Bright, 1996; Gal, 2002). This tradition is thousands of years old, and was common in the earliest alphabetic orthographies (Daniels & Bright, 1996). Even though Abjads predate the Latin alphabet, it is unlikely that participants in this study have had enough contact with writing in Semitic languages to influence their writing practices. It is more likely that representing words with only consonants is an effect of the relationship between language and writing.

Language does not depend on writing to exist, and many languages “borrow” writing systems to be recorded (i.e., any language without its own orthography). However, writing does depend on language to visually encode a spoken and signed system, but this does not mean that writing is bound to the way a language is spoken. Writing is free to encode words that could not be spoken by combining impossible to pronounce letters together (“xyz”), or omitting all vowels (“vdd”). The freedoms afforded by representing language through writing is not limited to abjads, but to all writing systems, including alphabets (where the shape of a letter represents the sound, but there is no inherent correlation between the two), and semanto-phonetic writing systems (such as that used to represent Chinese), which incorporate ideograms into writing as well. For semanto-

42 For example, isiXhosa (South Africa), Swahili (East African Countries), Luganda (Uganda) all utilize the Latin script. This stands in contrast to Ekoí (Cameroon and Nigeria), which uses the Nsibidi script, a writing system developed specifically to write the Ekoí languages, including Ekoí (Heine & Nurse, 2000, 2008).
phonetic systems, the meaning of the word is more significant than its spoken counterpart, resulting in characters that have multiple pronunciations in different environments. It could be argued that txt is also beginning to incorporate more meaning-based symbols, as the Oxford English Dictionary’s word of the year for 2015 is in Figure 17 as it comprised nearly 20% of all emoji usage in 2015 (Rosenthal, 2015).

Figure 17 – Face with tears of joy

A second area where vowels are reduced in both Spanish and English writing is in children’s spelling patterns. When children are first learning to read and write, they tend to use consonant-only patterns until they develop spelling mastery (Gentry, 1978). After kindergarten and in to first grade, this pattern persists on unstressed syllables – where children drop either the vowel or the entire syllable (Gentry, 1978). There are multiple explanations about why this is so, but one of the most popular theories is based on the phonological saliency of vowels and unstressed syllables (Vihman, 1996). Educators and researchers postulate that children who are learning to read and write are making a connection to the verbal language that they are familiar with (Robinson, McKenna, & Wedman, 2000; Strickland, 1998). Researchers suggest that during this process, children focus on stressed syllables and consonants because they are acoustically easier to perceive and therefore easier to spell (Vihman, 1996). When children write in this reduced way, most of the text is still recoverable

---

43 Emojis, of course, are not writing systems. They are not flexible enough at this stage to represent sounds. However, multiple emojis together may function more like early pictoral writing systems (NOT Chinese), where the emojis stand for meanings rather than sounds. Currently, Fred Benensen is doing a lot of work in this area, translating both *Moby Dick* into emoji (emojidick.com) and various documents and reports (Benensen, 2014). For an excerpt, see .
by naïve readers, possibly because of the similarities to the phonological representation of the words. If it is the case that nearly all abbreviations in txt follow this pattern as well, then it may be due to the phonological influence on writing.

The final writing form in which vowels are often omitted is in shorthand systems (Gregg and Pitman being the most common in the United States) (Gregg, Leslie, & Zoubek, 1955; Pitman, 1964) and telegraphic writing (Bryan & Harter, 1897). Both of these systems either omit vowels or indicate them with small symbols such as a dot or a dash, or omit as many letters as possible to insert as much information as possible. The connection between the language used in texting and telegraphic writing has been noted by many other researchers (Beasley, 2009; Crystal, 2009; Holtgraves & Paul, 2013; Thurlow & Brown, 2003), generally with the conclusion that texting mirrors telegraphic writing because texters are trying to insert as much information as possible into the smallest space possible. While this may be a reason that some texters abbreviate, researchers who have tested this hypothesis with bilingual texters have found that it is not the case (Carrier & Benitez, 2010; Deumert & Masinyana, 2008). They found that the abbreviations that texters use have very little if any correlation with intentionally shortening the messages (Carrier & Benitez, 2010). Deumert and Masinyana found, in fact, that while texters may be willing to abbreviate English words, they are not willing to abbreviate isiXhosa words even though there would be a financial incentive to do so (2008). So while Txt abbreviations may have the same shape as telegraphic representations, it is unlikely that the only motivation for abbreviation is to save space.

Finally, all three of these systems combined suggest that vowel-less writing is an easy to understandable form of writing that has a long tradition in the history of written language. This suggests that just as the semantic content must be recoverable from the context, the form of the word must be recoverable as well. The difference between Abjads, child writing, and telegraphic systems and txt, however, is that abbreviations in txt have a stylistic component that is important for
identity construction (Goldbarg, 2009; Spears et al., 2009). Since abbreviations in txt are chosen from a range of options and flouting the prescriptive spelling rules of Spanish and English, the options for where and when to use a reduced form carries communicative meaning for the texter. This communicative function may be an overlay of information that would be encoded in non-verbal features during face-to-face communication (Section 3.3).

4.2 Clusters

Clusters also exist throughout the corpus, however, there are much fewer types of clusters than types of abbreviations. Clusters are when two or more letters are represented by one letter. In Spanish, the most common cluster in the corpus is the change from ‘qu’ and ‘que’ to the letter ‘k’ (k-cluster) as in Example 6; in English, the most common form is ‘you’ to the letter ‘u’ (u-cluster) as in Example 7.

Example 6

Aki en el trabajo k m toca trabaja hoy
Aquí en el trabajo que me toca trabaja hoy

Here at work I get to work today

Example 7

I can't get tired of u but I work at 12 noon tomorrow remember

I can’t get tired of you but I work at 12 noon tomorrow remember

The k-cluster is a result of the complementizer, ‘que’ being pronounced the same way that the English letter ‘k’ is. In the Spanish alphabet, the letter, ‘k’ is traditionally pronounced as /ka/. In the English alphabet, however, ‘k’ is pronounced /keI/, similar to how “que” is pronounced. So rather than appealing to the sound that ‘k’ makes, this cluster is appealing to how the letter ‘k’ is
pronounced – in English, though. Another related form is the use of ‘x’ for ‘por’. Again, ‘por’ is a very common preposition in Spanish, though it also the symbol for multiplication, pronounced ‘por’. So, in this case, the grapheme, ‘x’ is being used to connote the word for multiplication, namely, ‘por.’

This combination makes ‘xk’ possible as an abbreviation for ‘porque’ which means ‘why’ versus ‘xk’ for ‘por qué’ which means ‘because’ (‘x k’ is not attested in this corpus, though ‘xk’ is). However, participants write ‘xk’ to mean both ‘why’ and ‘because’. In Spanish, the x-cluster is formed by substituting ‘x’ for ‘ch’ because in a Spanish word, ‘x’ makes the same sound as the ‘ch’ combination. On the other hand, the u-cluster in English does not involve any borrowing; it is formed by substituting the letter ‘u’ for the word, ‘you’, because in the English alphabet, the letter, ‘u’, is pronounced with the same sound as the word ‘you’.

After these two types of clusters, the most common letter involved in cluster abbreviations is ‘x’. Participants used ‘x’ to replace both ‘ch’ in Spanish (Example 8), and ‘ks’ in English (Example 9).

Example 8

Lo siento. Se termino mi saldo flz nxce. ;-) 

Lo siento. Se termino mi saldo feliz noche. ;-) 

*I’m sorry. I lost my balance good night. ;-)*

Example 9

And thx.

*And thanks*

The ‘ch’ combination was never replaced with ‘x’ in English, however, indicating that there is either a language-specific effect, or a phonological effect preventing ‘ch’ from being clustered into ‘x’ in English.
4.2.1 The k-Cluster

By far, the most common cluster is the k-cluster. The k-cluster appears both word-initially and word-externally (and in initialisms, as will be discussed in Section 4). The k-cluster can be found in everything from a stand-alone ‘k’ for ‘qué’ (Example 10) to ‘klk’ for “que lo que” (“what’s up?”) (Example 11) and ‘porke’ for ‘porqué’ (“because”) (Example 12) and ‘equivoke’ for ‘equivoque’ (“wrong”) (Example 13) to ‘kiero’ for ‘quiero’ (“like/want”) (Example 14).

Example 10

Hey primo quieres trabajar mañana a las 9 de la noche no se k hora termina

Hey primo quieres trabajar mañana a las 9 de la noche no se qué hora termina

*Hey cousin want to work tomorrow night at 9pm I don’t know what time it will end*

Example 11

A: klk

Qué lo que

*What’s up*

B: Watching a movie

Example 12

Te escribo luego porke estoy busy

Te escribo luego porque estoy busy

*I will write later because I am busy*

44 “Porke” is also used to represent “por qué” (why?). An alternative version (found in the corpus) is “pork.”
Example 13
Mmmmm ups lo siento quizas me equivoke losiento feliz noche
Mmmmm ups lo siento quizas me equivoque lo siento feliz noche.
_Mmmmm ups I'm sorry maybe I'm wrong Good Night_

Example 14
Yo kiero habla contigo
Yo quiero hablar contigo
_I want to talk to you_

Notably the ‘k’ in “porke” and “kiero” does not replace the shape ‘que’ but only ‘qu’. In this instance, the sounds are not even the same yet “k” stands in for both instances of “qu.” This cluster requires bilingual awareness to be produced and understood because texters must be aware that the letter ‘k’ can be pronounced the same way as ‘que’ is. At first, it may appear as something of a joke, a way to play with the intersection of Spanish and English, but it is so common that nearly 75% of participants use it. The k-cluster also has the benefit of being easier to type in the multi-press system that preceded smart phones and miniature keyboards, which may have helped it catch on. Students reported that the /k/ abbreviation is heavily present in Dominican text messages, more so than in Mexican, Puerto Rican, or Ecuadorian messages.

To better understand how this came about, the distribution of the k-cluster was looked at for Twitter users in the Dominican Republic, Mexico, Ecuador, and Spain. This was done through collecting 100 messages from users in each country during the week of December 14th, 2015. The

---

45 This feature deserves further consideration comparing phonological and morphological substitutions. Since phonological substitutions are not covered in this dissertation, this topic is set aside for future work.
Spain-based Spanish speakers rarely used the k-cluster, option instead to spell out the word, ‘que.’ The Mexican and Ecuadorian twitter users utilized the k-cluster in 62% of the available instances whereas the Dominican twitter users utilized the k-cluster in 96% of their messages. This is an informal survey of how the k-cluster is used, but the findings are disparate enough to justify a closer look at Dominican Spanish. What caused ‘k’ to go from just play with language into a full blown replacement for the ‘qu’ combination likely has to do with the rise of secondary education and the popularity of texting in the DR.

To be clear, ‘k’ was used as a substitute for ‘que’ and ‘qu’ as early as 2000 (based on a simple Google Search). However, its current status as a lexicalized form is the focus here. The status of the k-cluster may have been more like the status of ‘x’ for ‘por’, or ‘x’ for ‘ch’ (i.e., it occurs regularly, but only a few people use it in their message) without a few key events co-occurring in the Dominican Republic.

First it should be noted that using ‘k’ is a marker of bi-literacy, a status that is respected in the Dominican Republic. From 2005 to 2007, secondary school enrollment in the Dominican Republic rose from 65% to 77.5%, the largest reported increase in the Dominican Republic’s history. At this time, cellphone subscriptions rose from 39% to 58%, moving from an item used by the upper classes only into a domesticated item that could be obtained by a critical mass of people. Cellphone plans continued to rise until 89% of the population had one in 2010 (“World Development Indicators | World DataBank,” 2014) (since then, cellphone subscriptions have fallen off as Internet users have increased). This spike in English exposure combined with the rise in cellphone plans may be able to explain the rise of the k-cluster. By contrast, during this same time period, Puerto Rico, Mexico, and Ecuador all had higher rates of cellphone ownership and English education than the Dominican Republic. Yet in these three areas, English proficiency higher for a longer period of time, and English was already domesticated into the culture. In these countries,
cellphones entered into a world where English was being taught before cellphone ownership. In the Dominican Republic, there was a spike in both at the same time, leading to large groups of young people being exposed to English (namely learning the alphabet), and gaining access to cellphones simultaneously. See Figure 17 for a graph of how these events intersect.

Figure 18 – Mobile ownership, education, and the k-cluster in the Dominican Republic

It is unlikely the k-cluster emerged from Spanish-English contact in the United States since texting was not a common practice in the U.S. until 2008 (Hafner, 2009; Rainie, 2005; Reardon, 2008). People in the U.S. texted before 2008, though it was not nearly as widespread. By then, complex versions of the k-cluster (word-internal, as a part of initialisms) were already in use on
Twitter and MySpace. The first definition of “klk” “que lo que” or “What the what?” meaning “What’s up?” on Urban Dictionary appeared in 2006. At that point, it is unlikely that enough Spanish speakers were texting in the United States to form a speech community large enough to use the ‘k’ cluster so fluently. It must have come from a country where texting was already a domesticated practice, such as the Dominican Republic.

Finally, there is a clear influence of bi-literacy on this abbreviation since /k/ is not an acceptable substitute in the Spanish alphabet. It is therefore both playful and a marker of education. For this abbreviation to make sense, the texter and receiver must have some command of the English alphabet, marking users as educated and aware of the interplay of languages.

4.2.2 The u-cluster

One of the most commons respellings in the entire corpus, as well as in other large texting and chatting corpora is the use of the letter, “u” to signify “you.” This abbreviation was in use long before texting began, and is one of the most common abbreviations both in early chat rooms as well as in pop culture. It is impossible to know if this abbreviation emerged as a result of dropping the first two letters or because the letter “u” is articulated in English with sound of the word, “you.” Building on the fact that the k-cluster is so popular and it is only possible to have been derived because of how the letter “k” is spoken in English, it will be assumed that the process that allows “u” to represent “you” is because of the pronunciation of the letter, “u”, not because the first two letters of the word, “you” are being dropped.

The u-cluster is extremely common in all digitally mediated environments. It is likely less common than it initially appears, though, as instances of “u” may stand out as opposed to “you” or “tu” because it is a non-prescriptive way of writing it. In this corpus (and all Spanish-English corpora), it quickly becomes complicated to calculate a percentage of the use of “you” and “u” versus “tu”, “te”, and “ustedes” because in English, the 2nd person pronouns are indistinguishable
between indicative and accusative case as well as singular and plural whereas in Spanish, there are three distinct words. In this corpus there are 558 instances of “te”, 272 instances of “tu”, 383 instances of “you”, and only 104 instances of “u”, making it less than half as frequent as any other form. Even this tally does not capture the entire picture because “u” almost exclusively represents “you” in initialisms (except for “wya” (where you at), and “ily” (I love you)), whereas “te”, “tu”, and “you” rarely do. The u-cluster has the same distribution as the fully spelled out version of “you”, being used in both the indicative and accusative positions as in Example 15 and Example 16 (both of these examples are conversations between the same participants).

Example 15

A: What about me

A: That’s for talking not text

B: I wanna tell u and learn about u face to face

I want to tell you and learn about you face to face

Example 16

A: [PICTURE]

A: Can u send me one back? :) ha

B: Hahahaha if u say so

Hahahaha if you say so

This shows that “u” is fully lexicalized into the vocabulary of txt, and has the same distribution as the prescriptive form. It also shows that “u” offers no semantic difference from “you” even though

---

46 For this part of the analysis, it is necessary to remove one very large conversation because they text almost exclusively in English, and it is not representative of the larger corpus since they only use the word “you”, and never use “u.” The sub-corpus analyzed here is the same one used for the quantitative analyses in Chapter 2.
there may be a pragmatic difference. As for the pragmatic difference, “u” appears to be more stylized and conversational than “tu”, “te”, or “you.” The traditional Spanish and English forms are the unmarked forms in this case, occurring more often and used by more participants than the bare “u” is. Finally, while many participants use the u-cluster in their text messages, not all do whereas all participants use “te”, “tu”, or “you” at least once. The abbreviated form of “te” is most often found in “tk” for “te quiero,” and there is no clear instance of “tu” being abbreviated in the corpus. In the case of “tk” for “te kiero,” it may just be that the first letter is being used, similar to how all other initialisms are formed. Finally, the distribution of “u” as well as how participants use it illustrates that using the abbreviated form is a choice, and when texters choose the u-cluster over the spelled out version of “you, they are flouting the prescriptive norms of writing. As discussed in the introduction to this chapter, flouting a prescriptive norm most often carries communicative effect since it is an intentional respelling rather than an ignorance of the prescriptive rules.

4.3 Abbreviations - Pragmatic and Communicative Effect

So far, we have covered the distribution, history, and rules of formation for abbreviations in Txt. This section will cover the pragmatic roles that abbreviations take in txt exchanges. The messages will be analyzed from three perspectives (politeness, group membership, and as markers of illocutionary force). By this approach, each aspect builds on the previous, providing the necessary information for the receivers of a message to decode the intended meaning of the utterance. In the case of both reductions and clusters, texters are flouting the prescriptive grammar rules to create meaning by deviating from the prescriptive norms.

4.3.1 Politeness

The first level of interpreted meaning is in the form of politeness. The analysis presented here will use Brown and Levinson’s theory of positive and negative face to analyze the exchanges in this
corpus (1987). Brown and Levinson proposed a theory of face wants (P. Brown & Levinson, 1987) that governs much of the research on politeness. In their theory, everyone has both positive and negative face wants. According to Brown and Levinson, positive face wants are the desire for inclusion, group membership, and social closeness; negative face wants are the desire for social distance, respect, and adherence to hierarchy (P. Brown & Levinson, 1987). Simply put, in Brown and Levinson’s theory, speakers preserve positive face of the receiver by using familiar, colloquial language. Shortened phrases such as “hi” or “hey” for “hello”, and “what’s up” for “how are you?” are key elements that signal familiarity and closeness. Speakers preserve negative face of the speaker by using more formal language, and adhering to prescriptive grammar rules. Extended expressions, often involving modals (should, could, would) and request flags (please, thank you), formal message openings and closings, and an acknowledgement of the receiver’s time are key elements to signal social distance and preserve negative face of the receiver.

The difference between positive and negative face preserving language is readily apparent in participants’ messages. Example 17 is between individuals who are socially very close to each other. Contrast this with Example 18, where the receiver is unknown to the sender, and it becomes readily apparent that social distance is a key element taken into consideration during texting (Beasley, 2009; Khalifa, 2015; Ling, 2005).

---

47 Analysis of face threatening acts via text message is beyond the scope of this dissertation, but would make a fascinating investigation.
Example 17

This is the end of a long conversation between two friends where one friend is asking the other to lend him some money. Money has been lent between these two before and not paid back, but they are still friends.

xk aki no hay dinero bro
Porque aqui no hay dinero bro

Because we don’t have any money here bro

Example 18

"[NAME], we appreciate your interest in the position with [NAME]. To complete your registration please call ###-###-####"

Example 17 contains three respellings. The first is the use of “xk” to represent “porque” (because). This is a highly stylized respelling because it depends on the mathematical pronunciation of “x” for “por” and the k-cluster. The second respelling is the use of the k-cluster in “aki” to represent “aqui” (here). The third is the use of the word “bro,” which is an English term of endearment between friends, and signals social closeness (Spagnolli & Gamberini, 2007). This form indicates that even though the texter is denying the request for money, he is still attempting to preserve the positive face of the receiver, and ensure that they are still friendly. Example 18 is an automated message from a company. These messages tend to be the most formal, incorporating the maximum amount of negative face saving markers. The sender uses formal phrases such as “we appreciate” versus “thanks for”, and fully written sentences which adhere to prescriptive punctuation rules (i.e.,

---

48 While this message was not sent by a participant, it was retrieved from the company’s publicly available website. The company was not contacted regarding this study.
proper noun capitalization, the comma, and period), and incorporate the formal request marker, “please”. This form indicates that the negative face of the receiver is being maximally preserved.

Many of the participants do not often write in the formal, negative-face-preserving register of Example 18 (neither in Spanish nor in English). Instead, it appears that participants rely on English to indicate they are preserving the negative face of the receiver. In Example 19, the sender may not be aware of the (prescriptive) grammatical mistakes she makes, but it appears she is attempting to adhere to the negative face wants of the receiver. She includes an explicit greeting (“Hi”), indicates that she already sent the message, but provides an excuse for the receiver (“I think that you didn’t see”), uses a modal to hedge the question (“can”), gives a reason for her absence (fever and sore throat), and includes two markers of negative face preserving requests (“please” and “thank you”). Notably, this participant writes exclusively in Spanish except when making requests, indicating that, to this participant, English is a language that indicates preserving the negative face of the receiver.

Example 19

Hi I send you a message and I think that you didn’t see … can you say to the teacher that I am sick I have fever and a terrible sore throat please thank you

Hi, I sent you a message and I think you didn’t see it… can you tell the teacher that I am sick. I have a fever and a terrible sore throat. Please. Thank you.

Similarly, in some instances, participants use English to indicate social distance, formality, and to signal respect with language, and preserve the positive face of the receiver. Many texters preserve positive face through the use of abbreviations, nicknames, and other non-prescriptive textual features (i.e., emojis, emoticons, and punctuation) (Adams, 2009; Spagnolli & Gamberini, 2007; Tossell et al., 2012). In Example 20, the texter switches between English (in response to the
question) and Spanish to make a request, they use the k-cluster, attach the preposition, para to an emoji, and refer to the receiver as “niggah”. In their conversation, these participants continue to use these features, indicating and reinforcing the social closeness (Adams, 2009).

Example 20

The participants in this conversation are engaged in a new sexual relationship, and have been engaged in flirting throughout the conversation. This message is more than 20 messages in to the conversation between them.

Done and no te comente tu ig para k ningun niggah 🙋‍♀️❤️💦💦💦❤️كانخص

Done and no te comente tu Instagram para que ningun niggah 🙋‍♀️❤️💦💦💦❤️كانخص

Done and no. Explain your Instagram. Like any niggah 🙋‍♀️❤️💦💦💦❤️كانخص

Another example illustrating the subtextual features interlocutors use to signal a preservation of positive face is in Example 21. In this case, abbreviations are used that even if the receiver is not aware of, they can probably recover since the first abbreviation (“tabn” for “también”) only omits a nasal consonant and the second vowel sound. A duplicated letter is used in place of an accented letter (“ii” for “í”), and “b” is used in place of “v”, which is a common substitution in Spanish since the sounds they make are nearly identical. These factors combine to signal that this is an informal message (Adams, 2009; Khalifa, 2015; Sotillo, 2012; Spagnolli, 2012; Spagnolli & Gamberini, 2007), and the receiver is close enough to the speaker to use this informal writing style.

---

49 The term, “niggah” is used predominately by African American men as a highly intimate term of endearment among friends or in-group members, it is often used to signal “close friendship, cultural awareness, and fearlessness” (Smitherman, 2000)

50 Emojis will not be discussed in this dissertation as they are outside the scope of the project. However, they provide ample opportunity for further research.

51 Person raising both hands in celebration emoji

52 Two hearts emoji

53 Splashing sweat emoji

54 Pouting face emoji
Example 21

Lol tabn ahii. Ba trabaja

Lol también ahí. Va trabaja

*Lol also there. Go to work*

As discussed in Section 2, for an abbreviation to be felicitously used, it must be recoverable by the receiver. Therefore the receiver must either be aware of the shared information or they must both be members of the same texting community. If the recoverability condition is not met, there are two possible outcomes. In one case, there is a communication breakdown in the literal meaning of the message, and the receiver must ask for clarification or be willing to tolerate the ambiguity. In the other case, there is a breakdown of the non-literal intended meaning of the message, and the receiver is not aware of the breakdown and the exchange proceeds with the misunderstanding. Assuming that the abbreviation is understood, the texter has marked that they are socially close enough to use an alternative form, and the texter wishes to preserve the positive face of the receiver by indicating that closeness (Spagnolli, 2012).

4.3.2 Group Membership

The next step of analysis illustrates how texters use language choice to indicate group membership and signal the use of a we-code to achieve their goals. By utilizing the language that “belongs” to a texting community (rather than the language associated with the prescriptive grammar), the texter is signifying that they are in the same group, with the same language and vocabulary. Gumperz (1977) and Myers-Scotton (1993) have both done extensive work to show that the choice of linguistic code is a powerful signifier of group membership. Myers-Scotton in particular has worked on a variety of language combinations to illustrate the role that language choice plays in a given interaction. She concludes that language choice is not an accident of interaction, but rather a careful choice to
encode a sub-textual social message into the overt message (Myers-Scotton, 1993; Myers-Scotton & Ury, 1977). In Example 22 (Example 22 is Example 20 plus the context) the role of language choice to mark the serious versus the playful dimensions of the message is apparent.

Example 22

As in Example 20, these messages are part of an ongoing flirtation between two participants engaged in a sexual relationship.

A1: [PICTURE]

A2: Add this on your ig and say “good night from my wife 😊 😌 😘

B1: Good night for u ?

A3: - if you want

B2: Ok 😘

A4: 🚰❤️

B3: Done and no te comente tu ig para k ningun niggah 🙏❤️ 😍 😘

Done and no te comente tu Instagram para que ningun niggah 🙏❤️ 😍 😘

Done and no. You’ll comment on your Instagram for any niggah 🙏❤️ 😍 😘

A5: OMG baby relax I only want you ❤️

Oh My God baby relax I only want you ❤️

55 Face savoring delicious food emoji
56 Sleeping face emoji
57 Face with stuck-out tongue and winking eye emoji
58 Kissing face emoji
59 Closed lock with key emoji
60 Heart emoji
61 Person raising both hands in celebration emoji
62 Two hearts emoji
63 Splashing sweat emoji
64 Pouting face emoji
65 Heart with arrow emoji
This exchange not only switches between English and Spanish (both participants are bilingual), but also involves multiple features that indicate that they are engaged in a playful social relationship. Participant B’s third message includes the k-cluster (“k” for “que”), and code switching (English to Spanish). The choice of words in English versus Spanish is probably not accidental, while most of the conversation is in English, the instances where A and B are either expressing their distrust or identifying inconsistencies in each others’ stories are all in Spanish. It appears that for this couple, Spanish is the language of confrontation whereas English is the language of flirtation.

Both Gumperz and Myers-Scotton analyze switches such as these as conversational switches that have social information highlighted through the language change (Gumperz, 1977; Myers-Scotton, 1993). Myers-Scotton in particular analyzes these types of switches as a way for the texter to communicate the intention of their message. From this perspective, these switches are almost certainly indicating the intended meaning behind the message. The lighter, more playful messages, and requests are written in English, reinforcing that they are meant to be interpreted with the same light tone that the rest of the conversation (also in English) takes on. However, when serious matters need to be discussed (namely, the lack of trust in each others’ commitment), they both switch to Spanish. Spanish, for these participants, serves to both mark that the tone is more serious and distance that part of the conversation from the remaining flirtations. It is therefore marked as the “we-code”, and in this case used for potentially sensitive confrontations (Myers-Scotton, 1993; Myers-Scotton & Ury, 1977).

In their discussion of the role of “we-code” in communication, neither Gumperz nor Myers-Scotton approach the idea from a perspective of marking the illocutionary force; however, their ideas about the role of language choice on making meaning are fundamental to this idea. In the next section, I will extend the ideas Gumperz (1977) and Myers-Scotton (1993) have about language choice to the idea that texters signal the illocutionary force of an utterance with their language
choices – both with respellings and with choosing between Spanish and English. Because the extra-
textual features are so constricted when analyzing text messages, the language choices are significant
factors when identifying the illocutionary force of the utterances.

4.3.3 (Illocutionary) Forces

According to Speech Act Theory, every utterance has three dimensions: the locutionary force, the
illocutionary force, and the perlocutionary effect (Austin, 1975). The locutionary force is the most
basic, literal meaning of the words used. The illocutionary force is the intended meaning, and the
perlocutionary effect is how the receiver interprets the message (Austin, 1975; Yule, 2010). The
classic example of this is in Example 23, the same locutionary act has only one locutionary force
(i.e., a statement of reality), but nearly infinite illocutionary forces.

Example 23

It’s cold in here.

The speaker could be complaining about the room temperature, asking someone to close the
window, requesting information about why it is so cold, etc. The perlocutionary effect could be an
offer of sympathy, someone shutting the window, an explanation about the heat, etc. Most
researchers focus on the illocutionary force, as is the focus here as well.

The illocutionary force is particularly interesting because the speaker (or texter in this case)
indicates their intended meaning through the sub-textual features of the utterance (Dresner &
Herring, 2010; Searle, 1976; Sotillo, 2012). In face-to-face conversations, this is indicated with
intonation, pitch, rate-of-speech, body language, and other phonetic features (Burgoon, Guerrero, &
Floyd, 2010; Cassell, Nakano, Bickmore, Sidner, & Rich, 2001; Kendon, 1995; Knapp, Hall, &
Horgan, 2013). In the case of txt, the sub textual features are language choice, abbreviation style, and
other respellings (Baron, 2010b; Baron & Ling, 2011; Dresner & Herring, 2010; O’Neill, 2010;
In social conversations (whether face-to-face, or via txt), these features indicate the communicative intent of the speaker (Burgoon et al., 2010; Knapp et al., 2013). Previous research has focused on the use of features such as emojis and punctuation as signaling the illocutionary force of an utterance (Baron & Ling, 2011; Crair, 2013; Dresner & Herring, 2010), relating these features to the non-verbal meaning conveyed in face-to-face conversation. Examples from this corpus suggest that not only do punctuation and emoticons do the work of communicating the significance of the illocutionary force, but initialisms do so as well.

Example 24 is an exchange between two people who are romantically involved. Participant A is trying to convince B to come over tonight even though A’s sister will be in the house, and A’s sister does not like B. Participant A has a motivation to be both clear but playful in order to achieve her goal of getting B to come over. If she indicates that she is angry, upset, annoyed, or otherwise feeling negatively about Participant B’s reluctance to come over, Participant B is likely not to come over. She therefore has the difficult task of expressing her desire for him to come over without appearing aggressive, desperate, or otherwise undesirable all via a completely written medium. The language choices that she makes in her messages illustrate her attempt at flirtatiousness and playfulness in her request.

---

As well as extraneous punctuation, emojis, emoticons, and pictures.
Example 24

A1: Eso si k m va encotrea dormiendo si es vdd k viene

Eso si que me va encontrar durmiendo si verdad que viene

*She will find me sleeping if she comes.*

A2: Y ya t dijekiene van star aki

Y ya te dije quien van a estar aqui.

*And I told you who will be here.*

A3: So tu m avisa

So tu me avisas

*So tell me what to do*

A4: Aunk no t creo de k tu venga

Aunque no te creo de que tu venga

*Although I don’t believe that you are going to come*

A5: D vdd tu va a venir

De verdad tu vas a venir

*Are you really going to come*

B1: Si Estas sola Si voy

Si estas sola si voy

*If you are alone, I’ll come*

In this case, fully half of the words in A1-A5 are written in a non-standard way (this is uncommonly high for the corpus more broadly). In the first message, participant A uses the k-cluster and incorporates an abbreviation for *verdad*, and she left two mistakes unchanged: encotrea (which is not a word in Spanish, so they probably meant *encontrar*, which means ‘find or discover’), and *dormiendo*
for dormiendo\textsuperscript{67}. The message is easily recoverable by the receiver provided two conditions are met. First, they are aware that k-clustering is allowed. Second, they are either aware that ‘vdd’ is an abbreviation for verdad, or can easily recover it since only the vowels are omitted. The combination of these three elements (the k-cluster, the abbreviation, and the mistake tolerance) function to mark that the conversation is playful and intimate, and distinct from the traditional orthographic conventions (Khalifa, 2015; Spagnolli, 2012). She then continues to use the k-cluster when possible, abbreviates two letter words to one letter (i.e, “t” for te, and “d” for de), and throughout the exchange, used the stylized form of each word to express herself. All of these features are allowed in the Txt medium and in other corpora have been postulated as expressing creativity (Tagg, 2009), intimacy (Spagnolli & Gamberini, 2007), and playfulness/identity (Jaffe, Androutsopoulos, Sebba, & Johnson, 2012; Spears et al., 2009). Another interpretation of this example is that the respellings are simply a representation of the speaker’s native dialect (Dominican), in which word-final morphology is often omitted. Even if some of the respellings are the result of phonological representation and some are strictly limited to Txt, the result of the combination has the same effect of signaling the social closeness of the interlocutors. Similarly, Example 19 was also written by a speaker of Dominican Spanish and includes none of these features. Therefore, even if these features appear in the speaker’s dialect, it appears as though they have control over the spelling of them and are choosing to use colloquial spellings to indicate the social closeness.

A second example of communicating the social closeness through spelling choices can be seen in Example 25. Here, Participant A and B are male friends who met in school. This is a new

\textsuperscript{67} This is an extremely common Spanish misspelling (Rivas, 2015), since dormir is an irregular verb. The texter may not have been aware of the mistake, may have been intentionally playing with tense/aspect, or may have noticed but not felt strongly enough to change it. This is a situation where it is impossible to know the texter’s intent, so it cannot be analyzed as anything other than a misspelling.
friendship though they see each other in class every day, they have only been close enough to text each other outside of class for one week.

Example 25

A1: Kloq 2:56 PM

Qué lo que

What’s up

B1: Klk 2:56 PM

Qué lo que

What’s up

A2: Ctt 3:01 PM

Como tu ta

What’s going on

B2: Bn y tu 3:01 PM

Bien y tu

Good and you

A3: Bn k ases

Bien que haces

Good what are you doing

Every exchange in this conversation involves stylized speech, from abbreviations such as “bn” for *bien* (good) to the use of the k-cluster, it is apparent that they are using the most familiar, informal register of Txt. In spoken conversation, the use of highly stylized features (alternative word forms, reduced articulation, etc.) are hypothesized to to express social closeness and to preserve the positive face of the receiver (Adams, 2009; Horn, 2010; Levinson, 1983). The Txt correlate of these
appears to be the use of abbreviations and initialisms, therefore these features are understood here as signaling the social closeness of the interlocutors. Notably in this conversation, Participant A asks Participant B about his state of being twice. The first time B replies (in B1), he mirrors what Participant A said though in a different (but still highly stylized form). Participant A then tries again, using another initialism to inquire into B’s state (“ctt”). Then when B replies, A follows up for a third time asking what B is doing at the time. Before there is any information exchanged, A has asked B three times about his current state of being. This is taken as evidence that a primary function of this part of the conversation is to establish rapport and acknowledge B’s positive face rather than exchange information. If the information exchange was the only purpose of this conversation, A would have texted the message in A3 earlier in the conversation (as he does in other situations). A final key element of this exchange is the use of “kloq” and “klk.” These both refer to the spoken form, “qué lo que,” but are expressed differently. They both rely on the k-cluster and they are both initialisms, and while “klk” is the more common form, they are both stylized ways of expressing the same question. What this shows is the role that individuality and creativity play in text messaging. Participant B uses the more creative and novel respellings than Participant A does, indicating that the way a word is respelled is also meant to convey ones individual identity. Other researchers have focused on the role of spelling alternations and identity (Baron, 2004; Jaffe et al., 2012; Ling & Baron, 2007b; Spears et al., 2009; Tagg, 2009), and the finding that spelling choice is a key identifier for texters is confirmed here.

Taken together, Example 24 and Example 25 illustrate scenarios where the meaning being exchanged is not necessarily the literal meaning of the words, but rather function to establish the social relationship between the interlocutors on the way to exchanging information. By flouting the prescriptive rules, texters are signaling their social closeness and the importance of preserving
positive face as well as indicating that there is more meaning behind the message than being overtly stated.

4.4 Abbreviations Conclusion

Overall, abbreviations are a powerful tool texters use in order to convey meaning in the Txt register. Abbreviations operate in parallel to the role that non-prescriptive forms have in spoken communication by saving positive face and indicating social closeness (Adams, 2009; Austin, 1975; Horn, 2010; Levinson, 1983). Furthermore, these features are used to express that there is more to a message than the literal meaning of the words. As in other corpora (Jaffe et al., 2012; Spears et al., 2009; Tagg, 2009), participants in this study incorporate their own linguistic creativity in a complex and effective process of communication and identity presentation. In this corpus, abbreviations come in three different shapes, omissions, reductions, and clusters. While these three may have an even more specific meaning behind them (i.e., expressing identity or creativity), it appears that each type does the same functional work for the texter. The transmission ofTxt abbreviations is person-to-person, and casual in nature, in contrast to the transmission of abjads, which is through explicit instruction in an educational setting, an environment very different from socially-motivated colloquial communication.

For an abbreviation to be felicitously used, it must be recoverable by the receiver. If the abbreviation is not recoverable, the texter must define the meaning of the abbreviation in order to incorporate the receiver into his speech community. In this way, abbreviations function to both signal in-group membership or to create an in-group among the interlocutors. As far as the function of abbreviations is concerned, it appears that they have three primary purposes. First, they are used to save positive face with the receiver and indicating their social closeness. Second, they are used to signal that the receiver is part of the texter’s in-group. Third, they are used to indicate the illocutionary force of a message.
5 Initialisms

Initialisms are phrases or expressions that are represented by the first letter of each word. Colloquially these are called acronyms, but there is some disagreement as to whether acronyms must be pronounceable or not (Cannon, 1989; Grange & Bloom, 2000; Harley, 2004; Javarone & Armano, 2013; Khalifa, 2015; Tagg, 2009). Therefore the broader term, initialism, is used following the tradition of other computer mediated communication researcher (Crystal, 2001; Leonardi, Huysman, & Steinfield, 2013; Tagg, 2009; Uygur-Distexhe, 2014). Initialisms did not emerge just because people started typing their correspondence. Initialisms probably date back to the beginning of writing, though there are not very clear records to show that (Cannon, 1989; Grange & Bloom, 2000; A. W. Read, 1963). There is, however, a very meticulous and lengthy dictionary of initialisms dating back to 1475 that identifies and defines every commonly used initialism at the time (Cannon, 1989). Cannon goes into detail about the differences in types of abbreviations that use different numbers of letters. While there are a few examples in this corpus of reductions that are neither clearly an abbreviation nor an initialism (for example, “pork” for both *porque* (because), in Example 26 and *por qué* (why), in Example 27, the broader patterns of initialism emergence and usage are more informative for this project, so ambiguous instances will be identified, but not incorporated into the analysis.

Example 26

Pork estoy muerta en vida.

Porque estoy muerta en vida.

*Because I am the walking dead.*
Example 27

Y *pork* tu te va a matar

Y *por qué* tu te va a matar

*And why are you going to kill me?*

In the late 1800’s, there was a rise in the use of initialisms and acronyms, as they came into fashion both in written and spoken text (A. W. Read, 1963). Though Read never discusses it in his history of “O.K.”, this rise in initialism usage follows the development of the rotary printing press by only a few years, and the corresponding increase of weekly and daily newspapers in 1863-1865 (Daniels & Bright, 1996; Musson, 1958; A. W. Read, 1963). It could be argued that there are three major leaps in production of print media in “recent” history. The first was the development of the Guttenberg printing press in 1440, which made mass printing possible in Europe. The result was the widespread distribution of books and a dramatic increase in literacy rates over the following generations (Clair, 1976; Hoe, 1902). The next major development in printing technology was that of the steam powered rotary printing press, which allowed for fast and cheap production of print (Clair, 1976; Eisenstein, 1980; Musson, 1958). This is when newspaper production dramatically increases, the price drops, and suddenly the majority of people in Western Europe and the United States have access to information about current events (Clair, 1976; Eisenstein, 1980; Musson, 1958). This increase in newspaper production occurs roughly 10 years before a dramatic increase in the use of acronyms (A. W. Read, 1963). Again we see that in the years following a new text-based communication technology, the use of initialisms and acronyms increases as well (Cannon, 1989; 68 It must be acknowledged just how new writing is in human history. Humans have used writing technology for language for roughly 5,500 years, with the earliest forms originating in Mesopotamia and Egypt in 3200 BC (Daniels & Bright, 1996). If homo sapiens originated 100,000-200,000 years ago (Fleagle, 2013), writing has been present for less than 5% of human history. 69 Moveable type printing was invented in China in 1040 (Needham, 2004), but it is generally not considered to have the same influence on Western History as the Guttenberg press.)
Crystal, 2009; Javarone & Armano, 2013; A. W. Read, 1963; Tagg, 2009). It is likely that this rapid increase in text-based communication (i.e., newspapers, the Internet) catalyzed a corresponding spike in acronyms. After each of these innovations, the amount of print-based content people read and write increases dramatically, so the amount of print they come in contact with hits an always unprecedented level. This dissertation proposes that a byproduct of this to this increase of print-based language to play with the way it is written, and to express creativity and identity through initialisms and acronyms.

This phenomenon is occurring now as a result of the most recent development in communication technology: digital communication platforms. Beginning with message boards and the intranet at CERN in the 1980’s (Bush, 1992; Segal, 1995) and then the Internet in the 1990’s (Gromov, 1998; Hauben & Hauben, 1998), and the smartphone in the 2000’s (Cougnon & Fairon, 2014; Lenhart, 2012), print-based communication is used for both widespread (newspapers, blogs, magazines) and individual communication (text messages and sms’s, email, Instagram). With digital communication technologies, once a user has access to a computer or cellphone, there is a nearly limitless amount of media to consume at virtually no cost to the consumer. There is no paper, no material consumption involved70 (i.e., a screen can continually display novel text): digital communication technologies allow people to produce and consume print-based media almost endlessly.71 Correspondingly, individuals are reading and writing even more than before, resulting in another wave of initialisms to play with the language. Crystal points out that initialisms and acronyms do not arise from texting and other computer mediated communication (2001, 2011).

---

70 Setting aside things like electricity and time.
71 Both Yochai Benkler and Lawrence Lessing have written about the production of knowledge, information, and culture are vastly different from the production of physical objects in the world. They both argue that the rise of the Internet and digital platforms allow for production and consumption in ways never seen before and at a rate never before witnessed (Benkler, 2006; Lessig, 2005).
Rather, they come in and out of fashion over time, though the popularity of written communication has likely contributed to their current omnipresence (Crystal, 2001).

Dictionaries and analyses of initialisms have been maximally inclusive, treating those initialisms used only once with the same attention as frequent initialisms (S. Anderson, 2015; Cannon, 1989; Crystal, 2009; Dobbins, 2015; Godschalk, 2013; Khalifa, 2015; Maltais, 2012; McWhorter, 2013a; Thurlow & Brown, 2003). Unlike these research programs, the focus here is on finding the common patterns, attempting to identify what factors are involved that lead to an initialism becoming lexicalized into common usage. Therefore, uncommon initialisms\(^{72}\) (those that were listed by participants but only used one or two times) are not taken into account here, though the comprehensive list that participants developed to identify abbreviations and initialisms is included in Appendix 7. Most initialisms occur somewhere between 4 and 200 times each in the 44,597 message corpus (see Table 15 for the complete list of initialisms considered here).

The initialisms included in Table 15 were used by at least two participants. If only one participant used an initialism, it is included in the complete list in Appendix 7.

Table 15 – All initialisms

<table>
<thead>
<tr>
<th>Initialism</th>
<th>Translation</th>
<th>Main Group of Texters (excluding lovers)</th>
<th>Overall Frequency (including lovers)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ok</td>
<td>(All correct)</td>
<td>434</td>
<td>1980</td>
</tr>
<tr>
<td>lol</td>
<td>Laugh out loud</td>
<td>336</td>
<td>6215</td>
</tr>
<tr>
<td>wyd</td>
<td>What You Doing?</td>
<td>118</td>
<td>118</td>
</tr>
<tr>
<td>idk</td>
<td>I don’t know</td>
<td>41</td>
<td>195</td>
</tr>
<tr>
<td>lmao</td>
<td>Laughing My Ass Off</td>
<td>27</td>
<td>121</td>
</tr>
</tbody>
</table>

\(^{72}\) These are distinct from novel initialisms, which interestingly are exceedingly rare in the corpus (i.e., there is no identifiable instance).
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Text Description</th>
<th>Frequency</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>jk</td>
<td>Just Kidding</td>
<td>19</td>
<td>67</td>
</tr>
<tr>
<td>omg</td>
<td>Oh my God</td>
<td>17</td>
<td>73</td>
</tr>
<tr>
<td>klk</td>
<td>Que Lo Que  ( (What's up?) )</td>
<td>15</td>
<td>16</td>
</tr>
<tr>
<td>ctt</td>
<td>Como Tu Ta  (^{1})  ( (How are you?) )</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>tk</td>
<td>Te Kiero  ( (I love you) )</td>
<td>11</td>
<td>34</td>
</tr>
<tr>
<td>gm</td>
<td>Good Morning</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>wya</td>
<td>Where You At?</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>qtl</td>
<td>Que Tal  ( (What's going on?) )</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>hbu</td>
<td>How ‘Bout You?</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>brb</td>
<td>Be Right Back</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>wtf</td>
<td>What The Fuck</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>tkm</td>
<td>Te Kiero Mucho  ( (I love you a lot) )</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>bae</td>
<td>Before Anyone Else</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>idc</td>
<td>I Don't Care</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>wbu</td>
<td>What ‘Bout You?</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>ttyl</td>
<td>Talk To You Later</td>
<td>5</td>
<td>81</td>
</tr>
<tr>
<td>lml</td>
<td>Laugh Mad Loud</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>gn</td>
<td>Good Night</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>tbh</td>
<td>To Be Honest</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>otp</td>
<td>One True Pair</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>lmfao</td>
<td>Laughing My Fucking Ass Off</td>
<td>2</td>
<td>5</td>
</tr>
</tbody>
</table>

\(^{1}\) ¿Como estas tu?
Participants identified all twenty-six initialisms in Table 15 during the abbreviation activity, indicating that these initialisms are in students’ txt vocabulary and can be understood by most members of this linguistic community.

5.1 Shape of Initialisms

Initialisms appear to be even more restricted than abbreviations in terms of formation. There are no initialisms in this corpus that are used only once. In fact, when asked about making up initialisms, participants said that it was an unacceptable practice because the receiver would not understand what the texter was trying to convey. It appears that an initialism must be in common use before participants will adopt it. This may be an effect of students’ bilingual status, as newcomers to English, and in an English-centric setting, participants may not see themselves at the forefront of language change. However, it appears as though some initialisms such as “otp” (one true pair) have only been in mainstream use for one to two years (“Urban Dictionary: Textism,” 2015), though participants use it in their messages as in Example 28.

---

74 Even in a bilingual or dual language program, many of the tests are in English and the focus (at the time of writing) is on learning English, not maintaining bilingualism (Baker, 2011; Carlo et al., 2004; Menken, 2013).

75 This is unsurprising as in any linguistic situation, “ownership” of the language is by those with the most prestige or power. English Language Learners in an alternative high school are not the group with the most social power or linguistic prestige. In fact, they are regularly confronted with the fact that mainstream media largely excludes their language (i.e., phone settings are in either Spanish or English, but not both, tests are in either Spanish or English, but not both, and the English test is considered more prestigious by administrators of their school). Given these factors, they are understandably hesitant to take ownership over the language to the point where novel initialisms would be acceptable to them.
Example 28

This message is in response to a delayed silence from the receiver. The sender of this message doubts her status as the receiver’s girlfriend, and accuses him the next day of cheating on her. “Babe” refers to the receiver and “otp” in this instance is likely sarcastic.

Babe didn’t even want to talk to me otp 😒

Babe didn’t even want to talk to me one true pair 😁

Twenty-one of the twenty-six initialisms participants used are in English (80%), with the remaining five in Spanish. In this corpus, participants occasionally write a Spanish message with an English initialism inserted into it as in Example 29. It is more common, however, for English messages to have initialisms than Spanish messages.

Example 29

Babe mi telef esta en 2%. Lo voy a poner a cargar ttyl 😘

Babe mi telefono esta en 2%. Lo voy a poner a cargar talk to you later 😘

Babe my phone is at 2%. I’m going to charge it talk to you later 😘

The use of English initialisms in Spanish messages and students’ responses during the acronym exercise suggest that many initialisms are lexicalized into the txt vocabulary of bilingual youth as inseparable units. When asked about the meaning of these items, most participants did not know what the individual letters stand for. They were aware of the meaning for the Spanish items, but nearly all were unaware for items such as “otp” (one true pair), “ttyl” (talk to you later), “tbh” (to be

---

76 Unamused face emoji
77 Kiss mark emoji
honest). Yet students still reported using them and incorporating them into their txt vocabulary, even if they did not know the meaning of each letter they know the meaning of the initialism as if it were its own word. For example, one participant was able to use “ttyl” (talk to you later) to close a conversation and express that they will talk again soon, but she did not know what each letter stands for. When asked what it means, she replied “t-t-y-l means ‘I gotta go’ or ‘see you later’, I don’t know what the letters mean.” Example 30 illustrates this. A is using “tbh” as a noun (as evidenced by the use of the determiner, “un” (a)) meaning ‘an instance in which you tell the truth or are honest.’ “Tbh” is not being used as a tag in this instance, as the texter is not giving her honest opinion, but rather, asking the receiver to be honest with her.

Example 30

Para que quiere saber para decir me que es un tbh

Para que quiere saber para decir me que es un tbh

So that you know to tell me it’s a tbh

Having command of the usage without the meaning may initially appear contradictory. However, it is not that participants do not know what the words mean when they are used in a sentence. Rather, they know the meaning of the initialism as a concept or expression, but they may not know where the expression came from or how the initialism was derived. This is how most people use most words anyhow. For the vast majority of words we use everyday, it is unlikely that we know where each one came from as it became a part of the English language. Ultimately, it does not matter if a participant does not know what the letters of “otp” stand for if they are able to use it to describe a perfect relationship.

Most of the initialisms listed in Table 15 are three letters, with the exception of “ttyl” (talk to you later), “lmao”, “lmfao”, “ok”, “jk” (just kidding), “tk” (te kiero I love you), “gm” (good morning)
and “gn” (good night). Out of twenty-six abbreviations, three are made of four or more letters (12%), nineteen are three letters (73%), and four are made of only two letters (15%). There is no immediately obvious reason as to why three is a preferred number of letters for an initialism, but it may have to do with the length of utterances that can be initialized. Any phrase that is longer than three letters may be more difficult to recover and therefore less likely to be used among mainstream texters. Of course there are plenty of initialisms that are longer than three letters (i.e., “tnstaaf” (there’s no such thing as a free lunch), or “diku” (do I know you)), however, while they may appear in chat rooms or in specialist communities (Bush, 1992; Godschalk, 2013; Grange & Bloom, 2000), they are not used by the participants in this study.

There are many possible reasons for why three letter initialisms are preferred. One reason that three letter initialisms are so popular could be that three word phrases are more common in both Spanish and English. This is an unlikely explanation since many of the initialisms used here have four word or two word alternatives “wya” (where you at) could also be written as “waya” (where are you at), and “ctt” (como tu ta~78~) could also be written as “ct” (como tas). The three letter version is heavily preferred for both of these words.

Another reason could be that after three letters, there are too many competing possibilities for what each letter could mean. With the addition of each letter, there are nearly infinitely many more possibilities of what the initialism could represent. With each added ambiguity, it becomes more difficult to guess successfully at the meaning, and more difficult to remember what the meaning is (Grange & Bloom, 2000). Both of these difficulties make it harder for the initialism to gain popularity since transmission from one person to another reduces drastically. This would also explain the widespread usage of longer and more complicated initialisms in chat rooms and Multi-User Domains (games played on the internet). The communities that develop in these spaces are

---

78 ¿Como tu estas?
communicating exclusively through a chat interface. In that environment, it becomes increasingly important to establish initialisms and abbreviations to allow for faster typing since typing is much slower than speaking. It is also easier to teach newcomers the vocabulary of a particular chat room or Multi-User Domain since the community is self-contained to that environment (Backer, 1999; Seneff, Goddeau, Pao, & Polifroni, 1996). Finally, users may be more willing to learn the language of a given chat room or Multi-User Domain to establish a we-code that bonds the users together in a space where all of the bonding activity must take place through computer mediated interactions. Language and language choice may be among the most powerful ways to show group membership when communication is exclusively text based.

5.1.1 Longer Initialisms

As stated in Section 5.1, most of the initialisms are three letters long. There are, however, three common initialisms that are longer than three letters. Of the longer initialisms, two are expressions of laughter that involve profanity (“lmao” and “lmfao”), and the other (“ttyl”) emerged when chat rooms first became popular (Crystal, 1997). When chat rooms were first becoming popular, hundreds of abbreviations appeared as well as lists to help new users decode the respellings. Many of the initialisms that emerged during this time relate to the status of the author, either updating the recipient on their physical status (“brb” (be right back), “ttyn” (talk to you later), “gtg” (got to go)), or commenting on a statement (“imho” (in my humble opinion), “afaik” (as far as I know)) (Bush, 1992; del-Teso-Craviotto, 2008; Maness, 2008; Merchant, 2001; Sykes, 2005). Most of those initialisms disappeared as they either went out of fashion, or they became less useful as the mediums evolved. Of the early chat room abbreviations, “ttyl” has persisted with this population and has a function in text messaging among bilingual youth (see Example 29, repeated here).
Example 29

Babe mi telef esta en 2%. Lo voy a poner a cargar ttyl 🥰

Babe mi telefono esta en 2%. Lo voy a poner a cargar talk to you later 🥰

Babe my phone is at 2%. I’m going to charge it talk to you later 🥰

“Ttyl” is unique among many of these initialisms because it saves positive face of the receiver. Leaving the conversation open to be continued in the future allows the texter to indicate that he or she must leave the conversation, but in a way that does not completely end the conversation. Whereas “brb” or “gtg” may convey the same sentiment that the conversation has ended, “ttyl” leaves an opening for the future, and involves the receiver in the texter announcing their departure. Therefore, “ttyl” may be popular among this population because it has the benefit of preserving the positive face of the receiver. The other two long initialisms both have to do with laughter (“lmao” and “lmfao”), as in Example 31.

Example 31

A: I could sleep for you when u be working 😁😁😁 and u never gonna be tired 😁😁😁


Nearly 5% of all the initialisms, interjections, words, and emojis used in this corpus refer to laughter (including every instance of “haha” or “jaja” and its equivalents). The frequency that laughing types are used indicates that laughter is an extremely important part of Txt. “Lol” is rarely

---

79 Face with tears of joy emoji
used to signal laughter, possible because it is used for a variety of other purposes. In most cases, participants use “haha”, “jaja”, “jiji”, or “jeje” to signal the sound of laughter as in Example 32.

Example 32

A: Con un afro tu vera

With an afro on your side

B: Ok jajaja

The difference between these methods and the initialism is that the duplicated laughter syllables indicate the sound of laughter whereas the initialism is used to identify that the texter finds something to be very funny and is laughing. There are a variety of other initialisms that were developed during the height of chat rooms in the late 1990’s that could have just as easily persisted in their usage. The most common one may be “rotfl” (rolling on the floor laughing), or “limh” (laughing in my head). Yet these do not seem to be nearly as popular in this corpus or in other large text messaging corpora (Bernicot et al., 2012; Tagg, 2009; Tagliamonte & Denis, 2008).

There are three factors that may have combined to give rise to the popularity of “lmao” and “lmfao” from a common initialism to the most common initialism expressing laughter (Urban Dictionary, 2015). First, there was a need for an acronym that could express laughter and humor and was not “lol.” Second, using profanity is a way for youth to mark that they own their language, and have a register that is separate from teachers and other authority figures in their lives (Coats, 2015; Cressman, Callister, Robinson, & Near, 2009; Kaye & Sapolsky, 2004; Shankar, 2008). When participants where giving translations of the initialisms, they were very hesitant to say the words, as we were in a classroom setting, where profanity is not allowed. Finally, there is a very popular pop band, LMFAO, who had a hit single in 2012, ‘Party Rock Anthem’ (Billboard, 2015). Their
popularity has served to domesticate the initialism, adding another layer of familiarity to both “lmfao” and “lmao” since they are really versions of each other.

Finally, even though these long initialisms are commonly used throughout this corpus, they are exceptions to the general tendency towards the preference for three letter initialisms among this population as well as texters more broadly (Crystal, 2009; Dobbins, 2015; Khalifa, 2015).

5.1.2 Shorter Initialisms

There are four shorter initialisms that were commonly used by participants. By the very definition of initialism, the only phrases shorter than three words are two words since a one letter initialism is an abbreviation rather than an initialism of a phrase. The four initialisms are “tk” (te quiero I love you), “gm” (good morning), “gn” (good night), and “jk” (just kidding). These four initialisms are also exceptions to the three letter tendency for varying reasons.

First, “tk” is an alternative version of “tkm” (te quiero mucho I love you a lot). In Spanish, this is a type of love that is expressed between friends, or as a term of endearment rather than necessarily an expression of romantic love between two people (serious, romantic love is expressed with “te amo”). It most often occurs in a position of conversation closure as in Example 33.

Example 33

 Ok babe 😘😘 t80 tk ttyl

 Ok babe 😘😘 te quiero talk to you later

Ok babe 😘😘 I love you talk to you later

Throughout the messages, both “tk” and “tkm” are used in this way, as a tag to soften the end of a conversation, to ensure that the receiver knows that the texter cares about them, but still has to end

80 Face throwing kiss emoji
the conversation. Yet for many of these participants, they are in an ambiguously defined relationship with the recipient of their messages. They may be romantically involved, but may not be committed to that involvement. Dropping the “m” from “tkm” may serve as a way to temper the expression of feelings. That is, “tk” is less emotive than “tkm.” This explanation is reinforced by the presence of “tkm” versus “tk” on Twitter. On December 1, 2015, there were 18,543 instances of “tkm” in either a Spanish, English, or Spanish/English bilingual tweet. In the same time frame, there were 242 instances of “tk” in tweets in the same language set. This indicates that while “tk” is common in this corpus, it is much less common than “tkm” more broadly. One interpretation is that, “tk” may be used more by this population as a way to de-escalate the emotional content of “tkm.”

The next short abbreviations that will be addressed are “gm” (good morning, Example 34 and “gn” (good night, Example 35).

Example 34

A: Good morning 😊

B: Gm 😊

Good morning 😊

Example 35

M voy a dormi 😩 byee Gn 😊 k duermas cn dios

Me voy a dormi 😩 bye good night 😍 que duermas con dios

I am going to sleep 😩 bye good night 😍 sleep with god (sleep well)

---

81 Weary face emoji
Both of these are greetings – either openings or closings to a message. There are very few message openings or closings that are three words long (“wyd” (what you doing?) is the only one in this corpus). Combes, Volckaert-Legrier, and Largy found that in French and English SMS messages, the opening and closing to a message are the most often respelled or abbreviated (2012). They argue that because they are the most frequently used parts of a message, they are the abbreviated for maximal efficiency (Combes et al., 2012). It has been argued is this research that recoverability is a more significant factor than efficiency in terms of what can be abbreviated and what must be fully articulated. Building on this approach, “good morning” and “good night” are very easily recoverable, and almost exclusively occur in the opening or closing position of the message at a particular time of day or night. There are very few expressions that could be used in those positions in either Spanish or English, adding to the recoverability because they are expected; once the recoverability condition is met, they may be abbreviated for the sake of efficiency, as Combes et al. suggest (2012).

Finally, there is the case of “jk” (just kidding, Example 36).

Example 36

Jk lol I know you really wanted to see that

“Jk”, like “ttyl” emerged during the early years of chat rooms (Crystal, 1997; Godschalk, 2013), acting as a tag to ensure that statements were taken as jokes, not at literal interpretations of the preceding message. As discussed in Sections 4.4 and 5.4, this is the primary function of initialisms and abbreviations is to signal that the intended meaning of the message is not necessarily what the literal meaning may suggest. “Jk” is an anomaly in this respect since it explicitly tells the receiver that the message is intended as a joke. There is no guessing, no interpretation. The texter who uses “jk” is overtly identifying the illocutionary force of the message to the receiver. No other initialism does this as all of the other initialisms that serve this purpose allow the receiver to guess.
about the intended meaning, leaving some ambiguity. It may be because “jk” serves such a unique purpose that it has persisted so robustly.

5.2 Special Initialisms

While there are broader pragmatic functions that all initialisms perform (Section 5.3), certain initialisms have a special status. This section will trace the birth of a word in response to textisms through analyzing “bae” (before anyone else), identify the order of operations for initialism formation by identifying the key features of “hbu” (how about you?) and “wbu” (what about you?), and finally discuss how “klk” (que lo que? what’s up?) and “wya” (where you at?) illustrate the role of African American English on bilingual text messaging.

5.2.1 bae – the birth of a new word

“Bae” refers to the best, or most important person in one’s life (Roberts, 2014; Steinmetz, 2014). Participants use it as a term of endearment as in Example 37.

Example 37

A: Wyd baby?

What you doing baby?

B: In clases bae y tu

In classes before anyone else y tu

In classes before anyone else and you

Spelling out “bae” is awkward and grammatically does not make sense in context. This may contribute to the tendency for participants to use this initialism but not necessarily know what the letters stand for.
“Bae” does not have to refer to a significant other, but must be someone the speaker feels is important. Many people think it is an abbreviation for Beyoncé. It is more likely an initialism for “before anyone else” (Dobbins, 2015). It is commonly accepted that the first appearance of “bae” was in hip hop songs in early 2005 (Roberts, 2014), and became popular in mainstream culture because of the rise of the internet and spike in youth communication across the world. Usage of “bae” increased dramatically in 2012, as it quickly became a common pet name (Todd, 2015). “Bae” has recently been used in a wide variety of media, including the television show, “Master of None” (Anasari, 2015), and Pharell Williams’ song, ‘Come Get It Bae’ (Williams, 2014), indicating that it has taken hold in mainstream communication. What is surprising about the usage of “bae” in “Master of None” is that it was not highlighted even though the word is relatively new. That is, the characters continued their dialogue, and “bae” was not in the punch line.

“Bae” is unique because while it has become common in mainstream speech, it is not fully lexicalized into the vocabulary of English. Notably, within four years, it has gone from a nearly unknown hip-hop term to a common and widespread pet name. Still, its usage is restricted, and it functions as an initialism or a proper noun without the privileges of common nouns (such as the term of endearment, “boo”). In the written Acceptability Judgment Task, participants indicated which environment were acceptable and unacceptable for “bae.” Unlike “boo” (another common pet name that preceded “bae”) or “baby”, it cannot follow a possessive adjective (such as “my” or “her”). This is shown in Examples 38 and 39; the * means that the sentence is unacceptable to native speakers (so Example 38 is unacceptable whereas Example 39 is acceptable).

Example 38

*(unattested in the corpus or on Twitter)*

* Wanda’s bae is coming to the party.

* My bae is the most beautiful.
Example 39

*(attested on Twitter)*

Bae’s working late, gotta make TV dinner.

Beyoncé is bae.

This is in contrast to the distribution of “boo”, which is allowed under possession as in Example 40.

Example 40

*(attested on Twitter)*

Jay-Z’s boo is the best dancer ever.

My boo’s on her way home!

To illustrate how often “boo” and “bae” are used in rap song lyrics, Figure 18 illustrates the frequency of “boo” versus “bae” since 1990 (this chart was generated on RapGenius.com (Todd, 2015), developed by Todd (2014)).
This chart shows that even though “bae” has received significant attention and made its way into mainstream media, it is not yet as frequently used as “boo” was at its height. Given this situation, it is expected that while students can use “bae” felicitously without fully knowing where it comes from, it is still too young of a word to have the full distribution of a more established noun. This is an excellent example of how a word comes into a language as it first appeared over ten years ago, and even though it has experienced a sharp increase in usage, it has not been used long enough or consistently enough to take on the status of a fully incorporated noun in terms of the morphology it can accept and the environments it can be in. That is, it is restricted and functions as a proper noun.

In recent years, words such as “Google” and “Xerox” have undergone similar processes. Both emerged as proper nouns but can now be used as verbs in all cases where a standard verb has appeared. The difference between these and “bae” is that “bae” emerged from pop culture whereas “Google” and “Xerox” were companies. “Bae” began its existence as an initialism and is being
incorporated into the language as a noun whereas both “Google” and “Xerox” entered as nouns and are used to describe actions (Wolfram & Schilling, 2015). This makes “bae” a unique item that, should it persist for another generation, it may become lexicalized to the extent that “boo” is.

5.2.3 hbu & wbu – order of operations

Many initialisms have undergone a series of phonological and graphemic transformations from the standard, prescriptive way of writing each word. Assuming that the derivation of lexicalized and semi-lexicalized follows a series of set rules, the status of “hbu” can shed light on the order of rule application. Most participants report that ‘hbu’ is an abbreviation of “how about you,” as in Example 41.

Example 41

A: Bueno y que hace
Bueno y que haces

Good and what are you doing

B: Asiendo brownies hbu

Haciendo brownies how ‘bout you

Making brownies how about you

Two students stated that it is an abbreviation for “how be you.” The “how be you” interpretation represents the “b” accurately, but it is unlikely because it is not a commonly used expression in English and it is not a direct translation from Spanish. Secondly, if the translation is “how be you,” it could just as likely be found to open a conversation as in the second turn of a conversation whereas “how about you” can only be used in the second turn of a conversation, which reflects its distribution. Finally, if the translation is “how be you,” it could only be used in the greeting position, never as a tag question. “Hbu” is regularly found in the tag question position, shown in the
conversation in Example 41. So, the “how be you” translation participants offered is most likely an effect of decontextualized reflection and focusing on the shape of the abbreviation rather than the meaning. Based on usage and other participants' translations for “hbu”, the actual representation must be for “how about you.” If this is the case, the shape would be “hay”.

There are two distinct processes at work here first is the use of “u” to represent “you,” and the second is the reduction of “about” to “bout.” The first process is referred to as the u-cluster in Section 3.2.2 of this chapter, where its derivation is discussed. The second process, the reduction of “about” is common in spoken English. Therefore the ‘about’ in ‘hbu’ must lose the ‘a’ before it is abbreviated, indicating word-initial vowel drop precedes the process of abbreviation. It seems apparent that all reduction processes occur before abbreviation, que→k, you→u. The following order of operations is proposed to explain how these items are formed:

Figure 20 – Respelling order of operations

Step 1: Clustering (i.e., u-cluster, k-cluster, etc.)

Step 2: Phonetic Reductions (i.e., about → ‘bout)

Step 3: Remove all but first letter of each word

When a step is not relevant, it is omitted. This order accounts not only for the formation of “hbu” and “wya,” but also “ctt” and “klk.”

Both of the cluster reductions (k-cluster and u-cluster) are more commonly used in abbreviations than on their own. Likewise, similar to the u-cluster, while the reduction of “about” to “bout” might appear extremely common, it is actually far more common for participants to write out the entire word as in Example 42.

The “bout” reduction is the non-traditional form and therefore marked as an alternative spelling.
Example 42

A1: Hi, how’s ur day going?

B1: Good doing home the whole day. How about you

B2: Homework* (* is a common indication of repair referring to a previously made mistake.)

A2: I slept all day and rushed to work, and now I'm working, I was going to text you when I got home

In Example 42, line B1, the same question is being asked as in Example 41, but in a more fully articulated format. The result is that A answers with a detailed account of what he did all day. In the previous messages, B was wondering where A was (since he had not texted yet). Therefore, B was not trying to convey positive feelings, but rather make a very specific inquiry, and correspondingly, B spells out the tag question, “how about you.” This shows that B expects an answer from A. If B did not expect a full account of A’s day, B could have texted “hbu”, as she does in other instances.

The difference between Example 41 and Example 42, where the tag question is in the initialism form versus the fully articulated form combined with research on the role that articulation plays on semantic meaning (Klatt, 1976; Wright, 2004) suggests that the use of initialisms may indicate that the illocutionary force of the utterance is more significant than the locutionary force when an initialism is present. In face-to-face, spoken conversation, researchers have found that precise articulation, including phonetic stress, a decelerated rate of speech, and increased syllable duration of a word or utterance signals to the listener to focus on that part of the utterance (Klatt, 1976; Wright, 2004). This does not suggest that the non-emphasized part does not carry meaning, but it does suggest that the meaning is not the focus of the utterance. Translating this into a texting
environment, where it is impossible to aurally create these effects, the effects must be interpreted visually. One piece of this reinterpretation is the use of initialisms.

With an initialism such as “hbu”, there are two processes occurring. The first is that the overall shape of the initialism is heavily reduced, indicating that the reason for writing “hbu” is more for pragmatic effect, to indicate that the texter is inquiring about the receiver than it is to actually gather precise information from the receiver. To gather precise information, the texter would emphasize the meaning of the utterance, as in Example 42.

This means that the semantic focus of the message must be either be encoded in another part of the message or completely pragmatic, intended save the positive face of the receiver.

5.2.4 klk & wya – the role of AAE in Dominican Spanish

African American English (AAE) is a variety of English spoken predominately by working-class Urban Americans, many of whom are bi-dialectal (between AAE and General American English (GAE)). This dialect was formerly spoken mostly by Black and African Americans, but has recently been shown to be spoken by people from many different races for a variety of purposes (Bailey, 2002; Chun, 2001; Cecilia Cutler, 2003; Gooden & Eberhardt, 2007; Igoudin, 2011; Reyes, 2005). The varieties of Spanish and English spoken by Latinos in New York City is influenced by AAE (Bailey, 2002; Wolfram, 1974). Much of the earlier research focused on Puerto Ricans (Wolfram, 1974), but the findings can be extended to other Latinos, including Dominicans and Mexicans (the dominant groups in the present study) (Bailey, 2002; Slomanson & Newman, 2004). The influence of AAE on bilingual youths’ text messages is readily apparent as well. The surprising part is that syntactic features of AAE are readily apparent on both English and Spanish initialisms.

---

82 African Americans are not to be confused with African Immigrants to the United States, who generally do not speak AAE.
In their groundbreaking work on New York City Spanish, Otheguy, Zentella, and Livert find that the Spanish spoken in New York City is influenced by a wide variety of Spanishes from Latin America and the Caribbean (2007) as well as English. They find that while there is a split between Caribbean and Mainland origin Spanish, the groups are both influenced by the English and other Spanishes they come in contact with in New York City (Otheguy, Zentella, & Livert, 2007b).

Spanish and English (like all major languages) spoken in one geographic area is different from the variety spoken in another. This fact has a significant effect on the language spoken by a population as heterogeneous as the Spanish-speaking population in New York City. Compared to NYC Spanish more generally, the Spanish of Puerto Rico is under represented in this study, and the Spanish of the Dominican Republic is over represented. As mentioned in previous chapters, the participants in this study were largely born in the Dominican Republic, followed by Mexico and Ecuador. Therefore the Spanish represented in this study is largely Dominican influenced.

The influence of AAE on Dominican Spanish has been analyzed for spoken language (Satterfield & Alexander, 2006; Satterfield & Benkí, 2009) from a variety of perspectives. Researchers have investigated the role that establishing an “American” identity and stylization have on the incorporation of AAE into Spanish (Bailey, 2002; Pavlenko & Norton, 2007) as well as the role that contact in schools and communities has on these languages (Bailey, 2002; Otheguy et al., 2007b). The dialect that has emerged from this contact has been referred to as *Lengua Reggaetona* (LR) (Satterfield & Alexander, 2006; Satterfield & Benkí, 2009). Focusing on phonological features, Satterfield and Alexander point to spoken examples such as “*Tú no eres calle*” (You ain’t street), and “*Ni que fronteando*” (Wasn’t even frontin’).

---

83 This example is from Las Guanábanas’ 2004 song, “Bien Bellaca”. Las Guanábanas are a Reggaeton group from Puerto Rico who mix Dominican and Puerto Rican Spanish in their music and are considered one of the first Reggaeton bands to become popular in Latin America and the United States.
Following the work of other researchers on the influence of AAE on speech communities (Chun, 2001; Cecilia Cutler, 2003; Gooden & Eberhardt, 2007; Igoudin, 2011; Reyes, 2005; Satterfield & Benki, 2009; Satterfield & Alexander, 2006), some of the initialisms used by participants are most likely the result of contact between Spanish and AAE. Key features of AAE that are relevant to this discussion are described in Table 16, with comparisons to both GAE and Academic Latin American Spanish (ALAS).

Table 16 – Key features of African American English

<table>
<thead>
<tr>
<th></th>
<th>AAE</th>
<th>GAE</th>
<th>ALAS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Negation</strong></td>
<td>Negative concord is necessary. No Negative Polarity Items</td>
<td>Negation only marked once. Negative Polarity Items</td>
<td>Negative concord is allowed. N Negative Polarity Items</td>
</tr>
<tr>
<td></td>
<td>She ain’t got no money</td>
<td>*She doesn’t have no money</td>
<td>*Ella no tiene nada dinero</td>
</tr>
<tr>
<td></td>
<td>She ain’t got any money</td>
<td>She doesn’t have any money</td>
<td>*Ella no tiene alguno dinero</td>
</tr>
<tr>
<td></td>
<td>She ain’t got money</td>
<td>She doesn’t have money</td>
<td>Ella no tiene dinero</td>
</tr>
<tr>
<td><strong>Copula</strong></td>
<td>Copula is dropped</td>
<td>Copula is present</td>
<td>Copula is present</td>
</tr>
<tr>
<td>- progressive verbs</td>
<td>*She dancin’</td>
<td>She is dancing</td>
<td>Ella está bailando</td>
</tr>
<tr>
<td></td>
<td>What you eating</td>
<td>¿Qué estas comiendo?</td>
<td></td>
</tr>
<tr>
<td><strong>Copula – inflection</strong></td>
<td>Infinitive copula is allowed</td>
<td>Copula is inflected for person</td>
<td>Copula is inflected for person</td>
</tr>
<tr>
<td></td>
<td>How you be?</td>
<td>How are you?</td>
<td>Como estas?</td>
</tr>
<tr>
<td></td>
<td>She be working (regularly)</td>
<td>She is working</td>
<td>Ella estas trabajando</td>
</tr>
</tbody>
</table>

84 Negative Polarity Items are words such as “any” that only follow negative words and signal to intensify the effect of the negation.
There are many more features of AAE that are not listed here, however these are largely not relevant to the discussion at hand (for more information, see (Labov, 2006; Sidnell, 2012).

Significantly, a very robust feature of AAE is copula-drop. There are many languages that allow speakers to optionally drop the copula, including American Sign Language, Arabic, Berber, Luganda, Bengali, and many more. However, it is most likely that Latinos in New York City are in contact with more AAE speakers than speakers of any other copula-drop language. The variety of English that urban Latinos speak is heavily influenced by AAE since native AAE speakers and Latinos go to the same schools, live in similar neighborhoods, and come in contact across all life’s domains (Garcia & Menken, 2006; Satterfield & Alexander, 2006). Features of AAE are seen in the Spanish texted by this group as well. Contact has been influencing the Spanish spoken by New York City residents for a very long time (Otheguy & Zentella, 2012; Otheguy et al., 2007b; Slomanson & Newman, 2004; Wolfram, 1974; Zentella, 2002), but this is the first time it has been documented in the txt language form.

The influence of AAE can be seen in the Spanish initialisms participants use. Three initialisms exist in the Txt of New York City Dominicans that appear to be Spanish translations from AAE. The first is “klk” for “qué lo que” (Example 43), which is a way of greeting someone or asking “how are you?”.

Example 43

A: Hello!
B: Klk

This greeting is not present in conventional Latin American Spanish, and they are heavily marked as being Dominican in nature as reported by participants of Dominican, Mexican, and Ecuadorian origin.
Because the k-cluster has been investigated as a unique form of abbreviation resulting from language contact, “Klk” will be addressed first. “Klk” stands for “qué lo que” and loosely means “how are you?” or “what’s up?” The literal translation is “what the what” or “what what.” This is not an acceptable question in Academic Latin American Spanish (ALAS), and literally translated, it does not make sense in English either. However, taking into account the role that AAE plays in the English these students come in contact with, “klk” may have arisen as a result of linguistic contact between native AAE speakers and bilingual Dominicans. One possible origin of “klk” comes from the ALAS question, ¿qué es lo que esta pasando? (lit.: what is it that is happening, or “what’s happening”). Stepwise, it could be derived by dropping the second half of this utterance (a process that is extremely common in greetings (for example, “how are you doing?” is shortened to “how are you?”)), the result is ¿qué es lo que? On its own, this is grammatically unacceptable, and does not make sense. Literally translated, it means “what is the what?” or “what is it that?” Elided utterances such as “what is it?” are common in both AAE and GAE, especially when they are easily recoverable in the course of conversation. However, dropping ‘to be’ verbs is less common in GAE, though extremely common in AAE. So, the transition from ¿qué es lo que? to ¿qué __ lo que? could be the result of copula drop, and applying a syntactic rule from AAE to Spanish. From there, the k-cluster is applied, and lo is represented by its initial letter, resulting in “klk”, a highly stylized way to say, “what’s up.”

Finally, it is clear that participants are aware of how stylized “klk” is, and how marked this form is. Whenever it was discussed, participants indicated that it was an acceptable thing to say, but that it also implies that the speaker is joking or trying to be funny. This further indicates that initialisms, especially highly stylized initialisms such as “klk” signal to the receiver that the illocutionary force of the utterance is more significant than the literal meaning.
Finally, the influence of AAE on the English used by participants can be seen in the use of the initialism, “wya” for “where you at”. “Where you at” is a question in AAE, but not in GAE since copula-drop is not allowed in GAE. The copula must have been dropped before it became an initialism, “wya”, otherwise it would be “waya.” It could be argued that auxiliaries and other function words are omitted in the process of deriving the initialism. However, if that were the case, initialisms such as “fomo” (fear of missing out) would be writing “fmo”. While “fomo” does not appear in this corpus, it does appear in the list of abbreviations and acronyms that participants offered when asked to name all the abbreviations and acronyms they could remember.

Taken together, these three initialisms indicate that AAE is influencing the Txt of bilingual Dominican youth, and that phonological and dialectical rules must be applied before the process of initialization. More broadly, this reinforces other researchers’ findings that initialisms and stylization are influenced by spoken social language (Akbari, 2013; Crystal, 2009; Tagg, 2009). In the process of being adapted into txt, they are converted into initialisms as a way to play with already stylized language and capture the colloquial connotations associated with these utterances. It is no coincidence that the utterances used in initialisms are strictly colloquial, and were formerly restricted to spoken social communication. In spoken communication, these utterances set the “tone” of the conversation, by indicating that the communication will be social, there may be a light or humorous aspect to it (Attardo, 2001, 2008). If these utterances (qué lo que, como tu ta, where you at) were written in their full forms, that tone and playfulness with the language would be lost whereas the use of an initialism preserves the broader purpose of using these phrases.

5.3 Pragmatic Role of Initialisms

The majority of the initialisms that participants use are in English. Of the 24 initialisms used in this corpus (excluding “lol” and “ok”), only five are in Spanish, listed in
Table 17 with their respective frequencies. By contrast, 68% of the main corpus is in Spanish, 30% is in English, and 3% is a single message that combines both languages. Yet only 8% of the initialisms used are in Spanish (excluding “lol” and “ok”).

Table 17 – Spanish initialisms

<table>
<thead>
<tr>
<th>Initialism</th>
<th>Translation</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tk</td>
<td>Te quiero (I love you)</td>
<td>18</td>
</tr>
<tr>
<td>klk</td>
<td>Que Lo Que (What’s up?)</td>
<td>16</td>
</tr>
<tr>
<td>ctt</td>
<td>Como Tu Ta (How are you?)</td>
<td>15</td>
</tr>
<tr>
<td>qtl</td>
<td>Que Tal (What’s going on?)</td>
<td>12</td>
</tr>
<tr>
<td>tkm</td>
<td>Te Kiero Mucho (I love you a lot)</td>
<td>8</td>
</tr>
</tbody>
</table>

There is a clear discrepancy between the language of the messages more broadly, and the language surrounding initialisms. This discrepancy may be due to the fact that there are simply more initialisms lexicalized into English txt than Spanish txt. If this were the case, it would be expected that participants would text in Spanish and use an English initialism to illustrate that sentiment.

While this type of usage (an English initialism inserted into a Spanish text message) accounts for 19% of all initialisms used, the most common usage by far is fully English text messages with English initialisms. Sixty-three percent of all messages involving initialisms are completely English (i.e., the message and the initialism are both in English). Only 18% of the messages are completely in Spanish. This is almost exactly opposite the overall corpus.

Upon closer examination, it is clear that the people who use the initialisms in their text messages are those who use more English overall. Likewise, they have higher overall English scores and report having more English speaking friends. It is important to remember that English is a
second language for participants, and if participants can communicate in English, it is certain that they can communicate just as well socially in Spanish. That does not, however, suggest that a participant who is proficient with English initialisms is necessarily proficient with Spanish initialisms. It may be the case that they have been exposed to more English on the txt platform. In fact, it is most certainly the case that they have been exposed to more English txt since they report having more English speaking friends than other participants. Therefore, this dissertation proposes that exposure to and use of initialisms and stylized Txt language positively influence language ability.

Similar hypotheses have been put forth for the role of texting on children’s literacy skills. Researchers have found that texting fluency is positively correlated with academic literacy (Plester et al., 2008, 2009; Wood et al., 2011), and that texting frequency is negatively correlated with academic literacy (Kemp & Bushnell, 2011). In this study, it appears that the participants who text in English are more capable of using English txt, more flexible and playful with the language, and more proficient with academic English. There are six participants who produced more than half of the initialisms in this corpus. Other participants incorporated “lol”, “ok”, and the k-cluster, but did not rely on initialisms to the extent that those six participants did.

The majority of the initialisms fall into one of two categories: inquiries into the state of the receiver, and interjections as indicated in Table 18.

This basic set of categories have been used by Tagg (2009), Sotillo (2012), and Accorsi et al. (2014) to classify and analyze text messages. The remaining initialisms fall into one of three other categories: Relationships, laughter, and greetings.
Table 18 – Initialisms by word function

<table>
<thead>
<tr>
<th>Inquiries into the state of the receiver</th>
<th>Interjections</th>
<th>Relationships</th>
<th>Laughter</th>
<th>Greetings</th>
</tr>
</thead>
<tbody>
<tr>
<td>wyd</td>
<td>Idk</td>
<td>Bae</td>
<td>Lmao</td>
<td>Ttyl</td>
</tr>
<tr>
<td>klk</td>
<td>Omg</td>
<td>Tk</td>
<td>Lmfao</td>
<td>Gm</td>
</tr>
<tr>
<td>ctt</td>
<td>Jk</td>
<td>Tkm</td>
<td>Lml</td>
<td>Gn</td>
</tr>
<tr>
<td>qtl</td>
<td>Wtf</td>
<td>Otpl</td>
<td></td>
<td>Brb</td>
</tr>
<tr>
<td>way</td>
<td>Idc</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hbu</td>
<td>Tbh</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wbu</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Almost all of these initialisms are tags that either comment on an utterance or ask for non-essential information from the receiver. In face-to-face conversations in both Spanish and English, it is customary to ask about the state of the speaker. This is generally performed with some variation of “how are you?” and a reply that is a variation on “good, you?” The one exception to this is when the conversation begins with some variation of “hello” or “good morning.” Deviating from this pattern is considered very strange, and a violation of the pragmatic norms of a conversation (Horn, 2010; Yule, 2010). It comes as no surprise then that there are so many ways to abbreviate the first turn in this sequence. In fact, this is what is found in both this corpus as well as in French and English corpora (Combes et al., 2012; Tagg, 2009). This suggests that any of the items in the ‘inquiries into the state of the receiver’ and ‘greetings’ categories can be omitted without losing the locutionary force of the message. However, the illocutionary force may be unrecoverable without these items. These initialisms are therefore used not so much to convey the meaning of the message, but to
negotiate the social interaction between interlocutors. By using a more colloquial form, such as “ctt” or “wyd”, the texter is highlighting the receiver that he trusts that the receiver will be able to interpret the alternative form and that he considers him or her to be a member of their in-group (Holtgraves & Paul, 2013; Ling & Baron, 2007b; Sotillo, 2012). Likewise, the use of initialisms and alternative spellings is often read as playful, or as a light and less serious way to communicate (Tagg, 2009). Therefore using these initialisms indicates to the receiver that the conversation is taking on a lighter tone. In face-to-face conversation, this may be conveyed through body language, smiles, or tone of voice.

All except one of these five categories is used either to ask a question or as a message greeting or closure. In effect, this means that these items are not contributing any semantic content to the message, only serving to ask for information or mediate social interactions. The only exception to this may be the relationships abbreviations. It is possible to construct a message where the semantic content of the utterance is indicated by the initialism as in Example 44.

Example 44

*Not attested in the corpus* \(^{85}\)

**Bae** just walked in!

This statement indicates who “bae” is in relationship to the speaker and notes that “bae” is now in the room. Even though “bae” is an initialism and is almost never fully written out, it still carries the locutionary force of the utterance. However, it should be noted that there are many ways to express this content without the use of “bae,” for example, by stating the person’s name, using a pronoun, or another descriptive noun. However, the use of “bae” lets the receiver know that the texter

\(^{85}\) This example is taken from Twitter, and is only used for illustration purposes because uses of “bae” in the corpus are contextually dependent and therefore do not illustrate this point as clearly.
considers the person walking in to be their “bae,” indicating that Example 44 has an important meaning to it in addition to the literal meaning of the words. That is, had the texter written, “Wanda just walked in!” the receiver would not associate any feelings or pay any special attention to Wanda unless they knew how the texter feels about Wanda. However, by sending the message in Example 44, the sender has flagged this person as their “bae”. Consequentially, the receiver should know that the sender may be excited or nervous, and that the person walking in has a special meaning to the sender (and the receiver should pay extra attention to the person walking in). The sub textual meaning of this utterance is much richer than the literal meaning that a person just walked in.

Each category will be addressed separately to begin with to identify both the unique features as well as develop a theory of the role that initialisms play in Txt.

5.3.1 Initialisms of opening questions and greetings

Seven of the twenty-six initialisms in this corpus are inquiries into the current state of the receiver. Three of these are in Spanish: “klk” (qué lo que what’s up), “ctt” (como tu ta how are you), and “qtU” (qué tal how are you). The remaining four are in English: “wyd” (what you doing), “wya” (where you at), “hbu” (how about you), and “wbu” (what about you). The conversation in Example 45 illustrates three of these initialisms.
Notably, the Spanish inquiries are versions of “how are you” whereas there are no initialisms of the English version of “how are you.” One can imagine the correlates to the Spanish questions in English (“hru” (how are you), “wup” (what’s up), “hub” (how you be)), but none of these were used or offered by participants. These and similar initialisms are conspicuously absent from other large monolingual English corpora as well, and are extremely rare on Twitter (“hru”, “wup”, and “hub” were used between 1 and 4 times per week on Twitter in December 2015), suggesting that making an initialism out of the most common conversation starting question is not a common way to use English Txt.

One explanation for this builds on previous research on code switching in text messages and the implicit rules of respelling associated with each language. Deumert and Masinyana found that bilingual isiXhosa/English texters treated their languages very differently when writing text messages (2008). They found that texters were unwilling to abbreviate isiXhosa words though they would abbreviate English freely. Most isiXhosa words were embedded in otherwise English text messages,

---

86 Smiling face with sunglasses emoji
87 Hibiscus emoji
making the avoidance of respellings even more striking when compared to the neighboring words (which were respelled freely). They conclude that there is a language-specific set of rules that govern how a language is texted (Deumert & Masinyana, 2008). By extending this explanation to the English/Spanish language pair, the discrepancy between Spanish and English initialisms can be explained. Deumert and Masinyana say that the ability to abbreviate is a part of the pragmatics of the language (2008). Rather than having a strict binary (i.e., respelling is or is not allowed), it may be the case that the type of respelling allowed is language dependent. That is, it may be the case that English does not allow initialization of opening questions, but does allow initialization of tag questions and commentary. This would account for the absence of opening question initialisms, but the large amount of commentary initialisms (discussed in the next section). That is, part of the pragmatic txt rules of English is that opening questions must either be fully spelled out or omitted.

Comparing these forms with the fully spelled out forms, there are 52 instances of “how are you”, and 49 instances of “what’s up”. In the list that participants produced, inventoring all abbreviations and acronyms (initialisms) they could think of, there were no instances of English opening question initialisms. On the other hand, Spanish allows initialized opening questions, hence the presence of “ctt”, “qtl”, and “klk”. These abbreviations are not as common as the fully spelled out variations. “Como estas” (the most common way to ask ‘how are you?’) occurs 73 times (Example 46). Que tal (what’s up) occurs 16 times (Example 47), most of these instances are more accurately referring to how something is going rather than how someone is.

Example 46

Hola mi amor como estas mami.

Hello my love how are you mami.
Bien bien estoy estudiando historia y esta feio :( que tal el work :

*Good good I'm studying history and it's cold :( how's the work ;)*

The fully spelled out version of “como tu ta” occurs only once, and there are 4 instances of “que lo que” fully spelled out. Overall, there are 101 fully written out greeting questions in English and 94 in Spanish. Again, given that the corpus overall is Spanish dominant, this is a surprising ratio, and may indicate that English is used more often for greetings and social facilitation than it is for content.

Finally, it may be the case that “gm” (good morning, Example 48) is used to signal the connotations of an initialism to the receiver (i.e., a casual or playful tone) yet adhere to the convention that conversation opening questions are not initialized in English.

Example 48

Gm amor

*Good morning love*

There are very few of these type of abbreviations in English and Spanish txt (“gm” is the only one in this corpus), yet according to Combes, Volckaert-Legrier, and Largy, the conversation openings and closing are the most common type of respelling in French Txt (2012). Taken together, this is evidence that not every phrase can be respelled in every language, and there are language specific rules that determine how abbreviations and respellings are made, and what categories of words and phrases are eligible for respelling. The rules can be as extreme as no respellings are ever allowed (as in the case of isiXhosa (Deumert & Masinyana, 2008), or as lenient as nearly anything can be respelled (as in the case of Indonesian (Barendregt, 2008).
On the other hand, it appears that in English Txt, conversation closing phrases and expressions are eligible candidates for respelling. For example, participants regularly use “ttyl” (talk to you later) and “brb” (be right back) to close a conversation or pause it for a moment. Between these two, “ttyl” is used more frequently, as it is found appended to or inserted in otherwise Spanish (Example 49) and English (Example 50) text messages. It also serves as the intersection between Spanish and English (Example 51).

Example 49

Babe mi telef esta en 2%. Lo voy a poner a cargar ttyl

*Babe my phone is at 2%[charge]. I'm going to charge it talk to you later*

Example 50

Babe ttyl 🙃 I'm on the train

*Babe talk to you later I'm on the train*

Example 51

Ok! So ttyl- me voy a bailar, & then I'm going to the hair salon 😊😊

*Ok! So talk to you later I am going to [dance], and then I’m going to the hair salon 😊😊*

Conversely, “brb” is only found in otherwise English text messages, never in Spanish or bilingual messages. It appears as though “ttyl” has been adopted by this speech community as a productive way to indicate that one needs to step away from the phone or otherwise end a conversation.

---

88 Smiling face with smiling eyes emoji
89 Grinning face with smiling eyes emoji
Similar to the conspicuous absence of English conversation opening questions, Spanish expressions of conversation closure are never found in an initialism form. It is not that Spanish conversation closures do not exist. In fact, they are very robust, including “hasta luego” and “hasta mañana,” which could be initialized as “hl” and “hm,” respectively, yet they are never found this way in this corpus, and it is extremely rare on Twitter as well (there were zero instance the week of December 3-10, 2015). Again, this points to a convention of Spanish that closing expressions cannot be made into initialisms. This is, however, an important and frequent turn in a conversation. In spoken communication, this is often accomplished through lengthening words, and a change in voice frequency or tone (Condon & Cech, 2003; Jordan et al., 2012). Because this communicative function is so common, texters need a way to encode this into the messages. One way to encode it is to fully spell out when the next conversation will be. This is regularly written in both Spanish and English (versions of hasta luego and “hasta TIME/DAY” occur 45 times in this corpus). Another way to accomplish this conversation function is to use an initialism such as “ttyl” which both accomplishes the task and encodes a tone of playfulness and casualness to the message (as discussed in Section 5.3). Having access to both Spanish and English, participants in this study are able to rely on English for an initialism to convey the sub textual meaning while performing the function of closing the conversation and still adhere to the norms of Spanish respellings. The purpose of each message in this section is summarized in Table 19.
Table 19 – Greetings initialism summary

<table>
<thead>
<tr>
<th>Initialism</th>
<th>Translation</th>
<th>Positive or negative</th>
<th>Communicative function</th>
<th>Additional meaning(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>klk</td>
<td>Que lo que?</td>
<td>Positive</td>
<td></td>
<td>Open a conversation</td>
</tr>
<tr>
<td>ctt</td>
<td>Como tu ta?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gm</td>
<td>Good morning</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>qtl</td>
<td>Que Tal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ttyl</td>
<td>Talk to you later</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>gn</td>
<td>Good night</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wyd</td>
<td>What you doing?</td>
<td></td>
<td>Positive</td>
<td>Inquire into the</td>
</tr>
<tr>
<td>wya</td>
<td>Where you at?</td>
<td></td>
<td></td>
<td>current state of the receiver</td>
</tr>
<tr>
<td>hbu</td>
<td>How ‘bout you?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>wbu</td>
<td>What ‘about you?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**5.3.3 Initialisms commenting on a message & laughing**

Another function of initialisms is to provide commentary on the preceding statement. There are nine initialisms of this type. Three of these are meant to signal actual laughter, “lmao” (laughing my ass off), “lml” (laughing mad loud), “lmfao” (laughing my fucking ass off); the rest are common discourse expressions, “idk” (I don’t know), “omg” (oh my god), “jk” (just kidding), “wtf” (what the fuck), “ide” (I don’t care), “tbh” (to be honest). Even excluding “lol” and “ok”, initialisms of this type are the most common in the corpus, accounting for 75% of all initialisms used by participants. Initialisms expressing laughter account for 17% of the corpus, and commentary accounts for 58%. The percentage of laughter is shocking given the most common way to express laughter is ‘haha’ or
‘jaja’ and its variants, of which there are 1,988 instances in the corpus (about 1/3 as many instances as “lol”, but 16 times as many instances as all laughter acronyms combined). This shows that there is a lot of laughing occurring on text messaging, but the question is why is laughter so important to text messaging. People laugh in conversation but certainly not at this rate. This section will examine initialisms for laughter first and then show how other commentary initialisms serve the same communicative function as laughter initialisms.

5.3.3.1 Initialism Laughter

Researchers have established that laughter in conversation has a wide variety of different roles (Attardo, 2008), including signaling in-group membership (Archakis & Tsakona, 2005), establishing rapport and consensus (Adelswärd, 1989), constructing and maintaining positive relationships in a professional setting (Holmes, 2006) and signaling the power structure of relationships (Rees & Monrouxe, 2010). The range of uses for laughter in this corpus is much smaller since the messages are largely between friends, not between co-workers, so the professional setting is irrelevant to this study. Likewise, the type of laughter that signals a power differential is within an utterance, not at the beginning or the end of an utterance. Within this corpus, only utterance-initial and utterance-final laughter occurs, suggesting that in Txt, laughter is not a mechanism for signaling power relationships. That leaves establishing in-group membership and rapport and consensus. Both of these types of laughter are found throughout this corpus.

It has already been discussed that one reason to use an initialism rather than a fully spelled out expression is to indicate social closeness. In this case, “haha” and its variants is the more fully spelled out version of “lml”, “lmao”, and “lmfao.” In Example 52, both users have iPhones and use spelling suggestions in English. This means that in order to use any non-standard word form (such as “lmao” for “laughing my ass off”), the user either needs to program it in to their phone or use it regularly enough that the iPhone dictionary has learning that it is an acceptable form for the user. In
either case, the use of “lmao” is intentional as it takes more effort on this platform to type “lmao” than “ha ha” (the English iPhone dictionary has “ha ha” by default, but not “lmao”).

Example 52

A:  Okay 😊90 be careful in there! There's a lot of things that could happen.

B:  Lmao 😄😄😄😄😄91

In this example, A’s comment is genuinely humorous, playing on the fact that B is heading to school, and it is a relatively safe place. However, B probably is not laughing as hard as “laughing my ass off” might suggest. Therefore, while it does seem to be indicating laughter, it is also serving the function of both saving A’s positive face, and establishing rapport and consensus (as a function that laughter performs). In this example, “ha ha” would have a more assertive tone here, one that still signals consensus, but not necessarily saving A’s positive face (i.e., “ha ha” would be appropriate if B understood A’s comment as not trusting her, since they are in a new relationship). Similarly, “lol” is inappropriate because while it still establishes rapport and consensus, it does not signal that B interpreted A’s comment as funny. Therefore, the use of “lmao” serves 3 purposes:

1. Because it is an initialism, it saves positive face.
2. Because it signals real laughter, it indicates that B actually laughed and found the comment humorous.
3. Because it is indicative of any type of laughter, it established rapport and consensus.

These functions can, of course, be signaled in other ways in addition to using “lmao”. However, in every instance, “lmao” minimally serves these three purposes as in the conversation in Example 53.

90 Winking face emoji
91 Face with tears of joy emoji
Example 53

A: Hehehe that ugly niggah is like my bro 😂😂

Hehe that ugly niggah is like my brother 😂😂

A: Sorry again baby that wasn't for you

B: 😰😢😢😢😢

A: lmao 😂

In this example, A initially thinks he is talking to someone else. He uses the fully spelled out form of laughter, indicating that he has multiple forms in his Txt vocabulary. Then he realizes that he made a mistake and sent the message to the wrong person. B replies with a series of emojis to indicate that she understands, but to also make fun of him in a light-hearted way. A then responds with “lmao” again indicated that he wants to save B’s positive face, signal that he found her response funny, and establish that they understand each other. Similar examples occur throughout the corpus, where “lmao,” and “lmfao” are used to indicate laughter in situations where the interlocutors have established rapport yet need to continue saving positive face and signal that they find the situation humorous. This need is heightened on all-text platforms because of the lack of non-verbal cues common in face-to-face communication.

“Lml” is slightly different from the other two initialisms. It only occurs seven times in the corpus, but every time it indicates a contradiction. That is, “lml” signals laughter because there is a contradiction, or the texter is saying something that is opposite to the reality, as in Example 54 and Example 55.
Example 54

A: N its funny how yu asking for facebook but yu said yu cant use ya internet lml

No it’s funny how you asking for [my] facebook but you said you can’t use your internet laughing

mad loud

In this example, A points out that the contradiction is funny and proceeds to use the “lml” initialism to indicate that he is laughing at the contradiction the other person has made. A still wants to preserve positive face of the receiver, and uses the initialism to indicate that, but is also flagging the laughter as sarcastic, since A still gives his facebook profile information in a later message.
Example 55

A1:  Ven sola Si vas a venir 😊😊😊

Come alone if you are coming

A2:  And hurry let me know 😊  

B1:  Why 😊😊😊

A3:  Cuz ima kiss you 😊😊😊😊😊😊😊😊😊😊😊😊😊😊 😊😊

Because I’m going to kiss you 😊😊😊😊😊😊😊😊😊😊😊😊😊😊 lauhging mad 😊😊

B2:  😊😊😊😊😊😊😊😊😊😊😊😊😊😊 soooo kiss me infront of my bestfriend 😊😊😊😊😊😊😊😊😊😊😊😊😊😊

A4:  I was joking about come alone come with her 😊😊😊😊😊😊😊😊😊😊😊😊😊😊

In Example 55, A has been flirting with B, but has not been willing to show his affection in public. Therefore, both participants know that while he may want to kiss her, there is something preventing him from doing so in social situations. The “lml” indicates positive face and humor, but it also signals that the humor results from a contradiction. In this way, “lml” is more similar to “lol.” But here, he may actually be laughing. If he were to use “lol” he might not actually find the contradiction

94 Loudly crying emoji
95 Weary face emoji
96 Face blowing a kiss emoji
97 Face savoring delicious food emoji
98 Person raising both hands in celebration emoji
99 Splashing sweat emoji
100 Face with tears of joy emoji
101 Another interesting feature of this exchange is that “mad” is, itself, highly stylized, emerging from hip hop culture in New York City, and being strongly associated with the variety of English spoken in colloquial settings throughout New York City (Cecelia Cutler, 2008).
102 Two hearts emoji
103 See-no-evil-monkey emoji
104 Speak-no-evil monkey emoji
105 Eyes emoji
106 White smiling emoji
funny. If he were to use “lmao” it would appear more like he is laughing at her as opposed to
laughing at the situation where he has not kissed her in public, but he might kiss her now. Therefore,
even though it indicate laughter, “lml” is slightly different from “lmao”; being able to make a
distinction between types of laughter is crucial to communicating effectively in Txt since there are
no auditory or visual cues to signal to the listener how to interpret the message.

5.3.3.2 Initialisms to comment

In addition to initialisms for laughter, there are six\textsuperscript{107} initialisms that are commonly used as
commentary to a message that all serve to indicate rapport, and save positive face. Each one,
however, has a nuanced meaning in addition to establishing rapport and saving positive face. Unlike
initialisms for greetings, these initialisms encode meaning that is crucial to keep the conversation
going and ensure understanding. Many of these initialisms would be replaced by facial expressions,
body language, non-verbal aural responses, interjections, or other cues common in face-to-face
conversation (Cassell et al., 2001; Fraser, 1996; Kendon, 1995). In a completely text based platform,
the function that non-verbal discourse markers play in conversation and communication. Each of
these will be taken in turn, from most common to least common.

In their analysis of the “IDK, my BFF Jill” (a popular tag line in the colloquial speech of
teenagers originating from a YouTube video) Jones and Schieffelin note that part of the novelty of
the expression arose from Txt being articulated \{Citation\}. Their analysis underlines the extend to
which “idk” is indicative of the Txt language form. The initialism, “idk” (I don’t know) occurs 195
times in this corpus, accounting for 25% of all initialisms in the entire corpus (excluding “lol” and

\textsuperscript{107} There are actually seven of this type of initialism, however “tbh” (to be honest) occurs only four times in the corpus.
In each instance, it is in an otherwise Spanish text message, and is being used as a noun, written in all capital letters.
When participants identified this initialism when they were listing the abbreviations that they use, they indicated that it
means “to be honest” and that it is used on facebook. However, in this corpus it appears following an article. Since most
of the participants know each other and form one speech community among themselves, the instances of “TBH” are
most likely a novel in-group initialism signifying something other than a discourse marker.
“ok”), and is the most common single initialism (again, aside from “lol” and “ok”). In nearly 1 out of 3 instances, “idk” is used in the same message as “lol” as in Example 56. In slightly more than 1 out of 3 instances, “idk” is used by itself, Example 57. In the remaining one-third of instances, it is used in a sentence to stand for “I don’t know,” Example 58. These are three very distinct uses, and may account for why “idk” is so popular.

Example 56

A: Ah okay 😁😁 ¹⁰⁸ he got ur number?
B: Idk

* I don’t know

Example 57

Lol idk she just said she was thinkin about them

*Lol I don’t know she just said she was thinkin about them*

Example 58

So idk why you still love me

*So I don’t know why you still love me*

Previous research on the usage of “I don’t know” among bilingual populations shows that it is extremely common in discourse and provides a way to hedge both the content of the message as well as the language of the message (Baumgarten & House, 2010). This most likely accounts for the third type of “idk” (Example 58) as it is used within the context of an utterance. Within the corpus,

¹⁰⁸ Flushed face emoji
there are 27 instances of “I don’t know” and its variants (i.e., “I dont know”), and 67 instances of “no lo se” (I don’t know) and its variants (i.e., no se), suggesting that if a texter needs to explicitly state that she does not know something, she can use the fully articulated form in either Spanish or English. Interestingly, there is no initialism for “no lo se,” suggesting that it either cannot be initialized in Spanish for pragmatic reasons or that these initialisms are directly adapted from early chat rooms which were largely in English109 (Chapelle, 2003; Lam, 2004; Merchant, 2001).

There are a variety of situations in which “I don’t know” is communicated in face-to-face conversation that are not literal instances of ‘not knowing’. Rather, “I don’t know” can mean “I don’t want to say” (parallel to Example 56), “I don’t want to commit to an answer” or “I know, but I don’t want to tell you,” for example. In face-to-face conversations, these alternative meanings are often signaled in the surrounding non-verbal cues, for example, a shrug, hand gesture, facial expression, change of voice, etc. All of these may suggest that “I don’t know” actually has an alternative meaning. In Txt, this contextual and nuanced meaning must be written out and encoded in writing. It appears as though texters use “lol” to notify the receiver that “idk” has an alternative meaning and should be interpreted based on the surrounding information. In the three examples below (Examples 59-61) “idk” is combined with “lol” to indicate that “idk” should not be interpreted literally (more on this in Section 5). Likewise, similar to all other acronyms, “idk” is used here to save positive face of the receiver, indicating that the message should be interpreted with a positive sentiment.

---

109 As mentioned previously, this dissertation favors the pragmatic explanation for this phenomenon for languages such as Spanish, which were widespread and global before the Internet and are written in the Roman alphabet, factors which contribute to the development of Spanish CMC.
Example 59

A: Tu le dijiste a [NAME] que yo estaba en tu casa el lunes?

You told [NAME] that I was at your house on Monday?

B: Oh ok! Idk how did he knows lol he told my mom that lol

Oh ok! I don’t know how be knew that lol he told my mom that lol

In this example, B is attempting to diffuse the situation and signal to A that she is innocent regardless if she knows the information or not. The combination of “idk” with “lol” suggests that A should interpret the message as B claiming her innocence (further conversation shows that A was not at B’s house after all). If B was taking the conversation completely seriously and literally, she likely would have spelled out “I don’t know” rather than using the initialism. A, on the other hand, is sending a more serious tone since there are no respellings, emojis, or emoticons in the message, and he follows prescriptive writing conventions with capitalization and punctuation.

Example 60

A: Seguro estan hablando de deportes, mujeres 😒, tenis idk lol

Surely they are talking about sports, ladies 😒, tennis I don’t know lol

In Example 60, A is making fun of the receiver who has been complaining about sports on the television. He is probably using “idk” in a sarcastic manner, signaling that women (the receiver of this message in particular) do not know anything about sports, and are not interested. “Idk” alone would not convey that message, and neither would “lol” alone. Rather, the combination of the two conveys a message both of ignorance and of playfulness.

110 Unamused face emoji
Example 61

A: Tomorrow is the day babe lol [to cut your hair]

B: Idk lol

Finally, in this example, “idk lol” is used to signal uncertainly rather than actually not knowing. The uncertainty results from being unsure about allowing A to cut her hair. This is not a literal interpretation of “I don’t know” rather, B does know and is signaling to A that her message should be interpreted as uncertainty about the haircutting.

Looking across these examples, it becomes apparent that the combination of “idk” with “lol” serves a communicative purpose notifying the receiver that the sender means something other than situational ignorance, and is indicating an alternative meaning. This is not the only instance where the receiver is expected to interpret “idk” with one of its alternative meanings, however. In very many situations, the initialism, “idk” on its own suggests that the texter is uncertain, hedging their response, hiding information, or is being sarcastic as in Example 62.

Example 62

A: Idk yo no quiero hablar con ellos

I don’t know. I don’t like to talk to them

In this example, A does actually know that she does not like talking to them, but she is softening the message with “idk” and indicating that the receiver should interpret the message beyond the literal meaning of the words. In this situation, it would be inappropriate to use any initialism signaling laughter to tell the receiver to interpret the message non-literally since the feelings are not positive. Therefore, in this situation, “idk” has a similar communicative function to “lol” where it is indicating that the illocutionary force does not match the locutionary force of the utterance, yet there is a more
complex meaning behind it. Another example is found in Example 63 where B is acknowledging A’s message, but not actually indicating that she does not know, rather she is more likely confused and resigned to the situation she is in.

Example 63

A: His situation is complicated
B: Idk!
B: I think you're right! He should be alone for a while

Again, B is responding to A’s message, acknowledging the receipt of it, affirming the sentiment, and indicating that she has negative feelings.

This is in contrast to the type of “idk” that actually indicates that the texter does not know, this almost never occurs in response to a question, but rather is used at the beginning of an utterance in lieu of the phrase “I don’t know” as in Example 64.

Example 64

A: Calling my mother but she doesn't answer the phone
A: Idk where she's at

I don't know where she's at

In this example, A genuinely does not know where her mother is at, and rather than writing out the whole phrase, “I don’t know”, she has chosen to use the initialism. This is one of the few instances where the initialism may be used for efficiency purposes. It still appears to be only used with people in a texter’s inner circle, and the alternative, “I don’t know” and “no lo se” are equally common. It may be the case that the abbreviated form is used when rapport has already been established and the
locutionary force of the message is the key feature whereas in situations where the interlocutors are more socially distant, the fully spelled out version is utilized.

Ultimately, it seems that texters need a way to express acknowledgement and signal to a receiver to attend to the illocutionary force of a message without the positive feelings that laughter may convey. “Idk” fills this space, serving to signal a wide range of communicative functions including hedging, uncertainty, sarcasm, resignation. In this way, it serves a communicative purpose that the fully spelled out version is unable to fulfill.

Another initialism of this type is “omg” (oh my god). “Omg” accounts for just 9% of the initialisms used by this population even though it is one of the most popular initialisms in the United States (Crystal, 2009; Ling & Baron, 2007b). This may be due to a combination of age and Spanish as the participants in this study would have been between 11-13 years old when “omg” was popular, putting it squarely into the stylized language of an older generation. Likewise, participants in these studies were English monolinguals, and “omg” may just not be as popular among Spanish speakers. Yet, it is still used by the participants in this study. The fully spelled out version, “oh my god” occurs once in the entire corpus, and its Spanish correlate, “dios mio” only occurs six times, indicating that there is a strong preference for the initialized form (which, for comparison, occurs 73 times).

“Omg” may be one of the easiest initialisms to understand since throughout this corpus it only occurs in situations of excitement, surprise or disbelief, and to engage the receiver of the message in the texter’s state of disbelief. “Omg” generally expresses a heightened emotional state; a respondent may be surprised at the texter’s message (Example 65) or excited or surprised by an outside stimulus (Example 66 and 67 respectively).
Example 65

A1: Are you hungry?

B: Nop

A2: Whaat omg lol

What oh my god lol

Through his conversations, B is always hungry and mentions food in nearly every conversation with A, so when B tells A that he is not hungry, she expresses her shock at that information.

Example 66

A: OMG babe relax I only want you 😘

In Example 66, A is expressing her exasperation through the use of “omg.” This is slightly different from all of the other examples presented thus far as the emotional content being encoded in 66 is not heightened in a positive way, but rather, heightened with a negative emotion.

The third initialism of this type is “jk” (just kidding). Comparatively, “jk” is rare, accounting for just 6% of the initialisms used in this corpus. Similar to “omg”, its distribution is restricted and it only occurs as a clarification for when a texter is joking, explicitly telling the receiver that the illocutionary force is the only thing they should pay attention to. This occurs most often in situations where there is some social distance between the interlocutors (i.e., in new relationships, friends who do not text often, etc.), and is used to be sure that the intended meaning of an ambiguous message is understood properly. This is extremely useful as without intonation, body language, or the opportunity for immediate repair, a texter must be certain that a receiver will understand their message as in Example 67.
Example 67

A: You are. My sister told me all about your tweeting at the party.

B: Liar!

A: Lol jk I believe you I'm just teasing

Lol just kidding I believe you I'm just teasing

In this case, A and B are newly texting (this is text number 54-57 in the conversation, all sent over the course of one week), and have not built up the rapport between them for the exchange to definitely be playful. Therefore, A must repair the situation since he cannot be sure that B was not offended by his joke. In this way, “jk” serves as a mechanism to ensure that the intended meaning is understood while maintaining the positive face of the receiver through the use of an initialism. Likewise, it provides a safe way to joke in the context of text messaging (which itself serves to establish rapport (Attardo, 2008)), and be sure that the message is interpreted properly. Finally, “jk” serves as a pragmatic marker of humor when there is potential ambiguity. In this way, it saves positive face and also conveys positive feelings.

The last two initialisms are rare in this corpus, though they are common in monolingual English corpora. The first is “wtf” (what the fuck) and its variant, “tf” (the fuck), combined they occur nineteen times throughout the corpus, all of which are between angry former lovers and exasperated current lovers as in Example 68.

Example 68

A: you know I really like you so wtf 😓 you know how I feel about you and that pissed me off. 😐
In this example, A is expressing her anger at the receiver, and the use of “wtf” serves to signal both anger and disbelief. Since every instance of “wtf” in this corpus is between lovers, it likely has a similar communicative function as “omg” in that both are an attempt to engage the receiver in the heightened emotional state the texter is experiencing.

The next example is “idc” (I don’t care), which occurs a total of 7 times in this corpus, across three different people (Example 69).

Example 69

IDC about her playing around I take you serious

In all but two instances, it stands alone and the preceding messages are not shown, so it cannot be recovered as to how “idc” is being used. In the other two instances, the texter is expressing that they do not care and the wish that their conversation partner would stop bothering them about a topic (i.e., Instagram comments and what to wear outside). Since this is such a small sample, it is difficult to tell if there is a pattern here, and if “idc” also expresses frustration or exacerbation.

In conclusion, the communicative function and associated sentiment of each initialism discussed here is summarized in

Table 20. It is assumed that each initialism saves positive face since that is a fundamental function of all initialisms.
Table 20 – Initialisms to comment summary

<table>
<thead>
<tr>
<th>Initialism</th>
<th>Translation</th>
<th>Positive or negative</th>
<th>Communicative function</th>
<th>Additional meaning(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>idk</td>
<td>I don’t know</td>
<td>Negative, can be made positive with context</td>
<td>Acknowledgement</td>
<td>Hedging, confusion, sarcasm, attend to the illocutionary force of the utterance</td>
</tr>
<tr>
<td>lmao</td>
<td>Laugh my ass off</td>
<td>Positive</td>
<td>Acknowledgement, Agreement, establish rapport, build consensus</td>
<td>Indicates genuine laughter and humor</td>
</tr>
<tr>
<td>lml</td>
<td>Laugh mad loud</td>
<td>Positive</td>
<td>Acknowledgement, Agreement, establish rapport, build consensus</td>
<td>Indicates genuine laughter and humor</td>
</tr>
<tr>
<td>lmfao</td>
<td>Laugh my fucking ass off</td>
<td>Positive</td>
<td>Acknowledgement, Agreement, establish rapport, build consensus</td>
<td>Indicates genuine laughter and humor</td>
</tr>
<tr>
<td>omg</td>
<td>Oh my god</td>
<td>Positive</td>
<td>Engage the receiver in the experience of the stimulus</td>
<td>Express excitement, surprise, or disbelief</td>
</tr>
<tr>
<td>jk</td>
<td>Just kidding</td>
<td>Positive</td>
<td>Clarification, ensure mutual understanding</td>
<td></td>
</tr>
<tr>
<td>wtf (tf)</td>
<td>What the fuck</td>
<td>Negative</td>
<td>Acknowledgement</td>
<td>Surprise, disbelief</td>
</tr>
<tr>
<td>idc</td>
<td>I don’t care</td>
<td>Negative</td>
<td>Acknowledgement</td>
<td>Frustration, exacerbation, resignation</td>
</tr>
</tbody>
</table>
5.3.4 *Initialisms for relationships*

In total, there are four initialisms that describe relationships or feelings within relationships in this corpus. Though participants listed them when listing all of the abbreviations and acronyms they use, these are relatively rare throughout this corpus. The most common of these is “tk” (te quiero *I love you*) occurring 18 times, and its variant, “tkm” (te quiero mucho *I love you a lot*) occurring just 9 times, together accounting for just 3% of all the initialisms used. The vast majority of instances of “tk” and “tkm” occur in message closures as in Example 70.

Example 70

A: Ok babe 😊😊 tk ttyl

Ok babe 😊😊 I love you talk to you later

By contrast, “te quiero (mucho)” occurs 235 times, nearly ten times as often as “tk(m)”, and “I love you” occurs 379 times, though “ily” (I love you) is completely absent from the corpus. Participants listed “ily” as an acronym that they use, but it appears they are not using it in text messages. If participants are regularly exchanging this sentiment in Spanish or English, the initialism version should be easily recoverable, so its scarcity is not a result of complexity or obscurity. Likewise, as an initialism “tkm” and “ily” are allowed, since they appear on the lists, it just appears that they are pragmatically not allowed in most situations.

This may be related to the function of initialisms more generally. If initialisms are primarily used to save positive face and indicate the playful, casual nature of the message, using an initialism for something as strong as an emotion of love would seem contradictory. While an expression of love does signal social closeness and intimate feelings, it is also meant to be taken seriously, not as an expression to indicate friendship, but rather to indicate emotional involvement. When initialisms of love do appear in messages, it is exclusively within the closing sequence. The participants who use
this initialism also refer to each other as “mi amor” (my love), and freely use expressions of love and endearment. Therefore, it appears that “tk” is allowed as an addition to an expression of closure or goodbye, almost as a way to ensure that the expression of closure (i.e., “ttyl” or “good night”) also has an aspect of love encoded in it.

The other two romantically based initialisms are “bae” (before anyone else) and “otp” (one true pair). “Bae” has already been discussed in section 5.2.1, illustrating the birth of a new word. It appears throughout the corpus as an important term of endearment, as in Examples 71 and 72.

Example 71

A1: why somebody put dique " bae 😘" under your picture. Smh you have somebody else 😞

Why did somebody write " bae 😘" under your picture. Shaking my head you have somebody else 😞

B1: Good morning baby 😘😂-proxy and Lmaooo I don’t have nobody else 😞

B2: They comment baby a lot 😞

A2: yeah ok 😘😂-proxy

Example 72

A: In clases bae y tu

In classes bae and you

B: igual bby 😞

The same baby 😞
In Example 71, she knows he is lying, because while they may comment “baby” a lot, they do not comment “bae.” Whoever wrote “bae” on his Instagram wall believes that she is his significant other, and probably, like A, also believes that she is the only one. “Baby” is flexible enough to be used for friends or newly dating couples, “bae” is not. In Example 72, “bae” is a term of endearment, and B responds with another term of endearment, “bby.” By contrast, “bb” and “bby” as an abbreviation for baby or babe occurs six times as often as “bae” in this corpus. Both “bb”/ “bby” and “bae” carry a playful connotation to them that is different from “babe” or “baby.” Many participants report that “bae” is almost ironic; in all four classes, female students gave the advice that “if anyone calls you “bae,” you probably should not trust him.” There is clearly a disconnect between the language that is used between couples and what participants are reporting. However, based on these comments, it appears that “bae” carries an almost sarcastic tone that is either treated as hyperbolic or as not trustworthy.

Within this population, it appears that “babe” is the most popular term of endearment, occurring 4,967 times, far more often than “baby” (685 times), the Spanish form, bebé (42 times), or “mi amor” my love (682 times). Participants use “babe” in almost any situation, so long as the conversation partners are romantically involved.

Finally, “otp” (one true pair) is the least common expression of this type, occurring only 4 times throughout the corpus. It is used by two different participants, both in reference to relationships that have ended, as in Example 73.

Example 73

A: You probably lying you jus don’t wanna be otp

You probably lying you just don’t want to be one true pair.
Example 73, A has hope that they can work out their relationship, but the receiver does not want to be her boyfriend. The choice of “otp” is interesting especially because the other participant who uses “otp” uses it in the same context, trying to encourage the receiver to reconsider the possibility of being “one true pair.” In both of these cases, the texter may be attempting to confront the situation with humor, and allow the receiver to respond with the lightheartedness that the initialism suggests or respond in a serious tone. The meaning and function of relationship based initialisms is summarized in Table 21.

Table 21 – Initialisms for relationships summary

<table>
<thead>
<tr>
<th>Initialism</th>
<th>Translation</th>
<th>Positive or negative</th>
<th>Communicative function</th>
<th>Additional meaning(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>tk(m)</td>
<td>Te quiero (mucho)</td>
<td>Positive</td>
<td></td>
<td>Add love to a greeting</td>
</tr>
<tr>
<td>bae</td>
<td>Before anyone else</td>
<td>Positive</td>
<td>Attend to illocutionary force, hyperbole</td>
<td>Term of endearment, diffuse conflict</td>
</tr>
<tr>
<td>otp</td>
<td>One true Pair</td>
<td>Positive</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

5.4 Initialisms Conclusion

Overall, it appears as though initialisms serve an important purpose in text messages, encoding much of the information that is lost going from a synchronous, face-to-face conversation to an asynchronous, textually mediated conversation. Most initialisms preserve the positive face of the receiver, serving to show the social closeness of the interlocutors and the positive nature of the message being sent. Even when the tone of the initialism is not necessarily positive, the use of an initialism rather than a fully spelled out version indicates that there is a playful, light, or creative nature to the message, and the receiver should not interpret the message seriously or literally. Except
for “bae” and “otp”, the texter always has the option to fully spell out the initialism; evidenced by the presence of the fully spelled out form at least once in the corps.

6 Pragmatic Particles – “Lol” and “Ok”

As mentioned in Section 4, “lol” and “ok” are the two most common initialisms, accounting for nearly all of the instances of initialisms participants use (6086 and 1980 instances, respectively). The initialism, “ok” accounts for 1.3% of all the words used by all of the participants. It is readily apparent that “lol” and “ok” have a unique status in txt as together they occur in 15% of all messages (the total count of words does not take into account emojis, though it does take into account onomatopoeic words such as “haha” and nonce words such as “zobrat”). One factor that must be considered that differentiates “lol” and “ok” from other initialisms is that “lol” is almost exclusively part of the Txt register (Levenson, 2014) whereas “ok” is widely used in casual and professional conversations (J. Brown, 1986). This is even more significant as “ok” has widespread use outside of digital platforms, suggesting that it is an initialism that belongs to all language forms and would be used frequently regardless of the platform. On the other hand, “lol” is restricted to Txt, but clearly plays a significant role in this communication form. Previous research on college-bound adolescents found that initialisms are exceedingly rare, appearing in less than 2% of all messages (Tagliamonte & Denis, 2008). The corpus presented here was created by a very different population who use far more initialisms, as more than 15% of all messages include one or both of “lol” and “ok.” The linguistic status of “ok” will be investigated first as it has a longer, more established usage than that of “lol”

Read (1963) argues that “O.K.” is not so much an initialism as it is the result of a play on words and sounds during an initialism craze that swept the Eastern United States in 1830’s and 1840’s. He pinpoints the first usage of “O.K.” in print to be March 1839, in a newspaper article, with the tag “(all correct)”. This was probably the response to “n.g.” becoming popular as an initialism for “no go”, and the popularity of spoken initialisms referring to alcoholic beverages such as “W.B.” for “wine bitters” (A. W. Read, 1963). It has since become lexicalized to the point where not only is it

“lol” or /lawl/ is occasionally used in spoken communication, but it is clearly marked as a feature that belongs to the Internet and Txt (Levenson, 2014)
6.1 Ok

It can be agreed upon that “ok” has a long history in spoken language, dating back to at least 1865 (A. W. Read, 1963). At that time, the meaning was understood to be “all correct” (A. W. Read, 1963). In the intervening 150 years, researchers have found that this literal meaning is no longer the most prevalent and salient function of “okay.” This process of meaning shift is common in spoken language in particular, which is vulnerable to effects of language contact and creative use (Bolinger, 1977). Functional analysis finds that “okay” in spoken language usually occurs at transitional points in the conversation, where there is a change in turn or a floor transition (Beach, 1993), and perception studies show that the meaning of “okay” in spoken language is dependent upon the surrounding prosodic cues, and without matching prosody, the contextually appropriate meaning of “okay” is cognitively very costly to disambiguate (Gravano, Benus, Chávez, Hirschberg, & Wilcox, 2007). These findings are confirmed in a variety of English language discourse studies, ranging from the use of “okay” during police questioning (Gaines, 2011), academic discourse environments (Schleef, 2005), as well as controlled laboratory tasks (Hockey, 1993). English, however is not the only language that uses ok to both mark transitions and indicate affirmations. Studies of the use of “okay” among Spanish-English bilinguals in New York City (Nowotny, 2004), monolingual and bilingual speakers of Hebrew (Maschler, 2004), and its role in code switching (Toribio, 2004) all confirm the finding that “okay” functions as a context-dependent transition marker with an optional interpretation of affirmation or assent.

In 60% of the instances of “ok” in this corpus, it is either the only item in the message or it is combined with either an emoticon or an emoji and nothing else (Example 74). This reinforces the hypothesis that “ok” signals a transition of turn because the sending of a message itself indicates a change of floor.
Example 74

A: I'm ready I going to take the train in 10 minutes

B: Ok

In Example 74, when A sends his message, the information has been conveyed to B and B has the opportunity to reply. If B has nothing more to add, B must indicate this and transfer the conversational floor back to A. If B does not give the floor back, communication and turn-taking is likely to break down. If B does not send a message to indicate that she received A’s message, A will not know when it is appropriate to send the next message or that the message had been received. Once B has the floor, if he has nothing more to add, he has to find a way to turn it back over to A. “Ok” fulfills this purpose for B, by both acknowledging that the message was received and that A can take the next conversational turn if she wants to.

Example 75

A1: I want see you today night but I don't want that you get tired of me

B: Ok so tomorrow after work

A2: Ok

In Example 75, B uses “ok” to mark that she is moving forward with the conversation, transitioning from A’s concerns about getting tired of him to her desire for him to come over as soon as possible. A’s reply in A2 indicates that he has understood and it is her turn to make the next conversational move.
Example 76

**A1:** What [NAME] wants she is following me in Facebook

**A2:** But I want to know What she want so I don't want to block her

**B:** It's okay

In Example 76, “okay” is spelled out fully, in contrast to the examples above. This use of “okay” is also markedly different since it is not marking a transition point in the conversation, rather, the semantic meaning of the item (i.e., “all correct” or “good”) is intended, possibly to make a distinction between this usage of “okay” and the more functional usage, “ok.” This example is evidence that there are two lexical entries for “OK.” The first is the initialism form of “ok” which serves a pragmatic function largely divorced from its literal meaning. This is the function that researchers have found indicates a change of turn (Beach, 1993; Gaines, 2011; Maschler, 2004) and the second is a more fully written form of “okay”113 defined by Merriam-Webster as “fairly good.” In spoken discourse, it would be impossible to differentiate these two, but in the printed txt medium, it is easy to see that “ok” has a purely pragmatic discourse function that serves to transition between a change speakers, topics, or turns. By contrast, “okay” carries a semantic meaning to indicate that everything is good enough.

Overall, the use of “ok” as a pragmatic marker sets a precedent in txt that there are items that only do pragmatic functional work, and do not carry any semantic meaning of their own. As we will see, this notion of a pragmatic operator can help explain the distribution of “lol” and from there be extended to understanding initialisms more generally.

113 Interestingly, this spelling of “okay” evolved from the pronunciation of the initialism. So while in the modern usage, “okay” appears in more journal articles and other formal settings than “OK” does, the entry for this lexical item in Mirriam Webster is “OK”, and “okay” is a variation, but not the standard. This essentially means that the usage here that connotes the actual meaning of “OK,” not the discourse function relies on the non-standard spelling of it. This should come as no surprise, because again, in our modern way of communicating, “OK” is more often a discourse marker than a semantic element adding content to an utterance.
5.2 Lol

At this point, it will be exceedingly difficult to determine the origin of “lol.” Though there are a variety of competing explanations, this dissertation defaults to Ben Zimmer\textsuperscript{114} as an expert of lexicography and his explanation of “lol” as having originated in a post on FidoNet\textsuperscript{115} on May 8\textsuperscript{th}, 1989 (Brandon, 2008) (see Appendix 8 for the content of this post). Regardless if this is the earliest instance, or if there was an earlier one, this post predates both mobile phone technology and the Internet in its modern form.

Everyone who comments, from Ted Talks to CNN to the Wall Street Journal, have proclaimed that “Lol isn’t funny anymore” (McWhorter, 2013b; Zakrzewski, 2015). That understanding is confirmed in this corpus as in Example 77. There is nothing funny about being sick, and the texter probably is not laughing out loud because she is feeling ill, so “lol” surely does not indicate humor or laughter here.

Example 77

Si y aunque tambien estaba enferma del estomago lol

Yes, and I am a little sick to my stomach lol

In modern computer mediated communication, most researchers agree that “lol” has another meaning; however, exactly what “lol” means is not as widely agreed upon. Researchers and journalists have a variety of hypotheses, each capturing a sub-group of the usage types for “lol.” Table 22 functions as a timeline of understanding of “lol,” beginning with its first appearance as a CMC initialism to the most current understanding of its meaning (studies and articles that treat “lol”

\textsuperscript{114} This choice is made on Zimmer’s credentials as a lexicographer at mainstream publications such as Oxford English Dictionary and has specialized in the history of American English slang terms.

\textsuperscript{115} FidoNet was a worldwide computer network (a precursor to the internet) that connected bulletin board systems together where people could make public and private posts and through that send near instantaneous (at that time) messages around the world to anyone on the network. FidoNet was by far the most robust of these systems, and many of the conventions of the internet emerged from FidoNet (Bush, 1992).
as a simple acronym for “laugh out loud” and do not make any further contribution are omitted).

This will be the foundation for the hypothesis presented here that “lol” is a pragmatic particle signaling the significance of the illocutionary force of the utterance.

Table 22 – Timeline of “lol” research

<table>
<thead>
<tr>
<th>Author, Date</th>
<th>Description</th>
<th>“lol” is:</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Baron, 2004)</td>
<td>Analysis of Instant Message records from 22 college students</td>
<td>a phatic filler</td>
</tr>
<tr>
<td>(Provine, Spencer, &amp; Mandell, 2007b)</td>
<td>Study of initialism and emoticon use by 226 anonymous individuals who posted on publicly available website message boards.</td>
<td>emotional punctuation</td>
</tr>
<tr>
<td>(Tagliamonte &amp; Denis, 2008)</td>
<td>Analysis of Instant Message records collected from seventy-one 17-20 year olds in university mentorship program.</td>
<td>a signal of interlocutor involvement</td>
</tr>
<tr>
<td>(Crystal, 2009)</td>
<td>Book on the emergence and value of text messaging</td>
<td>attending to textees’ communicative needs</td>
</tr>
<tr>
<td>(Tagg, 2009)</td>
<td>Analysis of a corpus of 11,067 messages collected from friends and family</td>
<td>a response token to indicate laughter</td>
</tr>
<tr>
<td>(Varnhagen et al., 2009)</td>
<td>Analysis of Instant messages sent in one week by adolescents known to researchers</td>
<td>an emotion acronym</td>
</tr>
<tr>
<td>Reference</td>
<td>Methodology</td>
<td>Description</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>(Baron &amp; Ling, 2011)</td>
<td>Focus groups with 75 university students</td>
<td>“a lexical shortening to express personal sentiment”</td>
</tr>
<tr>
<td>(Markman, 2013)</td>
<td>Analysis of 104 IM conversations between previously unacquainted undergraduate students</td>
<td>a turn-taking device</td>
</tr>
<tr>
<td>(McWhorter, 2013a)</td>
<td>Ted Talk</td>
<td>a pragmatic particle to express empathy</td>
</tr>
<tr>
<td>(Uygur-Distexhe, 2014)</td>
<td>8,000 messages from SMS4Science</td>
<td>a semantically opaque discourse marker</td>
</tr>
<tr>
<td>(Tagliamonte, 2016)</td>
<td>Emails, instant messages, and text messages from college students</td>
<td>a variant of laughter with social meaning to show interlocutor involvement; a phatic filler</td>
</tr>
</tbody>
</table>

These analyses fall into three broad groups, split between laughter (Bush, 1992; Tagg, 2009), a variety of pragmatic functions (Baron, 2004; Crystal, 2009; Markman, 2013; McWhorter, 2013a; Tagliamonte, 2016; Tagliamonte & Denis, 2008; Uygur-Distexhe, 2014), and as an emotive element (Baron & Ling, 2011; O’Neill, 2010; Provine et al., 2007a; Varnhagen et al., 2009). Taking each of these in turn, first there are those who suggest that “lol” indicates laughter or humor (Bush, 1992; Tagg, 2009). The first item is included because in 1989, it is safe to say that “lol” probably did mean
laughter. Tagg contributes something meaningful to the discussion, suggesting that it is a “response token” (2009). In this way, her analysis edges on the pragmatic analysis that treat “lol” as a phatic marker.

Analyses of “lol” as a purely pragmatic marker date back as far as 2004, when Baron proposed that “lol” was functioning in Instant Messages as a “phatic fillers, for the equivalent of OK, cool, or yea in spoken contexts” (pg. 411). Other researchers also analyze it as a phatic filler, or an item to signal attention and involvement (Crystal, 2009; Markman, 2013; Tagliamonte, 2016; Tagliamonte & Denis, 2008), serving an essentially pragmatic purpose without adding any semantic content to the message. These five studies have in common that they treat “lol” in the same way that “ok” is treated in analyses of spoken conversations (Beach, 1993; Hockey, 1993; Nowotny, 2004).

The remaining two pragmatic approaches to “lol” come to very different conclusions. The first, proposed by Uygur-Distexhe in 2014 is that “lol” is a semantically opaque discourse marker. She determined this by looking at French corpora and comparing it with the use of other French initialisms for laughter. This explanation is the first to identify “lol” as a strictly pragmatic marker and interpret it as having a function that operates on the entire message. The next important explanation comes from McWhorter in his 2013 Ted Talk116 where he identifies “lol” as a pragmatic marker to express empathy. He is the first one to discuss empathy as being a key feature that “lol” contributes. Empathy, however, is a very complicated thing since it can only exist in relationship to another person. This will be revisited in the hypothesis this dissertation presents.

In addition to these two approaches, there is a group of researchers who treat “lol” as an emotional acronym or emotional element (Baron & Ling, 2011; O’Neill, 2010; Varnhagen et al., 2009). These researchers argue that “lol” serves to contribute the emotional overlay to a message, conveying positive feelings. Baron is among these researchers though she also analyzed “lol” as a

116 Where he also makes a joke that dissertations are being written on “lol,” lol!
pragmatic marker seven years earlier. Notably, this explanation was popular in a two-year span of time when it is very possible that “lol” was in a transitional state where it was being used to convey emotion, or these researchers are focusing on just one set of functions that “lol” performs. In any event, the emotive state does not capture the modern use of “lol.” Again referring to Example 75, repeated here, it is clear that the texter is not contributing an emotional overlay to the message with her use of “lol”. She is sick, she likely does not feel positively about it at this time.

Example 75

Si y aunque tambien estaba enferma del estomago lol

Yes, and I am a little sick to my stomach lol

There is one more study that will be significant in constructing a definition for “lol,” and that is Dresner and Herring’s 2010 analysis of emoticons. Their argument is that emoticons are not, in fact, emotion-icons because they do not express an emotion, nor do they align directly with any given emotion. Rather, they are contextually dependent elements that serve to indicate the intended meaning of an utterance by signaling that the receiver should interpret it with a particular context, not that they are contributing an emotional overlay as originally thought (Dresner & Herring, 2010)

This dissertation takes Dresner and Herring’s analysis for emoticons and extends it to “lol.” This builds off previous interpretations of “lol” as a strictly pragmatic marker by also acknowledging that it does not contribute any semantic meaning to an utterance. In fact, in instances where it appears, the actual words, “laugh out loud” would often seem ludicrous, Example 75 again illustrates this (though nearly any example in the corpus can illustrate that point).

Example 75

Si y aunque tambien estaba enferma del estomago lol

Yes, and I am a little sick to my stomach laugh out loud
Uygur-Distexhe makes the same point in her analysis, that “lol” is doing functional work, though she avoids defining it, indicating that it is “semantically opaque” (2014). Looking at it from another way, it is not that it is semantically opaque so much as it has no semantic meaning of its own. As a pragmatic particle, it derives its meaning from the surrounding context and does not contribute any meaning without the context. Of course it will appear semantically opaque if it has no semantic meaning to be discovered. The treatment of “lol” as a pragmatic particle is really her analysis taken one step further, divorcing “lol” from the idea that there should be any meaning at all. The strength that this argument has over the arguments that have come before is that it is not trying to explain what “lol” means, only what it does. McWhorter makes this connection in an almost off-handed remark in an opinion piece for CNN (2013b). He draws a connection between “lol” and grammatical markers such as tense, but never develops it or makes any claims about the role that “lol” is playing in discourse (McWhorter, 2013b). There is this one small comment and then he moves on to keep talking about what “lol” is not, there is no further analysis, no further discussion, and we are left still wondering what the function of this pragmatic particle is. But his Ted Talk comes after that remark and there is no further evidence that he has continued research in that direction. But that is the missing piece that takes analysis of “lol” from looking for a meaning to looking for a function.

As a purely functional element, “lol’s” meaning is completely contextually dependent, and it cannot exist without previous discourse. It can, of course, stand alone in a message, but even then it is doing pragmatic work for the conversation. It still remains to be understood exactly what that pragmatic work is, though. “Lol”, like emoticons appears in a wide variety of contexts, ranging from positive to neutral to negative. Also like emoticons, “lol” does not have just one meaning associated with it, but rather a range of contextually dependent meanings. Dresner and Herring (2010) argue
that because emoticons are doing functional work and are their meaning is contextually dependent and because they can appear in a wide range of situations, they are serving to signal the illocutionary force of the message rather than adding semantic content to the message. As we will see, this appeal to illocutionary force accounts for all of the environments that “lol” appears in within this corpus and can account for the wide disparity in analyses of “lol” that have been proposed.

In the mid-1970’s, theorists developing this theory of Speech Acts and illocutionary force could not come to an agreement about whether the construction of illocutionary force was conventionalized forces (Austin, 1975; Dummett, 1979; Searle, 1976) or if the landscape was freer (Sperber & Wilson, 1986, 1997) and dependent strictly upon the context. The taxonomy that Searle proposed includes five acts that people can perform with words: assertive illocutionary acts (e.g., statements), commissive acts (e.g., promises), directive acts (e.g., commands), expressive acts (e.g., avowals of emotion), and declarative acts (e.g., christenings). Without a great deal of creative reinterpretation, this taxonomy is too restrictive to capture acts such as giving or requesting empathy (which is different from expressing emotion since it is not the emotion of the speaker, but rather a joining in of emotion), flirting (which could be interpreted as expressing interest, and therefore another expressive acts, but flirting is different from saying “I like you”), or tagging an utterance as humorous or ironic. Yet, all of these functions are performed by essentially “reading-between-the-lines” of the utterance. In a formal linguistic setting, this is referred to as interpreting the illocutionary force. Therefore, this analysis breaks away from the conventionalized approach to Speech Act Theory and adopts the perspective that illocutionary forces are not conventionalized, but are still interpretable. Examples 79-84 illustrate a variety of uses for “lol” that illustrate its function as a pragmatic particle signaling the importance of the illocutionary force. This does not suggest that “lol” creates the illocutionary force, only that it is a particle to indicate that there is an

---

117 Much to the personal chagrin of the author.
illocutionary force to the message that the receiver should interpret based on the context of the conversation.

Example 78

A1: Te gusta vestirte de blanco hee!

You like to wear white hee!

B: Si 😊

A2: Lol se te ve bien igual el negro y el rojo :) una percha

Lol you look good in both, the black and the red :) a hanger

In Example 78, A is giving B a compliment, telling B that he looks great no matter what he wears, and she tags this compliment with “lol.” In a conventionalized sense, “lol” is used here to express emotion and indicate that A has positive feelings towards B. In an non-conventionalized interpretation of Speech Acts, the “lol” is serving to flag this message as flirtatious, and that B should interpret A’s message as her expressing her unconditional interest in him, not his shirt. This type of exchange is likely why so many researchers have interpreted “lol” as a marker of emotion. In this setting, her positive emotional feeling is clear, but “lol” is doing far more than just adding a positive feeling (that would be there with or without the “lol”). Rather, “lol” is marking her compliment as an act of flirting, and she must signal that she is flirting because it is not readily apparent without the “lol.” For example, if she wanted to just be friends with him and give him fashion advice without suggesting that she was flirting, she would likely send the exact same message without the “lol.”

118 Smiling face with heart-shaped eyes emoji
119 This example either does not exist in the corpus or could not be found. In either case, this scenario remains a hypothetical.
Example 79

**A:** Me alegro <3

*I'm happy <3*

**B:** *Lol* estoy haciendo un ensayo :'(  

*Lol I am writing an essay :'(

For Example 79, in a semantically-driven explanation, “lol” would be interpreted as possibly laughing about the contrast between A being happy and B writing an essay. However, with an appeal to the argument that “lol” is serving to mark the illocutionary force of the utterance, it makes more sense to interpret “lol” as B asking for empathy from A. This request for empathy is not clearly stated in the message. That is, there is nothing about writing an essay that obviously tells A that he should offer a statement of empathy to B. However, with “lol” present, it becomes clear that there is another, subtextual, meaning that A should attend to. The purpose of this message, then, is not to inform A about B’s current activity. Rather, it is to ask B to express empathy for her having to write an essay. This type of exchange is most likely where McWhorter (2013a) derives his empathy-based explanation for “lol” (the example he uses is actually eerily similar, just in English). In his explanation for the popularity of “lol,” he states that it is a pragmatic particle marking empathy (McWhorter, 2013a). This also may be what Crystal meant when he refers to “lol” as “attending to communicative needs of your textees” – that is a particle to express empathy, or that a texter is sharing the emotional state of the textee (Crystal, 2008). While this explanation does capture this type of usage, this is only one of the many shapes that “lol” can take.
Example 80

A: Lol yu always use the internet with imessage

Lol you always use the Internet with iMessage

To analyze the role that “lol” plays in Example 80, it is important to diverge slightly into the discussion of assertions as speech acts. Speech Act Theory is inherently social. There is no illocutionary force without at least two people. One to create the utterance, and one to be the hypothetical receiver and interpreted of the utterance. Text messaging is also highly social. There is almost always a recipient to the message (unlike Tweets or long-form blogs). There is a whole area of research on the role of assertions (Bach & Harnish, 1979; Pagin, 2004; Stalnaker, 2013) and their communicative effect, with particular attention paid to the social or non-social dimension of assertions (Pagin, 2004). Pagin argues, very convincingly, that even though assertions are one of Searle’s original categories in his taxonomy of speech acts, it is not social in nature, and is therefore different from all other speech acts (2004). In Example 80, without the “lol,” the message is simply an assertion. It is merely an observation, a statement that could be true or false about the world. With the “lol,” however it becomes a social speech act with an illocutionary force that is likely more significant to the texter than the locutionary force of the utterance. Regardless if the purpose of this message is to soften the accusation, make fun of the receiver’s iMessage habits, or offer empathy and criticism together, with the “lol,” the importance of the intended meaning has been identified, and the receiver has the conversational work of determining what the sender means by her remark. With the “lol,” it may be interpreted as merely an observation, and the receiver is not under any obligation to determine what she means by her message.
Example 81

A1: [PICTURE] (stock photo of very sexy girl leading a man into the ocean)

A2: *Lol* I wanna have a boyfriend in the summer and go to the beach with him

B: 😂 u got a beach in your houses Bathroom

A3: 😠 (

A4: No

The role of “lol” in Example 81 is similar to that in Example 80, the texter is flagging her message as meaning more than the locutionary force may suggest. The “lol” here could either apply to the image in A1, or it can apply to the message in A2. Since “lol” was not sent alone, it more likely applies to the message written in A2. The “lol” is doing functional work here to signal to the receiver that the illocutionary force of the utterance does not necessarily match the locutionary force. Throughout their conversation, A has been saying that she wants a boyfriend, but that B is not ever going to make time for her. This is clearly a flirtation strategy that is not really working. Example 81, line A2 is another instance where she is making a statement that is explicitly flirtatious and she expects B to interpret that she is overtly performing an act of flirtation. Clearly he rejects this act (like B1), which she is unhappy about (lines A3 and A4). Even though her speech act failed, he still interpreted the act as an act of flirtation, recovering the significance of the illocutionary force. “Lol” does not mean flirtation, per se, only that the non-literal meaning is more important than the literal, and the receiver should work to recover the contextual force of the speech act. This is taken as further evidence that “lol” is not just any pragmatic marker or a phatic filler. Rather, it is a signal

---

120 Flushed face emoji
121 Unamused face emoji
to the receiver that they need to do conversational work to recover the meaning. More formally, this is a marker of illocutionary force.

Example 82

[NAME] me quieres hacer.un super.hiper.mega favor.me puedes compras panties hoes o como.se escriban jajaja lol

[NAME] do you want to do me a super.hyper.mega favor.can you buy for me pantyhose or however you write that jajaja lol

Example 82 illustrates an instance where the texter is asking her sister for a favor, and she needs it quickly. This example is interesting because it includes both laughter (“jajaja”) and “lol.” It has already been determined that “lol” does not signal laughter, but this provides further evidence that it does not because laughter has already been expressed. The spelling of ‘pantyhose’ is certainly not funny enough to warrant two types of laughter. However, the texter is asking a rather urgent, last-minute favor in a completely written platform. She needs to soften that request somehow, and express that she is aware that the request may be considered a burden, but also that they are close enough that it is a reasonable thing to ask. “Lol” serves this purpose. First, she hedges the request with an appeal to her sister’s negative face wants through the use of a question as to if she can ask a favor (which her sister clearly cannot reply to). She further has to express that she is aware that she is asking a big favor, but also that she really needs them. In this case, “lol” helps signal to her sister to interpret the request as not too imposing, but still urgent. Again, it serves as a marker of the illocutionary force of the message – that the pantyhose are a critical item that she needs her sister to procure.
Example 83

A: 2000 para una lipo lol

2000 for lipo(suction) lol

B: no una lipo no

no liposuction no

Example 83 illustrates another case where the texter is signaling that the illocutionary force of the message is significant. Without the “lol” this is again a statement of fact. It costs $2,000 to get liposuction. With the “lol,” it can be read a number of different ways. One of which is that the texter is making a commentary on the price or the fact they know the price, it could be that they are asking for empathy (because they believe they need liposuction), or any range of other possibilities. However, B’s response indicates that A’s messages was successful in asking for a compliment. Namely, that A does not need liposuction. As with all of these examples, the illocutionary force may have been recoverable by the receiver, but the use of “lol” signals that the illocutionary force is the most important element in the message.

This interpretation is further reinforced by the types of messages where “lol” is conspicuously absent, namely messages overtly expressing earnest emotion or conveying literal information. Example 84 illustrates a case where the texter is expressing exactly what she wants from the receiver.

Example 84

Yo kiero habla contigo

I want to talk to you
In this example, there is no veiled meaning, no alternative interpretation, she simply wants to speak to the receiver. She does not use “lol,” and doing so would diminish the intended meaning of her message, indicating that it was less serious than it is. Since the locutionary force of her message matches the illocutionary, she does not use “lol.”

Another situation where “lol” would be expected if it were simply a phatic marker or an item to express emotion would be in the most emotional type of message of them all, messages expressing love or desire. Throughout this corpus “lol” never appears in expressions of love. Examples of this type of message (between romantic partners in order to keep a uniform type of exchange), and pragmatic features that do co-occur with expressions of love are found in Examples 86-90.

Example 85
Mi amor te amo mucho amor mío

*My love I love you so much my love*

Example 86
Mami espero que me responda es que entiendame yo te amo

*Mami, I hope that you respond to me and you understand that I love you*

Example 87
Ok gracias igualmente te kiero mucho

*Ok, thanks to you too I love you so much*
Each of these examples is an expression of love and intimate feelings\textsuperscript{124} towards the receiver. These messages are intimate and earnest, and none of the messages of this type include “lol.” Some, as in Example 85 and Example 86 are clear, deep expressions of love. Some, as in Example 87 include phatic markers (i.e., “ok”) in addition to the expression of love, indicating that “lol” does different work than a standard phatic marker. Some, as in Example 88 include iterations, which, while they have not yet been discussed, perform a very different function from “lol,” as evidenced by their presence in expressions of love. Some, as in Example 89 include emojis in addition to the expressed sentiment. Emoticons are rare in this corpus, though the do co-occur with expressions of love (love for a child, not a romantic partner, though). Overall, this goes to show that while “lol” appears throughout the corpus, it never appears in an expression of love even though other pragmatic features such as phatic markers, iterations, and emojis do co-occur.

In addition to the expressions of love, there are a range of different contexts where “lol” does not appear. The first class of contexts is the simple exchange of information, including phone

\textsuperscript{122} Heart with arrow emoji
\textsuperscript{123} Closed lock with key emoji
\textsuperscript{124} It actually does not matter if they really feel this way or not, they are trying to convince their receivers that they do, so the locutionary force and the illocutionary force match.
numbers, addresses, directions, etc. Because these have been identity masked, it makes no sense to include them here. However, it is noteworthy that “lol” never occurs in this exchange of practical information either. They are also never found in purely phatic greetings, such as “how are you”, “good morning”, etc. If it only serves as a marker of illocutionary force, it could never be found in these contexts since their entire meaning is either very literal (i.e., the exchange of information), or purely intended to adhere to social conventions.

Overall, Examples 79-84 illustrate that “lol” is a purely pragmatic marker that signals to the receiver to attend to the illocutionary force of the utterance. Examples 84-89 illustrate that it does not occur in contexts where the locutionary and illocutionary forces match, indicating that it is a particle to signal to the receiver to evaluate the illocutionary force of the message against the shared context. The speech acts performed in messages with “lol” are always contextually defined and dependent upon the texter and receiver having shared context to build from. “Lol” does not, itself, transform a message from a statement into a flirtation (as in Example 78 and Example 81), but it does serve to alert the receiver to the fact that they should look beyond the literal meaning of the message to interpret the illocutionary force as the texter intended it. This sets “lol” apart from other phatic particles in that while it serves a discourse function, it does not necessarily trigger a response (in the way that turn-taking devices like “ok” do, or English has a wide variety of functional particles, including prepositions (“of”, “at”, “in”), articles (“the”, “a”), tense and aspect markers (“-ing”, “-ed”), and a wide range of other function words. However, all of these function words serve a syntactic purpose, essentially serving as the particles that hold a sentence together.

Nowhere else in adult English are functional particles used for strictly pragmatic purposes (Edwards & Goodwin, 1986). This type of particle occurs throughout Chinese (Liu, Jin, Li, Li, & Wang, 2009), and is hypothesized to be present in the speech of children (Edwards & Goodwin, 1986), but has never been attested in adult English. There are, of course, expressions in English that
convey pragmatic information, such as “you know”, “yeah, yeah, yeah”, and similar expressions. However, these expressions also carry semantic meaning, adding content to the conversation as well as serving a pragmatic purpose.

In conclusion, “lol” and “ok” in Txt function as pragmatic markers that serve a discourse function within a texted exchange, but do not carry semantic meaning of their own. They are used in digital platforms to convey non-literal meaning to the receiver. The individual meanings of each are summarized in Table 23.

Table 23 – Special initialisms summary

<table>
<thead>
<tr>
<th>Initialism</th>
<th>Translation</th>
<th>Positive or negative</th>
<th>Communicative function</th>
<th>Additional meaning(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>lol</td>
<td>Laugh out loud</td>
<td>Positive</td>
<td>Affirmation, acknowledgement, building rapport and consensus</td>
<td>Attend to the illocutionary force of the utterance</td>
</tr>
<tr>
<td>ok</td>
<td>(All correct)</td>
<td>Positive</td>
<td>Affirmation, assertion, confirmation</td>
<td></td>
</tr>
</tbody>
</table>

7 Iterations

Iterations are word forms with a repeated letter or letters (i.e., “heyyyy”). This respelling type is very common both in this corpus and in text messaging more broadly, so much so that it is often identified as one of the primary indicators of the Txt form (Crystal, 2009; Tagg, 2009). Almost all letters can be iterated, but there is a pattern to what letters are most often repeated, with preference given to vowels over all over letters, for reasons to be discussed in Section 7.1. Furthermore, the
meaning of an iteration is distinct from that of an initialism. The argument is made in Section 7.2 that iterations are meant to encode the emotional state of the speaker. While this is not in complementary distribution with the illocutionary force of the message (as will be discussed in Section 7.3, which investigates the role of iterated initialisms), it is a distinct purpose from signaling the illocutionary force of the utterance.

### 7.1 Shape of Iterations

There are three ways that letters in words, initialisms, and interjections are iterated, resulting in three different shapes. The first is when the final letter of the word is repeated but the remainder of the word is spelled in the prescriptive fashion as in “casaa” (casa house, Example 90).

Example 90

**A:** Wyd

*What you doing*

**B:** Pa la casaaa

*At home*

These will be referred to as final-letter iterations. The average number of repetitions of the final letter is three to four. The second type is when all of the letters except the first letter are repeated, keeping the same letters and letter order, but repeating each letter following the first letter as in “maaaaammiii” (mami a term of endearment, Example 91).

Example 91

Maaaaammiiiii preciosa chiquita hermosa.

*Mami lovely beautiful little girl*
These will be referred to as word-final iterations. The final type is where all of the letters are repeated, including the first letter, following the prescriptive form of the word in terms of letters and letter order, but repeat each letter as in “qqqqquuuuiiiieeeerrrrooooo” (quiero I love, Example 92).

Example 92

Teeeee qqqqquuuuiiiieeeerrrrooooo. mucho mi Bioleta hermosa.

*I love you so much my beautiful violet*

These will be referred to as complete word iterations. The difference between these types of iterations may be affected by the way that an item would be spoken. However, this is not the only contributing factor since unvoiced sounds (such as /q/) cannot be lengthened in spoken language. Therefore, there must be an additional difference between word-final iterations and complete word iterations that is independent of how they sound when spoken.

In this corpus, some letters are more often repeated than other letters. The letter /o/ is the most often iterated, appearing in 92 items. The complete list of iterated letter frequency is in Table 24. Each word is counted individually, so “hellooo” is distinct from “helloooo.” These items were identified computationally by looking for instances where the letter appears at least 3 times in a row. Items with 2 letters may either be part of the traditional spelling of the word or may simply be a mistake. After manually looking through a list of all words with double letters in the corpus, the instances where there are double letters in the spelling but not in the prescriptive form are indistinguishable from mistakes, therefore the computational approach was sufficient. Finally, if a word has two or more iterated letters in it, each letter is counted. Therefore, “maaaaammmmiiiii” (mami mother) contributes 3 letters (/a/, /m/, and /i/).

---

125 Letters with accents are combined with unaccented forms; this was especially important because participants often omitted the accents on letters, and often, the accent was not preserved, so /é/ is combined with /e/.

126 *‘Items’* is used here rather than ‘words’ because many of the iterations occur in non-lexical words such as “oooh.”
Table 24 – Iterated letter frequency

<table>
<thead>
<tr>
<th>Letter</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>o</td>
<td>92</td>
<td>20%</td>
</tr>
<tr>
<td>i</td>
<td>65</td>
<td>14%</td>
</tr>
<tr>
<td>e</td>
<td>59</td>
<td>13%</td>
</tr>
<tr>
<td>a</td>
<td>56</td>
<td>12%</td>
</tr>
<tr>
<td>y</td>
<td>54</td>
<td>12%</td>
</tr>
<tr>
<td>m</td>
<td>43</td>
<td>9%</td>
</tr>
<tr>
<td>h</td>
<td>22</td>
<td>5%</td>
</tr>
<tr>
<td>u</td>
<td>14</td>
<td>3%</td>
</tr>
<tr>
<td>w</td>
<td>14</td>
<td>3%</td>
</tr>
<tr>
<td>l</td>
<td>11</td>
<td>2%</td>
</tr>
<tr>
<td>s</td>
<td>8</td>
<td>2%</td>
</tr>
<tr>
<td>d</td>
<td>6</td>
<td>1%</td>
</tr>
<tr>
<td>n</td>
<td>6</td>
<td>1%</td>
</tr>
<tr>
<td>r</td>
<td>4</td>
<td>1%</td>
</tr>
<tr>
<td>t</td>
<td>3</td>
<td>1%</td>
</tr>
<tr>
<td>b</td>
<td>2</td>
<td>&lt;1%</td>
</tr>
<tr>
<td>k</td>
<td>2</td>
<td>&lt;1%</td>
</tr>
<tr>
<td>q</td>
<td>2</td>
<td>&lt;1%</td>
</tr>
<tr>
<td>x</td>
<td>2</td>
<td>&lt;1%</td>
</tr>
<tr>
<td>f</td>
<td>1</td>
<td>&lt;1%</td>
</tr>
<tr>
<td>z</td>
<td>1</td>
<td>&lt;1%</td>
</tr>
</tbody>
</table>
Both /o/ and /m/ are used in non-lexical items such as “oooh” and “mmm,” which may contribute to their popularity as repeated letters on Txt.

There are multiple competing explanations as to why this may be with different implications for the role that iterations have inTxt, and the reason some letters are repeated more often than others. The first explanation has to do with the frequency of letters in the language more generally. The most common letters averaged evenly across Spanish and English is /e/, accounting for 13% of all written Spanish and English words. The second is /a/, followed by /o/, accounting for 10% and 9%, respectively. A complete table illustrating the frequency (in percent) of each letter (with accented forms combined), and averaged evenly across Spanish and English is in Table 25. Both the Spanish and English frequencies were derived from greater than 40,000 words from published books (Huntley, 2003; Trost, 2007).

Table 25 – Frequency of letters in standard written Spanish and English (averaged)

<table>
<thead>
<tr>
<th>Letter</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>e</td>
<td>13%</td>
</tr>
<tr>
<td>a</td>
<td>10%</td>
</tr>
<tr>
<td>o</td>
<td>9%</td>
</tr>
<tr>
<td>n</td>
<td>7%</td>
</tr>
<tr>
<td>t</td>
<td>7%</td>
</tr>
<tr>
<td>i</td>
<td>7%</td>
</tr>
<tr>
<td>s</td>
<td>7%</td>
</tr>
<tr>
<td>r</td>
<td>6%</td>
</tr>
<tr>
<td>d</td>
<td>5%</td>
</tr>
<tr>
<td>l</td>
<td>5%</td>
</tr>
</tbody>
</table>
If it is the case that more frequent letters in the language are the most frequently iterated, then the written form of the language has the most influence on which letters are repeated, and the letters and words that are commonly iterated is a matter of chance. As shown in Figure 21, frequency in the language may contribute to which letters are more often repeated, but it cannot explain the entire phenomenon. In Figure 21, the combined frequency of letters (between Spanish and English) is represented on the y-axis, where more common letters near the top, and the size of spheres indicates the frequency with
which the item is repeated. Sonority of the letter is represented on the x-axis, with more sonorous on the right (how this was derived will be addressed shortly). As shown here, while /o/ is the most commonly iterated letter, it is not the most common letter in written Spanish and English. In formal, written English/Spanish, /e/ is the most common letter, but it is the third most commonly iterated letter. This shows that writing is not the most salient factor in terms on Txt, and while there is an influence of written letter frequency on the rate of iteration, there must be another factor at play to cause the discrepancies in rate of iteration (for example, /q/ will never be the most commonly iterated letter since it is so rarely written).

Figure 21 – Iteration frequency by language frequency and sonority
Rather than an effect of writing, it may be an effect of the sound of language that contributes to which letters are more often repeated. By this approach, speech would be influencing Txt to determine which letters are iterated; more salient letters would be more often iterated. The sonority of letters is represented in Figure 21.

Figure 22 – Sonority scale (Hartmann, 2014)

<table>
<thead>
<tr>
<th>Oral Stops</th>
<th>Fricatives</th>
<th>Nasals</th>
<th>Liquids</th>
<th>Semivowels</th>
<th>Vowels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voiceless</td>
<td>Voiceless</td>
<td></td>
<td></td>
<td>i</td>
<td>u</td>
</tr>
<tr>
<td>p</td>
<td>b</td>
<td>f</td>
<td>m</td>
<td>l r</td>
<td>i j</td>
</tr>
<tr>
<td>t</td>
<td>d</td>
<td>θ</td>
<td>n</td>
<td>e o</td>
<td></td>
</tr>
<tr>
<td>k</td>
<td>g</td>
<td>s</td>
<td>z</td>
<td>a</td>
<td></td>
</tr>
</tbody>
</table>

Referring back to Figure 20, if this is the case, the larger circles would appear on the left. Yet, some less-sonorous letters are iterated frequently (i.e., /h/ and /m/), and some more sonorous ones are infrequently iterated (i.e., /u/). This shows that the sonority is not the most important factor in determining which letters will be more commonly iterated, which suggests that there is yet another reason some letters are repeated more often than others. Another possible reason could be the type of words that are iterated.

Table 26 shows the breakdown of word types across six categories (adverbs, (dis)agreements\textsuperscript{127}, greetings, interjections, nouns, verbs).

\textsuperscript{127} Instances of ‘yes’, ‘no’, ‘true’, ‘false’, in both Spanish and English.
Table 26 – Word types by category

<table>
<thead>
<tr>
<th>Word Type</th>
<th>Iteration Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>adverbs</td>
<td>20</td>
<td>10%</td>
</tr>
<tr>
<td>(dis)agreements</td>
<td>23</td>
<td>12%</td>
</tr>
<tr>
<td>greetings</td>
<td>25</td>
<td>13%</td>
</tr>
<tr>
<td>interjections</td>
<td>85</td>
<td>43%</td>
</tr>
<tr>
<td>nouns</td>
<td>21</td>
<td>11%</td>
</tr>
<tr>
<td>verbs</td>
<td>25</td>
<td>13%</td>
</tr>
</tbody>
</table>

This table shows that interjections are disproportionately iterated, nearly four times as often as other word types. This may be due to the nature of iterations such as “oooh” and “aaaaa,” and the fact that since they are largely created by lengthened vowel sounds and there is no definitive spelling for many of them, they are directly transcribed from spoken language into Txt, resulting in a large variety in repetitions. When all of the same types of words are combined (i.e., “mmm” and “mmmm” are counted as one rather than two), there are 30 distinct interjections that account for all 85 instances. The revised calculations are shown in Table 27, where all of the different spellings of a single word are combined.
Table 27 – Iteration frequency by word type and type:token ratio

<table>
<thead>
<tr>
<th>Word Type</th>
<th>Iteration Frequency</th>
<th>Percent</th>
<th>Type: Token Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>adverbs</td>
<td>18</td>
<td>17%</td>
<td>1.1</td>
</tr>
<tr>
<td>(dis)agreements</td>
<td>15</td>
<td>14%</td>
<td>1.5</td>
</tr>
<tr>
<td>greetings</td>
<td>12</td>
<td>11%</td>
<td>2.1</td>
</tr>
<tr>
<td>interjections</td>
<td>30</td>
<td>29%</td>
<td>2.8</td>
</tr>
<tr>
<td>nouns</td>
<td>13</td>
<td>12%</td>
<td>1.6</td>
</tr>
<tr>
<td>verbs</td>
<td>17</td>
<td>16%</td>
<td>1.5</td>
</tr>
</tbody>
</table>

This table shows that interjections are iterated in the most diverse ways, followed by greetings, meaning that there are more variations of the way interjections and greetings are spelled than any other types of words. For example, “hola” (*bello*) and its variants are spelled thirteen different ways, listed in Table 18 and “mmm” (a sound understood to mean contentment or thinking (“mmm,” 2016)) is spelled seven ways listed in Table 28.
Greetings (openings and closings) are so often spelled in alternative ways in a variety of corpora (Combes et al., 2012; Tagg, 2009). Likewise, almost every texter uses openings and closings. So, while interjections and other types of words may have a wide variety of spellings, greetings are used so frequently that the variation results in a higher type: token ratio. Table 29 illustrates the different spellings of the interjection, “mmm.”
Since there is no standard way of spelling interjections, the variation is inherently higher than words with a prescriptive spelling. While Urban Dictionary lists “mmm” as the way this interjection is spelled, it is not clear that there ever was a prescriptive form. Therefore, texters are at greater liberty to use alternative spellings for it than other words. On its own, this is provides no more explanation for why certain letters are iterated more than others. However, combining this information with the role of stress and emphasis has on conveying emotion may provide more informative. By lengthening the interjection, it is as if the speaker is holding the sound longer. This may not transfer directly to lexical words, but in the case of interjections, because they exist in a grey area between non-lexical sounds and lexicalized words, the intended meaning of the interjection can be intensified with the repeated letters. The difference between interjections and other words is that the sound itself carries the meaning of an interjection whereas the sound-meaning relationship in lexical words is almost always arbitrary.
Researchers who mention the use of repeated letters in texting and IM contexts tend to interpret them as conveying emotional content or affect (Bernicot et al., 2012; Herring, 2012; Mironovschi, 2007). In each of these studies, it is taken for granted that this is what it means without much inquiry into how the meaning is derived. While Herring does give a thorough explanation for how repeated punctuation marks convey affect and then mentions that repeated letters do the same, she does not explain the relationship in any detail (2012). While the same result is met in this dissertation, that repeated letters convey affective or emotional information, it is a non-trivial exercise to derive how iterations convey emotional content, providing evidence for the assumptions that Herring (2012), Bernicot et. al (2012), and Mironovschi (2007) make.

Often in spoken language, suprasegmental features\footnote{Suprasegmental features are not limited to this small selection, but these are the ones relevant here.} are utilized to convey emphasis. Suprasegmental features include an increase or decrease of stress (loudness), duration, and changes to the frequency (raised or lowered voice). These changes are made primarily to vowels and voiced consonants (Koike, Suzuki, & Saito, 1998; Yildirim et al., 2004), and vowel lengthening in spoken English often indicates stress and emphasis on a word or syllable (Klatt, 1976). Similarly, research on the acoustic features of emotion indicates that there are vocalized differences between anger, sadness, happiness, and neutral emotions; these differences occur within the frequency or tone of voice and length of the vowel(s) (Frick, 1985; Koike et al., 1998; Yildirim et al., 2004). Texters are limited in the features they have available to them to communicate emotional content. In face-to-face conversations, emotional content of a message is communicated non-verbally, either through suprasegmental features, facial expressions, body language (Burgoon et al., 2010; Frick, 1985; Halberstadt, 1986; Knapp et al., 2013).

With features such as stress and frequency unavailable to texters, they must find an alternate way to encode the emotional content. Taking these factors into account, it may be the emotional
weight that a word carries that may be the reason for the lengthening of a given letter. In Txt, any word can carry emotional content, but interjections are especially good candidates since they have very little semantic content of their own. That is, “aaah” does not have any prescriptively defined meaning, so when it is used in a text message, it is an attractive candidate to encode other information onto it. This accounts for the prevalence of iterated interjections, and the diversity of interjection tokens for each interjection type. If texters are repeating letters in order to convey emotion, this also explains why neither frequency nor sonority can completely explain the frequency of iterated letters, but both seem to have some influence in which letters are iterated and how often. The next section further supports this interpretation by investigating the contexts that iterations occur in and the role that they play in communication.

7.2 Pragmatic Function of Iterations

As shown in the previous section, there is some evidence from the shape of iterations to show that they are used to carry the emotional content of a message. The most commonly iterated words in this corpus are interjections, which, as mentioned previously have no lexical meaning of their own. In the same way that “lol” no longer carries the meaning of laughing, it is likely that the act of repeating letters is a functional action, operating on a message in the same way that tense might operate on a verb. In the same way, iterations may function to situate the message in the emotional space that the message can possibly occupy.

Figure 23 illustrates the frequency of words surrounding the iterated items and the words in the general corpus, for the ten most common words surrounding the iterated words. Figure 24 shows the same relationship for the ten most common words in the corpus. This shows what context the iterated items are situated in. Five words to either side of each collocation were collected up to the end of the message. Capitalization was ignored, but alternative spellings were not combined. So, “que” and “q” appear separately in this chart. The rational for not combining
alternative spelling is that they provide information about the context of a message and the register it is written in. As shown in Figure 23, this is significant here as both “que” and “q” as well as “siii” and “si” are found in the top ten words near iterations.

Figure 23 – Most common words near to iterations compared to corpus

![Frequency of most common words surrounding iterations compared to the corpus](image)

Figure 24 – Frequency of most common words in corpus compare to those near to iterations
As is illustrated in Figure 23, “que”, and “te” occur more frequently near iterations and “you” occurs less frequently than in the corpus more generally. Taking each of these in turn, first “que” is investigated.

Both “que” and its variant, “q”, are very closely associated with iterations, as the first and ninth most common words near iterations, respectively. Without the accent, “que” in Spanish functions similar to “that” and “what” in English. It functions to mark the boundary between subordinate and relative clauses, to ask questions, and in a variety of other colloquial expressions. It is a function word that serves a variety of purposes, and it is the third most common word in the corpus more generally. However, the discrepancy between the frequency with which it appears near to iterations and in the corpus more broadly is worth considering. This likely means that iterations appear in more multi-clause text messages, meaning they are either in the main clause or the embedded clause of the message, suggesting that these messages are comprised of two parts. One
explanation for this is that one clause of the message may be meant to encode the content, and the other clause intended to encode the emotional significance of the message.

Likewise, the presence of the variant form of “que”, “q”, in the most common collocations indicates that these messages are more likely very casual, social messages. The use of abbreviations indicates that the message is between socially close participants, and that the content of the message is lighter. This may give further evidence that iterations are used to convey emotional content since they are used with socially close interlocutors.

The comparative prevalence of “te”\textsuperscript{129} and absence of “you” in the words surrounding iterations suggests that more of the iterations are in Spanish than English because “you” in English is both nominative and accusative, meaning it can be both the subject and the object of a sentence whereas in Spanish, the accusative is “te” and the nominative is “tu.” “Tu” does not appear near to the iterations, though “te” does, indicating that the second person singular is more often an object than a subject when near to iterative forms. This may provide more evidence for the argument that iteratives are primarily used to encode the emotional content of the message since the subject of the message is not the other person. The other person is involved in the message as the receiver of an action or a feeling, not the agent of the action or the feeling, and it is more likely that a person is expressing their own feelings about someone else than articulating someone else’s feelings.

Finally, the presence of “siii” (si\textsuperscript{130} yes) and “eyyy” in close proximity to other iterations indicates that iterations occur near other iterations, and multiple items in the message are being respelled in this way. “Siii” is the most common iteration, occurring 49 times in the corpus (Example 93).

\textsuperscript{129} It is unlikely that any of these instances are actually misspellings of “té” (tea), since participants are rarely discussing tea. However, if they are, it likely does not add enough instances to significantly change the numbers.

\textsuperscript{130} It is possible that participants are using “si” (if) and “sí” (yes) interchangeably since accents are difficult to type (addressed in Section 6.3, this chapter). Since the iterated “si” as in “yes” is so common throughout pop culture, this is the interpretation that will be explored, bearing in mind that there may be instances of iterated “if” included as well.
Example 93

Siii a las 7 :) le.mando.un mega abrazo hermano lo quieroooo mucho:)

Yes at 7 :) I send a mega hug brother I love him so much:)

“Eyyy” is not nearly as common, occurring 20 times. By comparison, “lmaoo” occurs 32 times (Example 94), as does “nooo” (Example 95).

Example 94

A: Give me some 😏人も (referring to money)
B: Lmaooo nah

Example 95

A: Dejame verte ;-)  
   Let me see you
B: Nooo baby no ahora plz
   Nooo baby not now please

The frequency of iterated “sí” may be explained by the role that “sí” plays in confirmation and intensification of the emotional content of an utterance. “Sí” is often used in a positive sense as it is the most basic form of affirmation. Throughout both Spanish and English pop culture, including TV shows, films, and memes, people are shown in a state of excitement, exclaiming the word “Siii” or “yesss,” as in Example 96 and Example 97.

In both of these examples, the subjects are clearly very happy and excited.

Example 96
However, this is not meant to illustrate that in a text message, it carries positive connotations. Rather, it is emotional, and on its own, it may be an exclamation of excitement or positive feelings, but in a text message, it is context dependent, used to intensify the emotional value that is already understood by the receiver. Example 98 shows how “siii” is used in an exchange to confirm or intensify negative emotions.

Example 98
A: Malo es poco :( Mi Vida esta jodida

There's a little bit of trouble :( My Life is fucked

B: Es LA vdd?

It's THE truth?

A: Siii

Yes

In this example, A is saying that something terrible has happened, and B is questioning if it is actually the truth, to which B responds yes, with the iteration that expresses the emotional significance of his answer. In Example 99, “siii” is show to confirm positive emotions.

Example 99

A: Holaaaaaa

Hello

B: A acabo el amor

To just be in love

A: Siii baby

Yes baby

In this example, “siii” is used to convey that being in love is a positive emotional experience. This is not an exclamation of joy or surprise, rather a confirmation of B’s expression that just being in love is enough.

The evidence thus far shows that iterations are likely working to encode or intensify the emotional content. Another example of the emotional content being conveyed through an iteration
comes in the form of “no.” It should be pointed out that “no” is the same in both Spanish and English, though “si” is still more commonly iterated than “no” is.

Example 100

A: noooo de verdad?? yo voy a trabajar en la noche el domingo de 9pm asta las 5:00 am del lunes

Nooooo is it true?? I am going to have to work Sunday night from 9pm until 5am Monday

In Example 100, the iteration is used to express that A is extremely disappointed at this situation (A wanted to spend Sunday night at the recipient’s house). Without the “noooo”, this message is devoid of emotional content and the recipient may not understand that A is upset. In order to ensure that she knows that he is upset about this turn of events, A must include this iteration or some other expression to convey the negative emotions he feels about having to work and not being able to spend the night with her. The iteration serves this purpose for him, highlighting that he is upset while still conveying a very important message. Example 101 further reinforces that the iterated “no” serves to encode emotional content.

Example 101

A: Que yo te diga un dia no tiene que ver con que sea simper

I tell you one day that it will not be forever

B: Hell noooo

In Example 101, B is expressing their intense disagreement with A’s statement. The iteration here likely reinforces the emotional content of the basic form. The intensifier, “hell” already serves to convey the emotional content of the statement, so the iteration may be a way to signal that the emotional content is extreme. That is, B feels very strongly that A’s claim is incorrect.
Finally, this form of “no” with an iterated /o/ is also a part of mainstream pop culture, as illustrated in Example 102.

Example 102

(https://memesvault.com/nooo-meme-darth-vader/)

Both “si” and “no” are indistinguishable if they are word-final, or last-letter iterations. An iteration of “yes” would give clues into which it is since it is a three-letter word, but “yes” does not appear in the corpus in an iterated form. However, as shown in Example 96, in mainstream pop culture, only the final letter of “yes” is iterated. Since /e/ is a commonly iterated letter, it appears that this is not due to a restriction on /e/, but rather a method of iteration.

It is still unexplained if the different types of iterations convey different meanings. Taking the intensity approach, it could be that the more letters are repeated, the greater the intensity of emotion being encoded into the message. Three examples of (te) quiero (I love you) are in Examples 104-106.

Example 103

A: Como estas mi chiquita
   How are you my cutie

B: Teeeceee qqqqquuuuuuuuuuieeeerrrrroooooo. mucho mi Bioleta hermosa.
   I love you. A lot my beautiful violet
Example 104

A: Mi [NAME] teeeeeeqqqquuuuuuiiiiiiirrrrrrooooo mi amor te amo preciosa.

My [NAME] I love you my love I love you beautiful

Example 105

A: Que la pases bien te quieroooo mucho

All is good I love you a lot

Across these three examples, the expression of love is probably more profuse in Example 104 and Example 105 than in Example 103 based on the amount of work required for Example 104 and Example 105. Autocorrect will not help the texter write these words, and they cannot be completed with any automated system. Therefore, the texter must have consciously written out each of those letters. It is much more complicated to repeat each letter than to simply lengthen the ending. The authors of Example 104 and Example 105 must have intended to repeat those letters as often as they did, illustrating to the receiver how much emphasis they are putting on the words. Notably, all of the words are not iterated, only the word that conveys the feelings and emotions that the texter is trying to emphasize. Therefore, it seems that the number of different letters that are repeated is a clearer indicator of emotional emphasis than the number of letters added to the end. Which, while expressive, do not require the time or focus that repeating all or most of the letters does.

Finally, there is an effect of personal style on the appearance of iterations. Over 80% of the iterations in this corpus are from just four participants, and some participants never utilize iterations.

---

131 This [NAME] replaces a term of endearment.
This does not detract from the explanation that iterations are meant to convey emotional content however as it may also be the case that some people communicate more emotively than others.

### 7.3 Iterations of Initialisms

Texters also iterate initialisms to convey emotion on top of the sentiment conveyed by the initialism. These forms have the potential to uncover the syntactic rules governing iterations as initialisms are not restricted to the same rules that words are. Acting more like interjections, they have no prescriptive form and have therefore always been a part of Txt register. In much the same way that investigations of nonce words\textsuperscript{132} can illustrate underlying rules operating on words more generally, the rules that restrict the iteration of initialisms may illustrate the syntactic rules governing iterations more broadly. A list of the initialisms that undergo iteration is in Table 30.

**Table 30 – Initialisms that undergo iteration**

<table>
<thead>
<tr>
<th>Iterated initialism</th>
<th>Basic form</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>hbuuu</td>
<td>hbu</td>
<td>how about you</td>
</tr>
<tr>
<td>lm(f)ao</td>
<td>lm(f)ao</td>
<td>laughing my (fucking) ass off</td>
</tr>
<tr>
<td>omgggggg</td>
<td>omg</td>
<td>oh my god</td>
</tr>
<tr>
<td>wtffffff</td>
<td>wtf</td>
<td>what the fuck</td>
</tr>
<tr>
<td>wyaaaaa</td>
<td>wya</td>
<td>where you at</td>
</tr>
<tr>
<td>tkmmmm</td>
<td>tkm</td>
<td>te quiero mucho (I love/like you a lot)</td>
</tr>
</tbody>
</table>

There is one initialism that is never iterated, “ctt” (como tu ta, how are you). Two possible explanations follow for why these forms are iterated and others are not. The first explanation is

\textsuperscript{132} One of the most famous language acquisition studies was conducted by Jean Berko Gleason in the 1960’s and asked children to make a plural out of the word, “wug” in order to better understand children’s knowledge of syntactic rules (Ratner & Menn, 2000).
phonetically-motivated, appealing to the way the words and letters would sound if they were spoken. The second explanation relates to the role that writing and text play in the creation of iterations. There is not enough data in the corpus to provide all of the evidence needed for either of these theories. Since the investigation here is on the structural rules of creating these forms, not their role in communication, evidence is collected from Twitter as well.

7.3.1 Phonetic Explanation

The reason that some initialisms are iterated and others are not could be phonetically based. It appears in the corpus that all of the initialisms that are iterated end with either a vowel, a voiced consonant, or a fricative. These elements from a natural class since they are the only elements that are created by making noise with the air moving through the throat, mouth, and nose. This means that they can be stressed whereas sounds that result from air stopping (i.e., /p/, /t/, /k/) cannot be stressed. It has already been shown that the effect of iteration is to give emotional emphasis to the expression. As already stated, in spoken conversation, this process is a result of the conversational practice of lengthening an emphasized word. If the process of iteration is a direct result of this conversational practice, then the words that can undergo iteration will be restricted by those that can be spoken. Regardless if anyone ever says a long /t/, it is possible to say it whereas it is not possible to say a long /f/, making ctt (como tu ta how are you) → cttttt impossible. It could be argued that this is semantic – that the reason cttttt is not allowed is because it never carries the emotional emphasis that an expletive like ‘wtffftf’ or ‘omgggg’ does. However, ‘hbuuu’ (how be/about you) is found as a variant of hbu even though ‘hbu’, like ‘ctt’ is a question, not an exclamation. With more information, it becomes clear that the phonetic explanation is incorrect since on Twitter, both ‘idc’ I don’t care’ is spelled as ‘idccce’ and ‘klk’ que lo que (how are you?) is spelled as ‘klkkk’.
7.3.2 Written Form

It could also be argued that ‘ctt’ does not undergo iteration because the abbreviation itself has two consonants at the end, so the shape of the abbreviation appears like an iterative form already since the double /t/ is at the end, in effect blocking any further duplication of the final letter. This is the most likely explanation since even on Twitter, it is extremely rare to find ‘ctt’ respelled as ‘cttt.’ Assuming that processes that are allowed on initialisms would also be allowed on abbreviations, one would expect to find common abbreviations such as “bro” (brother), “mñn” (mañana tomorrow), and “vdd” (verdad true) respelled as “brooo”, “mñnn,” and “vdddd.” However, which each of these abbreviations appears in the corpus, only “bro” appears in an iterated form (actually with four variations on the number of /o/’s at the end). The other two abbreviations, “mñn” and “vdd” did not appear in the corpus. Furthermore, they did not appear on Twitter from December 3-10, 2015.

This means that items which end in a double consonant, regardless of the voicing status are unable to undergo iteration, showing that the written form influences iteration rules more than the spoken form does.

7.4 Iterations Conclusion

Based on the types of words that are most often iterated, the context that they occur in, and the shape that they take, this dissertation adopts the perspective that iterations are used to encode and intensify the emotional content of the message. The focus of the emotional content appears to be the speaker’s emotions, though they may be about or directed at the receiver of the message. While they are not the only way to encode emotions, they provide a clear and efficient way to add emotional content without explicitly stating how the texter feels.

While the mechanism that allows iterations to be effectively used and interpreted by receivers likely stems from the way that emotion is phonologically encoded, it appears that the rules for how iterations are created is related to the shape of the item being iterated. That is, the
phonological features of an item do not determine if it can or cannot be iterated; the orthographic features of the item have more influence as to which items can be iterated. This provides further evidence that whileTxt is influenced by spoken and written language, it is not dependent upon either modality.

8 Mistakes

Throughout this corpus there are, of course, mistakes. There are multiple ways to approach mistakes, and the focus of this section is on the relationship between texters and their machines, not on spelling or grammatical mistakes that participants make in the course of texting. The primary reason for this is that it is impossible to determine what is a mistake (the texter knows the prescriptively correct form of the target item) versus an error (the texter does not know the prescriptively correct form of the target item) without explicitly assessing a rule. Secondly, in the context ofTxt, there is no reliable way to determine what is a mistake (unless it is corrected by the participant in the conversation) since there is no definitive inventory of what is right and wrong inTxt.

Still, most texting follows the prescriptive grammar rules. In order to account for this, commonly misspelled words from both Spanish and English are not considered respellings. These lists were derived from lists created by teachers for pedagogic purposes, and are found in Table 31 and Table 32.

Table 31 – Spanish Spelling Mistakes (Werbeck, 2015)

<table>
<thead>
<tr>
<th>¿cómo?</th>
<th>Catorce</th>
<th>fácil</th>
<th>restaurant</th>
</tr>
</thead>
<tbody>
<tr>
<td>¿cuándo?</td>
<td>Chaqueta</td>
<td>fruta</td>
<td>seis</td>
</tr>
<tr>
<td>¿cuánto,a,os,as?</td>
<td>Comida</td>
<td>fuí</td>
<td>sí</td>
</tr>
<tr>
<td>¿dónde?</td>
<td>Cuarto</td>
<td>generoso</td>
<td>sociable</td>
</tr>
<tr>
<td>¿por qué?</td>
<td>Cuatro</td>
<td>hamburguesa</td>
<td>tarea</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>¿qué?</td>
<td>Día</td>
<td>inteligente</td>
<td>teléfono</td>
</tr>
<tr>
<td>¿quién?</td>
<td>Diciembre</td>
<td>interesante</td>
<td>tenis</td>
</tr>
<tr>
<td>abuela</td>
<td>Difícil</td>
<td>matemáticas</td>
<td>trece</td>
</tr>
<tr>
<td>aburrido</td>
<td>Dinero</td>
<td>necesito</td>
<td>treinta</td>
</tr>
<tr>
<td>agosto</td>
<td>Doce</td>
<td>pantalones</td>
<td>vacacion</td>
</tr>
<tr>
<td>atlético</td>
<td>Dólares</td>
<td>parque</td>
<td>veinte</td>
</tr>
<tr>
<td>atrevido</td>
<td>Dos</td>
<td>pelo</td>
<td>zapatos</td>
</tr>
<tr>
<td>azul</td>
<td>Enseñar</td>
<td>pequeño</td>
<td></td>
</tr>
<tr>
<td>básquetbol</td>
<td>Escuela</td>
<td>piscina</td>
<td></td>
</tr>
<tr>
<td>bolígrafo</td>
<td>Excelente</td>
<td>quince</td>
<td></td>
</tr>
</tbody>
</table>

Table 32 – English Spelling Mistakes (Wrenn, 2015)

<table>
<thead>
<tr>
<th>again</th>
<th>dropped</th>
<th>looked</th>
<th>their</th>
</tr>
</thead>
<tbody>
<tr>
<td>all right</td>
<td>every</td>
<td>many</td>
<td>then</td>
</tr>
<tr>
<td>always</td>
<td>February</td>
<td>money</td>
<td>there</td>
</tr>
<tr>
<td>an</td>
<td>First</td>
<td>morning</td>
<td>they</td>
</tr>
<tr>
<td>and</td>
<td>for</td>
<td>mother</td>
<td>they're</td>
</tr>
<tr>
<td>animals</td>
<td>friend</td>
<td>name</td>
<td>things</td>
</tr>
<tr>
<td>another</td>
<td>friends</td>
<td>named</td>
<td>thought</td>
</tr>
<tr>
<td>around</td>
<td>frightened</td>
<td>off</td>
<td>threw</td>
</tr>
<tr>
<td>asked</td>
<td>from</td>
<td>once</td>
<td>through</td>
</tr>
<tr>
<td>babies</td>
<td>Getting</td>
<td>our</td>
<td>to</td>
</tr>
</tbody>
</table>
Any word from this list that was respelled was disregarded and not included in any analysis.

Throughout the corpus, there are three types of mistakes considered here. First is the practice of mistake repair and how these conventions may have emerged. Second is the omission of accents throughout the corpus and how bilingual texters deal with autocorrect and language choice on their machines. Finally, student awareness of error handling and the way that smart phones “learn” new spellings will be addressed.
8.1 The Asterisk

With Txt being a written medium that is often written very quickly or while doing something else (Combes et al., 2012; Holtgraves & Paul, 2013; Rainie, 2005; Tagg, 2009), mistakes occur very often. Texters have developed a series of conventions to deal with this; the first is the use of the asterisk. Participants in this study almost exclusively put the asterisk after the correction, as in Example 106.

Example 106

A: I know you like to **tweet** idk about drinking

*I know you like to **tweet** I don’t know about drinking*

A: **Twerk**

In this example, the receiver knows that the intended word to be replaced is “tweet” first because it is the only verb in the same tense, second because it is logical that autocorrect would mistake a word beginning with “twe” for another beginning with “twe.” Finally, the interlocutors had been discussing dancing up to that point, and while “twerk” is a dance move, “tweet” is not.

Sometimes the clarification makes no difference in the meaning of the utterance, so there must be another reason why texters make corrections as in Example 107.

Example 107

A: I don’t really get to watch it though I'm either at the gym or forcing myself to sleep whenever it comes **inc**

A: **On**

The meaning is easily recoverable without the correction, the receiver probably knew what she meant since “inc” does not make sense in the context, and the only thing that does make sense in that context is “on.” This shows that the reason for the correction is not necessarily to ensure that
the receiver understands, but for the texter to illustrate features about herself. By correcting the message, she is showing that she is aware of the correct version, that she is paying attention to the conversation, and that she wishes to present herself as someone who texts carefully. There may be other features that she is intending to convey through making this correction, but ultimately, she is also showing that she does not wish to make misspellings in this conversation.

Example 108 comes from the same conversation as Example 106 and Example 107, and is a correction that does significantly change the meaning of the utterance even though it is not a correction of a word.

Example 108

A: Oh I like [NAME]?  
A: !*

In this case, if A leaves the mistake uncorrected, the meaning of the first utterance is completely different. It is in her best interest to be sure that the receiver understands that the “?” is not what she intended, and she intended to use “!” This is the first case where the correction is not a word, but rather a punctuation mark that completely changes the meaning of the utterance. Here, the receiver is expected to know that the punctuation marks are substituted rather than any other item in the message. Again, even though the meaning is completely different (i.e., going from a question to an exclamation), the meaning would have been recoverable from the context of the conversation, so there must be another reason why texters are correcting their mistakes.

There are hundreds of these corrections throughout the corpus in both Spanish and English. Most of these are not necessary for clarification, and even if they do change the meaning of the message, they are mostly recoverable. Ultimately, this is evidence that texters are considering the way that they text as an extension of their identity. Someone who pays attention to prescriptive rules
would not want to be considered someone who is not meticulous about them, or someone who
does not pay attention to their text messages. Therefore, in an effort to ensure that they are
presenting themselves accurately, texters tend to correct their mistakes.

8.2 Teaching the Dictionary

The final way that mistakes are regularly made is through learning a respelling. Most Android and
Apple phones have a mechanism for learning words that the phone user will regularly be typing.
One way is through the contacts list. Depending on the phone, they may also learn through
repetition, selecting an alternative spelling while typing, searching the internet, or using it in another
app. This effectively means that one wrong spelling in any of these functions will result in the
dictionary “learning” that respelling. In this corpus, there are multiple instances of this. For example,
one participant often types the word, “worlk” in lieu of “work” as in Example 109.

Example 109

A:   A worlk tomorrow

At work tomorrow

This occurs ten times in her text record. Apparently this is not an issue for her or the receiver of her
messages since it is never corrected, never spelled correctly, and the receiver never mentions the
error either. Ultimately, this shows that tolerance for repeated errors is high as participants adapt to
each other’s texting styles, regardless of how unconventional they may be.

This further reinforces that the purpose of correcting mistakes is to accurately express
identity since receivers are able to recover the meaning of messages with highly unconventional
spellings in them.
8.3 Accents and Bilingual Spellcheck

Throughout the corpus, the use of accents appears nearly random, sometimes accents are written on words, sometimes they are not, even if the accent is obligatory and changes the meaning of the word. For example, “qué” usually means “what” and is primarily used in questions, whereas “que” means “than” or “that” and is primarily used in statements. Looking first at participants, it appears that some participants completely omit all accents, and some use them regularly though not always. Going back to the reported phone behavior, it turns out that those who never use accents also text the most in English and have both their phone’s language settings and spell check in English. This then makes sense that they would not use accents.

Furthermore, some words never get accented in the corpus. This suggests either that the unaccented form is so common and recoverable that it is a proxy for the accented form or that accenting is too difficult for that word so it has become normal in Txt to omit the accent. Of course, there are other explanations, but with the goal of understanding the norms in Txt, these two explanations will be discussed. There are a few words in Spanish that are only differentiated by the accent, listed in Table 33.

Table 33 – Spanish words only differentiated by accent

<table>
<thead>
<tr>
<th>Words</th>
<th>Meanings</th>
</tr>
</thead>
<tbody>
<tr>
<td>de, dé</td>
<td>of/from, 3rd person singular subjunctive give</td>
</tr>
<tr>
<td>el, él</td>
<td>the, he</td>
</tr>
<tr>
<td>mas, más</td>
<td>but, more</td>
</tr>
<tr>
<td>que, qué</td>
<td>that, what</td>
</tr>
<tr>
<td>se, sé</td>
<td>reflexive indirect object, I know</td>
</tr>
<tr>
<td>si, sí</td>
<td>if, yes</td>
</tr>
<tr>
<td>te, té</td>
<td>you, tea</td>
</tr>
</tbody>
</table>
Taking “qué” as an example, it does not appear a single time in the corpus even though prescriptively, the accent is required. The unaccented form, “que” is exceedingly common in colloquial Spanish as it fills a variety of functions. See Example 110 and Example 111, written by the same author.

Example 110

A: que ases?

¿Qué haces?

What are you doing?

Example 111

A: que ases mi vida

Que hace mi vida

That [You] make my life

It is so versatile that there are multiple abbreviations for it in Spanish Txt (both “q” and “k” are acceptable respellings). Automated spelling suggestions therefore prioritize “que” (no accent) over “qué” unless there is a leading question mark. However, the leading question mark is very difficult to type. On a smart phone, it requires going to another keyboard and from that keyboard, holding down the question mark until the inverted question mark appears. This all takes extra time and attention, which most users do not do since there is not a single instance of the inverted question mark used in this corpus. Therefore, among bilinguals, it appears that the inverted question mark has become obsolete in the Txt register. This, in turn, makes the accented “qué” rare as well since
autocorrect never receives the trigger that would signal the accented form. In effect, this means that the physical layout of the keyboard and the use of autocorrect influences which words are spelled correctly and which words are not.

Moreover, the direction that accents appear on the keyboard may also be playing a role in which accents are used and which are not. The most commonly accented letter is /í/. In nearly every instance, /i/ receives its accent whereas other letters are only occasionally accented. While this may be a completely random effect related to the words which have an /i/ and those that have other accented letters (/i/ often appears in the middle of verbs). This is highly unlikely, but always possible. More likely is either the fact that /i/ on a smartphone keyboard is unique in its directionality. For most phone models, to access any accented letter while in an English keyboard, the user must hold down the letter and then choose the correct accent. See Figure 25 for iPhone 5 examples. Most participants use the English keyboard. When asked why, they said that the dictionary was better, even if they send more text messages in Spanish.

Figure 25 – Accents on the iPhone keyboard

---

133 This is not true for the Galaxy S, as accented letters are only available in language-specific keyboards. Only one participant hand this phone, and her settings were in Spanish, so the keyboard is nearly the same (with the addition of ñ).
This may seem like a small difference (selecting an item that is one space to the right versus one that is all of the way to the left), but it appears to be the only difference that makes /í/ unique unless it is simply in more words that autocorrect recognizes. Regardless which explanation is correct, there is a clear influence of the design of technology on the way that Txt is evolving. Ultimately, the norms and conventions of Txt are inseparable from the technology it is emerging on, resulting in the selective omission of accent marks and punctuation because of the barriers to prescriptive writing.

8.4 Mistakes Conclusion

Overall, there are two conclusions that can be drawn from the prevalence of mistakes on Txt. The first is the role that identity plays in how a person texts. Some participants are more comfortable with mistakes and typos in their messages than others are. For some participants, even if there is no clarification needed, they correct any mistakes that they catch in their messages. This is further evidenced by the mistakes that participants never correct.

The second conclusion is the role that the technology is playing in the evolution of the Txt language form. When the meaning is clearly recoverable and the technology stands in the way of
writing the prescriptive form of a text, it appears that texters are willing to omit features such as accents and punctuation that may simply be too difficult to type. Ultimately, the result of this is the omission of these features in the conventional form of Txt. As the technologies evolve and adapt, it may become the case that accents are easier to type, but by that point, texters may no longer be incorporating them into their messages because that is the norm of this language register.

9 Conclusion

Overall, the features of Txt documented in this chapter have shown that Txt is a rapidly evolving medium that adapts to the communicative needs of those who text. This research shows that respellings serve a variety of different functions for texters, including setting a light, casual tone to the message (all respelling types), saving positive face (abbreviations and initialisms), indicating in-group membership (abbreviations and initialisms), establishing rapport (laughter), managing discourse functions (“Ok”), signaling a mismatch between the locutionary and illocutionary forces (“lol”), encoding emotional content (iterations), and constructing identity (mistake tolerance and correction). Because text messages are lacking many of the non-verbal cues found in face-to-face communication, texters have adapted the resources they have to recover the information and possibly encode more information.

These effects can be grouped into two categories: those that focus on the state of the receiver of the message (conversational tone, politeness and discourse markers), and those that signal information about the texter (emotional states and identity construction). These will be called outward functions and inward functions. These functions are not mutually exclusive as texters often combine respelling types.

Outward pragmatic functions are those that attend to how the receiver understands the message. Texters signal that they wish to save the receiver’s positive face and incorporate them into their in-group through the use of abbreviations and initialisms. Both of these functions are meant to
affect how the receiver feels about the message. Texters appear to use pragmatic particles such as “lol” and other abbreviations and initialisms to signal to the receiver that they should consider alternative or contextually-motivated interpretations of the message. These markers signal the significance of the illocutionary force, thereby affecting the cognitive state of the receiver. In face-to-face conversations, these functions are often accomplished through other means such as changes in voice quality, body language, facial expressions, or other non-verbal mechanisms. This is not meant to diminish the role that verbal mechanisms have in accomplishing this effect (i.e., the use of nicknames, and jokes), only that abbreviations and initialisms enhance the communicative ability ofTxt.

The inward functions convey something about the texter. These are either intended to illustrate their creativity or identity or encode their current emotional state. Texters accomplish this function through iterations and attention to texting style. Iterations serve to either convey or intensify the emotional content of a message. When texters repeat the letter(s) in a word, they are conveying the intensity of the their feelings about the overall message. These feelings may or may not be related to the recipient, but they always originate from the texter, serving as a way to illustrate their emotional state.

The second inward function is the performance of identity largely done through attention to texting style. While some texters will allow mistakes to remain uncorrected in the conversation, others correct them even when it is not needed for clarification. Their best explanation of this is that texters feel that their style of texting is a reflection and representation of themselves; texters show different sides of themselves to different people, so while a texter may correct all their mistakes with one person they may tolerate them with another person.

These functions taken together are significant because this is the first time in English that lexical items have been used to mark pragmatic meaning without semantic content. As long as
English has been called English, there have been syntactic markers that operate on sentences to change the meaning of the sentence without having semantic meaning of their own. This function is now available for pragmatic functions as well, as abbreviations, initialisms, and iterations operate on the pragmatic meaning of a message without having any semantic meaning of their own.
Chapter 4 – The Lovers

1 Introduction

Young people all over the world are communicating more and more via text messaging. This inevitably means that they are also dating, establishing intimacy, and navigating new relationships via text messaging. Within this corpus, young adults are setting up dates, establishing intimacy, building relationships, negotiating sexual encounters, and breaking up all in ‘writing’. With text messaging (as well as WhatsApp, Facebook, and other messaging services) at the center of their social world, there is a written record of participant relationships from inception to conclusion. The pragmatic functions that participants rely on for communication more generally are intensified in their romantic relationships, and will be explored here.

The purpose of this chapter is to apply the hypotheses developed in the preceding chapter and the analyses presented in previous research to a romantic conversation. This conversation spans the first four months of a relationship, beginning in late January until the time of collection in April. There are nearly 30,000 messages in this corpus. Since both members of this pair were participants in the study, the conversation is completely intact, and there is some demographic information both about them individually and about their relationship. For the purposes of this discussion, they will be called Jasmine and Michael, or “the lovers”.

The chapter is organized by the topics addressed in the respellings chapter, with examples to illustrate how these interlocutors use each feature to communicate and build their relationship. The first section gives some background on their language skills, academic experiences, phone use.

---

134 These are not their real names. Names were chosen based on popular Hispanic names for 1996, which is the median birth year of participants.
behaviors, and their relationship. The second section examines how they use respellings to enhance their communication. This section is divided into four parts. The first illustrates how they use abbreviations as a way to signal closeness and playfulness. The second shows how they use initialisms, including novel initialisms, “lol,” and “ok.” The third investigates iterations, which are used disproportionately by Jasmine compared to Michael, and the fourth identifies the mistakes Jasmine and Michael choose to repair and how that relates to identity. The third section addresses Jasmine and Michael’s use of both Spanish and English, both as code switching and as language choice. It should be made clear that this is a case study; every individual texts differently, and part of the Txt register is the ability and freedom to express one’s individual identity. Finally, the results of applying the hypotheses developed in Chapter 3 will be summarized in light of this conversation.

2 Background

The data reported here is a combination of assessment scores and self-reported information. The self-reported information reflects how participants thought and felt about their relationships, language, and their phones at the time of data collection. Both Michael and Jasmine completed the language background questionnaire, though Michael did not complete the survey on colloquialisms or phone usage. Both participants were born in the Dominican Republic, though they came at different ages and had different schooling experiences in both the United States and the Dominican Republic. Their academic goals and motivations are different and may play a role in both their academic English and Spanish levels as well as their texting behavior.

Jasmine is 20 years old and has been in New York City for 8 years. She was born in the Dominican Republic and went to private school there before moving to the United States. She took English classes in the Dominican Republic, and throughout her childhood, she expected to move to New York City at some point. Her family knew Michael’s family when they were children in the
Dominican Republic, but they were not close and only saw each other a few times in the Dominican Republic. Jasmine says that she wants to earn her bachelor’s degree in Criminal Justice, and that is her motivation to finish her High School Equivalency. She likes that her boyfriend is proficient in English, but can also understand her in Spanish. She also likes that he has a connection to the Dominican Republic.

Michael is also 20 years old, and is the only English-dominant participant in this study. Spanish is his first-learned language, but he came to the United States when he was five years old and started school in the United States. He returned to the Dominican Republic often during his childhood. He lives with his parents in New York City, and his family owns a store; he works a separate job and helps at the store occasionally. He wants to go to college, but feels he has a responsibility to his family to make sure the store is successful.

2.1 Language Background

Both participants took the English Literacy Evaluation for Newcomer SIFE. Their self-assessment on the language questionnaire indicates that Jasmine is representative of the broader population in terms of language and phone usage, though on some measures (i.e., English proficiency), Michael is an outlier and does not represent the larger population even though he is bilingual.

Their self reported ratings suggest that they both can communicate in Spanish and English. The scale is from 1-5, where 1 is “Novice”, 3 is “Intermediate” (can give an opinion or read directions), and 5 is “Advanced” (can converse fluently or read anything). Jasmine rates herself a 5 in Spanish speaking and writing and 3 in English speaking and writing. Michael rates himself a 5 in

---

135 Michael was enrolled in the school when data collection began. However, because he is English-dominant, he was not enrolled in the dual language program, though he is also bilingual. He attended class occasionally to visit Jasmine (though he was not enrolled in her class); he participated in the study on those days.
Spanish speaking and 3 in Spanish writing (an indication that most of his schooling has been in English). He rates himself 5 in English speaking and writing. The median rating for Spanish speaking and reading was 5, and the median for English speaking and reading was 3, indicating that Jasmine is representative of the participants more broadly in terms of their own perception of their language skills. These numbers are summarized in Table 34.

Table 34 – Summary of Jasmine and Michael’s self reported language skills in Spanish and English

<table>
<thead>
<tr>
<th></th>
<th>Spanish Speaking/Writing</th>
<th>English Speaking/Writing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jasmine</td>
<td>5/5</td>
<td>3/3</td>
</tr>
<tr>
<td>Michael</td>
<td>5/3</td>
<td>5/5</td>
</tr>
<tr>
<td>Group Average</td>
<td>4.95/4.70</td>
<td>2.57/2.34</td>
</tr>
</tbody>
</table>

Jasmine’s overall academic literacy level (the average of her performance across all the diagnostics she took) is approximately equivalent to grade 5.4 and Michael’s is approximately grade 8.0. These scores are detailed in Table 2. The average for all 50 participants in this study is approximately equivalent to grade 5.5, and the highest performing participant was at grade 9. Therefore, Jasmine’s academic language skills are representative of the participants more broadly. Michael, however, is more academically proficient than the majority of participants. Individual scores on each test are in Table 35. These scores are aligned to New York State grade level.

---

136 These scores are compiled by aligning all scores to a common scale (grade levels) and averaging across them.
137 The TABE returns a number that is then correlated to a grade. The creators of the TABE indicate that the grade level could be off by as much as two grade levels in either direction.
Jasmine and Michael have differing levels of Spanish and English Academic Proficiency, and very different academic experiences. In addition to being more academically prepared, Michael works more than he goes to school, so on a daily basis, he speaks more English and Spanish than he reads.

### 2.2 Relationship Background

Jasmine and Michael went on their first date in mid-January, 2015. They knew each other as kids, but met as young adults through a mutual friend (Michael is Jasmine’s friend’s neighbor); their friend arranged the date at his house by inviting both Michael and Jasmine over to watch a movie with him and his girlfriend. This was a blind date set up by a mutual acquaintance. Their first texted conversation is in Example 112.

---

138 This, in and of itself, is interesting. According to recent research on partnership and dating, this type of arrangement is becoming increasingly rare among young adults today (Ansari & Klinenberg, 2015; Maas & Zijdeman, 2010).
Example 112

**Michael:** Gonna be able to sleep tonight?

**Jasmine:** I don't think so 😴

**Michael:** Maybe watching that movie wasn't such a good idea

**Jasmine:** I like scary movies but I didn't like that one

**Jasmine:** Why are you texting me when you r driving 😡

**Michael:** You're right I'm sorry

The text messages span the first 99 days of their relationship, during which they sent 31,577 messages, which is almost 320 messages per day (combined). They reported seeing each other once a week, occasionally twice a week, which is approximately 15-20 times over this time period. They reported using FaceTime nearly everyday, though their primary method of communication is via text messaging. While both indicated that their *preferred* method of communication with their significant other is in person, it appears that in reality, most of their communication is digitally mediated. Therefore, text messaging and FaceTime play a significant role in how they are establishing intimacy and building their relationship. The features of Txt that convey non-literal meaning and emotional significance are an important component of this.

### 2.3 Phones and the Internet

In terms of technology usage and access to digital tools, all three participants mirror the broader trends among participants in this study. Both Jasmine and Michael have smartphones that they do

---

139 Tired face emoji
140 Unamused face emoji
not share with anyone else (i.e., they do not have co-ownership of their phone), and the interface settings (i.e., the date, meta-language, etc.) are in English. They both report using Google Translate regularly for looking up words in English and Spanish, and text in both languages as well.

Both participants say that they use the Internet everyday, and send text messages everyday. Michael however, does not use Facebook, though he has an account. He said that he does not like how it made him feel but he cannot delete the account because he uses it for planning events. Jasmine does have a Facebook account that she posts to and checks regularly. Michael prefers communicating in-person with most groups of people in his life (significant others, friends, parents, siblings, and grandparents). Jasmine said she prefers in-person communication for Michael, phone calls for friends and text/chat for her parents and grandparents. Jasmine’s preference for phone calls is unique, as most participants were split between preferring text/chat and in-person for most of their communication.

Both participants use social media outlets. The favorite is Instagram, and they, like other participants, say that it is a primary way that they keep in contact with friends, but not family. They post pictures and comment on friends’ pictures, though Jasmine comments much more frequently than Michael. Both use WhatsApp everyday, and Skype occasionally, though they use Skype mostly with people in the Dominican Republic. Jasmine and Michael use FaceTime when communicating with each other, and they do not use FaceTime with anyone else. They both use Snapchat, but not very frequently, since it is hard to have a conversation there. Ultimately, participants said that they

---

141 The need to abstain from or remove oneself from Facebook is becoming a common phenomenon. It appears that many young adults find that they experience an increase in negative emotions when the spend too much time on Facebook (Baumer et al., 2013; Portwood-Stacer, 2012; Steiger, Burger, Bohn, & Voracek, 2013).

142 FaceTime is a video chatting application that comes pre-installed on iPhones.

143 Snapchat is primarily designed as a way to send images. The messages (image, text, video, etc.) are erased after a set time, 30 seconds by default. Users can “screen capture” the Snapchat to preserve it, but as far as the platform is concerned, it is no longer accessible to the sender or receiver.
only use certain platforms with certain people, and it is a very good way of keeping different groups of people separate. This is one way of preventing a phenomenon known as “context collapse” where different worlds come together in ways that may be undesirable in certain situations (Boyd, 2014; Marwick & Boyd, 2011; Vitak, 2012).

3 Messages

Jasmine and Michael are prolific texters and sent 31,577 messages over the course of their conversation (nearly 320 per day). Jasmine sent 14,811 (149 messages per day), averaging 3.44 words per message and Michael sent 16,766 (169 messages per day), averaging 5.37 words per message (Michael sent 76% more words than Jasmine did). They sent messages at all times of day, from early morning until late at night, often texting each other “good morning” or “good night” (or some variation on that message).

Both Jasmine and Michael generally adhere to prescriptive writing rules, including features such as periods and capital letters. Both of them use iPhones, so they are communicating with iMessage, which autocorrects for these features. This is significant because in order to employ alternative spellings and non-prescriptive grammatical features, they must intentionally override autocorrect. Therefore, while there are autocorrect mistakes in their conversation that are later repaired (see Section 3.4), if they use an alternative spelling, they have most likely done so on purpose as they had to intentionally change it back from whatever autocorrect suggested. During data collection, Jasmine and another participant were asked to show how they would text “heyyy”

---

144 iMessage appears like a text message to the end user, though users often state that it is “more seamless” of an experience. iMessage has a symbol to indicate to a receiver when a sender is typing whereas text messaging does not do this. iMessage also keeps conversations intact across platforms (i.e., phone, computer, iPad). Finally, iMessage also tends to be faster in terms of sending and receiving messages (though this is measured in milliseconds). From the developer side, these are completely different systems. Text messages operate as an SMS (short message service) and iMessage operates as an OTT (over-the-top) configuration. The details of how this works is far beyond the scope of this dissertation.
(neither of these participants actually iterated “hey” regularly, though they both use iterations).

When they typed “h-e-y-y-y” and then “[space]” the iPhone changed the word to “Getty.”

Apparently since “heyyy” was not in the dictionary, the closest approximation for that letter combination is “Getty.” When asked what they would do then, they showed that they would use the override ‘autocorrect’ selection. That is, for the vast majority of respellings, there was an extra step of going back to the word, tapping on it, and choosing the original spelling. The only exceptions to this process are “lol” and “ok,” which are recognized by the dictionary in Apple products.

3.1 Abbreviations

As stated in Chapter 3, Section 3, abbreviations are often used to save positive face, mark the casual tone of the message, and signal the significance of the illocutionary force of the message.

Interestingly, neither Michael nor Jasmine uses abbreviations in their text messages. Most of their words are spelled out completely; words that are not spelled out completely (such as “gotcha” or “wanna”) are lexicalized into the English vocabulary and are commonly used in spoken registers.

They do not use the k-cluster or the u-cluster, preferring to spell out items such as “te quiero” (I love you) with the /qu/ fully written out, and the word, “you.” Because there are a variety of factors contributing to users’ texting styles, the lack of evidence does not suggest that Jasmine and Michael are not performing the functions that abbreviations perform, but that they choose not to when communicating with each other.

Jasmine uses abbreviations in her messages with other people as in Example 113 (a text message sent to a female friend) for items that she generally spells out with Michael, as in Example 114.
Example 113

**Jasmine:** Aki tranquilo! El jefe tuyo trabajara en la tarde mañana?

**Aqui tranquilo! El jefe tuyo trabajara en la tarde mañana?**

*Quiet here! Does that boss of yours work in the afternoon tomorrow?*

Example 114

**Michael:** Text me when you're done then

**Jasmine:** Aquí estoy! I was cleaning the bathroom

*Here now! I was cleaning the bathroom*

This shows that she does employ both the k-cluster as well as clipping when communicating with her friends, just not when communicating with Michael. One explanation for this is that Michael does not show any evidence of using abbreviations, so she may be mimicking him. This is unlikely since Jasmine uses other features (i.e., iteration of /s/) that Michael does not use.

A more likely hypothesis I would like to pursue is that they do not use abbreviations with each other because of the connotations associated with abbreviations. First, they wish to date each other, not become friends, and using abbreviations would suggest a casual, friendly, and possible inconsequential tone rather than new romantic involvement. As people who are dating (who also knew each other as children), they do not wish to signal that they are members of each other’s in-groups. Secondly, abbreviations indicate that the messages are casual in nature. If Jasmine and Michael indicate that their messages are casual, it could be interpreted that they are not taking the conversation and the exchanges seriously, and by implication, that they are not taking their relationship seriously. Finally, the identity that they are constructing is carefully attended to, and they most likely want to present themselves as more formal than the use of abbreviations would indicate...
(I will return to this in Section 3.4). Without careful examination, these are speculations as to their motives, but they serve to indicate that abbreviations are not appropriate for the tone that Jasmine and Michael wish to establish in their conversation at this time.

3.2 Initialisms

Whereas Jasmine and Michael largely do not use abbreviations, they do employ initialisms throughout their messages. The frequency with which Jasmine and Michael use the initialisms addressed in Chapter 3 Section 4 is in Table 36.

Table 36 – initialisms used by Jasmine and Michael

<table>
<thead>
<tr>
<th>Initialism</th>
<th>Translation</th>
<th>Jasmine</th>
<th>Michael</th>
</tr>
</thead>
<tbody>
<tr>
<td>lol</td>
<td>Laugh out loud</td>
<td>3732</td>
<td>2147</td>
</tr>
<tr>
<td>ok</td>
<td>(All correct)</td>
<td>1287</td>
<td>1273</td>
</tr>
<tr>
<td>idk</td>
<td>I don’t know</td>
<td>89</td>
<td>65</td>
</tr>
<tr>
<td>lmao</td>
<td>Laughing My Ass Off</td>
<td>0</td>
<td>117</td>
</tr>
<tr>
<td>wyd</td>
<td>What You Doing?</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>omg</td>
<td>Oh my God</td>
<td>49</td>
<td>7</td>
</tr>
<tr>
<td>ttyl</td>
<td>Talk To You Later</td>
<td>35</td>
<td>41</td>
</tr>
<tr>
<td>jk</td>
<td>Just Kidding</td>
<td>12</td>
<td>36</td>
</tr>
<tr>
<td>tk</td>
<td>Te Kiero (I love you)</td>
<td>21</td>
<td>2</td>
</tr>
<tr>
<td>tkk</td>
<td>Que Lo Que (What’s up?)</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>ctt</td>
<td>Como Tu Ta (How are you?)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>gm</td>
<td>Good Morning</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>qtl</td>
<td>Que Tal (What’s going on?)</td>
<td>0</td>
<td>2</td>
</tr>
</tbody>
</table>
Two features of this data will be analyzed here. First, the initialisms that either participant uses much more than the other, including “lmao”, “omg”, “jk”, and “tk.” Secondly, the absence of “wyd”, “klk”, “ctt”, “wya”, “hbu”, and “wbu” will all be addressed.

### 3.2.2 One of us is Laughing My Ass Off

As stated in Chapter 3, Section 4.3, “lmao” (laughing my ass off) is used to indicate actual laughter (as opposed to “lol”, which does not generally refer to laughter). Jasmine never uses this initialism, though Michael uses it regularly (117 times during their 3 month conversation). Jasmine instead uses “hahaha” to indicate that she is laughing or finds something funny. This difference is most likely a...
matter of personal preference and identity construction on a digital platform. That is, while Michael uses “lmao,” Jasmine conveys the same meaning and the same sentiment with “hahaha.” This may be one of the clearest cases that speech community does not completely predict how someone is going to text. There may also be an influence of gender here as the letter “a” refers to a swear word, and she may not be willing to say it because of the image she wishes to present of herself. Over the 99 days of their conversation, Jasmine never adopts the “lmao” initialism, it is simply not a part of her Txt idiolect even though it is a part of Michael’s. It is clear that she can understand it, however, as she responds with laughter, indicating that she is laughing along with Michael as in Example 115.

Example 115

Michael: Lmao you're in trouble
Jasmine: Babe hahaha

Most instances of “lmao” and “hahaha” in their conversation are single word messages (i.e., “lmao” or “hahaha” are the only word in the message) where a joke was made or something funny occurred, then Jasmine replies with “hahaha” and Michael replies with “lmao.” This occurs more than half of the times “lmao” is used in the corpus. Michael only uses “haha” and its variants (i.e., “hahaha”) forty-one times throughout the conversation, accounting for 26% of the times he indicates laughter.

Over the course of their conversation, however, there is a peak in the humor that comes shortly after the peak in “lol”, shown in Figure 26. Michael uses “lmao” nearly twice as often during the week of February 28th. This may be due to both Jasmine and Michael making more jokes at this time since there is a greater degree of trust and familiarity established. This does not, however, explain why there is a sudden fall off of “lmao” immediately after, and a return to levels of laughter below the beginning of their conversation.
Overall, it appears that “lmao” is Michael’s preferred way to express laughter whereas “hahaha” and its variants is Jasmine’s preferred way, showing that Txt allows individual users to express their texting identity while simultaneously conveying their intended meaning.

3.2.3 OMG JK TK (Oh my god, I was just kidding, I love you)

The initialisms, “omg”, “jk”, and “tk” also highlight how individual texting patterns are, and that texters have a Txt idiolect that distinguishes them from other texters. The patterns of usage of these items also suggests that users begin adopting each others’ Txt features. While both Jasmine and Michael use all three of these initialisms, Jasmine relies on “omg” and “tk” more than Michael does (Jasmine uses “omg” and “tk” 49 and 21 times, respectively whereas Michael only uses them 7 and 2 times, respectively). Michael’s first usage of “omg” occurs at February 18th, a full month after they had started texting with each other. He then slowly increases his usage over the period of their relationship until he is using it more than once a week near the end of their conversation. While this
is still speculative, it does give some suggestion that certain features of Txt can be adopted into the vocabulary of texters in contact.

The discrepancy in usage of “tk” is likely for different reasons, though. Jasmine is more proficient in Spanish than Michael is, and “tk” is an initialism representing “te quiero.” Michael uses it only after Jasmine has used it five times. Both times he uses it, he is stating that he has to pause the conversation and he is in a hurry. When she uses it, it is embedded in her message, often in a playful or creative way that could be interpreted flirtatiously as in Example 116.

Example 116

Jasmine: Pero yo tk+

Pero yo te quiero mas

But me, I love you more

Jasmine: 😛145

Jasmine: Ahaha you don’t get it 😛

Michael: Lol yes i do

In this example, she explicitly points out that she is creatively using the “+” symbol to indicate “mas” (more), which is how addition is articulated in Spanish. She uses this same construction one other time, even though she also explicitly states “I love you more” nine times throughout the corpus. It appears that, for Jasmine, “tk” is a more playful and creative equivalent way to express that she loves Michael. However, for Michael, it appears that “tk” is a way to quickly express that he loves Jasmine even though he cannot spend time conversing with her at the time as in Example 117.

145 Face with stuck-out tongue and tightly closed eyes emoji
Example 117

**Michael:** I'm good babe but we'll talk later I'm super busy. 😘😘

**Jasmine:** Ok babe 😘

In this case, he explicitly states that he is busy and cannot talk, and uses “tk” to indicate that even though he is setting the boundary and is unable to talk, he still loves her and wants her to know that he is feeling affectionate towards her. She responds with affection as well, likely indicating that his love for her was communicated even though he is in a hurry.

The final initialism in this set is “jk.” Michael uses “jk” three times as often as Jasmine does. Throughout their conversation, it appears that Michael does more teasing and needs to ensure that Jasmine understands that he is teasing. This is likely a result of their personalities and texting style. In Example 118, Michael is playing the jealous boyfriend, though he needs to be sure that Jasmine does not think that he is actually jealous.

Example 118

**Michael:** Have you made any friends in class?

**Jasmine:** Yeah

**Michael:** What's his name? 😘

**Michael:** Jk that's good babe

Without the “jk” it may not be clear that Michael is only pretending to be jealous. With the “jk”, he makes the joke explicit, thereby informing Jasmine that he is not jealous and it is not in his character to be jealous. This is in contrast to how “lol” is used, where the receiver must interpret the exact

---

146 Face throwing a kiss emoji
147 Pouting face emoji
meaning. With “jk”, there is no ambiguity, and the exact meaning of the message is identified explicitly.

Overall, the use of these three initialisms show that the decisions that texters make to use or not use certain Txt features is highly individual, and while transmission may occur from one person to another, not all features are necessarily transferred, and certain types of usages may continue to be separate for one person or another.

### 3.2.4 AAE influenced initialisms

The dialect of English that Jasmine and Michael use is very formal and academic in nature even though Jasmine and Michael both are Dominicans and speakers of the New York City Dominican variety of Spanish, they do not utilize many of the features that are found in this language variety when they are texting with each other. There seems to be a discrepancy between the initialisms that they use and the ones that they avoid.

In the larger corpus that this conversation came from, and in Jasmine’s messages with her friends, there are many AAE influenced initialisms, including “klk” (que lo que what’s up), “wyd” (what you doing) and “hbu” (how about you) as in Example 119.

Example 119

**Jasmine:** Psst wyd  
**A:** Eating lol  
**Jasmine:** Want to go with me and [NAME] and [NAME] to [PLACE]??  
**A:** I’m tired lol

In conversations with friends, Jasmine is clearly willing to use these features. However, these initialisms do not appear in the conversation between Jasmine and Michael’s. The question is then
why she does not use them with Michael, and why is there no evidence that Michael uses them in other situations.

The rationale for this may be found in the Match.com report. In their annual report on the dating behavior of singles, Match.com found, and multiple sources have confirmed, that singles judge potential dates’ grammar and punctuation very harshly (Ansari & Klinenberg, 2015; Fisher & Garcia, 2015; Forgays et al., 2014). It may be that both Jasmine and Michael intuitively know that they are being judged not only on what they write, but how they write it. Consequentially, they may be avoiding reference to less prestigious dialects such as AAE. It may also be that even though Jasmine’s Spanish and English are influenced by AAE, Michael’s is less influenced by AAE, making her less likely to use features of AAE when communicating with Michael. Of course, it is unlikely that either of them are explicitly aware of these features. It appears, however, that Jasmine is at least subconsciously aware that the connotations associated with initialisms such as “klk” do not match the identity that she wishes to present to Michael.

3.2.4 The Lovers’ Initialisms Conclusion

Overall, it appears that both Jasmine and Michael use initialisms in much the same way that other participants do. Throughout, it appears that initialisms are used to save positive face and indicate that the tone of the message is light or positive. They use “lol” to signify that the illocutionary force of the utterance is important and “lmao” to indicate laughter and positive feelings. Overall, initialisms in their conversation serve to save positive face, highlight the illocutionary force, and express empathy, just as they do among participants more generally.

The only difference from the wider trends that Jasmine and Michael exhibit is with initialisms that are influenced by AAE. Unlike the rest of the participants, and unlike in other conversations Jasmine has, they careful avoid these initialisms, opting instead for fully articulated
forms and other ways to indicate that the message they are sending is colloquial and positive in nature.

3.3 LOL - Lots of Love?

Throughout their conversation, Michael and Jasmine use “lol” (*laugh out loud*) so often that it accounts for nearly 7% of all the words Jasmine uses and 2.5% of all the words Michael uses. More significantly, it appears in 25% of all of Jasmine’s messages, and 13% of all of Michael’s. Since “lol” is postulated to be a function that applies to an entire message, it is more appropriate to consider the percentage of messages it appears in. Likewise, it rarely appears more than once in the same message (this occurs only nine times in the entire corpus), further confirming that each instance of “lol” operates at the message level rather than the individual word level. Unlike abbreviations, which can appear casual, and other initialisms which indicate creativity and stylization, “lol” is used to indicate the significance of the illocutionary force of the message and that there is more to the message than the literal meaning.

Previous research has posited “lol” as a marker of emotional content (McWhorter, 2013b; Provine et al., 2007b), fake laughter (O’Neill, 2010), attention (similar to *mm-hm* in spoken contexts) (Tagliamonte & Denis, 2008), and a discourse marker (Dresner & Herring, 2010; Uygur-Distexhe, 2014). “Lol” was hypothesized in Chapter 3 to be a pragmatic marker that signals the receiver to attend to the illocutionary force of the message, as the texter has an intended meaning for the message beyond the literal meaning of the words. In effect, this interpretation allows “lol” to take on many of the meanings proposed by previous researchers while offering a unified explanation for the role that “lol” plays in text messages. Consequentially, the meaning of “lol” is contextually defined and dependent upon the situation it is written in. For Jasmine and Michael, this often means that it is used to express empathy, positive feelings, or signal flirtations. The function “lol” performs is especially important in the early stages of romantic texting as interlocutors must ensure that their
messages are being interpreted accurately without having to state explicitly that they are interested in each other.

The importance of “lol” is apparent based on Jasmine and Michael’s usage of it through the beginning months of their relationship; in mid January (the beginning of their relationship), their text messages relied heavily on “lol”, with nearly 25% of their messages containing “lol” per week. This slowly falls off, though, and by mid-April (3 months later), only 13% of their messages contain “lol.” The only exception to this is during the week of February 14th, when the number of “lol” peaks at nearly 25% again. This may be explained by Valentine’s Day, which the couple celebrates. This trend is shown in Figure 27.

Figure 27 – Percent of Michael and Jasmine’s messages containing “lol” per week

It appears that in the early stages of building intimacy, it became increasingly important to use “lol”, possibly to signify that they had things they wanted to say but could not say without appearing too forward or aggressive. Essentially, they had more to express than exactly what they were saying. The
spike around Valentine’s Day can also be explained by the romantic feelings they were feeling, but still too early in their relationship to overtly state.

While “lol” is used to express feelings that should be interpretable, “lol” is not used in messages explicitly stating emotions such as love or desire. There are 42 instances of Michael or Jasmine expressing that they cannot wait to see the other, 85 instances of missing the other, and 581 instances of expressing love (in Spanish and English combined), for a total of 127 messages explicitly expressing longing and 581 explicitly stating love. None of these messages involve “lol.” In these cases, the illocutionary force matches the locutionary force, as they are both explicit statements of feelings towards or about the other person. To use “lol” in these messages would make the message less powerful, taking away from or even negating, the meaning of the message, rather than enhancing it (e.g., “I love you, lol” just does not convey love in the same way that “I love you 😊” does). If the purpose of a message is to express an emotion and that emotion is explicitly stated in the message, “lol” would be inappropriate as there is no sub-textual meaning to be conveyed. Therefore, if “lol” is signaling to the receiver to attend to the non-explicit meaning of the utterance (as hypothesized in Chapter 3), it would not be expected in messages with overt emotional content. This is evidence that the meaning of “lol” is beyond simply positive feelings or empathy, and its use conveys that the non-literal meaning of the message is important and the entire message should be interpreted relative to the surrounding context.

There are many uses of “lol” during the first two months of Jasmine and Michael’s relationship that rely on this function of signaling the mismatch between locutionary and illocutionary force. The following examples illustrate how Jasmine and Michael construct meaning with their messages with an appeal to the functional work that “lol” performs. Many of these

148 It seems fairly clear that “I can’t wait to see you, lol” has a very different meaning than someone in a new romantic relationship might want to express.
messages are not explicitly positive, but calling attention to the significance of the illocutionary force allows reconstruction of abstract conversational features such as empathy (McWhorter, 2013b), flirtation, and tone. Example 120 illustrates both a request for and offer of empathy.

Example 120

**Michael:** I don't know lol I know how to do other stuff I just can't think of them haha

**Jasmine:** Hahahaha

**Michael:** Okay I lied. I just don't want you to know about how much I don't know how to cook.

**Jasmine:** It's okay if you don't know how to cook lol a lot of people don't know how to cook

In this example (which follows a list of items that Michael knows how to cook, including multiple types of sandwiches), Jasmine is empathizing with Michael, and reinforcing that her statement should be taken as empathetic with the use of “lol.” Without the “lol” it may appear that she is disappointed in him, but normalizing his inability to cook. With the “lol”, it marks that there is something more to the message, likely an offer of empathy to him (possibly regarding his embarrassment over not being able to cook) (McWhorter, 2013b). Furthermore, without the “lol” Jasmine’s message may appear judgmental or condescending, or otherwise have a harsh tone to it. The “lol” signals that the intended meaning is positive and the illocutionary force (in this case, empathy) is important. In contrast, “lol” is never used in the conversation in Example 121. In this conversation, Jasmine is acknowledging her lack of knowledge, and disclosing that she does not know something, just as Michael was in Example 121. Even though this conversation does not take
on a condescending or belittling tone, the exchange has a problem-solution structure rather than a vulnerable disclosure-reassurance structure.

Example 121

Jasmine: The thing is that I don't know the neighborhood
Jasmine: I just moved here like three weeks ago
Michael: Oh okay yeah that's pretty fresh but the only way to get to know it is to get out there and walk around
Michael: You'll be an expert in no time

Michael is not offering her empathy, but rather a solution to remedy her unfamiliarity with the neighborhood. This solution is meant to be taken directly, not interpreted based on any contextual information, he is simply offering a suggestion and reassurance. This conversation has a more earnest, literal tone where rather than an exchange of empathy, there is an exchange of reassurance, normalization, and advice. Jasmine is stating the reason for her lack of knowledge, not necessarily asking Michael to provide emotional support. Likewise, Michael responds be giving her practical advice. Together Example 120 and Example 121 illustrate the difference between exchanges with emotional, empathetic content and those without, highlighting the role that “lol” has in signaling an exchange of empathy.

In the next example, Jasmine is pointing out that watching ‘Telemundo’ is rather strange for someone as young as Michael. Her use of “lol” here marks that she is not simply gathering information about Michael’s television preferences, but also either flirting with or teasing him (or both).
Example 122

**Jasmine:** So you watch Telemundo lol

**Michael:** At the barbershop the guys love to watch caso cerrado

*At the barbershop the guys love to watch 'case closed'*

In Example 122, the “lol” is necessary to mark the message as more empathetic and intimate than insulting. “Telemundo” is the type of television station that is easily made fun of, and most people around age 20 do not watch it (or would not admit to watching it). Without the “lol”, Jasmine’s message either does not make sense as it is only a statement of fact, or could be interpreted as Jasmine passing judgment on Michael. Still, it is strange that he would watch it, so he follows up with a justification of why he is watching it. Under both of these interpretations, it is important to include “lol” in order to ensure that he understands that she is doing more with that message than simply gathering information about Michael. This stands in contrast to Example 11, where Michael is incredulous about Jasmine’s preference for Mississippi\(^{149}\), and gathering more information about her and her preferences. In the conversation leading up to this, they had discussed the places they had lived. Jasmine lived in Mississippi and mentioned that she liked it there. Michael does not like Mississippi, so he is clarifying if she liked it better than New York City. The tone is serious and earnest rather than light, and there is no “lol” or other respelling feature to indicate that she should take the message as anything other than a direct question.

\(^{149}\text{The state has been changed to protect their privacy.}\)
Example 123

**Jasmine:** Mi tia estaba en el Army & su esposo sigue en el Army todavia

*My aunt was in the Army and her husband still is in the Army*

**Michael:** So you like it there better than here?

**Jasmine:** Nooooo

In Example 123, Michael does not follow up his question with “lol” as it could be seen as offensive. Therefore, he is not indicating a light tone with Jasmine or signaling that he has a non-literal meaning behind his question, and correspondingly, there is no “lol.” This is taken as evidence that “lol” signals that the texter intends something more than what is written, resulting in “lol” being absent from serious or straightforward situations.

Another example of “lol” as a pragmatic marker can be seen in Example 124. Here “lol” takes on softening function. Here, “lol” is a key element in making sure that the intended meaning behind each message is understood, and that Michael is not insulting Jasmine, but flirting with teasing her.

Example 124

**Jasmine:** yeah! I also watch judge Judy

**Michael:** Ugh I hate judge Judy lol I think she's so mean

**Jasmine:** That's why I like it

**Jasmine:** Lol

**Michael:** Haha I see. You wanna be mean just like her.

**Jasmine:** No lol I like her attitude\(^{150}\)

---

\(^{150}\) The original text involved an autocorrect mistake that Jasmine corrected in the next message, it is repaired here.
In the course of this exchange (six messages), there are three instances of “lol.” Without the first instance of “lol,” the message could be interpreted as Michael judging Jasmine for watching Judge Judy. Instead, the “lol” mitigates the statement of a contradictory opinions and softens the word, ‘hate.’ The other two instances are written by Jasmine. The first instance may be used to indicate that she understands what he means, the equivalent of a smile or a wink in a face-to-face conversation. Another interpretation is that it may be taking scope over the preceding message to soften her defense of Judge Judy. As the only element in the message, it is impossible to determine which line of the exchange it applies to. Jasmine may also be playing on this ambiguity and using this one “lol” for both purposes. In the second instance, she is defending herself, but in order to soften the defense and maintain the light, intimate tone they are in, she uses “lol” to mark that while she is disagreeing, there is more to the message than just the disagreement. Namely, she is telling Michael that her self defense is not strictly a self defense, there is an intimate quality to it that is not otherwise conveyed by the literal meaning of the words. This is in contrast to Example 125 where Michael asks Jasmine if she likes Mark Cuban (a celebrity billionaire). In this instance, she replies that she does, but neither conversation partner uses “lol” since there is nothing to defend or soften. Liking Mark Cuban is less controversial and emotionally charged than liking Judge Judy.

Example 125

**Michael:** What you up to?

**Jasmine:** Watching Shark Tank

**Jasmine:** One of my favorite

**Michael:** It's a good one

**Michael:** You like mark Cuban?

**Jasmine:** Yup

**Michael:** I wish he was the Knicks' owner
Furthermore, liking Mark Cuban is not stigmatized in any way and is not a sensitive subject (as opposed to Judge Judy). Rather, this exchange is in earnest, and both Jasmine and Michael agree in their affinity for Mark Cuban, and respect each others’ opinion. The lack of “lol” here and its presence in Example 125 reinforces the hypothesis that “lol” is used to indicate to the receiver that the texter is softening the message and signaling that the literal and intended meanings do not match.

Another instance where Jasmine and Michael are using “lol” to signal that the unwritten content is just as important as the written is in Example 126.

Example 126

Jasmine: You like that I worry about things? Lol why?
Michael: That means you care about
Jasmine: About?
Michael: About everything lol
Jasmine: Lol oh okay
Jasmine: Well I do care

This exchange is significant because the tone is overtly flirtatious. Michael is essentially exposing Jasmine as someone who cares about him, the repeated use of “lol” signals that he knows and she knows, but neither one of them are ready to express their affection even though there is a message of affection being exchanged covertly. At this point, they had not overtly expressed their feelings for each other, but only done so in this highly flirtatious way. This exchange without the “lol” would not have the same flirtatious tone as it would be a statement of fact rather than a statement that implies there is more to the story. This is in contrast to Example 15 where there is still emotional
content being expressed, and Michael is still expressing how much he appreciates her, but it is in earnest, and there is no use of “lol.”

Example 127

**Michael:** I’m happy you came over babe

**Michael:** And that you got to be around my family

**Jasmine:** Tienes una familia Hermosa Dios la bendiga

\*You have a wonderful family God bless*

**Michael:** Thank you babe

**Michael:** We got our problems but we always try to stick together

In this instance, using “lol” would take away from the sincerity of Michael’s message, and therefore it is not present. While this is not a flirtatious conversation, it does serve to show the difference between earnest expressions of appreciation and affection and more playful, casual expressions of the same.

In addition to recovering the flirtatious tone of a message, “lol” can be used to indicate the intended meaning behind the message is that empathy is desired, as in Example 128.

Example 128

**Jasmine:** Yeah! I hate taking trains 😔 tomorrow is going to be a long day for me lol

ima get lost

In this case, the use of “lol” is not to express empathy for Michael, but rather to ask Michael to express empathy for her, which he does in the following message. After this, they have a discussion

---

151 Confounded face emoji
of getting lost with Michael reassuring her that she will be fine, indicating that he responded to this request for reassurance. Without the “lol”, this may be interpreted more seriously, as if she knows she will get lost, or that she is stating a fact about the day. With the “lol” she is likely indicating that there is more than the literal meaning to the message. Namely, she just wants reassurance and emotional support from Michael about her day tomorrow. Again, this usage informs the receiver that the illocutionary force does not match the locutionary. Example 17 provides a contrasting situation where Jasmine is expressing her dislike for something, but not asking for empathy from Michael. In this situation, they are discussing an event that is happening in a factual manner. She is not experiencing a negative emotion or event and has no reason to want Michael to empathize with her. Rather, they are discussing the smell of burning hair (Michael’s sister is straightening her hair), and Jasmine is stating her dislike for the smell. She is not experiencing the smell at the moment and therefore has no need to ask for empathy from Michael.

Example 129

**Jasmine:** I don't like that smell at all

**Jasmine:** That's why I don't do my hair like that

**Michael:** Well then that's another thing I like about you

Example 18 illustrates another instance where the illocutionary force of the message does not match the literal meaning of the words. This is another example of flirting, but it is also a clear instance of implied meaning.
Example 130

**Michael:** What you guys doing?

**Jasmine:** Hablando *lol*

*Talking lol*

**Michael:** About me I bet. 😁

**Jasmine:** *Lol* un poco

*Lol* a little

In this case, “lol” is used first to indicate that they are talking about Michael¹⁵³. Michael explicitly states this (without an “lol”), making the subtextual meaning explicit, shared knowledge. Jasmine confirms his identification, including an “lol” likely to signal flirtation or that she it talking about her feelings about him, or both. The nature of the discussion is further articulated in the conversation, that all the talking is good and they love each other, but it is important for Jasmine to include this “lol” to indicate that and assure Michael that the discussion is nothing he should be concerned about. Research on “lol” has suggested that it is used for conveying positive emotions (O’Neill, 2010; Varnhagen et al., 2009), and while it does capture this usage of “lol” as a marker that the conversation should be interpreted positively, clearly “lol” signals more than just positive emotions.

Both participants also use “lol” to soften confrontations and indicate that a confrontation should be interpreted in the context it is written in. This allows each partner to confront behaviors they do not like while hedging the confrontation and softening it. An example of this is in Example 131. Here, Michael is confronting Jasmine on being late. She is often late, and when this is noticed,

¹⁵² Smirking face emoji
¹⁵³ The role that the switch in to Spanish plays will be discussed.
she sends kiss emojis. He likely wants her to know that he does not like this behavior, yet he wants to save positive face and he probably does not want to express that he is angry.

Example 131

**Michael:** It’s not funny babe

**Jasmine:** 😜😜😜😜

**Jasmine:** I know

**Michael:** Lol you always send me kisses to get out of trouble.

**Michael:** I’m on to you 😹

The “lol” in this case does a lot of work. It serves to soften the confrontation and save Jasmine’s positive face wants. Without the “lol”, this confrontation could be a direct criticism of her behavior. The next message he sends (“I’m on to you”) with the emoji serves to reinforce that even though he does not like the behavior, he still has positive feelings for her, and may even be masking this confrontation in a flirtation. Compare this to when Michael expresses serious concerns and wants Jasmine to know that he is disappointed and explicitly wants her to change her behavior as in Example 132.

---

154 Cat face with wry smile emoji
Example 132

Michael: I need to tell you something. I wouldn't feel good not telling you.

Jasmine: Yeah

Michael: I didn't like what happened yesterday at [NAME]'s house. I don't care about what it was about what I didn't like was that I don't think the time was appropriate.

Michael: And then I think that you should have at least waited for me and [NAME] to leave before you made that comment to your mother. I don't ask you to get along with them but they are my family and I think that you talking about them like that in front me wasn't appropriate either.

Michael: That's all I have to say and I hope you don't take it the wrong way. It's just my opinion.

Michael: You can talk to me one on one about whatever you want and however you want. I just think that things take a whole different aspect when it's in front of other people.

Jasmine: I know babe and I'm sorry if I made you feel uncomfortable at the moment!

This conversation is a serious and sensitive confrontation where Michael is trying to be as honest and open as possible, while acknowledging that he is explicitly confronting Jasmine on her behavior. There are no “lol”s, no iterations, or other respellings. The meaning is literal and serious, and Michael is indicating that through the use of explicitly formal writing features (i.e., using all the prescriptive punctuation, and avoiding any respellings or playful language). Jasmine responds with an apology, taking his confrontation seriously as well. This stands in stark contrast to the instances
where “lol” is used to soften potential confrontations and ensure that the non-literal meaning of the messages is being conveyed.

Overall, Jasmine and Michael use “lol” in a variety of contextually dependent and highly specific ways, but all of these serve to indicate that the illocutionary and locutionary forces of the message do not match and the receiver should work to recover the intended meaning. While this dissertation does not take a conventionalized approach to how these messages are recovered, it is clear that there is a group of illocutionary forces that these users are signaling. These include exchanging empathy (Examples 8 and 16)\textsuperscript{155}, flirting (Examples 10, 12, 14, 18, and 19), teasing (Examples 10 and 12), confronting (Example 19), and veiled meanings (Example 18). This range shows that because it has no meaning of its own, but signals something about the meaning of a message in context, “lol” is remarkably versatile.

3.4 Iterations

In Chapter 3, iterations (instances of repeated letters) were hypothesized to be instances of expressing the emotional state of the texter, with the option of attempting to engage the receiver in the same emotional state. In this corpus, as in many corpora, women text differently from men (Baron, 2004; Çubukçu & Kutlu, 2013; Forgays et al., 2014). Iterations are one area where women and men text differently, and this is played out between Jasmine and Michael. Jasmine repeats letters in her text messages 541 times whereas Michael only uses this strategy 157 times.

In addition to overall repeating letters a different number of times overall, Jasmine and Michael repeat different letters. Similar to the remainder of the participants in the corpus, Jasmine repeats /o/ the most often (168 times), and Michael repeats it the second most often (23 times).

\textsuperscript{155} The role it can play in signaling empathy could probably be translated as something similar to “this message is sent with lots of love.” Interestingly, this is a much older meaning for “lol”, going back to the burst of initialisms in the late 19\textsuperscript{th} century (Cannon, 1989; A. W. Read, 1963).
Figure 28 shows which letter Jasmine, Michael, and the other participants prefer to repeat in their iterations.

![Iterations by Letter (percentage)](chart)

Omitted letters were either never used by participants (c, g, k, p, and q), or used only once by one person (j, and z).

This chart shows that with the exception of repeating the letter /a/, Michael and Jasmine’s preferred letters to repeat are different from the participants more broadly. This indicates that while there may be some significance to the letters that are repeated, the actual act of repeating and the words or interjections that get iterated is more significant than the individual letter.

Similar to the rest of the participants, the words that they iterate are predominately interjections and greetings as in Example 133. In this case, they have been talking about Jasmine having a cold and staying home from school. Michael is concerned that she will not be healthy enough to see him on Sunday. When she clarifies that she will still be able to come, he is
understandably very happy and in a heightened emotional state as evidenced by his use of the iteration.

Example 133

Jasmine: Don't worry I will be fine for Sunday!! 😅
Michael: Yaaayyyyy

In this example, Michael is expressing his happiness and relief; if Michael were to simply write “yay”, it may not convey just how excited he is and it may be interpreted as though he is happy about it, but not necessarily excited, or excited, but not necessarily happy. Michael needs to do something to reinforce to Jasmine that he his very happy and excited that she will be healthy by the time they are scheduled to see each other. In this case, he uses an iteration to convey that emotion. This is in contrast to an unenthusiastic or somber expression of affirmation as in Example 134. Here, Jasmine is unhappy about the dog, so while she is confirming Michael’s statement with “yeah” (underlined), she does not iterate it. This is not an expression of joy, excitement, or an otherwise heightened emotional state. Rather, this is an instance of a depressed emotional state.

Example 134

Michael: Poor dog
Michael: Hope he found a good home
Jasmine: Yeah
Jasmine: I think he was crying

156 Face with stuck-out tongue and tightly closed eyes emoji
The contrast between these examples is taken as evidence that iterations are allowed in situations of heightened emotion, and the iteration itself is the process that encodes the emotion into a written register. Jasmine and Michael, unlike the rest of the participants, iterate terms of endearment such as “babe” as in Example 135.

Example 135

Jasmine: No deal then
Michael: Baaabbbbee
Jasmine: Okay fine you can use whatever picture you want it's up to you.

In this example, Michael is essentially using the iteration to express his displeasure in Jasmine’s refusal to let him use a picture of her on his phone. He is unhappy with her response and shows this through the use of an iteration of the term of endearment he most often uses for her (“babe”) 157. However, the use of an iteration indicates that he is only playing angry. 158 Clearly this is effective since Jasmine consents to Michael’s request. Without the iteration, the emotional content is not present, and his response may be misinterpreted by Jasmine. If he were to simply write “babe”, she could interpret it as him being more annoyed than he actually is, since it would be a more serious, prescriptive form. By contrast, they often refer to each other as “babe” without the iteration. In Example 136, they both use “babe”, but the emotional content is expressed in the literal meaning of the words as opposed to the iteration of the letters in the words. Here, using an iteration would not be appropriate since the emotional content is explicitly stated by both people.

157 This is an uncommon iteration of babe, as most instances of iterations of “babe” occur in sexting contexts, where Jasmine and Michael are exchanging photos and describing sexual acts. These are still serving the same purpose by encoding emotional content, yet those examples require many turns to parse.

158 Real anger is not expressed in this conversation, though in testing, participants stated that capital letters read as the most angry features, and other researchers have found the period to indicate anger (Crair, 2013).
Example 136

Michael:  Babe
Jasmine:  Yeah babe
Michael:  Miss you
Jasmine:  I miss you too babe so much
Michael:  😘😘

In Example 136, an iteration would take away from the emotional content being expressed. This illustrates that while iterations can be used to signal that there is an emotional overlay to the message, if the emotions are overtly stated, the iteration is inappropriate.

These findings reinforce the hypothesis proposed in Chapter 3, that iterations are independent of the sound the letter represents since if the sounds of the letter determines which letters are repeated, the lovers would behave the same as the rest of the participants since they are texting in the same languages. Whereas if the type of word contributes more significantly to which letters are repeated, it is expected that the lovers would repeat different letters since their conversation is different from the other types of conversations found in the corpus more generally.

3.5 Mistakes

Recent research based on records kept by the online dating service, Match.com, identified patterns in texting behavior that anthropologists correlate with dating behaviors such as sex and courtship (Fisher & Garcia, 2015). Most of the data relates to non-linguistic aspects, but they do find that people looking for romantic partners care about the style their potential partners text in. Fisher and Garcia found that “both single women (54%) and men (36%) find misspellings and incorrect

159 Face throwing a kiss emoji
grammar to be the biggest text message turn-offs” (Fisher & Garcia, 2015, p. 1). This implies that individuals expect that their potential partners will not only know the rules of prescriptive grammar, but also be able to use them in text messaging.

Mistakes occur regularly throughout participants’ messages. Mistakes are items that participants do not intend to type (as opposed to deviations from prescriptive grammar rules). Often, these items go uncorrected in the corpus, but occasionally, participants send a follow-up text that includes a correction. Both Jasmine and Michael rely heavily on these correction messages. In addition to primarily adhering to prescriptive spelling conventions and grammar, they are careful to show that they know how to write a text message. As mentioned previously, recent research on the communication patterns of people who are dating shows that single adults are “turned off” by someone who uses “incorrect” spelling and grammar (Fisher & Garcia, 2015). In a text message, it has not been explicitly established what “incorrect” means although within a conversation it appears that interlocutors have a sense of what is acceptable and what is not. For example, it is not incorrect to omit the period from the end of a message even though omitting periods is incorrect according to prescriptive grammar. Jasmine and Michael’s messages give more insight into what is allowed and what must be corrected in their conversation.

3.5.1 Punctuation

In writing, the broad purpose of punctuation is to make the writing more comprehensible (Christin, 2002). The most dominant explanation for punctuation in non-academic settings is that it marks prosody and utterance breaks (Chafe, 1988; Squires, 2012). This idea is used in many computational approaches to language, and is often taught as a way to “proofread” punctuation in school. Another theory of punctuation is that there is a hierarchy of punctuation in English, from the period to the semicolon to the comma, and the longer an utterance is, the more necessary it is to include one of these marks to separate the content (C. Meyers, 1987). In text messages, the purpose is complicated,
ranging from identifying the type of message (i.e., a question, exclamation, etc), to encoding emotional content (Baron & Ling, 2011; Crair, 2013), or to mark prosody (Baron, 2010b). The punctuation in this corpus is most likely used for both purposes (i.e., to mark prosody and to separate complex ideas into comprehensible units). 160 Txt is notorious for having extraneous punctuation that does not fit in to prescriptive norms (Crair, 2013; Crystal, 2009; Provine et al., 2007b; Squires, 2012). In the absence of a research project specifically on the meaning of punctuation in text messaging, the only data available is what participants have used in their messages. This means that when punctuation is omitted, it cannot be taken as evidence. With this obvious limitation, punctuation will be investigated only as it is either corrected or deviates from the participant’s normal texting patterns 161. In almost all cases of mistaken punctuation, the intended item is a question mark. Both commas and exclamation points are substituted for question marks, as well as other, extraneous marks such as letters.

3.5.1.1 Commas

Commas are most often used by these participants after a transition word (such as “anyway” or “still), or at a clause boundary (especially when going from Spanish to English or vice versa) as in Example 137.

Example 137

Jasmine:     Yo quiero hacer tantas cosas & ima ended up doing something that I don’t like

I like doing lots of things, and I’m going to end up doing something that I don’t like

160 There is evidence that the punctuation participants use has a third purpose, to encode non-literal meaning into the message. However, this usage will be set aside for future work.

161 There are so many fascinating and potentially illuminating patterns of punctuation use that another entire research project could be devoted to the role of punctuation on text messaging, especially looking at how exclamation points encode emotional content and potentially provide an invitation to the receiver to be excited with the texters, and the role of question marks as stand-alone messages, and how they are different than explicitly constructed questions.
These are traditional, prescriptive uses of commas. The use of commas after transition words accounts for the prosodic motivation of punctuation use, and their use at clause breaks accounts for the conceptual motivation. Both participants use commas frequently (Michael uses 226 of them, and Jasmine uses 203 of them). Some researchers would say that this is because both Jasmine and Michael regularly send multi-clausal messages or messages that involve transition words (Squires, 2012). Even though they both use them frequently, it appears that they are not always necessary for comprehension as mistakes involving commas are frequently left unresolved. The punctuation in Example 138 should be a question mark by all prescriptive measures, since Michael is asking a question and Jasmine responds. Yet, Michael ignores the mistake, does not correct it in the second line, and moves on with the conversation even though in many other instances, he corrects the mistakes he makes.

Example 138

**Michael:** How'd you sleep,

**Jasmine:** Great

**Michael:** Me too

It is clear that this is most likely a mistake since at the end of other questions, he uses a question mark as in Example 139.

Example 139

**Michael:** You work all week?

**Jasmine:** Idk yet! Mira my boss is so Cara dura, he just told me to came tomorrow at the same time

*I don’t know yet! Look, my boss is so tough, he just told me to come tomorrow at the same time.*
It appears as though this mistake is a regular issue for Michael as he makes this same substitution of a comma for a question mark again in a later message and it also goes unrepaired as in Example 140.

Example 140

**Michael:** Y ahora que haces,

*And now what are you doing,*

**Jasmine:** Aquí arrecostada

*Lying down here*

The most important thing to note about this mistake is that the meaning could actually change based on using a comma in place of a question mark. The comma indicates that there may be more to the utterance whereas a question mark indicates that the message is final and Jasmine should answer the question. However, it is clear that Jasmine knows that she should answer the question rather than wait for the remainder of the message since she provides a response within a few seconds.

Apparently both from context and the norms of their conversation, it is clear to her that the comma can take the place of a question mark. She is able to repair the conversation and recover the meaning without the appropriate punctuation. However, if there is any possible ambiguity, it appears that repairing the mistake is necessary as in Example 141.
Example 141

Michael: Who do you wanna be like? 1:28:58 PM

Jasmine: Como nadie babe! 1:30:09 PM

Like no one babe!

Jasmine: Quiero ser yo misma 1:30:40 PM

I want to be myself

Michael: The best version of you, 1:30:53 PM

Michael: ? 1:30:53 PM

In this example, the comma could actually mean that Michael wants to complete his thought in the next message. Instead, he repairs the comma to indicate that the message is a question, not the first clause of a more complex thought.

Since these messages were sent so close together, it is unlikely that he explicitly considered whether it could be ambiguous or not. Rather, he likely sent the question mark to repair the question immediately upon noticing the mistake in a subconscious realization that the message is ambiguous. This happened so quickly that there was likely little time for rereading and reflection, meaning that Michael likely is aware that some messages require a question mark and others do not. It appears that, for Michael, repair is only necessary if there could reasonably be ambiguity, not if there could possibly be ambiguity. This pattern exemplifies one way that Jasmine and Michael correct mistakes: to clarify in the situation where there is reasonable ambiguity.

3.5.1.3 Exclamation points

The other common mistake is to substitute an exclamation point for a question mark, which Michael frequently does, and usually repairs the error in the following utterance. There is not a clear
case in the corpus where an exclamation is used yet the message is a question and it is not repaired, suggesting that he always repairs this error as in Example 142 and Example 143.

Example 142

Michael: Te acuerdas de las promesas que te pedí! 11:28:48 PM
           [Do] you remember the promises you made!

Michael: *? 11:28:48 PM

Jasmine: Si 11:29:59 PM
          Yes

In this example, the exclamation point signals that the message may be a direct command rather than a question since there is nothing about the syntax that indicates that it is definitely a question. Therefore, if Michael does not immediately repair the exclamation point, it may be interpreted as a confrontation and an affront to Jasmine’s positive face wants. Even though within the context of the conversation, it does not make sense that Michael would be directly commanding Jasmine to remember her promises, he still appears to be obligated to repair the mistake. This is likely due to the nature of text messaging. Since it is decontextualized and there are no verbal cues, the only information in being conveyed through what is written. The exclamation point in this message could reasonably be ambiguous taken out of the context of their conversation, so he must repair it immediately. Second, the exclamation point construction has a meaning that is an affront to Jasmine’s positive face wants.
Example 143

Michael:  Got your bikini packed!  1:20:21 PM
Michael:  ?*  1:20:22 PM
Jasmine:  Yeeaaap  1:21:14 PM

In this case, Michael probably is not commanding Jasmine to pack her bikini so there is little ambiguity, but again, Michael repairs the mistake. The alternative interpretation is actually a humorous situation in which Michael packed Jasmine’s bikini and is enthusiastic about it. This is clearly not the situation and therefore there is no reasonable ambiguity.

To account for this, the original explanation should be revised to state that when there is a sensible ambiguity, repair is necessary. That is, when there is any possible ambiguity, and the message could be interpreted another way that makes sense, the mistake must be repaired in order to clarify the ambiguity. For example, if things were another way, Michael could be exclaiming that he packed Jasmine’s bikini. Because that is a world that could reasonably exist even though it is not the world we are living in, he must repair the mistake.162 This accounts for the unrepaird commas, since there is no alternative that has any logical meaning, that is, there is no alternative world in which the comma makes sense at the end of those utterances, and it accounts for every instance of repairing question marks.

The explanation for which mistakes are repaired and which are not at this stage is that those which are not repaired must be easily recoverable for the intended mark, and are largely non-sensical in that they are grammatically and semantically wrong in any hypothetical world. Those that must be

162 An alternative explanation is that exclamation points and question marks have pragmatic meaning of their own in the same way that “lol” or an iteration does. These explanations are not mutually exclusive, though a full treatment of the meaning and features of punctuation required for this explanation is beyond the scope of this dissertation.
repaired either propose a different meaning in this world, have an unintended pragmatic meaning, or propose an alternative world, in effect proposing a different meaning.

3.5.2 Spelling

Whereas Michael makes more grammatical mistakes, Jasmine makes more spelling and auto correct mistakes. Just as with punctuation mistakes, spelling mistakes do not always have to be corrected, though they often are. Again, similar to punctuation, there are many instances where it is impossible to determine if the non-prescriptive usage is an alternative usage or an actual mistake. Therefore, for this discussion, only corrected mistakes will be addressed because those are overtly identifiable as mistakes.

The first example (Example 144) is an extreme example of misunderstanding that Jasmine must correct in order to be understood by Michael. The mistake resulted in a long repair exchange even though Jasmine repaired the mistake very quickly (as above, times are included to identify the length of time the mistake is allowed to persist.

Example 144

**Michael:** So you tell me. Are you scary when you're mad? 10:53:43 PM  
**Jasmine:** All I can say is that I'm violent 10:55:11 PM  
**Michael:** Uh oh. 10:55:29 PM  
**Michael:** You would hit me?!?! 10:55:32 PM  
**Jasmine:** All I can say is that I'm not* violent ***** 10:55:53 PM  
**Jasmine:** Lol 10:56:02 PM  
**Michael:** Lmao I'm confused now 10:57:24 PM  
**Jasmine:** Yo no doy golpe lol 10:57:33 PM  

* I never hit lol  

**Michael:** So you're NOT violent? 10:57:41 PM
In this example, Jasmine wrote the exact opposite of what she meant and the mistake was allowed to persist for 40 seconds, during which, Michael sent two messages drawing attention to the mistake. Since this mistake conveyed the exact opposite of what Jasmine intended to write, to emphasize that it was an extreme mistake, she repeated the asterisks that is most often used to signal that a message is a correction. Even with the emphatic correction, Michael continues to either tease her or genuinely ask for clarification of her intended meaning. This is an obvious case of necessary repair, most instances are not this extreme.

In Example 145, Jasmine repairs a spelling mistake that could potentially be ambiguous, but Michael likely would have been able to recover the meaning without the repair.

Example 145

Jasmine: o se, porque la segunda pantalla medio rata tambien 11:09:48 PM

Or because the second screen also means rat

Jasmine: *media rota 11:10:01 PM

Half rotates

In this example, there is no reasonable alternative reading other than that the screen rotates, yet Jasmine feels the need to repair the mistake. In this case, the world would have to look very different from the one Michael and Jasmine are in for that statement to make sense, although it is possible. This is taken as evidence that the motivation for mistake repair is not necessarily for comprehension, but to portray a certain version of oneself. She takes thirteen seconds to make this repair, in contrast to Michael, who generally repairs mistakes in under two seconds. This may be because she consciously proofreads her messages or because it takes her that long to write them or because she is

163 Again, this indicates that the asterisk itself carries a meaning beyond a simple typographical convention. A full analysis of the role and meaning of asterisks is, of course, beyond the scope of this dissertation.
simultaneously writing to friends and family members as well. In any case, she has time to
consciously consider the message when she is correcting spelling mistakes.

Example 146 is another, even more extreme example of a case where there is no possible
ambiguity as to the intended meaning, yet Jasmine still repairs the mistake.

Example 146

Jasmine:  Lo boy a secar & despues la plancha  7:49:22 PM
           I am going to dry it and then iron it

Jasmine:  *voy  7:49:31 PM
           Will

In this example, there is no other meaning she could have other than “voy,” and “boy” is not even
in Spanish, and she never code switches between an article and its noun, making it highly unlikely if
not impossible that “boy” is the intended word. Yet, she still repairs the mistake. There are a variety
of types of mistakes throughout that both Jasmine and Michael repair that do not make any
difference in terms of meaning yet they still repair them. This is evidence against the initial
explanation of why some mistakes are corrected and some are not, and argues for an explanation of
mistake repair that takes into account the identity that a texter is trying to construct as much as the
meaning they are trying to convey.

3.5.3 Mistakes conclusion

Considering all of these types of mistakes together, it appears that while ambiguity is certainly an
important factor in determining which mistakes are repaired and which are not, it is not the only
contributing factor. It may be that presentation and identity construction is a more important factor
than resolving ambiguity. For example, it may be that prescriptive punctuation is not considered part
of the “rules” of text messaging while prescriptive spelling is. Punctuation in text messaging allows
users to convey meaning in a non-formal way and express non-linguistic features of their messages (Baron & Ling, 2011; Crair, 2013; Provine et al., 2007b). There is a long tradition going back to FidoNet of non-prescriptive punctuation being used in digital communication (i.e., :c pouting or … thinking (Bush, 1992). Spelling conventions, on the other hand, can be adhered to (i.e., using prescriptive spelling), flouted (i.e., using an allowed respelling to convey subtextual meaning), and violated (i.e., a mistake). According to the Match.com survey, “improper” grammar and spelling is an undesirable quality in a romantic partner. However, if the partners are close enough, it is common practice to use respelled words and phrases throughout romantic text messages (Ansari & Klinenberg, 2015; Baron, 2004). It is either that texters are unaware of this preference, or some types of respellings are allowed and others are not. This dissertation suggests that respellings that carry subtextual meaning are therefore flouting the prescriptive spelling norms and are allowed whereas violating the norms is not. This explanation allows for non-prescriptive spelling and punctuation use while still encouraging romantic partners to repair mistakes in their messages.

4 Code Switching

Throughout this conversation, there are three types of language mixing, of which two types will be considered here. The first type is traditional code switching, where both Spanish and English are incorporated into the same message. The second type is the use of an English initialism in an otherwise Spanish message. The third is switching between the language of conversation (i.e., one message in Spanish and the next in English), but this third type will not be addressed here. The two types of code switching will be analyzed separately, though they can both give insight into the particular techniques bilinguals have access to in text messaging.

---

164 This type only exists in discourse contexts, and while the lovers’ conversation is a case study of one conversation, making discourse analysis possible, it is not possible for the corpus more broadly and will therefore be set aside for another project.
Before discussing code switching in detail, some terms must be defined. First, code switching refers to mixing Spanish and English into the same text message by starting a message in one language and then switching to another language. This is a widely accepted definition at the sentence level (Poplack, 1980) that has been used in other studies of code switching on computer mediated communication platforms (Dorleijn & Nortier, 2009; Montes-Alcalá, 2007) and is used in this study at the text message level. Borrowing is a different phenomenon where one word from the other language is incorporated into the message without any observable use of the grammar of the other language. There has been considerable discussion about what constitutes code switching especially on digital interfaces (Alvarez-Cáccamo, 1998; Auer, 2011; Dorleijn & Nortier, 2009, 2009; Morel, Bucher, Pekarek-Doehler, & Siebenhaar, 2012; Paolillo, 2011), and while the working definitions used for this study may initially appear to distinguish between the two types, there are ambiguous instances such as in Example 147.

Example 147

**Michael:** Alright! No te me pierdas 🧡

*Alright! Don’t miss me 🧡*

In this case, based on this definition, it is nearly impossible to determine if Michael is code switching or borrowing. He starts the message in English, switches at the sentence boundary and then finishes in Spanish. However, the first clause/sentence is only one word, which creates the ambiguity. Further investigation of Michael’s texting patterns reveal that this should most likely be classified as an instance of borrowing because all of the instances in which Michael is code switching (5

---

165 🧡 Face throwing a kiss emoji
instances) are ambiguous between code switching and borrowing. He does, however, often borrow both words and initialisms into his Spanish and English messages as in Example 148.

Example 148

**Michael:** Y tu *baby* que haces?

*And you *baby* what are you doing?*

Jasmine, on the other hand, regularly starts a message in one language and finishes in the other, clearly switching both words and grammar. Jasmine does this approximately 125 times, as in Example 149.

Example 149

**Jasmine:** Que tu uvieras dicho *[if I would say no] lol* 166??

*What would you have said *[if I would say no] lol*??

In this example, the first half of the message is in Spanish and the embedded clause is in Spanish, and is a clear case of code switching since Jasmine uses English words and structure (semantics and syntax) and then switches into Spanish semantics and syntax.

Research on spoken code switching finds that it is only allowed in places where it will not violate syntactic rules of either language (Myers-Scotton, 1997; Poplack, 1980). Clause breaks are therefore ideal locations for language changes since each clause is syntactically and semantically well formed without depending on the matrix clause (‘matrix clause’ is also referred to as the ‘main’ or ‘superordinate’ clause). Jasmine engages frequently in texted code switching that occurs at clause boundaries. Michael, however, does not engage in this type of code switching.

166 Flushed face emoji
There are a variety of speculative reasons for why Michael uses less code switching and more borrowing. One that has been proposed before is that code switching requires that speakers have command of both languages (Alvarez-Cáccamo, 1998; Auer, 2011; Poplack, 1980); it may be the case that Michael's Spanish is not strong enough for him to go between the languages. Alternatively, the social connotations with code switching may prevent Michael from engaging in it. In many communities, code switching is stigmatized and viewed as “bad” or “broken” language (Alvarez-Cáccamo, 1998; Gumperz, 1977). For this reason, Michael may avoid code switching in general as he is trying to portray himself as someone who is educated enough to keep the languages separate. In face-to-face conversations, he was only recorded in English, and his tests are all in English, so it is impossible to know which explanation is more accurate.

Researchers who have investigated the role of code switching and language mixing on digital platforms have found that identity construction is a primary reason for users to mix language in chat rooms, public forums, and on long-form blogs (Dorleijn & Nortier, 2009; Montes-Alcalá, 2007; Paolillo, 2011). While these researchers acknowledge the distinction between code switching and borrowing, they do not make a distinction between the two types in terms of identity construction. It appears that, for these formats, code switching serves the same purpose as language mixing for establishing a bilingual identity. All three of these studies have in common that the platforms they are interested in are public. A private platform, such as text messaging is different because both interlocutors are aware that the other is bilingual and a member of a bilingual speech community. Therefore, the identity-motivation is reduced. One study did investigate the role of language mixing on text messages between isiXhosa and English (Deumert & Masinyana, 2008).  

This is a very different study because the social and technological relationship between these languages is vastly different from the relationship between Spanish and English. For example, in 2007, when they were collecting their data, there was no such thing as an isiXhosa keyboard and very few pages written in isiXhosa. Spanish, on the other hand is
texters treat the languages very differently and mixed isiXhosa into English matrix messages in order to convey emotional content or approach sensitive topics (Deumert & Masinyana, 2008).

Jasmine uses code switching regularly, and she may be doing so to add affect to her messages (though both Jasmine and Michael discuss sensitive or emotional topics in both languages168).

Example 150 is at the end of a conversation about how Jasmine gave Michael a cold and the way she is choosing to describe it. Michael just asked if *catarro* is her favorite word. *Catarro* is an impolite way of describing a cold.

Example 150

**Jasmine:** For the winter….. Yeah lol my mom se encojona quando digo asi lol

*For the winter……. Yeah lol my mom was annoyed when I said it lol*

This message is unique for a variety of reasons. First, there is code switching. Jasmine answers Michael’s question in English (“for the winter”), then indicates a pause with the use of ellipses, then continues with an interjection in English (“yeah”), and then the first “lol”. This first “lol” is also the turning point in the utterance when she switches into Spanish to make the commentary on her answer. The English portion would suffice to answer the question, but the Spanish is used to change the tone slightly from the answer to her commentary on the answer. She moves from answering Michael’s question to telling an anecdote related to the answer. The change of language is likely a way to distinguish between the two parts of the conversation and indicate to Michael that she is moving to the next topic (Gumperz, 1977; Paolillo, 2011). “Lol” is at the clause boundary here, and is used to end the English clause since “lol” can only occur once per clause.

---

168 Sexting (having a sexual conversation via text message) may be considered among the most intimate of conversations yet Jasmine and Michael sext mostly in English and occasionally in Spanish, indicating that they are not distinguishing between intimate topics and casual topics with language choice.
It should be noted here that “lol” is used twice. “Lol” is used twice in the same text message only nine times in the entire corpus. In six instances, it starts and ends the message, so the first and last items are “lol.” However, in three instances, it appears at the end and at the clause boundary as in Example 38 (and Example 39 below). The first “lol” likely indicates that she is aware of how derogatory the word she has chosen is and that she is indicating to Michael she knows that he knows that, too. As mentioned before, Jasmine and Michael are establishing their relationship, and taking it very seriously. She may be embarrassed by the word she is using and indicating to Michael that she wishes to soften the harshness of the word, and conveying to him that she is aware of how harsh it sounds. The second “lol” likely applies to the second clause, showing that even though her mother was annoyed at her use of the word, she is still taking the situation lightly, and Michael should, too. It is interesting that one “lol” would not suffice since there are plenty of multi-clausal messages with only one “lol” in it that takes scope over (applies to) the entire message. It appears that there is more than just a difference between the words here. As it was argued in Chapter 3 (to explain why some expressions are initialized in one language but not another), pragmatic markers and processes are language specific and the pragmatic markers can only take scope over the language that they are used with. So while “lol” is an element that belongs to both Spanish and English, it can only be used to apply to one message at a time. This text message is evidence that pragmatic mechanisms are language dependent and confined to where that language is being used. That is, when a language switch occurred, not only are the syntactic and semantic marker confined to that clause, the pragmatic markers can only operate on that clause as well.

The way that both Michael and Jasmine employ borrowing is similar to Example 38 above in that it is often used to add playful commentary or an interjection (Paolillo, 2011; Tagg, 2009). Again, this type of switch often occurs at a clause boundary, making it ambiguous between code switching and borrowing. This may be evidence that borrowing is best used when it is in a syntactic place.
where code switching is allowed. However, without a much larger study, that is difficult to determine. In the following example, Michael and Jasmine have been trying to figure out when they are going to meet next. At the beginning of the conversation, they had said Sunday, Example 151 is the 10th exchange in the conversation.

Example 151

**Jasmine:** Entonces Sunday is better

*So Sunday is better*

Jasmine switches at the clause boundary though she is only using one word. Many of both Jasmine’s and Michael’s one word switches are of this type or it is something that exists in Spanish but not really in English (i.e., *telenovela*). In this way, Jasmine is again adding a commentary to the message. *Entonces* can have a “finally” type of reading to it, suggesting that she may be tired of trying to figure out which day to meet, and just wants to go back to the original plan of meeting on Sunday. The switch into Spanish (the rest of the messages are in all English) may be a signal that she is adding commentary that doesn’t necessarily fit into the immediate conversation, but that she wants Michael to be aware of. Likewise, one of the common functions of language switching is to add emphasis (Gardner-Chloros, 2009; Gumperz, 1977). Therefore, this “entonces” may be serving to add emphasis to her statement here.

5 Lovers Conclusion

Overall, Jasmine and Michael’s conversation reinforces the findings proposed for the corpus more broadly. The Txt features that they use (initialisms, iterations, and corrections) highlight how important these features are to communicating non-literal and non-verbal meaning on a completely text based platform. Furthermore, the regular correction of mistakes as well as the avoidance of abbreviations and AAE-influenced initialisms suggests that identity construction is more important
in a romantic relationship than in a platonic one. However, while the standard they adhere to is
driven by prescriptive spelling and punctuation norms, the conventions of Txt allow them to deviate
from this without appearing as though they are ignorant of the prescriptive conventions.

Both Jasmine and Michael use initialisms throughout their messages, though they carefully
select which initialisms to use. In every instance, the initialism is used to convey the growing social
closeness between them and the positive feelings they share for each other. They both use “lol”
more regularly than any other participant in the study. This is evidence for the conclusion that “lol”
is used to signal that a message is intended positively and to signal empathy (either displaying
empathy, or asking the receiver to be empathetic). Furthermore, their use of “lol” supports the
proposal that “lol” is used to indicate a sub-textual meaning, and that the illocutionary force of the
utterance (or of the lack of the “silence” in a message that only contains “lol”) is important to be
attended to.

Likewise, they use iterations to encode their emotional state and repeat letters based on their
position in the word rather than the sound of the letter. The extension of this behavior to terms of
endearment further points to an emotional motivation since terms of endearment (such as “babe”)
are inherently emotionally driven.

In conclusion, by behaving as other participants do (with initialisms and iterations), and at
the same time deviating from what other participants do (with abbreviations), Jasmine and Michael
confirm the findings in Chapter 3. They have shown that abbreviations are largely used to indicating
in-group membership and mark a casual tone to the message, which is something that is either not-
true (they are becoming in-group members, but they are not yet) or something they wish to avoid (as
the casual tone would signal that they either were not taking the messages seriously or that they
weren’t taking their relationship seriously). They have shown that initialisms such as “lol” and “jk”
are useful in signaling to the receiver that there is positive feelings and a sub-textual meaning to the
message. Finally, they have confirmed that iterations are effective in encoding emotional content onto a message and signaling to the receiver that the texter has an emotional meaning as well.
Chapter 5 - Conclusion

This study set out to describe the academic, social, and txt literacy skills among an under-studied population, emergent bilingual youth. The results of this study show that students have communication skills in academic, social, and txt forms that are not traditionally assessed in academic settings, and may have stronger receptive than productive academic skills. The second objective of this study was to identify communication norms and patterns emerging in the Spanish/English bilingual texting register as well as document this language form for historical purposes. Through this investigation, it has become clear thatTxt has a highly systematic set of conventions that texters rely upon to create meaning and communicate on a completely text-based platform. Furthermore, even though the average academic reading level for participants in this study is at approximately fifth grade, they are adept at manipulating the linguistic features of Txt to communicate on a completely text-based platform.

Part of the data collected for this study are 44,597 text messages collected directly from participants’ phones. These messages form the first bilingual part of speech tagged, publically available corpus of bilingual text messages (Spanish-English BYTs – Bilingual Youth Texts). This corpus is unique for many reasons. First, it was collected directly from participants’ phones and contains extremely rich metadata as well as demographic information about the participants and the relationships between texters and receivers. Second there is one intact conversation of over 30,000 messages that spans to first four months of a relationship; to the best of my knowledge, such a candid and intimate look at the establishment of intimacy between two people through text messaging has never been made available before. Furthermore, it presents the first bilingual Part-of-Speech tagged corpus available for public use with full documentation of the code used to implement the tagging. As more communication moves onto digital platforms and more individuals
become multilingual, the ability to work with and analyze multilingual texts on digital platforms will only increase. This project provides a preliminary framework for how to work with multi-lingual texts.

While there has been extensive research on the role of text messaging on the literacy skills of children and young adults (Drouin, 2011; Drouin & Davis, 2009; Plester et al., 2008; Wood et al., 2011), there has been little research that takes into account academic literacy in aural and digital modalities as well. Furthermore, this is the first study to look specifically at bilingual GED students, and document the linguistic skills of this understudied population. It is clear that these students are academically underprepared for the grade level they are in, even if they may not have had large temporal gaps in their academic careers. Most research discusses emergent bilingual students in terms of their linguistic deficiencies (Bartlett & García, 2011). Rather than pointing out exclusively what they are unable to do, this study has taken a holistic approach to students’ literacy skills with a comprehensive documentation of their academic, social, and Txt literacy skills across written, aural, and digital platforms. The results are important in developing an understanding of how language and communication norms are changing as more conversations and interactions take place in a computer-mediated format.

The results from this study can also be applied to classroom settings by supporting students’ complete linguistic development. Students come to the classroom with a wide range of linguistic skills. Often their skills in one language or area may be stronger than their skills in another language or area (Shaughnessy, 1976, 1979). By recognizing and valuing the linguistic skills that students do bring to the classroom, it is possible to use those as a foundation to build academic literacy skills. For example, a student may come to the classroom with very limited ability to express her ideas in
academic writing in English, but may be able to understand spoken academic English very well. If this student uses the writing skills that she does have to communicate what she understood, she is using writing for communication of an academic idea and illustrating that she has some academic English language skills. However, if this student must develop her academic writing skills before completing such a task, she may never be able to express that she understands academic English. A holistic approach and recognition of separate competencies allows students to engage with academic content before they have fully mastered academic writing in English (or Spanish).

1 Relationships between academic, colloquial, andTxt literacy skills

Through assessment of three types of literacy across multiple modalities, this research sought to isolate factors that contribute to student achievement on different language tasks. In this process, it was discovered that participants in this study perform at approximately a 5.5 grade level averaged over all of the Spanish and English diagnostics, with a range from 3rd grade to 8th grade on certain diagnostics. This level of academic under-preparedness is similar to another population of students, SIFE (Students with Interrupted Formal Education) even though they are not identified as such. SIFE are students who have at least a two-year gap in their education. These students are typically academically underprepared in their first language, which makes developing academic language and content mastery in their second language even more of a challenge (Klein & Martohardjono, 2006). SIFE Students tend to perform academically at least two grade levels (often more) below their age-grade level. These populations (SIFE and the emergent bilinguals in this study) are similar in that they are both underschooled. They are dissimilar in that the emergent bilinguals in this study have not had a break in their schooling, as most have been in school consistently since age five or six.

169 Many students in this study had exactly this combination of skills.
Characteristic of all GED student populations (not just this population), attrition rates were very high in this study (Cook et al., 2014) due to scheduling conflicts and inconsistent attendance. One hundred students participated in at least one activity though only half participated in more than one assessment. There was a significant change in attendance after Spring Break, many students who had been coming dropped out, and another group enrolled in school and joined the study. The demographic background of the students who joined was very similar to the group who left. Teachers reported that this happens every year and they are not sure why, but the enrollment overall stays about the same even though students regularly come and go.

There are two major take-aways from the relationships between the different types of language competencies. First, students with higher literacy levels overall do tend to text more in English. However, the lexical density of the English messages was much lower than that of the Spanish, indicating that even though students may be more capable in English than other students, they are still more capable in Spanish than English. This finding is particularly important for its implications for the classroom. If the most English proficient students in the study are sending more complex messages in Spanish than in English, it is clear that to convey complex ideas, students use their first language more than their second. Clearly, in a classroom setting, it would then be more pedagogically sound to have students first express their ideas in their first language and then craft the same message in their second rather than force production of a complex message in a second language when they are still developing linguistic skills in that language (in social or academic registers).

The second finding from this section is with respect to the relationship between different testing modalities. There was a significant correlation between the Academic Chatting Game and the Aural Comprehension Task. Both of these tests were unique in that students had to understand academic content, but they did not have to produce academic language. Likewise, these were the
only tasks involving a multimedia platform. Students were free to use any language form that they could, provided that they could convey their ideas. They were assessed based on comprehension, not production. The results indicate that students have stronger receptive than productive skills and that they perform better with multimedia than with traditional written formats. However, it is not clear from the findings that this freedom benefits all students, as the average performance on the Aural comprehension was at approximately grade level 5.2 (slightly below the 5.5 overall average), and 5.69 on the Chatting Game (slightly above the average). So, while the freedom to use both languages may benefit some students, it does not benefit all students.

2 Spanish/English Bilingual Txts

The development of the Spanish/English BYTs corpus allowed for an investigation into the texting conventions of urban bilingual youth. The participants who donated their messages differ from most text message donors as they are not college educated nor are they in a college preparatory class. Many corpora have been challenged for providing a skewed perspective on Txt because the message donors disproportionately value academic writing. The results show that participants in this study adhere to conventional spelling norms in both Spanish and English at rates comparable to those found in other corpora (Bernicot et al., 2012; Tagg, 2009).

Analysis of participant respelling patterns showed a high degree of systematicity throughout and that different types of respellings convey different messages. The use of abbreviations and most initialisms conveys social closeness, saving positive face of the receiver, or a casual, playful tone. The use of iterations (repeated letters such as “hellooooo”) adds emotional information to the message, largely serving to convey a heightened emotional state to the receiver. Furthermore, building on previous research on the pragmatics of code switching in text messaging (Deumert & Masinyana, 2008), it was shown that pragmatic processes and respelling rules are language-specific, and both
Spanish and English adhere to different conventions of respelling. Finally, analysis of mistakes highlights the ways that texters use language form to construct their identity in a digital format.

Building on previous work on “lol”, “ok”, and emoticons (Beach, 1993; Dresner & Herring, 2010; McWhorter, 2013b; Uygur-Distexhe, 2014), this dissertation proposed a hypothesis for the function of “lol” in text messaging. Throughout the corpus, “lol” is used as a purely pragmatic marker that serves to signal when the locutionary force of a message does not match the illocutionary force (i.e., the intended meaning does not match the literal meaning). Previous work on “lol” showed conclusively that it is not meant to signal something funny, but none of the proposed definitions were able to fit all use cases. However, by reinterpreting “lol” as a purely pragmatic marker that alerts the receiver to the mismatch between the illocutionary and locutionary forces, all of the use cases for “lol” are captured, and its overwhelming presence in text messaging is explained by the lack of non-verbal cues in the construction of meaning from language.

Some of the respellings participants use are the result of contact both between Latin American Spanish and Academic English as well as between Latin American Spanish, African American English (AAE), and Academic English. The first case where this is obvious is the use of the k-cluster; the k-cluster is when the letter, ‘k’, is used in place of “qu” or “que.” This most likely derived from contact between Spanish and the American English alphabet. Likewise, initialisms such as “klk” (qué lo que, what’s up) are likely the result of Dominican Spanish speakers coming in contact with AAE speakers as these Spanish initialisms have syntactic features that do not exist in other varieties of Spanish. Similarly, participants use a variety of initialisms and abbreviations in English that are clearly influenced by AAE such as “wya” (where you at), and “ay”(a’ight, alright), illustrating the colloquial nature of texting and the power that linguistic contact has on language change.

Finally, in addition to the features resulting from linguistic contact, participants mix languages through both code switching (starting a sentence in one language and finishing it in
another) and borrowing (inserting one word or initialism from another language) in their messages. Most instances of borrowing occur when participants insert one English word or initialism into an otherwise Spanish message. This may be due to the fact that most participants are Spanish-dominant bilinguals and the majority of initialisms participants use are initialisms of English phrases rather than Spanish (even excluding universal initialisms such as “lol” and “ok”). This may also be due to the emergence of a new variety of Spanish that has borrowed extensively from English textisms.

Languages in contact worldwide borrow from each other, incorporating vocabulary and adapting to fit the needs of the matrix language. Languages such as Japanese are so adept at this that there is an entire character system\(^{170}\) devoted to “foreign” words that have been borrowed into Japanese.

English is another language that frequently takes words from the languages it comes in contact with such as Luganda in Uganda, and Chinese in Singapore, to create varieties of English that are distinct from each other (Bamgbose, 1998; Yano, 2001).

Overall, spelling provides a flexible domain for texter to add pragmatic features to their messages as well as express individuality and creativity. The role that “lol” plays in constructing meaning is symbolic of the Txt domain more broadly as texters work to add extra-semantic content to their messages.

### 3 The Lovers

One significant part of the Spanish-English BYTs corpus is a conversation spanning the first four months of a relationship between two participants (named Jasmine and Michael in this dissertation). These two participants only see each other once a week yet they send hundreds of text messages a day, suggesting that texting is a significant part of their relationship, and through this conversation,

---

\(^{170}\) Japanese writing consists of Kanji, hiragana, and katakana. Kanji generally carries semantic meaning, hiragana is used for particles and inflection, and katakana is derived from hiragana to write words that have been borrowed into the language.
they are doing the emotional work to establish trust and intimacy. Through their conversation, they confirm many of the findings from the previous sections, both by using and avoiding features of Txt.

Both Jasmine and Michael use initialisms; ‘lol’ is so common that it appears in 25% of Jasmine's messages. Given the kind of conversation they are having (i.e., one that is full of flirtation and seeks to establish intimacy and trust), it is clear that much of what they are saying needs to be interpreted from the context. Therefore, the illocutionary force is more significant than the locutionary and the exchange of empathy is crucial to their conversation. Michael uses initialisms to express laughter and to ensure that Jasmine knows when he is joking. Without the initialisms to flag both of these conversation events, their messages appear too formal even for a romantic setting.

Jasmine uses more iterations than Michael does, and chooses to iterate different letters and words than Michael. The messages and words they choose to iterate reinforce the hypothesis that iterations encode emotional content. The discrepancy in their iterations may be because Jasmine overall expresses more emotional content in her messages than Michael does. Both Jasmine and Michael correct mistakes even when it has no effect on the meaning of the message and when the message is not ambiguous. Furthermore, mistake repair is very fast, generally occurring within 3 seconds. These behaviors reinforce the importance of texting form in self-presentation and identity.

4 Implications

The results from this study show that the features associated with Txt are used systematically to ensure accurate and clear communication between interlocutors. These features are defined by the community of users themselves (that is, Txt is a bottom-up language form rather than top-down like academic language, which is defined by the institutions that teach it). In many ways these features serve to encode information that, in face-to-face communication is encoded through prosody, speech volume, rate of speech, gestures and facial expressions. The result is a response to Hymes’s
appeal to linguists to systematically document the social features of communicative codes (languages), and approach the pragmatic or supra-conversational features of communication with the same academic rigor as the syntactic and semantic features (Hymes, 1967). Txt provides an ideal format to do this as pragmatic content must be explicitly encoded in a permanent way (i.e., written into a text message).

In a practical sense, there are immediate implications and applications for bilingual education. First, this research shows that while the students with the highest English proficiency levels also text in English the most, they also tend to write more simply in English than in Spanish. While they may be communicating with English monolinguals or using English for stylistic or pragmatic reasons, the messages they produce in their second language are less syntactically complex than those in their first. Likewise, when students were given a receptive task such as watching an educational video and asked to respond to it in any language in any form, they performed much better than when they had to respond in academic English. This shows that while students may not be able to produce academic content in English, they are able to understand aural academic content at a level higher than their production would indicate. Other researchers have also found that students’ receptive skills are better developed than their productive (N. F. Davies, 1976; Lee & Lyster, 2016). Correspondingly, an approach that allows students to utilize all of their linguistic skills has been argued for in the translanguaging literature (Celic & Seltzer, 2011; Wei & García, 2013). By confirming this finding for multiple literacy types, it becomes clear that the effect is not register-dependent, and students are better able to express themselves in their first language across registers (Txt, academic, and social). Therefore, the effect that teachers witness in their classrooms, where students can express themselves better in their first language than their second is not classroom-specific.
The second major implication from this dissertation is for our understanding of text messaging. Digital communication technologies are becoming increasingly omnipresent, and more communication is occurring in a printed form than ever before. The effect of these technologies on every area of modern life cannot be underestimated as it ranges from changing the life of fishermen in Uganda (Buys, Dasgupta, Thomas, & Wheeler, 2009) to teenagers in New York City (Boyd, 2014). Therefore a deep understanding of how language is evolving to meet the communication needs of users is crucial to successfully communicating via Txt in the future. Through investigation of a wide range of respelling types, this research proposes a variety of ways that texters construct meaning through texting features.

5 Directions for Future Research

There are several areas that this research could be developed in the future. First, this study only looked at the language being used by Spanish/English bilinguals in New York City. Even though Spanish is the most commonly spoken language after English in the United States, many other languages (i.e., Chinese, Haitian, Arabic, and Bangla) have a significant presence both in New York City and the United States, and it would be beneficial to understand how other language combinations interact in the Txt register as well. Furthermore, the development of a Part-of-Speech tagged text-messaging corpus for another language combination would both provide an opportunity for comparison and allow for better understanding of bilingual text messaging more broadly.

A second direction for future research is to investigate the role that emojis play in texted communication. While there has been research on emoticons (Dresner & Herring, 2010), there is still little if any research on the role of emojis in text messaging among bilingual or monolingual texters. The relationship of emojis to words is a productive and increasingly significant area of research as emoji use continues to increase (Benenson, 2011; Bennett, 2014; Blagdon, 2013).
Understanding the role that these images have in creating multimodal meaning in text messages is important for better understanding how texters are communicating.

A third area for future research is on the features texters use to establish intimacy and trust as well as those to cover deception in a texted platform. Within this corpus, there is the conversation between The Lovers as well as one side of over 12 conversations between one young man and 12 young women. An in-depth comparison of these two message types may reveal differences texting behavior designed to establish intimacy versus that designed to promote casual dating. Future research on this corpus could prove productive for understanding the emotional nature of text messaging and dating in an increasingly digitally mediated culture.

A fourth area for future research is to look more closely at pragmatic particles and conduct an experiment by asking participants’ judgments on various uses of markers such as “ok” and “lol” or features such as iterations, abbreviations, and other initialisms. An experimental project along these lines would be able to build on this by confirming or rejecting the hypotheses presented here for the function of these orthographical elements.

Overall, this research sought to better understand the literacy practices of bilingual youth in modern society, where adolescents and young adults communicate in colloquial, academic and digitally mediated settings every day. The result is a composite understanding of both the relationship between each of these literacy types as well as the features used by texters to create meaning on digital platforms.
Appendices

Appendix 1– Survey

Answer the questions below.

I. General

1. Gender:
   - Male
   - Female
   - Transgender
   - Other ___________________
   - Prefer Not to Answer

2. Age
   - Under 14
   - 14-16
   - 17-19
   - 20-22
   - 23+
   - Prefer Not to Answer

3. Grade:
   - 9-10
   - 11-12
   - P2G
   - Other ___________________
   - Prefer Not to Answer

4. Where were you born? ____________________________

5. How old were you when you came to the U.S.? ________________
II. School

6. Did you go to school in the U.S. before Mary Mitchell? ______________
   • What grade did you start school in the U.S.? ______________

7. Did you go to school before you came to the U.S.? __________
   • What country? __________________________________________
   • What grades? _________
   • Did you study English? _________
III. Languages

1. What is your native language? __________________________

2. Please list all the languages you speak and rate how well you speak it.
   1 = Novice. Limited knowledge.
   2 = Beginner. Some ability to speak.
   3 = Intermediate. Good ability to speak (can give an opinion).
   4 = Advanced. Fluently speak.
   5 = Native. I have been speaking this language fluently since birth.

   Example:
   Spanish_____, level 1 2 3 4 5 , I was _1__ years old when I started speaking.
   Spanish_____, level 1 2 3 4 5 . I was _____ years old when I started speaking.
   English_____, level 1 2 3 4 5 . I was _____ years old when I started speaking.
   __________________, level 1 2 3 4 5 . I was _____ years old when I started speaking.

3. Please list all the languages you read and rate how well you read it.
   1 = Novice. Limited knowledge
   2 = Beginner. Some ability to read
   3 = Intermediate. Good ability to read (can understand directions).
   4 = Advanced. Easily read (magazine, newspaper).
   5 = Expert. Can read almost anything (books).

   Example:
   Spanish_____, level 1 2 3 4 5 , I was _5__ years old when I started reading.
   Spanish_____, level 1 2 3 4 5 . I was _____ years old when I started reading.
   English_____, level 1 2 3 4 5 . I was _____ years old when I started reading.
   __________________, level 1 2 3 4 5 . I was _____ years old when I started reading.

4. For each of language, describe where and how you learned it.
   Example: __Spanish__, learned it at home
              __English__, learned it in school from 2nd to 6th grade

   __________________
   __________________
   __________________

5. What language do you use the most? __________________________
6. For each person, what language or languages do you **mostly** use?

<table>
<thead>
<tr>
<th>Language</th>
<th>Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>(example)</em></td>
<td></td>
</tr>
<tr>
<td>Friends</td>
<td><em>Spanish and English</em></td>
</tr>
<tr>
<td>Friends</td>
<td></td>
</tr>
<tr>
<td>Significant Other (bf/gf)</td>
<td></td>
</tr>
<tr>
<td>Parents</td>
<td></td>
</tr>
<tr>
<td>Brothers/ Sisters</td>
<td></td>
</tr>
<tr>
<td>Grandparents</td>
<td></td>
</tr>
<tr>
<td>Teachers</td>
<td></td>
</tr>
<tr>
<td>Boss</td>
<td></td>
</tr>
<tr>
<td>Co-Workers</td>
<td></td>
</tr>
<tr>
<td>Other:</td>
<td></td>
</tr>
<tr>
<td>Other:</td>
<td></td>
</tr>
</tbody>
</table>
III. Phone Use

1. Do you have a cell phone?
   - ☐ Yes
   - ☐ No

2. Type of cell phone:
   - ☐ Blackberry
   - ☐ Smartphone
     - ☐ iPhone
     - ☐ Android
     - ☐ Other / I don’t know
   - ☐ Other (No Internet)

3. Do you share your phone with anyone?
   - ☐ Yes
   - ☐ No

4. What kind of plan do you have?
   - ☐ None – only wifi
   - ☐ data + talk/text
   - ☐ data
   - ☐ talk/text
   - ☐ Other ________________

5. Where do you access wifi? Check all that apply.
   - ☐ School
   - ☐ Home
   - ☐ Public Network (library, park, bus station, restaurants, stores)
   - ☐ Work
   - ☐ Other ________________

6. How often do you go on the internet with your phone?
   - ☐ Multiple times a day
   - ☐ Every day
   - ☐ Most days, but not always
   - ☐ Sometimes
   - ☐ Never

7. How often do you send messages with your phone (text/chat/snapchat/etc.)?
   - ☐ Multiple times a day
   - ☐ Every day
   - ☐ Most days, but not always
   - ☐ Sometimes
   - ☐ Never
8. How often do you make phone calls?
   - Multiple times a day
   - Every day
   - Most days, but not always
   - Sometimes
   - Never
IV. Texting and chatting

1. For each person, what language do you mostly use to send messages (texts, chats, tweets, etc.), and how often do you communicate with them? If you often mix languages please list both languages.

<table>
<thead>
<tr>
<th>Language</th>
<th>Multiple times a day</th>
<th>1-2 times a day</th>
<th>Almost everyday</th>
<th>Sometimes</th>
<th>Never</th>
<th>Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friends</td>
<td>Spanish &amp; English</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friends</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Significant Other (bf/gf)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parents</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brothers/</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sisters</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grandparents</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teachers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boss</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Co-Workers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2. How do you **like** to communicate with people?  
*Please choose your favorite 3.*

In this example, my first choice for my parents is face-to-face, my second choice is texting, and my third choice is a phone call:

<table>
<thead>
<tr>
<th></th>
<th>In person</th>
<th>Facetime or Skype</th>
<th>Text or Chat</th>
<th>Email</th>
<th>Phone calls</th>
<th>Picture messages</th>
<th>Other</th>
<th>Not Applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parents</strong> (example)</td>
<td>1</td>
<td></td>
<td>2</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Friends</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Significant Other (bf/gf)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parents</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brothers/Sisters</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grandparents</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teachers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boss</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Co-workers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3. Which apps do you use?

<table>
<thead>
<tr>
<th></th>
<th>Everyday</th>
<th>3-5 times/week</th>
<th>1-5 times/month</th>
<th>Never</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dubsmash</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facebook</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fb messenger</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Facetime</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Instagram</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kik</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pheed</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Skype</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Snapchat</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tango</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tumblr</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Twitter</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WhatsApp</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YouTube</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MySpace</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
4. Outside of school, do you read books, magazines, blogs, etc.?
   □ Yes
   □ No
   • In what language? _____________________________

5. Do you use search engines (Google, Yahoo, Bing, etc.)?
   □ Yes
   □ No
   • In what language? _____________________________

6. Do you use translation services (Google Translate)?
   □ Yes
   □ No
   • In what languages? _____________________________

7. My phone’s setting are in:
   □ Spanish
   □ English
   □ Other _______________________

8. Do you use abbreviations when you text?
   □ Yes – in Spanish
   □ Yes – in English
   □ Yes – in Spanish and English
   □ No
   □ Other _____________________________

Thank you so much for your help!!
Appendix 2– Acceptability Judgment Task

**Part 1**

**Directions:** Read each sentence. If it sounds like something you or your friends would say in a text message, Facebook Messenger, WhatsApp, etc., put an X in the Good box. If it sounds weird or awkward, or like something people just don’t say, put an X in the Bad box. If you don’t know what the sentence means, put an X in I don’t know.

<table>
<thead>
<tr>
<th>Good</th>
<th>Bad</th>
<th>I don’t know</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Example

<table>
<thead>
<tr>
<th>ex</th>
<th>Good</th>
<th>Bad</th>
<th>I don’t know</th>
</tr>
</thead>
<tbody>
<tr>
<td>what do u wanna do tonight?</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Who do you wanna win the election?</td>
<td></td>
<td>X</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>1</th>
<th>Good</th>
<th>Bad</th>
<th>I don’t know</th>
</tr>
</thead>
<tbody>
<tr>
<td>#muchacosabuena De eso se trata esto, siempre sonreír</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2</th>
<th>Good</th>
<th>Bad</th>
<th>I don’t know</th>
</tr>
</thead>
<tbody>
<tr>
<td>OFICIALMENTE hoy es viernes de #muchacosabuena</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3</th>
<th>Good</th>
<th>Bad</th>
<th>I don’t know</th>
</tr>
</thead>
<tbody>
<tr>
<td>#icecream4dinner Ben &amp; Jerry's giving out free samples</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4</th>
<th>Good</th>
<th>Bad</th>
<th>I don’t know</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ben &amp; Jerry's is giving out free samples #icecream4dinner</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5</th>
<th>Good</th>
<th>Bad</th>
<th>I don’t know</th>
</tr>
</thead>
<tbody>
<tr>
<td>mi corazon es tuyo 100pre tqm</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>6</th>
<th>Good</th>
<th>Bad</th>
<th>I don’t know</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thanks for the movie Gloria! TQM</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>7</th>
<th>Good</th>
<th>Bad</th>
<th>I don’t know</th>
</tr>
</thead>
<tbody>
<tr>
<td>xp no te gusta facebook</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>8</th>
<th>Good</th>
<th>Bad</th>
<th>I don’t know</th>
</tr>
</thead>
<tbody>
<tr>
<td>xp you don’t wanna go to the party</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>9</th>
<th>Good</th>
<th>Bad</th>
<th>I don’t know</th>
</tr>
</thead>
<tbody>
<tr>
<td>My life is complete when Wanda’s my bae.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>10</th>
<th>Good</th>
<th>Bad</th>
<th>I don’t know</th>
</tr>
</thead>
<tbody>
<tr>
<td>bae’s working late, gonna make TV dinner</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>11</th>
<th>Good</th>
<th>Bad</th>
<th>I don’t know</th>
</tr>
</thead>
<tbody>
<tr>
<td>Julio’s bae is coming to the party</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
12 W: I need an iPhone 6!
M: Why?
W: Because iPhone 6!

13 W: I need an iPhone 6!
M: Why?
W: Because awesome!

14 dil tu respuesta!

15 dim tu repuesta!

16 everybody elses in school. im in Costa Rica #fomo

17 My friends are at Beyonce and I'm at work #fomo

18 S: u finish the hw?
W: No, half done
S: its so hard, lol

19 W: You see Mr. Ryley’s new haircut?
S: omg lol too much hairspray
W: What? I like it!

20 Bus came right on time! Smh

21 senior year schedule: 3 lunches smh

22 tbh I didnt like the movie

23 tbh the school is on 2nd ave

24 tbt that time we found $50 on the beach!

25 last year there wasn't school on Columbus Day #tbt

26 ¡Agradecer de todo corazón a Nany por ser una gran defensora de Fedel! ¡eres la mejor! te queremos bb

27 bb va a cantar "Right now I wish you were here with me"
<table>
<thead>
<tr>
<th></th>
<th>Went to the gym, ate a salad, and cleaned the kitchen #yolo</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Just spent $500 on Beyonce tickets #yolo</td>
</tr>
<tr>
<td></td>
<td>Claro, ntp! Espero k no sea malo</td>
</tr>
<tr>
<td></td>
<td>Ntp, you'll do great!</td>
</tr>
<tr>
<td></td>
<td>Xp no te gusta facebook</td>
</tr>
<tr>
<td></td>
<td>Xp you don’t wanna go to the party</td>
</tr>
<tr>
<td></td>
<td>Wtf! Congratulations on your test 😊</td>
</tr>
<tr>
<td></td>
<td>You took my lunch wtf</td>
</tr>
<tr>
<td></td>
<td>Look out, mrs. F tryna be sly again</td>
</tr>
<tr>
<td></td>
<td>I’m tryna out for soccer tonight</td>
</tr>
<tr>
<td></td>
<td>Quiero ser cool tbn</td>
</tr>
<tr>
<td></td>
<td>Good</td>
</tr>
<tr>
<td>---</td>
<td>------</td>
</tr>
<tr>
<td>1</td>
<td>haha he chirped you</td>
</tr>
<tr>
<td>2</td>
<td>haha you got chirped</td>
</tr>
<tr>
<td>3</td>
<td>iOS 8 is cray! I feel like I got a new phone</td>
</tr>
<tr>
<td>4</td>
<td>Wanda is cray! It's snowing and she's wearing sandals</td>
</tr>
<tr>
<td>5</td>
<td>I spent a lot of guap on those shoes!</td>
</tr>
<tr>
<td>6</td>
<td>Wakin up at 5am to make this guap!</td>
</tr>
<tr>
<td>7</td>
<td>i'm fleek for the first day of school</td>
</tr>
<tr>
<td>8</td>
<td>first day of school eyebrows on fleek</td>
</tr>
<tr>
<td>9</td>
<td>they make a nice ship</td>
</tr>
<tr>
<td>10</td>
<td>do you ship us?</td>
</tr>
<tr>
<td>11</td>
<td>why do guys gotta catcall my sister? swerve!</td>
</tr>
<tr>
<td>12</td>
<td>i got 100% on the test!! swerve!!</td>
</tr>
<tr>
<td>13</td>
<td>he shaded me in class</td>
</tr>
<tr>
<td>14</td>
<td>how you gonna throw shade at a toddler?</td>
</tr>
</tbody>
</table>
Part 2

In each item, person A sends a text message. Match the response with how person B is feeling. One feeling can have more than 1 response.

<table>
<thead>
<tr>
<th>Text message</th>
<th>Response</th>
<th>Feeling</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Broccoli for dinner.”</td>
<td>yum!</td>
<td>***happy</td>
</tr>
<tr>
<td></td>
<td>😊</td>
<td>***don’t care</td>
</tr>
<tr>
<td></td>
<td>fine.</td>
<td>other: <strong>unhappy</strong></td>
</tr>
<tr>
<td></td>
<td>yuck</td>
<td></td>
</tr>
<tr>
<td></td>
<td>y</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Text Message</th>
<th>Response</th>
<th>Feeling</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ok</td>
<td>***fine</td>
</tr>
<tr>
<td></td>
<td>ok.</td>
<td>***fine</td>
</tr>
<tr>
<td></td>
<td>OK</td>
<td>***annoyed</td>
</tr>
<tr>
<td></td>
<td>O.K.</td>
<td>***angry</td>
</tr>
<tr>
<td></td>
<td>ok ;)</td>
<td>***happy</td>
</tr>
<tr>
<td></td>
<td>OK ;)</td>
<td>other: ____________</td>
</tr>
<tr>
<td>“going to Maria’s”</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Text Message</th>
<th>Response</th>
<th>Feeling</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>ur 30 minutes late</td>
<td>***fine</td>
</tr>
<tr>
<td></td>
<td>ur 30 minutes late ;)</td>
<td>***annoyed</td>
</tr>
<tr>
<td></td>
<td>you are 30 minutes late</td>
<td>***angry</td>
</tr>
<tr>
<td></td>
<td>UR 30 MINUTES LATE</td>
<td>***happy</td>
</tr>
<tr>
<td></td>
<td>YOU ARE 30 MINUTES LATE</td>
<td>other: ____________</td>
</tr>
<tr>
<td>“sorry, running late!”</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| 3 | "How was the date?" | :) 
;)...
;) !!!
ok
ok...
ok!!!
… | ***great
***good
***fine
***terrible
***it's a long story
other: ______________ |
| 4 | "r u going to the party tonight?" | no
no...
no.
maybe
maybe.
maybe… | ***no
***no and they're annoyed
***maybe
***maybe and they're annoyed
***it's a long story
other: ______________ |
| 5 | "Be there in 10" | Kk
kk.
k
ok
ok. | ***fine
***annoyed
***angry
***happy
other: ______________ |
Appendix 3– Chatting Game Vocabulary Lists

Group A List

1. Nutrir
2. Intricate
3. Correr
4. Read
5. Predecir
6. Promotion
7. Discernir
8. Feasible
9. Amor
10. Hope
11. Coherent
12. Juego
13. Media
14. Smile
15. Música

Group B List

1. Enhance
2. Alejarse
3. Perish
4. Descifrar
5. Modify
6. Libro
7. Table
8. Escalar
9. Hybrid
10. Paseo
11. Classroom
12. Elusive
13. Oscuro
14. Lucid
15. Copious
## Appendix 4 – Sixth Grade LTELL Writing Rubric

<table>
<thead>
<tr>
<th>Level</th>
<th>Topic, Focus, &amp; Making Claims</th>
<th>Content Development</th>
<th>Organization</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 Starting</td>
<td>*Does not respond to the prompt at all&lt;br&gt;*The topic is indecipherable or nonexistent&lt;br&gt;*Does not indicate understanding of the topic</td>
<td>*Does not use any examples, descriptions, definitions, etc.</td>
<td>*No paragraphs&lt;br&gt;*No evidence of prewriting&lt;br&gt;*No conclusion</td>
</tr>
<tr>
<td>1 Emerging</td>
<td>*Responds to some parts of the prompt&lt;br&gt;*Topic is present, but unclear&lt;br&gt;*Indicates little understanding of the topic</td>
<td>*Examples, descriptions, definitions or other evidence are irrelevant or disconnected from the topic</td>
<td>*Ideas are scattered&lt;br&gt;*No paragraph structure, but evidence of prewriting, either graphical or written&lt;br&gt;*Conclusion does not relate to the topic</td>
</tr>
<tr>
<td>2 Advancing</td>
<td>*Responds to most of the prompt&lt;br&gt;*Presents weak or poorly defined topic&lt;br&gt;*Indicates some understanding of the topic</td>
<td>*Examples, descriptions, definitions or other evidence are vaguely related to the topic</td>
<td>*Similar ideas are sometimes grouped together, though some may be out of place&lt;br&gt;*Paragraph organization is present, though it either does not follow a logical sequence OR is missing key components&lt;br&gt;*Conclusion is weak or may be only vaguely related to</td>
</tr>
<tr>
<td>3 Accomplishing</td>
<td>*Responds to all parts of the prompt&lt;br&gt;*Presents a clear topic&lt;br&gt;*Indicates good understanding of the topic</td>
<td>*Examples, descriptions, definitions or other evidence is related to and supports the topic</td>
<td>*Similar ideas are grouped together&lt;br&gt;*Paragraphs have internal structure and are presented in a logical order&lt;br&gt;*Conclusion relates to and supports the topic</td>
</tr>
<tr>
<td>4 Excelling</td>
<td>*Responds to all parts of the prompt in an interconnected way&lt;br&gt;*Presents a clear, well defined topic&lt;br&gt;*Indicates insightful understanding of the topic</td>
<td>*Examples, descriptions, definitions or other evidence is related to and supports the topic</td>
<td>*Similar and related ideas are grouped together to support the main topic&lt;br&gt;*Paragraphs have internal structure and are logically ordered, building on each other and connecting to main idea&lt;br&gt;*Effective</td>
</tr>
<tr>
<td>Language Use: Sentence variety</td>
<td>*Uses very simple vocabulary</td>
<td>*Simple or inaccurate vocabulary</td>
<td>*Occasionally varies vocabulary/sentence structure though it does not improve understanding</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Word choice Transitions</td>
<td>*Sentences lack structure</td>
<td>*No variation in sentence structure</td>
<td></td>
</tr>
<tr>
<td>Conventions of Standard Academic English: sentence structure (run ons/fragments, verb tense, spelling, punctuation, capitalization, formatting (indentation; genre-specific parts, ie salutation, titles, in-text citations)</td>
<td>*Does not resemble Standard English; the text is incomprehensible</td>
<td>*Makes numerous and repeated errors in Standard English; the text is difficult to read</td>
<td>*Makes frequent and distracting errors in Standard English that occasionally impede comprehension</td>
</tr>
</tbody>
</table>

topic conclusion that reinforces the points made in the text

Appendix 5 – Corpus ReadMe
BYTs_Corpus_ReadMe

Messages

The messages in this corpus were downloaded directly from participants’ phones by creating a history of all of their text messages and downloading the archive. Participants signed two consent forms for this and had to enter their passcode to allow access. The messages have been anonymized first computationally and then they were hand masked by two independent researchers. There are 44,500 messages in the corpus, including spam, automated messages, and mass messages from participants’ cellphone carriers.

There is one conversation in this data set that spans the first four months of a relationship. Both contributors were participants in the study, allowing for recovery of the complete conversation. When possible, other conversations have been kept intact and the conversation has been assigned a code. The intended purpose is for discourse analysis.

• Msg_ID: Each message has a unique identifier that begins with the letter, M

• Usr_ID: Each user has a unique identifier that begins with the letter, U. This indicates who donated the messages. See the Send information to determine who wrote the message.

• Conv_ID: Each conversation has a unique identifier in order to keep continuous exchanges by two people intact. The purpose for this is conversation analysis. C000 is not a conversation, this code was used for mass messages and alerts. C999 is unidentifiable as a conversation though it was sent by an individual.

• Rel_ID: When possible, the relationship between the people is indicated.
  o friend: friend
  o casual: engaged in a casual sexual relationship
  o significant other: engaged in an ongoing romantic relationship
• rel: non-immediate relative
• sib: sibling
• parent: parent

• Gender: When possible, the gender of the receiver is indicated

• Message: The message contents.

• Date: When available from the metadata, the date the message was sent is included. When the date was not recoverable, 1/1/00 is used.

• Time: When available from the metadata, the time the message was sent is included. When the time was not recoverable, 12:00:00AM is used.

• Sent: Whether the donator sent or received the message.

Masked Items

'd' indicates how many characters, including spaces and punctuation were in the masked item. For example, 234-567-8900 is masked as [NUMBER, 12] whereas 2345678900 is masked as [NUMBER, 10]

• [NUMBER, 'd'] numbers that could be linked to one individual. Phone numbers, heights, weights, shoe sizes, were removed, but time of day was preserved.
• [EMAIL, 'd'] email address
• [PLACE, 'd']: place, including addresses or partial addresses, street names, directions, train stations, sports and music venues, names of libraries, schools, community centers, and event spaces
• [NAME, 'd']: proper names, including first and last names, diminuitive forms of names, pet names and relationship-specific terms of endearment
• [WEB, 'd']: website
Emojis

Emojis are encoded as [EMO, 'd'], there is another table in this package that replaces these characters with either unicode or bytes or the actual image. Because of the encoding and decoding processes, there are mistakes in this correlation. Therefore the emoji data should not be treated as preliminary data. Please contact Michelle McSweeney if you plan to work with emojis.
CHAPTER 1

Loomings

Call me Ishmael.

Some years ago—never mind how long precisely—having little or no money in my purse, and nothing particular to interest me on shore, I thought I would sail about a little and see the watery part of the world.

It is a way I have of driving off the spleen and regulating the circulation.

Whenever I find myself growing grim about the mouth; whenever it is a damp, drizzly November in my soul; whenever I find myself involuntarily pausing before a6 coffin warehouses, and bringing up the rear of every funeral I meet; and especially whenever my hypochondria get such an upper hand of me, that it requires a strong moral principle to prevent me from deliberately stepping into the street, and methodically knocking people's hats off—then, I account it high time to get to sea as soon as I can.

This is my substitute for pistol and ball.

With a philosophical flourish Cato throws himself upon his sword; I quietly take to the ship.

There is nothing surprising in this.

If they but knew it, almost all men in their degree, some time or other, cherish very nearly the same feelings towards the ocean with me.

There now is your insular city of the Manhattenes, belted round by wharves as Indian isles by coral reefs—commerce surrounds it with her surf.

Right and left, the streets take you waterward.

Its extreme downtown is the battery, where that noble mole is washed by waves, and cooled by breezes, which a few hours previous were out of sight of land.
### Appendix 7 - All Initialisms and Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation/Acronym</th>
<th>Translation</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>100pre</td>
<td>for ever</td>
<td>Spanish</td>
</tr>
<tr>
<td>2much4u</td>
<td>for ever</td>
<td>English</td>
</tr>
<tr>
<td>4ever</td>
<td>forever</td>
<td>English</td>
</tr>
<tr>
<td>4life</td>
<td>forever</td>
<td>English</td>
</tr>
<tr>
<td>5comentatios</td>
<td>sin comentarios</td>
<td>Spanish</td>
</tr>
<tr>
<td>aki</td>
<td>here</td>
<td>Spanish</td>
</tr>
<tr>
<td>b2w</td>
<td>back to work</td>
<td>English</td>
</tr>
<tr>
<td>bae</td>
<td>before anyone else</td>
<td>English</td>
</tr>
<tr>
<td>bb</td>
<td>baby</td>
<td>Bilingual</td>
</tr>
<tr>
<td>bbff</td>
<td>behind beautiful, ugly from front</td>
<td>English</td>
</tr>
<tr>
<td>bby</td>
<td>baby</td>
<td>English</td>
</tr>
<tr>
<td>bf</td>
<td>best friend/boyfriend</td>
<td>English</td>
</tr>
<tr>
<td>bff</td>
<td>best friend forever</td>
<td>English</td>
</tr>
<tr>
<td>bibl</td>
<td>might be back later</td>
<td>English</td>
</tr>
<tr>
<td>bn</td>
<td>bien</td>
<td>Spanish</td>
</tr>
<tr>
<td>bno</td>
<td>bueno</td>
<td>Spanish</td>
</tr>
<tr>
<td>bnx</td>
<td>buenos noches</td>
<td>Spanish</td>
</tr>
<tr>
<td>Brb</td>
<td>be right back</td>
<td>English</td>
</tr>
<tr>
<td>bro</td>
<td>brother</td>
<td>English</td>
</tr>
<tr>
<td>bs</td>
<td>besos</td>
<td>Spanish</td>
</tr>
<tr>
<td>bx</td>
<td>besos</td>
<td>Spanish</td>
</tr>
<tr>
<td>cc</td>
<td>cute couple</td>
<td>English</td>
</tr>
<tr>
<td>como tas</td>
<td>how are you</td>
<td>Spanish</td>
</tr>
<tr>
<td>cray</td>
<td>crazy</td>
<td>English</td>
</tr>
<tr>
<td>cuz</td>
<td>because</td>
<td>English</td>
</tr>
<tr>
<td>d</td>
<td>de</td>
<td>Spanish</td>
</tr>
<tr>
<td>da=</td>
<td>da igual</td>
<td>Spanish</td>
</tr>
<tr>
<td>dime ab</td>
<td>dime a ver</td>
<td>Spanish</td>
</tr>
<tr>
<td>dime ave</td>
<td>what's up/tell me</td>
<td>Spanish</td>
</tr>
<tr>
<td>dimela</td>
<td>what's up</td>
<td>Spanish</td>
</tr>
<tr>
<td>dio mio</td>
<td>OMG</td>
<td>Spanish</td>
</tr>
<tr>
<td>dm</td>
<td>dime</td>
<td>Spanish</td>
</tr>
<tr>
<td>dm</td>
<td>dios mio</td>
<td>Spanish</td>
</tr>
<tr>
<td>dm ave</td>
<td>dime a ve</td>
<td>Spanish</td>
</tr>
<tr>
<td>fb</td>
<td>facebook</td>
<td>Bilingual</td>
</tr>
<tr>
<td>ftn</td>
<td>fuck that noise (shut up)</td>
<td>English</td>
</tr>
<tr>
<td>ga</td>
<td>Garcia</td>
<td>Spanish</td>
</tr>
<tr>
<td>gf</td>
<td>girlfriend</td>
<td>English</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
<td>Language</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
<td>----------</td>
</tr>
<tr>
<td>gm</td>
<td>good morning</td>
<td>English</td>
</tr>
<tr>
<td>gn</td>
<td>good night</td>
<td>English</td>
</tr>
<tr>
<td>gnight</td>
<td>good night</td>
<td>English</td>
</tr>
<tr>
<td>gno</td>
<td>girls night out</td>
<td>English</td>
</tr>
<tr>
<td>grax</td>
<td>thank you</td>
<td>Spanish</td>
</tr>
<tr>
<td>gtg</td>
<td>got to go</td>
<td>English</td>
</tr>
<tr>
<td>gws</td>
<td>get well soon</td>
<td>English</td>
</tr>
<tr>
<td>gx</td>
<td>gracias</td>
<td>Spanish</td>
</tr>
<tr>
<td>h-bday</td>
<td>happy birthday</td>
<td>English</td>
</tr>
<tr>
<td>hbd</td>
<td>happy birthday</td>
<td>English</td>
</tr>
<tr>
<td>hbu</td>
<td>how about you/how be you</td>
<td>English</td>
</tr>
<tr>
<td>hdp</td>
<td>hijo de puta</td>
<td>Spanish</td>
</tr>
<tr>
<td>hru</td>
<td>how are you</td>
<td>English</td>
</tr>
<tr>
<td>hw</td>
<td>how</td>
<td>English</td>
</tr>
<tr>
<td>hwu</td>
<td>hey, what's up</td>
<td>English</td>
</tr>
<tr>
<td>hyd</td>
<td>how are you doing?</td>
<td>English</td>
</tr>
<tr>
<td>I &lt;3 u</td>
<td>I love you</td>
<td>English</td>
</tr>
<tr>
<td>icq</td>
<td>I seek you (chat program)</td>
<td>English</td>
</tr>
<tr>
<td>idc</td>
<td>I don't care</td>
<td>English</td>
</tr>
<tr>
<td>ig</td>
<td>instagram</td>
<td>Bilingual</td>
</tr>
<tr>
<td>ii</td>
<td>okay</td>
<td>English</td>
</tr>
<tr>
<td>ikd</td>
<td>I know, dude</td>
<td>English</td>
</tr>
<tr>
<td>ilu</td>
<td>I love you</td>
<td>English</td>
</tr>
<tr>
<td>ily</td>
<td>I love you</td>
<td>English</td>
</tr>
<tr>
<td>ily (like)</td>
<td>I like you</td>
<td>English</td>
</tr>
<tr>
<td>imy</td>
<td>I miss you</td>
<td>English</td>
</tr>
<tr>
<td>K</td>
<td>OK</td>
<td>English</td>
</tr>
<tr>
<td>kiero</td>
<td>quiero</td>
<td>Spanish</td>
</tr>
<tr>
<td>klk</td>
<td>que lo que</td>
<td>Bilingual</td>
</tr>
<tr>
<td>klk chapi</td>
<td>what's up, bitch</td>
<td>Spanish</td>
</tr>
<tr>
<td>klk manin</td>
<td>what's up, homeboy</td>
<td>Spanish</td>
</tr>
<tr>
<td>klk wa</td>
<td>what's up, what's going on</td>
<td>Bilingual</td>
</tr>
<tr>
<td>klkl</td>
<td>cool cool</td>
<td>English</td>
</tr>
<tr>
<td>kml</td>
<td>kill myself laughing</td>
<td>English</td>
</tr>
<tr>
<td>ldk</td>
<td>I don't know</td>
<td>English</td>
</tr>
<tr>
<td>lmao</td>
<td>laughing my ass off</td>
<td>English</td>
</tr>
<tr>
<td>lmfao</td>
<td>laughing my fucking ass off</td>
<td>English</td>
</tr>
<tr>
<td>lmk</td>
<td>let me know</td>
<td>English</td>
</tr>
<tr>
<td>lml</td>
<td>laughing mad loud</td>
<td>English</td>
</tr>
<tr>
<td>lmt</td>
<td>let me think</td>
<td>English</td>
</tr>
<tr>
<td>lol</td>
<td>laugh out loud</td>
<td>English</td>
</tr>
<tr>
<td>m</td>
<td>me</td>
<td>Spanish</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Meaning</td>
<td>Language</td>
</tr>
<tr>
<td>--------------</td>
<td>---------</td>
<td>----------</td>
</tr>
<tr>
<td>mae</td>
<td>maje (friend)</td>
<td>Spanish</td>
</tr>
<tr>
<td>mkt</td>
<td>market</td>
<td>English</td>
</tr>
<tr>
<td>mmg</td>
<td>mama gueno</td>
<td>Spanish</td>
</tr>
<tr>
<td>mán</td>
<td>mañana</td>
<td>Spanish</td>
</tr>
<tr>
<td>net</td>
<td>Internet</td>
<td>English</td>
</tr>
<tr>
<td>ntp</td>
<td>no te preocupes</td>
<td>Spanish</td>
</tr>
<tr>
<td>ok</td>
<td>okay</td>
<td>Bilingual</td>
</tr>
<tr>
<td>omg</td>
<td>oh my god</td>
<td>English</td>
</tr>
<tr>
<td>omw</td>
<td>oh my word</td>
<td>English</td>
</tr>
<tr>
<td>porq</td>
<td>porque</td>
<td>Spanish</td>
</tr>
<tr>
<td>pp</td>
<td>profile picture</td>
<td>English</td>
</tr>
<tr>
<td>prp</td>
<td>Puerto Rican Picnic</td>
<td>English</td>
</tr>
<tr>
<td>q</td>
<td>Question/Queer</td>
<td>English</td>
</tr>
<tr>
<td>qp2</td>
<td>que puedes</td>
<td>Spanish</td>
</tr>
<tr>
<td>qpz</td>
<td>que puedes</td>
<td>Spanish</td>
</tr>
<tr>
<td>s^</td>
<td>what's up</td>
<td>English</td>
</tr>
<tr>
<td>sf</td>
<td>sexually frustrated</td>
<td>English</td>
</tr>
<tr>
<td>sis</td>
<td>sister</td>
<td>English</td>
</tr>
<tr>
<td>smh</td>
<td>shaking my head</td>
<td>English</td>
</tr>
<tr>
<td>snu</td>
<td>what's new</td>
<td>English</td>
</tr>
<tr>
<td>sos</td>
<td>save our souls</td>
<td>English</td>
</tr>
<tr>
<td>sr</td>
<td>Senor</td>
<td>Spanish</td>
</tr>
<tr>
<td>syl</td>
<td>see you later</td>
<td>English</td>
</tr>
<tr>
<td>syt</td>
<td>see you tomorrow</td>
<td>English</td>
</tr>
<tr>
<td>t</td>
<td>te</td>
<td>Spanish</td>
</tr>
<tr>
<td>t'amo</td>
<td>I love you</td>
<td>Spanish</td>
</tr>
<tr>
<td>tato</td>
<td>we're cool</td>
<td>Spanish</td>
</tr>
<tr>
<td>tb</td>
<td>tambien</td>
<td>Spanish</td>
</tr>
<tr>
<td>tbh</td>
<td>to be honest</td>
<td>English</td>
</tr>
<tr>
<td>tbm</td>
<td>tambien</td>
<td>Spanish</td>
</tr>
<tr>
<td>tbt</td>
<td>throw back Thursday</td>
<td>English</td>
</tr>
<tr>
<td>thx</td>
<td>thanks</td>
<td>English</td>
</tr>
<tr>
<td>tk</td>
<td>te quiero</td>
<td>Spanish</td>
</tr>
<tr>
<td>tkm</td>
<td>te quiero mucho</td>
<td>Spanish</td>
</tr>
<tr>
<td>tlvz</td>
<td>tal vez</td>
<td>Spanish</td>
</tr>
<tr>
<td>tmr</td>
<td>tomorrow</td>
<td>English</td>
</tr>
<tr>
<td>tmw</td>
<td>tomorrow</td>
<td>English</td>
</tr>
<tr>
<td>tq</td>
<td>te quiero</td>
<td>Spanish</td>
</tr>
<tr>
<td>tqm</td>
<td>te quiero mucho</td>
<td>Spanish</td>
</tr>
<tr>
<td>tranki</td>
<td>tranquilo</td>
<td>Spanish</td>
</tr>
<tr>
<td>ttm</td>
<td>talk to me</td>
<td>English</td>
</tr>
<tr>
<td>ttyl</td>
<td>talk to you later</td>
<td>English</td>
</tr>
<tr>
<td>tuh</td>
<td>tsk + duh</td>
<td>Bilingual</td>
</tr>
<tr>
<td>---------</td>
<td>------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>twu</td>
<td>that's what up</td>
<td>English</td>
</tr>
<tr>
<td>txt</td>
<td>text</td>
<td>English</td>
</tr>
<tr>
<td>ty</td>
<td>thank you</td>
<td>English</td>
</tr>
<tr>
<td>tyl</td>
<td>talk to you later</td>
<td>English</td>
</tr>
<tr>
<td>u</td>
<td>you</td>
<td>English</td>
</tr>
<tr>
<td>ud</td>
<td>usted</td>
<td>Spanish</td>
</tr>
<tr>
<td>usd</td>
<td>usted</td>
<td>Spanish</td>
</tr>
<tr>
<td>vdd</td>
<td>verdad</td>
<td>Spanish</td>
</tr>
<tr>
<td>vip</td>
<td>Very Important Person</td>
<td>English</td>
</tr>
<tr>
<td>wa</td>
<td>what's going on (comes from wa gwan)</td>
<td>English</td>
</tr>
<tr>
<td>wa gwan</td>
<td>what's going on</td>
<td>English</td>
</tr>
<tr>
<td>wasupp</td>
<td>what's up</td>
<td>English</td>
</tr>
<tr>
<td>wbu</td>
<td>what about you</td>
<td>English</td>
</tr>
<tr>
<td>wgo</td>
<td>what's going on</td>
<td>English</td>
</tr>
<tr>
<td>ws</td>
<td>work-safe</td>
<td>English</td>
</tr>
<tr>
<td>wsp</td>
<td>what's up</td>
<td>English</td>
</tr>
<tr>
<td>wtf</td>
<td>what the fuck</td>
<td>English</td>
</tr>
<tr>
<td>wth</td>
<td>what the hell</td>
<td>English</td>
</tr>
<tr>
<td>wua</td>
<td>where you at</td>
<td>English</td>
</tr>
<tr>
<td>wud</td>
<td>what you doing</td>
<td>English</td>
</tr>
<tr>
<td>wug</td>
<td>what you got?</td>
<td>English</td>
</tr>
<tr>
<td>wuw</td>
<td>what you want?</td>
<td>English</td>
</tr>
<tr>
<td>wya</td>
<td>where you at?</td>
<td>English</td>
</tr>
<tr>
<td>wyd</td>
<td>What you doing?</td>
<td>English</td>
</tr>
<tr>
<td>wyg</td>
<td>Where you goin?</td>
<td>English</td>
</tr>
<tr>
<td>wym</td>
<td>what you mean?</td>
<td>English</td>
</tr>
<tr>
<td>x-marry</td>
<td>merry x-mas</td>
<td>Spanish</td>
</tr>
<tr>
<td>x?</td>
<td>por</td>
<td>Spanish</td>
</tr>
<tr>
<td>x100pre</td>
<td>por siempre</td>
<td>Spanish</td>
</tr>
<tr>
<td>xd</td>
<td>lol</td>
<td>Bilingual</td>
</tr>
<tr>
<td>xk</td>
<td>porque</td>
<td>Spanish</td>
</tr>
<tr>
<td>xoxo</td>
<td>hugs &amp; kisses</td>
<td>Bilingual</td>
</tr>
<tr>
<td>xq</td>
<td>porque</td>
<td>Spanish</td>
</tr>
<tr>
<td>y?</td>
<td>why?</td>
<td>English</td>
</tr>
<tr>
<td>yo</td>
<td>hello</td>
<td>English</td>
</tr>
<tr>
<td>yolo</td>
<td>you only live once</td>
<td>English</td>
</tr>
</tbody>
</table>
Appendix 8 – FidoNews (First record of “lol”)

FidoNews 6-19 Page 10 8 May 1989
L. Edel
151/2

MO_ICONS_PLEASE

For some reason, the ICONS in a past Fido Newsletter, were not the icons I have seen in use the past several years!!! Where did the nose come from? Originally, the set goes as follows, excluding some that can only be used on IBM, ATARI, COMMODORE, etc.

:) - smile/happy :D - big smile or laugh :c - pout
:> - mischievous smile :'( - cry
;D - wink ;> - sly wink
:* - kiss [] - hug
:o - shout/yawn O:) - angel/innocent
]:> - devil/guilty :1 - content
:# - grimace/frustrated :/ - disfavor/baffled
:P - sticking out tongue :X - not talking
:d - tastes good :9 8) - wide-eyed surprise
B) - glasses [% - mug
c% - coffee cup U - glass
u - shot glass Y - wine/cocktail glass
---<--<@ long-stemmed rose
--------- drink sliding down bar

Also worth considering are the following:

OLM - On Line Message OTW - On The Way
OIC - Oh I See H - HUH???
BTW - By The Way LOL - Laughing Out Loud
ROTF - Rolling On The Floor RAO - Rolling All Over
LMTO - Laughing My Tush Off BRB - Be Right Back
AFK - Away From Keys BBL - Be Back Later
BAK - Back At Keys WLCM - Welcome
BCNU - Be Seeing You L8R - Later
ODM - On De Move OTB - Off To Bed
LTNT - Long Time No Type TTFN - Ta Ta For Now
FE - Again (Greetings, as in "re-hi")
LTNS - Long Time No See
M/F - Male or Female (also known as 'MORFING', as in "Oh no! I've been morfed!!")

I hope this make for more "colorful communicating". Just remember the quote from Alex Bell (no relationship to the famous Bell) "Anything that can be said in a few words, isn't worth saying and should be forgotten". Don't blame me I didn't say it.
References


http://doi.org/10.1017/S0332586500002018


Anasari, A. (2015). Master of None - Episode 2 - Parents - 3D or 2D. Retrieved from https://www.youtube.com/watch?v=jQTgXq9-arY


http://doi.org/10.1080/10862960903340165


http://philpapers.org/rec/BACLCA


Retrieved from
http://works.bepress.com/cgi/viewcontent.cgi?article=1043&context=benjamin_bailey


Language Acquisition, 15*(03), 279–304. http://doi.org/10.1017/S0272263100012122

Barendregt, B. (2008). Sex, cannibals, and the language of cool: Indonesian tales of the phone and


http://doi.org/10.1177/0261927X04269585


http://doi.org/10.1093/applin/amt056


http://doi.org/10.1016/j.sbspro.2013.01.116


http://doi.org/10.1525/jlin.2003.13.2.211


http://doi.org/10.1177/13670069080120010201


http://doi.org/10.5539/ijel.v1n2p222


http://doi.org/10.1016/j.soscij.2012.05.003


Heller, R., & Greenleaf, C. L. (2007). Literacy Instruction in the Content Areas: Getting to the Core of Middle and High School Improvement. Alliance for Excellent Education.


Proceedings of the Second Workshop on Spanish Sociolinguistics (pp. 102–8). Retrieved from

Proc. of the Third Workshop on Spanish Sociolinguistics (pp. 162–170). Retrieved from

Morel, É., Bucher, C., Pekarek-Doehler, S., & Siebenhaar, B. (2012). SMS communication as
plurilingual communication: Hybrid language use as a challenge for classical code-switching

Retrieved from
https://books.google.com/books?hl=en&lr=&id=cdk5CgAAQBAJ&oi=fnd&pg=PP1&dq
downfall+of+literacy&ots=FyLmFs_u39&sig=_JjlfxTjbkF8ijGqHu3qCOqAG_8

achievement. The Elementary School Journal. Retrieved from
http://psycnet.apa.org/psycinfo/1984-27183-001

Mossberger, K., Tolbert, C. J., & McNeal, R. S. (2008). Digital citizenship the internet, society, and
participation. Cambridge, MA: MIT Press. Retrieved from

Motel, S., & Patten, E. (2012, June 27). The 10 Largest Hispanic Origin Groups: Characteristics,
Rankings, Top Counties. Retrieved from http://www.pewhispanic.org/2012/06/27/the-10-
largest-hispanic-origin-groups-characteristics-rankings-top-counties/


http://doi.org/10.1016/j.pragma.2003.10.004


https://www.youtube.com/watch?v=UfGMj10wOzg


https://books.google.com/books?hl=en&lr=&id=vPdgBgAAQBAJ&oi=fnd&pg=PP10&dq=new+words+added+to+English+&ots=YdI33InFD1&sig=Otjg0m7ppjuJVncUZ_YpzyS3ae8


