DIVERGENCE IN CEBUANO AND ENGLISH CODE-SWITCHING PRACTICES
IN CEBUANO SPEECH COMMUNITIES IN THE CENTRAL PHILIPPINES

A Thesis
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Master of Arts
in Linguistics

By

Glenn Abastillas, BSN

Washington, DC
April 1, 2015
The Philippines is a diverse linguistic environment with more than 8 major languages spoken and a complicated language policy affected by its colonization history. With this context, this research investigates Cebuano and English code-switching (CS) in the Central Philippines and Mindanao. This research draws from prior studies placing multilingual and code-switched language practices at the center of an individual’s identity rather than at the margins (Woolard, 1998; Stell, 2010; Eppler, 2010; Weston, 2013). Code-switching is defined to be the hybrid of multiple languages and, subsequently, multiple identities (Bullock & Toribio, 2009). I expand on these ideas to examine the homogeneity of Cebuano identity across four Cebuano speaking provinces in the Central Philippines and Mindanao through their CS practice in computer mediated communication (CMC) on Twitter. I demonstrate that the Cebuano speech community is divergent in their CS practices split into two general groups, which are employing CS practices at significantly different rates.

Using computational tools, I implement a mixed methods approach in collecting and analyzing the data. My data consist of short manually tagged messages called tweets from
the social media platform Twitter. Tweets were collected at various times during the day and night over a period of 3 months from the Cebuano speaking provinces of Cebu, Negros Oriental, Misamis Oriental, and Davao del Sur. Collectively, there were 2,652 users, tweeting 7,729 times, who contributed to this corpus, representing language from all four provinces in both rural and urban contexts. A chi-square ($\chi^2$) analysis on CS with respect to province found that the four provinces employ CS at significantly ($\chi^2 = 84.75$, $p < .001$) different rates. A chi-square analysis also showed that there was a strong relationship between CS and population density ($\chi^2 = 3.47$, $p < .1$). Lastly, a T-test analysis showed that longer tweets are significantly more likely to have CS than shorter tweets (one-sample $t(105) = 6.7963$, $p < .001$).

The results of the chi-square analysis demonstrate a divergence in the Cebuano speech community in the Philippines. That is, the southern provinces of Misamis Oriental and Davao del Sur (Southern Group) adopt CS significantly more than the northern provinces of Cebu and Negros Oriental (Northern Group), which were less likely to adopt CS. Because of a strong pro-Cebuano sentiment in Cebu, I reason that the Northern Group adheres more strongly to the Cebuano identity resulting in less CS. Conversely, the Southern Groups may be identifying less with Cebu and the Cebuano identity, which results in more CS. In summary, the Cebuano speech communities in the Philippines express their differentiating identities through adoption of CS.
ACKNOWLEDGEMENTS

The research and writing of this thesis is dedicated to everyone who helped along the way. I want to thank my immediate and extended family here in the United States and in the Philippines for their linguistic and cultural insights with regards to my thesis. I especially want to thank Lucas Heinkel and Dawn Moller who have provided much needed support throughout this whole research and writing process.

Many thanks,

Glenn Abastillas
### TABLE of CONTENTS

**TABLE of FIGURES & TABLES** ................................................................. ix

**TABLE of ABBREVIATIONS USED IN LINGUISTIC EXAMPLES** .................... xi

1. INTRODUCTION ....................................................................................... 1

2. BACKGROUND .......................................................................................... 4
   2.1 Demographics .................................................................................... 5
   2.2 Current Status of Cebuano ................................................................. 6
   2.3 Ethnic Origins .................................................................................. 8
   2.4 The Spanish Period .......................................................................... 9
   2.5 The American Period ...................................................................... 11
   2.6 Anti-Tagalog Attitudes ..................................................................... 12
   2.7 Language History ........................................................................... 13
   2.8 Technology in the Philippines ......................................................... 15

3. LITERATURE REVIEW .............................................................................. 16
   3.1 Code-Switching ................................................................................. 17
   3.2 Language Contact ........................................................................... 19
   3.3 Sociolinguistic Reasons for Code-Switching .................................... 20
   3.4 Code-Switching and Community ..................................................... 23
   3.5 Code-Switching and Age .................................................................. 25
   3.6 Code-Switching and Social Media .................................................. 26
   3.7 Grammar of Code-Switching .............................................................. 27

4. DATA & METHODS ................................................................................... 29
4.1 Data Collection ................................................................. 32
4.2 Unit of Analysis .............................................................. 32
4.3 Encoding ........................................................................... 33
4.4 Criteria ............................................................................. 33
4.5 Adaptation of Prior Linguists’ Coding Schemas ................. 37
4.6 Post Processing ................................................................. 38
4.7 Analysis Methods ............................................................. 38

5. RESULTS ............................................................................ 39

5.1 Code-Switching and Province ............................................ 40
5.2 Code-Switching and Population Density ............................. 43
5.3 Code-Switching and Tweet Length ..................................... 45
5.4 Emerging Structures ......................................................... 45
5.5 Linguistic Innovation: Pa More ......................................... 45
5.6 Detached Prefixes ............................................................. 48
5.7 Lexical Observations ......................................................... 50
5.8 Verb Stems .................................................................... 50
5.9 Changes in Lexical Category ........................................... 51

6. DISCUSSION ..................................................................... 52

6.1 Code-Switching and Province ............................................ 53
  6.1.1 The Northern Group ...................................................... 53
  6.1.2 The Southern Group ..................................................... 55
6.2 Code-Switching and Density ............................................. 59
6.3 Code-switching and Tweet Length ..................................... 61
6.4 Limitations ........................................................................................................ 62
6.5 Suggestions for future research ................................................................. 64

7. CONCLUSION .................................................................................................... 66

REFERENCES ........................................................................................................ 70
TABLE of FIGURES & TABLES

MAP 2.1 CEBUANO speaking areas (box) in Central and Southern Philippines in

MAP 4.0 THE CENTRAL PHILIPPINES AND MINDANAO ................................. 30

TABLE 4.3 CODES USED IN CATEGORIZING TWEETS ........................................... 33

TABLE 4.4A COORDINATE BOUNDING BOXES BY PROVINCE ................................. 34

TABLE 4.4B EMOTICONS AND UNICODE ................................................................. 35

TABLE 4.5 STRUCTURAL MORPHEMES AND SYNCRETIZED CS PHRASE ...................... 37

TABLE 5.1A CHI-SQUARE TABLE. CS IN TWEETS PER REGION. (OBS. = OBSERVED COUNTS;
EXP. = EXPECTED COUNTS) (OVERALL MEAN CS RATE = 0.26834 [26.8%]) ............... 42

TABLE 5.1B CHI-SQUARE TABLE. CS FREQUENCY PER REGION AND GROUP AVERAGE.
(MEAN FREQUENCY = 26.83%; STANDARD ERROR = 0.5%) ......................................... 43

TABLE 5.2 CHI-SQUARE (x²) CONTINGENCY TABLE OF CS IN TWEETS BY POPULATION

DENSITY. (OBS. = OBSERVED COUNTS; EXP. = EXPECTED COUNTS) (OVERALL MEAN CS
RATE = 0.26834 [26.8%])........................................................................ 44

TABLE 5.5 FREQUENCY OF "PA MORE" IN CORPORA FROM EACH PROVINCE

FIGURE IN PARENTHESIS (%) IS THE PERCENTAGE OF THAT PROVINCE’S TOTAL CORPUS ...... 47

GRAPH 5.5 PERCENTAGE OF EACH PROVINCE’S TWEETS CONTAINING “PA MORE”. (C =
CEBU; N = NEGROS ORIENTAL; M = MISAMIS ORIENTAL; D = DAVAO DEL SUR) ....... 47

GRAPH 5.6 PERCENTAGE OF EACH PROVINCE’S TWEETS WITH SEPARATED MAG- AND NAG-
PREFIXES. (C = CEBU; N = NEGROS ORIENTAL; M = MISAMIS ORIENTAL; D = DAVAO
DEL SUR; DARK BARS = MAG; LIGHT BARS = NAG) ........................................... 50
MAP 6.1.1 THE NORTHERN GROUP. SHOWN: CEBU CITY (LEFT) AND DUMAGUETE CITY

(right). ([CS+] = BLUE; [CS –] = RED)................................................................. 55

MAP 6.1.2 THE SOUTHERN GROUP. SHOWN: DAVAO CITY (LEFT) AND CAGAYAN DE ORO

(right). ([CS+] = BLUE; [CS –] = RED)................................................................. 59
### TABLE of ABBREVIATIONS USED IN LINGUISTIC EXAMPLES

<table>
<thead>
<tr>
<th>PREFIXES</th>
<th>GLOSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>C-PREFIX</td>
<td>causal prefix</td>
</tr>
<tr>
<td>N-PREFIX</td>
<td>nominal prefix</td>
</tr>
<tr>
<td>M-PREFIX</td>
<td>imperative prefix</td>
</tr>
<tr>
<td>F-PREFIX</td>
<td>future prefix</td>
</tr>
<tr>
<td>I-PREFIX</td>
<td>incepted/incepting prefix</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>PERSON</td>
<td></td>
</tr>
<tr>
<td>1S</td>
<td></td>
</tr>
<tr>
<td>2S</td>
<td>first person singular</td>
</tr>
<tr>
<td>3S</td>
<td>second person singular</td>
</tr>
<tr>
<td>1P-INCL.</td>
<td>third person singular</td>
</tr>
<tr>
<td>1P-EXCL.</td>
<td>first person plural exclusive</td>
</tr>
<tr>
<td>2P</td>
<td>second person plural</td>
</tr>
<tr>
<td>3P</td>
<td>third person plural</td>
</tr>
</tbody>
</table>
1. INTRODUCTION

Code-switching (henceforth CS), has been thoroughly investigated by linguists and non-linguists from various viewpoints such as (a) grammatical analysis (Gardner-Chloros, 2009; Myers-Scotton, 1993; Myers-Scotton, 2003; Sankoff & Poplack, 1981; Joshi, 1985; Woolford, 1983), (b) language acquisition (Edwards, 2003), (c) pragmatic analysis (Auer & Roberts, 2011), (d) sociolinguistic analysis (Gumperz, 1982; Gardner-Chloros, 2009), and (e) computational analysis (Barman et al., 2014; Elfardy et al., 2014; Elfardy & Diab, 2012; Ang et al. 2014). Oftentimes, CS is misunderstood in its meaning and its scope.

Linguists tease apart the nuances of CS to mean the mixed-language product of a multilingual person’s linguistic repertoire (Milroy & Muysken, 1995; Bautista, 2004; Bailey, 2007). In a program called Code Switch broadcast by the radio station NPR, CS is defined to be “… [a] mixing of languages or patterns of speech in conversation” (NPR, 2013). The program expands the meaning of CS to include the myriad of ways that individuals act in order to express their identities. Although this definition conveys a linguistic connection of CS to identity, there is prior research showing that CS plays an important role in the identity of individuals using it (Gumperz, 1982; Gardner-Chloros, 2009; Milroy & Muysken, 1995; Gross, 2006; Bullock & Toribio, 2009; Mishra & Rahman, 2013; Weston, 2013).

This thesis investigates CS with Cebuano and English. Cebuano is a Bisayan language influenced by centuries of language contact with both the Spanish and English languages through colonization. Consequently, features from these languages became incorporated
into Filipino languages\(^1\) and identity (Thompson, 2003; Bautista, 2004). Cebuano historically played a role as the lingua franca in the Central and Southern Philippines and maintains that role today. Currently, Cebuano and English CS is considered a feature of the upper-classes and well-educated (Wolff, 2006)\(^2\). Although this study did not examine demographic information, the medium in which the data was found (the Internet) as well as the language domain (writing) suggest that the individuals in this study have some access to technology typically associated with middle and upper classes, as well as some education to produce comprehensible language on Twitter.

Within the last several decades, researchers were able to investigate CS in the Philippines in various modes of communication including speech, print, and TV and radio (Thompson, 2003; Bautista, 2004; Erwin-Billones, 2012). However, these studies focused on CS as it appeared in formal genres and did not capture streaming CS as it appears on the Internet, a rapidly evolving environment. Within the last decade alone, introduction of new social media applications allowed for new, rapid means of communication transcending national borders such as Instagram (2004), Twitter (2006), and Snapchat (2014). Furthermore, most research on CS in the Philippines focuses only on the Tagalog and English language pair. The research on CS between Cebuano and English is sparse.

I investigate CS in computer mediated communication (henceforth CMC) utilizing a mixed methods approach drawing from computational methods and a sociolinguistic framework to collect, process, and analyze the data. My analysis is informed by my

---

\(^1\) Different Spanish words are more salient as originating from Spanish than others. For example, banyo (báño) and kasilyas (casillas) are more salient and less salient to Cebuano speakers.

\(^2\) This is similar to the utilization of Tagalog and English CS among educated Tagalog speakers.
membership in both the English and Cebuano speech communities, and knowledge of both languages as a native bilingual speaker.

For this reason, I question whether or not CS patterns on the Internet across the Cebuano-speaking world are homogeneous or heterogeneous, which may relate to the identity of speakers from these locales. Cebu City in the province of Cebu holds much prestige in the region. Deviations in the presence of CS on Twitter may be representative of a particular province’s alignment with Cebu City. My thesis is based on the assumption that, in addition to social network strength, the Internet may facilitate standardized language (Paolillo, 2001). I also dispute Erwin-Billones (2012)’s implicational doubts of the existence of CS in print from Mindanao, a large ethnically and linguistically diverse island in the Southern Philippines. Because of the varied environments that the Cebuano-speaking world encompasses, I hypothesize that there will be a difference in the frequency of Cebuano and English CS on Twitter from Cebuano-speaking areas.

I argue that there is a difference in the frequency of CS in CMC across these areas and that some areas employ CS features at different rates than others. I also argue that there is a difference in the frequency of CS in CMC between urban and rural areas and that urban areas employ CS at a higher rate than rural areas. Lastly, I argue that there is a relationship between CS and tweet length and that longer tweets are more likely to contain CS than shorter tweets.

---

3 The author also suggests that this is due to the area’s low exposure to English, to which I disagree.
In the following sections, I introduce the history and background of the modern day Cebuano speakers and prior linguistic influences on the languages in the Philippines. I also introduce the launch of the Internet in the Philippines and their entrance into the world of social media. Next, I establish the framework through which I base my research study and analysis. Afterwards, I present the results of my analyses followed by a discussion of the findings. Lastly, I expound on some limitations of the study, connections to the greater discussion on CS, and propose suggestions for future research in this area.

2. BACKGROUND

The Philippines is a Southeast Asian island nation made up of more than 7,000 islands (BBC, 2014; CIA, 2014). It is subdivided into 17 regions, which are further divided into 80 provinces. The nation gained independence twice – the first time from Spain in 1898, and the second time from the United States in 1946. Over 92 million people live on the islands and belong to various ethnic groups (Philippine Statistics Authority, 2010). Tagalog, Cebuano, and Ilocano compose the three largest groups (BBC, 2014; CIA, 2014; Philippine Statistics Authority, 2014). There are two official languages, Filipino and English, and more than 8 other languages with official status at the regional level, such as Cebuano (Lewis et al., 2015). Filipino is the standardized form of the Tagalog language.
2.1 Demographics

Cebu is composed of nearly all ethnic Cebuanos totaling around 97% of the province (Philippine Statistics Authority, 2002). The remainder consist of other Filipino ethnic groups such as Bisaya, Kankanai, Boholano, Tagalog, Ilonggo, and Bicolano (Philippine Statistics Authority, 2002). Like much of the Philippines, the Cebuano population is relatively young with a median age of 23 years (Philippine Statistics Authority, 2013).

Cebuano is spoken across several regions in the Central Philippines and Mindanao. These regions vary in ethnic make-up, and political, social and linguistic influences. Whereas Cebuano speakers in Cebu and Negros Oriental constitute a very strong majority of the population on their respective islands, Cebuano speakers in the provinces of Misamis Oriental and Davao del Sur share their island with multiple ethnic groups.

Misamis Oriental and Davao del Sur are located on the island of Mindanao in the Southern Philippines, which is shared by various languages, such as the Manobo, Su’banon and Magindanao (Map 2.1), and autonomous regions, separatist groups, and terrorist groups such as the Moro National Liberation Front (MNLF), Moro-Islamic Liberation Front (MILF), and the Abu Sayyaf Group (ASG) (Banlaoi, 2012; United States Institute for Peace, 2011; U.S. Department of State, 2014).
Cebuano is named after the language variety spoken on the island of Cebu and is known for its characteristic of /l/ deletion between vowels, e.g., *sulod* $\rightarrow$ *suod* (inside) (Wolff, 1969; Zorc, 1977). Cebuano is also known as Bisaya or Visaya\(^4\) by its speakers.\(^5\)

### 2.2 Current Status of Cebuano

There has been contention with regards to the legal definition and recognition of the Cebuano language. Cebuano is a language variety that is mutually intelligible to varying

\(^4\) As with many other languages of the Philippines, Cebuano does not make a distinction between the /v/ and /b/ phonemes. Cebuano speakers pronounce both sounds as /b/.

\(^5\) Bisaya is a term that many Cebuano speakers employ to refer to themselves. Binisaya is derived from this term and is used to reference the language. Although Bisaya is also used by speakers of other languages such as Hiligaynon to refer to their language, it is widely known to refer to Cebuano (cite).
degrees with other Visayan variants, yet is classified as a language in its own right. On the other hand, other languages are grouped under the single title of “Bisaya”. Thompson (2003) proposes political motivations behind this classification system. He posits that solidification of power drove those in power to separate out Cebuano from the Bisaya language group while allowing Tagalog and its variants to unify into one group.

This categorization had a significant effect on the official numbers of Cebuano speakers in the Philippines, which has hovered around 10 million from the 1980s to the most recent census in 2010⁶. The Filipino census reports that the Cebuano language population reduced by about 800,000 speakers from the decade of 2000 to 2010. In contrast, the population of Tagalog speakers nearly doubled since the 1980s from around 13 million to around 22 million speakers in 2010 (Philippine Statistics Authority, 2013; Dow & Fishman, 1991). These classifications have been perpetuated by publications over time (Ammon et al., 2006:2018). The categorization definition and qualifying parameters seem to be in constant flux and can be seen in the reduction of the Cebuano speaking population by about 800,000 speakers from the decade of 2000 to 2010 whereas the Bisaya language group nearly doubled from 5 million to 10 million in the same period (Philippine Statistics Authority, 2014; Mansueto, 2013).

The classification problem may have affected the language ideology where Tagalog is a language, and Cebuano and the other Filipino languages are dialects. This ideology is particularly prevalent in informal discussions about language in the Philippines. On an

⁶This number is used to illustrate how the Tagalog hegemony can successfully maintain its argument that having the most speakers positions Tagalog to be accepted as the national language in the Philippines (Dow & Fishman, 1991; Thompson, 2003).
online forum called Word Reference, one user writes, “Tagalog is a language, Cebuano and Bikol are dialects” (Wordreference.com, 2011). On Yahoo! Answers, another user writes, “dialect: a particular form of a language that is peculiar to a specific region or social group. So it is dialect. Waray is synonymous with Leyte as is Cebuano with Cebu” (Yahoo!, 2014).

This paper will use Cebuano to refer to the language as it is spoken in the four provinces of Cebu and Negros Oriental in the Central Visayas Region\(^7\), and Misamis Oriental in Northern Mindanao, and Davao del Sur in the Davao Region. Cebuano spoken among these four provinces is relatively the same.

2.3 Ethnic Origins

The original inhabitants arrived in the Philippines through an ancient land bridge that disappeared long ago. This land bridge connected Southeast Asia to the surrounding islands in the southeastern waters, but disappeared at the end of the ice age as water levels rose globally.

According to the 2015 edition of the Ethnologue, there are 15.8 million Cebuano speakers in the Philippines, who make up about 17.2% of the total population\(^8\) (Paul et al., 2015). These speakers are largely distributed in the Central Visayas region, and

\(^7\) Regions are large administrative divisions of Filipino government. The 17 regions are composed of 80 provinces, and further into municipalities and cities. These are further divided into the barangays, the smallest administrative division of Filipino government. (pg. 27)

\(^8\) As I will illustrate, the current constitution designates Filipino and English as the only official languages, while allowing provisions for other regional languages to be recognized.
farther south on the island of Mindanao (Lewis et al., 2015). The only language to have more speakers than the Cebuano speaking community is Tagalog (Philippine Statistics Authority, 2014; CIA, 2014).

Despite having such a large community, the origin of the Cebuano speaking people is nebulous (Zorc, 1977). There are multiple hypotheses as to how the people arrived to their island home and how they became known by their general name – Visayan or Bisayan⁹. The leading hypotheses on the origin of the term “Visaya” examine various perspectives. Linguistically, it has been proposed that the name is derived from the word *sadya* “happy.” Historically, the name could either be derived from an older word meaning “slave”, or a word derived from Malay or Sanskrit for *vijaya* “victorious” The word ‘slave’ was proposed because ancient Visayans were purportedly attacked, captured, and sold into slavery by marauding pirates. The suggestion for “*victorious [people]*” relates the origins of the name and people to an ancient kingdom based in Sumatra, farther south from the Philippine archipelago, known as Sri Vijaya (Narva, n.d.).

**2.4 The Spanish Period**

By the 16th century, Spaniards started to colonize the Philippines. Spanish rule lasted until the end of the 19th century, when they were defeated by the United States who took control of the islands until 1946 (LOC, 1991).

---

⁹ See Footnote 1
Cebu City, the capitol of the island and province of Cebu, garnered prestige over the centuries after first contact with the Spanish (Cebu City, 2010). The city’s list of accolades includes being the first and oldest city in the Philippines, the first Spanish settlement, and the first to have aggressively resisted the Spaniards within weeks of first contact (Thompson, 2003; Dumdum et al., 2004). Portuguese explorer Ferdinand Magellan was purportedly killed by Lapu-Lapu, ruler of Mactan. Eventually, the Spaniards were able to gain control of the region and established a military, missionary, and cultural base in Cebu City. As a result, Cebu City became the cradle of Christianity in the Far East, which is evident in the numerous relics found in the city today (Cebu City, 2010).

Despite over 300 years of Spanish rule, there were very few instances in which the Spanish would educate the indigenous people in the Spanish language (Thompson, 2003). They feared an over-educated underclass and potential uprisings. However, several centuries of Spanish rule still left a large linguistic footprint on the Filipino languages. For example, the Spanish number system is still used alongside the Cebuano number system. Additionally, terms for everyday objects and abstract ideas such as *banyo* (bathroom; Spanish: baño), *baso* (glass; Spanish: vaso), and *relihyon* (religion; Spanish: religión) are still utilized in modern Cebuano.

With these achievements and accolades under its belt, Cebu City and the associated language of Cebuano became prestigious in the region (Philippine Statistics Authority, 2002; Cebu City, 2010). It replaced and displaced local vernaculars as it spread

---

10 Mactan is a small island located right off the coast of the cities of Cebu and neighboring Mandaue.
throughout the area — a small-scale parallel to the effects of globalization and the spread of English today.

2.5 The American Period

In 1898, the United States established control over the Philippines acquiring the former long-standing Spanish colony in the Treaty of Paris. However, Filipino forces led by Emilio Aguinaldo who had been fighting for independence from the Spaniards beforehand, were unsatisfied with solely a change in colonizer and continued to fight for independence during U.S. control (Thompson, 2003:19; U.S. Department of State, 2013). Unlike the Spanish colonizers before them, Americans employed a pacification campaign designed to quell the insurrection. Led by Governor William Howard Taft, the pacification campaign aimed to convince Filipino elites and the local population to join the U.S. side. In addition to allowing certain degrees of self-governance to those who did not embrace Aguinaldo’s insurrection, Taft implemented plans for economic development and introduced social reforms, which included educating the elites and local Filipinos in reading, writing, and speaking English (Thompson, 2003:20; U.S. Department of State, 2013). In the following years, the Philippines held its first general assembly in 1907, and in 1916, the Jones Act promised eventual independence for the Philippines (U.S. Department of State, 2013). During this time, much attention had been focused on Manila, acting as the nation’s capital, and Tagalog and English became the languages of power, government, and, eventually, unification.
With the national capital in Tagalog speaking regions to the North, and Cebu City’s role as national capital barely a memory, Cebuano has been relegated to less than equal status to Tagalog, English, and the newly minted Filipino language\textsuperscript{11}.

2.6 Anti-Tagalog Attitudes

In opposition to a Manila centered government, based out of an ethnically Tagalog area, separatist sentiment pervades throughout the Cebuano speaking world. A quick Google search for the terms “cebu” and “secede” returns over 400,000 results. Many of these results are online forums, blogs, or news articles that talk and debate about the topic. The idea is very well known and widespread. For example, on the website \textit{PinoyExchange.com}\textsuperscript{12}, a discussion spanning over 100 posts discusses the repercussions of Cebuano separation. A user writes, “\textit{If Cebu does declare secession, other provinces in Visayas and Mindanao will follow too}” (Pinoy Exchange, 2005).

These public attitudes and sentiments have not only affected conversations, but have also affected speech itself. Although Tagalog and subsequently, Filipino are widely used languages in the Philippines\textsuperscript{13}, Cebuano speakers are more than eager to communicate in English instead of Tagalog. In his article on the state of the English language in Cebu, Ceniza (2014) writes, “Cebuanos might prefer English than Tagalog as a way to communicate to non-Cebuanos” (Ceniza, 2014). Paul, a member of the online forum for

\textsuperscript{11} “Filipino” has garnered much criticism since its inception as a unifying national language. Many of the non-Tagalog ethnic groups saw this action as one of centralizing power with the Tagalog ethnic group since Filipino and Tagalog are effectively identical languages (Thompson, 2003; McFarland, 2004).

\textsuperscript{12} Pinoy is the colloquial term for Filipino

\textsuperscript{13} Filipino or Tagalog is a compulsory school subject, while local languages typically are not taught or even banned from the classroom.
Cebu residents, expresses that, “[even] while watching Tagalog shows, Cebuanos still speak Cebuano” (Living in Cebu Forums, 2014). Lastly, in response to the question asking if one can get by in Cebu with only “Filipino (Tagalog)”, a user responds that, “...Language politics in Cebu can get tense, and the Cebuanos are very proud of their language” (Quora, 2015).

These anti-Tagalog attitudes are deep rooted and exemplified in the Cebuanization of what would typically be Tagalog language domains, such as the national anthem. On the community forum Asia’s Finest, a user attests to this resistance and writes, “... Cebuanos even still insist in singing the National Anthem in Cebuano and using English rather than Filipino (based in Tagalog)” (Asia’s Finest Forums, 2010).

2.7 Language History

The official languages of the Philippines today are Filipino and English. In most contexts, Filipino is synonymous with Tagalog, or the Metro Manila dialect of Tagalog (CIA, 2014). The language underwent a couple of name changes in 1959 and 1973 between Filipino and Pilipino in an effort to dissociate it with the Tagalog ethnic group. Eventually, the Philippine Constitution under Article XV, Section 3(2) mentions development of Filipino, with no mention of Pilipino or Tagalog associations. The change in the spelling from Pilipino to Filipino attempted to dissociate the language from Tagalog since the phoneme /f/ does not naturally appear in Tagalog (Thompson, 2003). Of course, the attempt to dissociate Pilipino from Tagalog ethnicity through name change only garnered criticism from the other ethnic groups in the Philippines. In 1987, the new
The constitution made provisions for other languages in addition to Filipino such as allowing for official recognition as regional languages. The distinction between Filipino and Tagalog is important politically, since the former connotes nationality, while the latter connotes regionalism and ethnicity.

The second official language of the Philippines is English. It dominates and permeates all aspects of Filipino life. From governmental and academic institutions to commercial television and radio broadcasts, English is used as the language of wider communication (Thompson, 2003). With such wide and constant exposure, English vocabulary, phrases, and idioms easily made their way into everyday parlance of the languages in the Philippines. New vocabulary such as kompyuter (computer), nars (nurse), and taym (time) have become commonplace. English also introduced new lexicon for already existing Cebuano lexical items, such as honorifics. Sir and ma'am are used quite frequently and are analogous to po, an honorific form of personal address, in Cebuano borrowed from Tagalog.

With over 100 years of exposure to English, Cebuano also shows effects of language contact with English. However, with the introduction of mass media, mobile technology, and the Internet, these languages can more easily come into contact with each other, catalyzing the process of language contact and change.
2.8 Technology in the Philippines

The Philippines first connected to growing Internet via a 64 kbps link to Sprint in the United States on March 29, 1994 around 1:15 A.M. at the University of San Carlos in Cebu City (Francesca de Guzman, 2004). The first message over the new connection was, “Subject: The Philippines is In!” (Santos, 2014). The event was low-key without any fanfare or major celebrations. Adoption of the Internet took a while to gain traction in the subsequent years.

According to the World Bank (2015), by the year 2000, only 1.98% of the country was reported to have used the Internet. However, in the following decade, Internet use exploded. By the year 2012, the percentage of individuals using the Internet rose to 36.24% of the population, which is a growth rate of over 1,700%. In contrast, the percentage of growth of individuals using the Internet in the United States had only grown by 88.08% in the same period.

In addition to a growing number of users accessing the Internet through personal computers, smart phone usage is also growing quickly. Mobile phone usage is projected to grow by 233% by 2015. Filipinos are already avid users of the Internet on their mobile phones, especially among the younger generations.

In the Philippines, social media penetration is very high, with 94% of the Internet using population connecting to Facebook. In contrast, social media penetration in the United States reaches only 54% of its Internet using population connecting to Facebook. In
addition to this, the Philippines is the most engaged in the Asia Pacific region when it comes to time spent on social media, spending an average of 4 hours a day on social media.

The Cebuano speech community in the Philippines cares strongly about its identity and the Cebuano language. However, with the growth and spread of Cebuano speakers throughout the Central Philippines and Mindanao, I argue that the greater Cebuano speech community is diverging with respect to the Cebuano identity that is supported by Cebu. The following section will explore prior literature that I use to support my thesis, providing previously discussed mechanisms, factors, and functions of CS and speech communities.

3. LITERATURE REVIEW

In this section, I establish the theoretical framework with which I structure my analysis of the data. I introduce the current definition of CS, and current studies that examine factors affecting CS patterns and frequencies in tandem with identity. These studies investigate CS as it is affected by language, community, age, and gender. Then, I introduce CS grammar models proposed by Myers-Scotton (1993) and Sankoff & Poplack (1981) to universally describe CS. Of these models, I will introduce Myers-Scotton (1993)’s *Matrix Language Frame* (MLF), which is an extensive model describing the production of CS speech.
Researchers examined CS from various perspectives such as (a) grammatical analysis (Gardner-Chloros, 2009; Myers-Scotton, 1993; Myers-Scotton, 2003; Sankoff & Poplack, 1981; Joshi, 1985; Woolford, 1983), (b) language acquisition (Edwards, 2003), (c) pragmatic analysis (Auer & Roberts, 2011), (d) sociolinguistic analysis (Gumperz, 1982; Gardner-Chloros, 2009), and (e) computational analysis (Barman et al., 2014; Elfardy et al., 2014; Elfardy & Diab, 2012; Ang et al., 2014). Grammatical analyses attempt to explain rules, constraints, and frameworks in code-switched utterances (Myers-Scotton, 1993; Sankoff & Poplack, 1981; Woolford, 1983; Joshi, 1985). Recently, researchers have explored CS with regards to its acquisition and deployment in bilingual individuals and L2 language learners\(^\text{14}\). Pragmatic analyses investigate the meanings constructed through the act of CS itself. Sociolinguistic analyses examine the relationship and function of CS and the communities that utilize it.

### 3.1 Code-Switching

The most common definition of CS is the alternation between two languages in an utterance. Linguists disagree on a clear definition for CS and proposed terms such as code-mixing to describe different types of CS (Bautista, 2004; Gardner-Chloros, 2009; Milroy & Muysken, 1995; Myers-Scotton, 2003; Poplack & Walker, 2003). This disagreement stems from the difficulty of clearly defining what constitutes CS.

On a fundamental level, there are two overarching types of CS, *intrasentential* and *intersentential*. Traditionally, intersentential CS is achieved through the alternation

\(^\text{14}\) L2 stands for “second language”. L2 language learners refers to those studying their second language in addition to their mother tongue.
between two or more languages for each phrase uttered (1). That is, one language is used for utterance A, while another is used for utterance B. Intrasentential CS is a bit more difficult to demarcate as it can resemble non-CS language contact phenomena (Bullock & Toribio, 2009). Typically, intrasentential CS involves a base language A with language B interspersed throughout (2) (3). Other types of CS involve tag-line CS (Bullock & Toribio, 2009; Myers-Scotton, 2003), in which the preceding sentence is composed of language A, while the tag phrase is in language B (4).

(1) “I don’t have second thoughts about including PWDs in the automatic PhilHealth coverage. I commend Dizon on this noble legislative initiative for the welfare of PWDs,” pada-yag ni Abellanosa.”

“I don’t have second thoughts about including PWDs in the automatic PhilHealth coverage. I commend Dizon on this noble legislative initiative for the welfare of PWDs,” reported Abellanosa.” (Cerojano, 2015)

(2) “[Lalake] nga motambong unta sa 1st death anniversary sa iyang kanhi live-in partner gipusil sa igsuon niini niadtong milabay nga gabii.”

“A man who was hoping to attend the one-year death anniversary of his live-in partner was shot by their sibling yesterday night.” (Lopez, 2015)

(3) “Sa pagkabasura sa motion, ang prosekusyon mahimo nga mosang-at og motion for re-investigation …”

“In trashing the motion, the prosecution can file a motion for re-investigation …” (Gravinez, 2015)

(4) “Yup, in English talaga.”

“Yup, in English, really.” (Banzon, 2015)

Despite the difficulty in determining certain types of CS activity to be true CS, researchers agree that employing CS is definitely not a result of language deficiencies. Rather, CS is used by proficient multilingual individuals (Milroy & Muysken, 1995;
Bullock & Toribio, 2009). Bullock & Toribio (2009) succinctly say that, “[code-switching] requires an advanced level of bilingual proficiency.” CS is a resource with which speakers can access a range of “… social and rhetorical meanings …” (Milroy & Muysken, 1995; Gumperz, 1982; Thompson, 2003; Bautista, 2004).

Lastly, language contact phenomena include lexical borrowings, loan translations, and language attrition, just to name a few. These phenomena resemble CS in that their lexicon and grammar also originate from language B rather than solely from language A entirely.

3.2 Language Contact

Many researchers agree that CS is the result of some form of language contact. Language crossing is a related phenomenon, in which an individual of language A appropriates vernacular from language B, without being a speaker of language B. Language contact phenomena considered when analyzing CS are lexical borrowings and loan translations.

Lexical borrowing, better known as borrowing, is the process by which a word from language B is incorporated into the vocabulary of language A. This process typically involves vocabulary of cultural nature. For example, English is rife with borrowed words such as fiesta, quesadilla, and barrio from Spanish, and strudel, blitzkrieg, and verboten from German. The phonological properties of these lexical items adapt to English’s phonological system, but still retain their orthographic shape from their original languages. Borrowing tends to be unidirectional in that lexical items from the more
dominant language, or acrolect, is adopted by less dominant ones, or the basilects (Odlin, 2009).

*Loan translation* is similar to lexical borrowing in that it involves movement of vocabulary from one language into another. However, unlike lexical borrowing, where vocabulary is typically appropriated wholesale by the language, *loan translation* borrows the semantic element of the word modeled by language B, utilizing native vocabulary to complete the lexical entry. For example, *skyscraper* is loan-translated into many other languages such as French *gratte-ciel*, Spanish *rascacielos*, and German *Wolkenkratzer*15.

### 3.3 Sociolinguistic Reasons for Code-Switching

The sociolinguistic approach to studying CS is one of the most diverse (Bullock & Toribio, 2009). There are many extralinguistic factors that affect CS, which include, identity (Bautista, 2004), social networks, attitudes, age, gender, class, and community norms (Bullock & Toribio, 2009). There are two levels CS research covers – micro and macro. Micro-level CS studies focus on the individuals motivations for CS (Gumperz, 1982; Le Page & Tabouret-Keller, 1985; Mishra & Rahman, 2013; Eppler, 2010). Macro-level CS studies focus on situating individual CS behavior within various social contexts and social norms (Weston, 2013; Poplack, 1993; Bullock & Toribio, 2009; Stell, 2010). I draw from literature from both of these perspectives in framing my analysis.

Gumperz (1982)’s early work investigated in various contexts around the globe from a sociolinguistic perspective. This included examining the historical genesis, conversation functions, and significance to speakers (Bullock & Toribio, 2009). His work provided us with a framework to study CS on both a micro and macro levels. His notions of *we-code* and *they-code*, in addition to *situational* and *conversational CS*, helps to lay the groundwork for examining CS in a variety of situations, contexts, and languages.

The notion of *we-code* and *they-code* explains that in a diglossic situation, the minority language assumes the role of the *we-code*, while the majority language assumes the role of the *they-code* (Gumperz, 1982; Bullock & Toribio, 2009). Additionally, usages of these codes are complex and dependent on the situation in which the speaker finds themselves. This is important to the speaker as it allows for them to possess a fluid identity, which they can evoke using the proper code (Nilep, 2006; Heller, 2001).

Related to the *we-code* and *they-code* distinction is the distinction between *situational* and *conversational CS*. Situational CS is associated with changes in interlocutor, context, or topic and is directly influenced by the distribution of languages in diglossic context (Bullock & Toribio, 2009). On the other hand, conversational CS occurs when languages are used to evoke a metaphorical meaning or connotation. This function is similar to Blom & Gumperz (1972)’s *metaphorical CS*, in which formality of the situation was effected by CS (Blom & Gumperz, 1972; Nilep, 2006). The distinction between the two is that external changes – such as changes in interlocutor, context or topic – effect situational CS whereas internal changes without external prompting effect conversational
CS (Bullock & Toribio, 2009; Gardner-Chloros, 2009; 106 – 107). In other words, although CS can be affected by the situation, the individual ultimately has the authority to determine the type of CS they would employ.

Some researchers question the discreteness of the languages in CS. They argue that languages are flexible, which is especially true in language contact situations. Le Page & Tabouret-Keller (1985) propose that the notion of languages as discrete codes that individuals can alternate between should be regarded as a myth. Eppler (2010) follows this approach in her study on Austrian refugee communities in the United Kingdom who code-switch between German and English. The resulting code-switched vernacular is referred as *Emigranto*. Stand-alone code-switched languages are not only found in immigrant communities. In Mexico, an indigenous Nahuatl and Spanish speaking community utilize a syncretized version of the code-switched languages, called *Mexicano*, as a single language in the community (Messing, 2007). Therefore, although CS can be used as a form of fluid identity between *we* and *they* groups, CS can also be viewed as a language in its own right rather than a deviation of it.

In addition to these pragmatic functions of CS, researchers examined extralinguistic factors that affect CS production. These include factors such as age, gender, and community.
3.4 Code-Switching and Community

Milroy & Milroy (1985) posit that community has a profound effect on the standardization or vernacularization of an individual’s language choices. They mention that, “… the closer the individual's ties to a local community network, the more likely he is to approximate to vernacular norms”. Milroy (2002) expands this definition and explains that close-knit networks are conducive to the vernacular variety, while loose-knit networks are conducive to standardization. Additionally, Milroy (2002) discusses that movement from rural into urban areas loosens close-knit networks allowing for standardized forms and linguistic innovations to be adapted into that community’s vernacular. I extend this effect to CS in that CS can be either vernacular or standard, depending on the strength of the network of the community that employs it.

Supporting CS research finds that the appearance of CS is not bound to either vernacular or standard forms. In Poplack (1993)’s CS study comparing CS patterns in Spanish and English speech communities in New York and French and English speech communities in Ontario and Quebec, she found that while CS in Spanish and English speech communities was attributed to stylistic or discourse functions by skillful bilinguals, French and English speech communities in Ontario and Quebec did not engage in the same CS behavior. Instead, French and English speech communities engage in mostly intersentential CS and acknowledge their performance of CS through various means such as immediate translation, metalinguistic commentary, and English bracketing (Poplack, 1993). It is clear from this study that CS is not automatically assigned to either vernacular or standardized speech.
Stell (2010) confirms that CS is assigned to vernacular or standard language categories in his investigation of Afrikaans and English CS in South Africa. Stell (2010) studied CS as it pertained to race, which he categorized as whites and “coloureds”. Stell defined coloureds according to the colonial British administration’s definition from 1806, which subsumes all “non-White Cape Dutch-speakers and acquired the trappings of an ethnic identity” (Armstrong and Worde, 1979: 122; van der Ross, 1979 as cited in Stell, 2010). He found that while CS was more prevalent among the non-white group, both groups used monolingual language to show affiliation with new South African values. In short, Stell (2010) found that CS is a marker of community to which the speaker belongs and can be used to express that membership.

Eppler (2010) finds similar patterns with Austrian immigrants who fled to London as refugees in the 1930s. She finds that these immigrants were more likely to code-switch in German and English in close-knit situations amongst each other and less likely to do so when interacting outside their community. This pattern also can be seen in Kontra (2001)’s study on Hungarian and Serbian CS where Hungarian individuals opted for the CS variety when they became a local minority in Hungary, where they are typically the majority.

These studies make clear the relationship between CS and community. However, their approaches involve comparison of split communities. Poplack & Walker (2003)’s study compared two linguistically different speech communities while Stell (2010) compared two racially different speech communities.
3.5 Code-Switching and Age

Age is an important factor to consider when studying language, especially since certain age groups are more inclined to adopt linguistic innovations than others (Alfonzetti, 2005; Weston, 2013; Eppler, 2010). In examining Spanish and English CS patterns among different age groups in Gibraltar, Weston (2013) found that younger age groups used more CS than the older age groups. Weston (2013) argues that each group had different access to linguistic resources, and, therefore, adopted the linguistic innovation at different rates.

In general, the older age groups were monolingual and had less access to Spanish. Among the younger age groups, CS innovations were facilitated by access to the two languages. Furthermore, Weston (2013) found that factors such as “… social network construction, residency patterns, and attitudinal orientations towards Spain and the UK” had an effect on their CS patterns. In other words, coupled with their attitudes to the countries where Spanish and English came from, individuals on Gibraltar would engage in CS depending on whom they knew and where they lived.

These factors also played an important role in Eppler (2010)’s study on German and English CS in the Austrian immigrant population in a northwest London neighborhood. The Austrian immigrants first arrived in London as refugees in the 1930s. They had a positive attitude towards English, resettled in a neighborhood where Standard British English was spoken, and formed a close-knit community as Jewish refugees from the conflict in Eastern Europe (Eppler, 2010). Unlike the younger age groups in Gibraltar
who were accepting of CS in practice, the younger age groups among the refugee groups in London shunned it. Eppler (2010) mentions how, in the context of speaking in English and German in public, one refugee youth would be “…terrified when [their] parents spoke German”, and that they would “…crawl away…” as a “…‘I don’t belong to them’ sort of thing” (pg. 82).

Gender also plays a role in CS. In monolingual settings, women have traditionally used more standard language varieties than do men (Labov, 1972; Trudgill, 1974; Chambers, 2003 as cited in Bullock & Toribio, 2009). Milroy (2002) discusses that women generally have wider networks, which is receptive to standardized forms instead of reinforcing vernacularity.

3.6 Code-Switching and Social Media

Researchers have also investigated CS with respect to CMC quite extensively finding that written language closely approximates spoken language in non-formal situations (Marley, 2011; Delamotte & Desoutter, 2011). In formal situations, researchers discovered that individuals tended toward monolingual practices, composing documents within one language (Delamotte & Desoutter, 2011). For some cultures, such as the Spanish and English speaking communities in New York (Poplack, 1980), CS is a normal part of communication, which holds true in CMC where it can be utilized to show in-group affiliations, as well as allow for individuals to express fluid identities (Paolillo, 2001; Marley, 2011).
Within the last several years, CS has caught the attention of linguists and computer scientists focused on the task of natural language processing (NLP) (Elfardy et al., 2014). In their article on the challenge of detecting CS language in social media, (Barman et al., 2014) note that this trend is particularly popular in language-dense regions around the world. Although English is currently the most dominant language in social media, its presence is receding (Barman et al., 2014).

3.7 Grammar of Code-Switching

There has been plenty of research exploring CS models, triggers, and structures. Most notably, the free-morpheme and equivalence constraints proposed by Sankoff & Poplack (1981), and the matrix language frame model (MLF) proposed by Myers-Scotton (1993).

Although they were not the first researchers to attempt to describe the rules through which CS is realized, Sankoff & Poplack (1981)’s model for describing the inner workings of CS speech became very well known. Their proposal expanded on Poplack (1980)’s earlier work on CS in Spanish/English speaking communities on the island of Manhattan in New York City. The first model explains CS as a function of the free morpheme constraint. This constraint stipulates that CS can occur as long as the constituent consists of at least one free morpheme\(^\text{16}\) (5) (6) (Sankoff & Poplack, 1980: 5, 13).

\[^{16}\text{A free morpheme does not need to be bound to other morphemes. That is, they can be used as complete words.}\]
Myers-Scotton (1993) expanded on prior linguistic research in formulating a CS grammar with her Matrix Language Frame (MLF) model based on her CS work in Africa. In this model, Myers-Scotton posits that there are two roles that are created in a multilingual production scenario. One language provides the frame for the utterance supplying the syntactic and, at times, the phonological properties of the utterance evident by the system morphemes in the utterance. Myers-Scotton (1993) defines system morphemes to be mainly inflection and function words and a closed class. She calls the language providing these system morphemes the *matrix language* (ML). The other language or languages provide content morphemes, which can be assimilated to varying degrees in their morphophonology. She calls this the *embedded language* (EL). She posits that extended stretches of EL can inhibit the ML, allowing the EL to provide the utterance with system morphemes.

With this model in mind, I introduce the data I analyze. I first describe the data with respect to its selection, collection, storage, and processing of the data utilizing several computer programs. Next, I explain the encoding process and coding systems that I adapted for my particular dataset. Lastly, I describe the pre- and post-processing of these data and the methods of analysis.
4. DATA & METHODS

In order to investigate CS in CMC, I examine short messages (henceforth tweets) on the popular micro-blogging platform Twitter. Twitter is a free service on the Internet that allows any user to register and publish digital messages provided they have access to the website. This dataset included over 10,000 tweets, 8,699 of which fell within the geographic bounding boxes corresponding to each of the 4 provinces in the Central Visayas, Northern Mindanao, and Davao regions. Regions are large administrative divisions of Filipino government. The 17 regions are composed of 80 provinces, and further into municipalities and cities. These are further divided into the barangays, the smallest administrative division of Filipino government. I use the term group to refer to groupings of two provinces throughout this study.

This type of data was chosen for analysis for various reasons: (i) semblance of digital communication in this format to natural speech (Marley, 2011; Delamotte & Desoutter, 2011; Crystal, 2008), (ii) flexibility and feasibility of data collection, and (iii) relevance to the increased Internet access that the Philippines has been experiencing in the past two decades and its symbolic relationship to power.
The four provinces I examine are Cebu, Negros Oriental, Misamis Oriental, and Davao del Sur (Map 4.0). I chose these provinces because they fall within the Cebuano speaking region of the Philippines and they are home to four well known Cebuano speaking metropolitan areas, which are Cebu City, Dumaguete, Cagayan de Oro and Davao City. These provinces are physically and politically separated as they are found on different islands situated far apart from each other and belong to different political districts and associate with different local centers of power (Thompson, 2003).

With regards to demographic data of the Twitter users definite data was particularly difficult to obtain since Twitter does not publish this information with each tweet. There were 2,652 users who collectively produced the 8,699 tweets that were collected for this
investigation. I further filtered the corpus down to 7,729 tweets after finding tweets containing CS employing Tagalog to provide the grammatical structure for the sentence, or as the ML, and not Cebuano.

Twitter did not explicitly provide demographic information with each tweet as collected through their streaming API. Additionally, Twitter did not provide any socioeconomic data. Therefore, I did not included these factors in my analysis of the data. However, I assume that the majority of the individuals publishing tweets are within the teenager to young adult age range. I support this assumption through research by Tagliamonte & Denis (2008:24), Crystal (2008:90), and Paolillo (2001), who find that adolescents and young adults compose a great majority of individuals who are involved in CMC. I also assume that the individuals publishing tweets range in socioeconomic status accessing the Internet through various means, such as personal computers, Internet cafes, and many through mobile telephones.

The data I collected were initially stored in the hard drive of my personal password protected laptop and were transferred to a secure 1 terabyte external hard drive for protection. The raw data totaled 66.7 megabytes (Mb) and contained each tweet as a separate row in JSON (Java-Script Object Notation) format in a comma separated value (.csv) file. The JSON format is a popular format to transfer large amounts of information through a network and takes the form of \{category: data, category-2: data-2, etc.\}. This format in particular allowed for Python scripts to parse out relevant information and clean the data for analysis.
4.1 Data Collection

The data were collected over a period of three months in the year 2014 from the end of August to the beginning of November every day of the week throughout various times of the day and night. I used this method to ensure variability in the number, types and topics of the tweets collected for analysis. Additionally, I alternated the daily collection periods between various times throughout the day to ensure that I created a corpus of tweets authored by different types of users. For example, I strove to gather tweets from consistent users as well as users who tweeted only during special occasions, such as birthdays, political events and natural phenomena, such as the blood moon in October. To collect this data, I interfaced with Twitter using Python scripts and their Application Programming Interface (API)\textsuperscript{17} to automatically collect tweets simultaneously as they were being published.

4.2 Unit of Analysis

I use the tweet as my unit of analysis when running statistical analyses and performing feature counts. Each tweet has a limit on how much can be said with a character limit of 140 characters. For perspective, users can publish messages with 35 words averaging 4 characters to reach this limit. I coded each tweet manually according to Table 4.3 in the following section. Tweets categorized as CS+ were converted in CS- if the ML of the tweet was actually Tagalog, evident through Tagalog system morphemes in the Tweet.

\textsuperscript{17} Many websites allow users to interact with their functions and data by provided APIs. For example, Google maps allows users to embed their maps into the user’s websites and allows the user to enable or disable functions through the API.
4.3 Encoding

Tweets are coded for their (a) geographical location, (b) CS status, (c) average length in words, and (d) presence of features (Table 4.3). Features that I examined were the syncrretized phrase “pa more” and the detached prefixes “mag” and “nag”. Geographical location contains four subcategories, each corresponding to the metropolitan areas at the center of each region. CS status codes “+” or “−” depending on the presence of one or more instances of CS. Lastly, tweet length is the count of the number of words in the tweet. Because of the genre, words are defined to be a collection of characters prefixed and suffixed by a white-space character.

<table>
<thead>
<tr>
<th>(a) Geographic Location</th>
<th>(b) CS Status</th>
<th>(c) Length</th>
<th>(d) Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>(c) Cebu</td>
<td>(+) CS present</td>
<td>(n) any number greater than zero</td>
<td>(+) present</td>
</tr>
<tr>
<td>(d) Negros Oriental</td>
<td>(−) CS absent</td>
<td></td>
<td>(−) absent</td>
</tr>
<tr>
<td>(o) Misamis Oriental</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(v) Davao del Sur</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 4.3** Codes used in categorizing tweets

4.4 Criteria

With regards to coding criteria for each of the coding categories mentioned above, the following section discusses how each meet the criteria for each of the categories’ codes. Geographical location was contingent upon the particular tweet’s coordinate information – in the format \((latitude, longitude)\) – as it is provided by twitter. Although the Python script I used to automatically collect live streaming tweets set limits on which geo-coded tweets would be collected, there were a sizeable minority of tweets that were collected
outside of these limits (Table 4.4a). The following coordinate restrictions for each of the subcategories in geographical location (C, N, M or D) are as follows:

<table>
<thead>
<tr>
<th>Province</th>
<th>Latitude 1</th>
<th>Longitude 1</th>
<th>Latitude 2</th>
<th>Longitude 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cebu (C)</td>
<td>123.3678</td>
<td>9.8033</td>
<td>124.0778</td>
<td>11.3111</td>
</tr>
<tr>
<td>Negros Oriental (N)</td>
<td>122.587742</td>
<td>9.005823</td>
<td>123.330715</td>
<td>9.583283</td>
</tr>
<tr>
<td>Misamis Oriental (M)</td>
<td>123.2412</td>
<td>8.2339</td>
<td>125.2561</td>
<td>9.1125</td>
</tr>
<tr>
<td>Davao del Sur (D)</td>
<td>125.2052</td>
<td>6.9312</td>
<td>125.7038</td>
<td>7.6192</td>
</tr>
</tbody>
</table>

Table 4.4a Coordinate Bounding Boxes by Province

Tweets were grouped into rural (R) or urban (U) depending on their location within or outside city boundaries. I utilized boundaries that I adopted from a population density map overlaid on a Google map[18] that divided the provinces into “barangays,” which are the smallest administrative division of government in the Philippines.

Tweets meet CS requirements with the occurrence of EL constituents in an otherwise ML only tweet. Tweets with EL constituents that may occur in hashtags[19] were not included. Hashtags are defined as special word units in a tweet headed by the hash tag symbol (#) for the purposes of archiving and indexing tweets so that others can find them. For example, #taurus would archive and index the tweet along with other tweets that contain the same hash tag. CS does not include EL constituents that may appear in personal references or calls to other Twitter users. References are achieved by the special word units in a tweet headed by an @ symbol. Since the creation of these special word

---

[18] The population density data can be found online. For example, the population density for the barangays in the province of Cebu are located here: [http://www.citypopulation.de/php/philippines-admin.php?adm1id=0722](http://www.citypopulation.de/php/philippines-admin.php?adm1id=0722)

[19] Some authors prefer to separate hash and tag. I choose to combine them as hashtag.
units is not under the control of the author of the tweet in particular, I exclude them from the analysis.

Average tweet length is counted by the number of constituents in a tweet including both the ML and EL, if present, as well as the special words and hyperlinks that may be present in a tweet. These figures are calculated through the use of the `.split(' ')` method in the Python script. This method splits the tweet at the white spaces within a tweet. For example, ‘this is a tweet #hashtag @user’ contains five white spaces and will return the following six words: ‘this’, ‘is’, ‘a’, ‘tweet’, ‘#hashtag’ and ‘@user’.

There were times when this method did not accurately capture the amount of words in the tweet. These cases typically dealt with Unicode representations of emoticons due to the way that Twitter encoded the tweets upon collection. Unicode representations are alphanumeric words that are presented on the screen as smileys and other emoticons when presented on Twitter. In the streaming process of collecting the tweets, two emoticon codes would sometimes be merged into one long code, which would count as one word using the `split` method described above. To remediate this issue, I excluded non-word tokens from the count. Non-word tokens include hashtags (#), personal calls (@), links (http(s)://…), and emoticons (\u…) (Table 4.4b).

![Table 4.4b Emoticons and Unicode](image)

---

20 For a fuller list of emoticons and their Unicode equivalents, you can visit this list online: [http://apps.timwhitlock.info/emoji/tables/unicode](http://apps.timwhitlock.info/emoji/tables/unicode)
In addition to combined Unicode words, the widespread use of *Jejemon* made classification of tweets particularly difficult when present. *Jejemon* is a type of word game similar to *1337 speak* in that its goal is to obfuscate and to encode the original typically English message. *1337 speak* is also known as “leet speak”\(^{21}\), which is a colloquial term for *elite speak* used in the online community supposedly because only those who know who to interpret the alphanumeric combinations are able to read messages written in it. Generally, the encoding process draws from Cebuano phonemes to affect the lexical product. For example, the English words ‘*love love*’ would become ‘*lablab*’ following the rules of *Jejemon* given that /b/ and /v/ and their corresponding written forms are phonologically and graphically identical in Cebuano and that Cebuano lacks the mid central vowel /ʌ/ or /ə/. In this sense, it would be difficult to determine whether the word is borrowed or a CS occurrence. The methodological difficulty arises because of the visual assimilation of the word into the phonology of the Cebuano language, which is one of the criteria Myers-Scotton (1993) discusses as determining lexical borrowing and CS.

To combat this obstacle, I used surrounding language context to determine whether or not the word processed through *Jejemon* rules was CS. For example, if the lexifier for all of the words in the sentences was Cebuano, except for the appearance of ‘*lablab*’, then ‘*lablab*’ was considered part of the EL. Likewise, if the lexifier is English and ‘*lablab*’ appears, then the word was also considered part of the EL since it was marked with

\(^{21}\)This is named *1337* because the numbers represent letters. Number one represents lower case “l”. The number three represents upper case “E”. Lastly, the number seven represents upper case “T”.
regards to standard English spelling. In cases where entire sentences were written in Jejemon, I classified the tweet as CS.

4.5 Adaptation of Prior Linguists’ Coding Schemas

Much of the CS and language coding scheme is adopted from the coding schema employed by Thompson (2003:179) in his analysis of Tagalog and English CS\textsuperscript{22} in Filipino media. I adapted the schema to fit Cebuano and English CS data rather than Tagalog and English CS data. I also adapted Paolillo (2001)’s methods in investigating particular tokens pertinent to the online community in his study of Hindi and English CMC on Internet Relay Chat (IRC). The tokens I examined in the corpus consisted of matrix language structural morphemes that marked grammatical aspect in Cebuano verbs (Mag- and Nag- columns in Table 4.5), and a syncretized phrase consisting of a Cebuano and English element (Pa more column in Table 4.5). The following table shows what these affixes and the syncretized phrase in Cebuano and English CS are.

<table>
<thead>
<tr>
<th>Meaning</th>
<th>Mag-</th>
<th>Nag-</th>
<th>Pa more</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example</td>
<td>Magluto ko niini</td>
<td>Nagkita ko niya</td>
<td>Eat pa more</td>
</tr>
<tr>
<td>Gloss</td>
<td>Will-cook I this</td>
<td>Saw I this</td>
<td>Eat still more</td>
</tr>
<tr>
<td>Translation</td>
<td>I will be cooking this</td>
<td>I saw/was seeing them</td>
<td>Keep eating</td>
</tr>
</tbody>
</table>

\textbf{Table 4.5} Structural Morphemes and Syncretized CS Phrase

\textsuperscript{22} This is known as Taglish.
4.6 Post Processing

Originally, I analyzed 8,699 tweets and found significant figures with regards to the association between CS and region as well as CS and population density. However, upon reexamination of the data, I noticed that a sizeable number of tweets consisted entirely of word-like entities that could not be verified as produced by the user. These word-like entities included hashtags and other Twitter users’ names. Although CS did occur within the hashtags, I found it difficult to determine whether or not the hashtag was already well established in the social media community or novel by the user.

4.7 Analysis Methods

I analyze these data using a mixed methods approach. To find out whether or not CS frequencies across provinces and between population densities were statistically significant, I calculated chi-square scores. I also calculated T-scores as part of the T-test to statistically compare differences in tweet lengths between the code-switched and non-code-switched tweets. I chose a 95% level of confidence ($\alpha = .05$). Although statistical significance was important in supporting my thesis, a non-significant result was also useful in detecting trends in CS, which was the case with the comparison between urban and rural areas. In addition to statistical analyses, I analyze the data according to Myers-Scotton (1993)’s MLF model and extract certain syncretized collocations that appeared to be the result of CS interaction between Cebuano and English.

Throughout this investigation, computational tools have provided assistance in collection and analysis of the data. The programming language Python was used to create scripts
(i.e., short programs) that automatically collected data from each of the four locations examined in the Philippines. I also used scripts to clean and analyze the tweets to produce descriptive statistics. For further statistical analysis, I used a combination of R and Microsoft Excel to manage, sort, and analyze my data. I also used a combination of R and Excel to prepare graphs, charts, and tables that I include in this study. I created maps with the help of Microsoft Excel’s map plotting module that allows for coordinate points to be plotted onto a Bing service supported map. Additionally, I obtained maps from free online resources and with permission from Lewis et al. (2015) of the online version of the *Ethnologue*.

In the next section, I explain the results of the analysis with respect to extralinguistic factors that may have affected the frequencies of CS among these four provinces and between urban and rural areas. Then, I refer back to the socio-political and linguistic background of the Philippines in order to frame my argument. I combine extralinguistic information with the results of this analysis in order to support my thesis that Cebuano speaking areas engage in CS at varying rates, which further predicts the particular Cebuano speaking area’s alignment with Cebu City.

5. RESULTS

To conduct my analysis, I categorized each of the 7,729 tweets according to the presence or absence of CS. There were 2,296 total users who contributed to the corpus at the rate of about 3.3 tweets per user. The average tweet length overall was 11.2 words long. The longest tweet was 70 words long, and the shortest tweet was zero words long. A count of
zero is possible because I excluded emoticons and widely used interjections (e.g.: \textit{omg}, \textit{lol}, etc.), hashtags, hyperlinks, and personal references (e.g., \texttt{@}mentions), invoked by the prefix ‘\texttt{@}’ in the form \texttt{@name} from the count.

Figures used in statistical analyses assumed the exclusion of these features. In general, about 26.83\% or 2,074 tweets exhibited CS to varying degrees across all regions. This mean is derived by dividing the number of CS+ tweets (2,074) by the total number of tweets in this data set (7,729). With regards to individual words, 38.42\% or 32,877 words were involved in the CS tweets across all regions.

5.1 Code-Switching and Province

With respect to province, with the exception of Negros Oriental, individuals from each region contributed over 1,500 tweets to the corpus. Negros Oriental contributed 378 tweets in total. This relative disparity in the data from Negros Oriental could be contributed to the higher population densities of Cebu, Misamis Oriental, and Davao del Sur. Each region employed CS from 18.8\% to 31.4\% of the time. The rest of the tweets were monolingual in English, Tagalog, Cebuano, and other minority languages\textsuperscript{23}. In general, individuals from northern Cebuano speaking provinces produced about 1 to 2 fewer tweets than individuals in the southern areas.

The number of words per tweet as well as the length of words differed across the regions. There were no significant differences across the regions (p > .05). Negros Oriental and

\textsuperscript{23} Other minority languages include Hiligaynon, Spanish, Arabic and Japanese
Cebu tweeted one and two more words on average than Misamis Oriental and Davao del Sur which averaged around 10 words per tweet. Word length across all regions generally averaged to 4.5 characters per word. Davao del Sur generally tweeted shorter words than the other regions averaging 0.18 fewer characters per word.

In analyzing users per region and the frequency of CS, I found that users from Davao del Sur employed CS the most. Of the 1,602 users from this province, 31.43% code-switched, which was nearly 4% more than the next highest province. In descending order, the provinces of Misamis Oriental, Cebu, and Davao del Sur employed CS 27.44%, 23.63%, and 18.78% of the time. Comparing these rates to overall mean rate of 26.83%, we find that Davao del Sur code-switches well above the average across all regions\(^{24}\). In contrast, Cebu and Negros Oriental code-switch well below the average across the regions\(^{25}\).

To measure the probability (p-value) of an association between CS and region as opposed to no association to the regions, i.e., due to chance, I used the chi-square (\(\chi^2\)) statistical analysis technique (Table 5.1a). I ran the analysis twice comparing each province to number of code-switched tweets, and compared each province to number of users who code-switched. The first analysis showed that the differences in CS frequencies in tweets across all provinces were highly unlikely to occur by chance alone (\(\chi^2 = 84.75\), p < .001)\(^{26}\). The second iteration was done on the number of users who code-switched versus province, which also showed that the differences in users who code-switched across

---

\(^{24}\) Standard deviation of 9.12.

\(^{25}\) Standard Deviations of -6.36 and -15.97

\(^{26}\) 3 degrees of freedom
provinces was highly unlikely to occur by chance alone ($\chi^2 = 74.05$, $p < .001$). Thus, we can conclude that there appears to be an association between region and the frequency of code-switching on Twitter between Cebuano and English.

In particular, the data show that the provinces form two general groups based on the CS frequencies relative to the overall mean rate (Table 5.1b). The first group consists of the Northern provinces (henceforth Northern Group) of Cebu and Negros Oriental with a CS rate within 1 standard deviation above the overall mean rate and more than 3 standard deviations below the overall mean rate respectively. The second group consists of the Southern provinces (henceforth Southern Group) of Misamis Oriental and Davao del Sur with a CS rate of more than 3 standard deviations above the overall mean rate. These two groups engaged in CS significantly differently from each other. That is, the Southern Group code-switched well above the mean frequency. In other words, the farther south the tweet came from, the more likely it was to contain code-switched elements\(^\text{27}\). The next step involves analyzing the relationship between CS and population density.

<table>
<thead>
<tr>
<th></th>
<th>Cebu</th>
<th>Negros</th>
<th>Misamis</th>
<th>Davao del Sur</th>
<th>Row Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS+</td>
<td>714</td>
<td>810.92</td>
<td>71</td>
<td>101.43</td>
<td>494</td>
</tr>
<tr>
<td>CS-</td>
<td>2308</td>
<td>2211.08</td>
<td>307</td>
<td>276.57</td>
<td>1306</td>
</tr>
<tr>
<td>Totals</td>
<td>3022</td>
<td>378</td>
<td>1800</td>
<td>2529</td>
<td>7729</td>
</tr>
</tbody>
</table>

Table 5.1a Chi-square table. CS in Tweets per Region. ($\text{Obs.} = \text{Observed Counts}; \text{Exp.} = \text{Expected Counts}^{28}$) (Overall mean CS rate = 0.26834 [26.8%])

\(^{27}\) This accounts only for Cebuano and English CS. There were other languages that were involved in CS as well, such as Arabic, Spanish, and Tagalog. If they were included, these figures would increase.

\(^{28}\) The expected count is the hypothetical count if there was no relationship between the two variables. This figure is calculated by multiplying the row total of the observed data by the column total of the observed data and dividing the product by the total number of observations. For CS+ tweets in Cebu, the expected count is calculated by $(2074 \times 3022)/7729$ to get 810.92.
5.2 Code-Switching and Population Density

Prior to statistical analysis, I reorganized the tweets according to the density of the municipality from which they originated and I analyzed the tweets according to their population density classification. The regions studied represent two groups in the Philippines, urban and rural. Barangays, the smallest administrative division in the Philippines, are classified as urban if they have over 1,000 people per square kilometer. The urban group data consisted of 7,480 tweets and the rural data consisted of 249 tweets. The results show that the urban group code-switched 27% of the time and the rural group code-switched 21.68% of the time.

The number of words per tweet as well as the length of words used in tweets differed slightly between the two groups. On average, the urban group tweeted one extra word compared to the rural group. Tweets ranged from around 5 to 17 words in length. There was also relatively little difference with regards to word length. Both groups generally averaged about 4.4 characters per word with the rural group tweeting marginally longer words by a count of .04 characters per word.
The next chi-square analysis examines the probability of an association between CS and population density categorized according to urban or rural classification (Table 5.1). The results of this analysis showed that the differences in CS frequencies in tweets across urban and rural areas were not significant ($\chi^2 = 3.47$, df = 1, $p > .05$).²⁹

In summary, although the chi-square statistic and associated p-value show that there is likely a relationship between the provinces and the frequency of code-switching, it is still unclear whether or not there is a causal relationship between the two variables. Furthermore, the regional effect on CS is unclear. The chi-square statistic only shows that the CS frequencies with respect to region is associated, but does not give indication as to which variables are affecting each other. This is to say that it is difficult to ascertain and to figure out, for example, whether the region variable is bolstering or dampening the CS frequencies on Twitter.

<table>
<thead>
<tr>
<th></th>
<th>Rural</th>
<th>Urban</th>
<th>Row Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs.</td>
<td>Exp.</td>
<td>Obs.</td>
<td>Exp.</td>
</tr>
<tr>
<td>CS +</td>
<td>54</td>
<td>66.82</td>
<td>2,020</td>
</tr>
<tr>
<td></td>
<td>195</td>
<td>182.18</td>
<td>5,460</td>
</tr>
<tr>
<td>Col. Totals</td>
<td>249</td>
<td>7480</td>
<td>7,729</td>
</tr>
</tbody>
</table>

Table 5.2 Chi-square ($\chi^2$) contingency table of CS in Tweets by Population Density. (Obs. = Observed Counts; Exp. = Expected Counts) (Overall mean CS rate = 0.26834 [26.8%])

²⁹With 1 degree of freedom and an actual p-value of $p < .059$
5.3 Code-Switching and Tweet Length

The third research question inquires about the relationship between the length of the tweet and CS. To analyze tweet length, I split the tweets into words and counted the number of words. Words were split by white space. Then, tweets were grouped based on presence of Cebuano CS. I ran a T-Test, which showed that there was a significant difference between the two groups (one-sample t(105) = 6.7963, p < 0.001). The relationship between these two variables makes sense in that either longer tweets allow individuals more opportunities for CS, or that it allows CS to become more salient since there are more context clues as to what the ML and EL are in the tweet.

5.4 Emerging Structures

There were a couple of structures that emerged from the corpus during analysis that I discovered were distributed differently among the four provinces. These structures were the syncretized phrase “pa more”, and aspectual suffixes mag- and nag-. In general, both of these structures were used more frequently among the Southern Group than they were among the Northern Group. In other words, each of these structures was associated with areas with higher CS frequency.

5.5 Linguistic Innovation: Pa More

The syncretized phrase pa more consists of two words, which originate from both Cebuano and English. The Cebuano element pa carries the meaning of yet or still, and the English element carries the meaning of more. If the initial word is a verb, the phrase
carries a new meaning signifying continuation of an action (7). This meaning is similar to
the meaning of the English verb *keep* as in *to keep doing something*. If the initial word is
a noun, the phrase carries a meaning of continued request of the noun (9). The
syncretized phrase triggers a fixed structure, which consists of an initial word or phrase
followed by *pa more* (7) (8).

(7) Study pa more
    Study still more

*Keep studying.*

(8) Pa sign up ug Clearance pa more!
    C-PREFIX sign up ACC clearance still more!

D pa ako naka pag enroll!
    Not still I 2P N-PREFIX enroll!

*Allow me to keep signing up for clearance! I am still not able to enroll!*

(9) Kape pa more!
    Coffee still more!

*Keep more coffee coming!*

There were 19 instances of the syncretized phrase *pa more* detected in the corpus. There
were two duplicates resulting from one user retweeting or relaying the tweet in which *pa
more* occurred. I analyzed 17 of the 19 phrases after removing the duplicate tweets. The
occurrence of *pa more* across all four provinces by different users suggests that the
phrase is not unique to one province. A majority of the occurrences were from Misamis
Oriental (58.8%) (Table 5.5).

The equivalent of this phrase utilizing non-English elements is simply *pa*. There is no
equivalent syncretized phrase consisting of Cebuano and Spanish. In other words, *pa*
mas is not used. Because of the absence of the Cebuano and Spanish syncretized phrase, the occurrences of pa more suggests the adoption of a novel linguistic innovation in Cebuano that may be a direct result of Cebuano and English CS.

<table>
<thead>
<tr>
<th></th>
<th>Cebu</th>
<th>Negros Oriental</th>
<th>Misamis Oriental</th>
<th>Davao del Sur</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>3</td>
<td>1</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>Percentage</td>
<td>17.6% (0.10%)</td>
<td>5.9% (0.26%)</td>
<td>58.8% (0.56%)</td>
<td>17.6% (0.12%)</td>
</tr>
<tr>
<td></td>
<td>23.5% (0.12%)</td>
<td></td>
<td></td>
<td>76.5% (0.30%)</td>
</tr>
</tbody>
</table>

**Table 5.5** Frequency of "pa more" in corpora from each province

Figure in parenthesis (%) is the percentage of that province’s total corpus

**Graph 5.5** Percentage of each province’s tweets containing “pa more”. (C = Cebu; N = Negros Oriental; M = Misamis Oriental; D = Davao del Sur)
5.6 Detached Prefixes

There were 113 tweets containing instances of the detached Cebuano verbal prefixes *mag*- and *nag*-. In formal Cebuano writing, verbal prefixes are attached to the verb stem (10) (11). Detached prefixes occurred with both Cebuano and English verb stems within a Cebuano ML (12) (13). In 42% of the cases, individuals did not code-switch out of Cebuano after the detached prefix. In 54% of the cases, individuals code-switched into English. The remaining 4% was unaccounted for. The stems in the remainder consisted of numerals, symbols, and uncommon abbreviations or indecipherable combinations of letters\(^{30}\). In addition to detached verbal prefixes, instances of detached verbal suffixes were also observed, although not as frequent.

\[
\begin{array}{cccccc}
\text{Ang} & \text{matag} & \text{usa} & \text{nga} & \text{nagtrabaho} \\
\text{TOP} & \text{every} & \text{one} & \text{REL} & \text{I-PREFIX work} \\
\text{adunay} & \text{katungod} & \text{sa} & \text{maangayon} & \text{ug} \\
\text{exists} & \text{privilege} & \text{to} & \text{appropriate} & \text{CONJ} \\
\text{mapuslanon} & \text{nga} & \text{suhol} & \text{nga} & \text{maghatag} \\
\text{useful} & \text{ADJ-LK} & \text{wage} & \text{REL} & \text{I-PREFIX give} \\
\end{array}
\]

“Everyone who works has the right to just and favourable remuneration ensuring for himself and his family an existence worthy of human dignity, and supplemented, if necessary, by other means of social protection.” (United Nations, 2015)

\[
\begin{array}{cccccc}
\text{Aron} & \text{siya} & \text{magpuyo} & \text{nga} & \text{malinawon} \\
\text{So that} & \text{s/he} & \text{I-PREFIX live} & \text{REL} & \text{peaceful} \\
\text{Kanusa} & \text{kaha} & \text{gamma} & \text{mag} & \text{remake} \\
\text{When} & \text{IRR} & \text{3P} & \text{I-PREFIX} & \text{remake} \\
\text{sa} & \text{It} \\
\end{array}
\]

\(^{30}\) This is possibly due to Jejemon, which has an objective to obfuscate written language as much as possible
ACC  It

*When are they going to remake “It”?*

(13)  

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ayay</td>
<td>dugay</td>
<td>kaayo</td>
<td>mag</td>
<td>post</td>
</tr>
<tr>
<td>INTJ</td>
<td>long</td>
<td>very</td>
<td>I-PREFIX</td>
<td>post</td>
</tr>
</tbody>
</table>

bay  !!
friend  PUNC

*Jeez! It’s taking so long to post buddy!!*

There were between 1 to 33 tweets containing each prefix from each area, which shows that there are differences in written language patterns between the Northern Group and Southern Group. Multiple users produced the split-prefix in their tweets, and considering that the tweets were gathered over a period of three months, could point to a generalized difference in at least electronically written Cebuano.

Out of the two prefixes examined, there seemed to be general agreement in the distribution in Cebu, Misamis Oriental, and Davao. Although this may be a frequency distribution of the appearance of the prefix itself in the language, the written language data corroborates that this data shows a trend.
5.7 Lexical Observations

Lastly, there were several other features that I observed occurring throughout the corpus that showcased CS and genre-specific innovations. These features include phrasal verb insertions, lexical category changes through CS, and orthographic behaviors pertinent to written media. These orthographic behaviors consisted of utilizing abbreviations, 1337 speak, and Jejemon in tweets.

5.8 Verb Stems

Another pattern that emerged across the regions with regards to code-switched verbs was that of the multi-word verb stem, which does not occur in Cebuano. For example, *to look up* meaning *to search* occurs as the single morpheme *tan-aw* or *ngita* in Cebuano. Multi-
word verb stems from English were switched into wholesale rather than in parts. The corpus provided evidence of these verbs produced as one word (15) and as separate words (14).

(14) Dli ko ka log out
NEG 1S P-PREFIX log out
aq sa aqng messenger acct.
1S from my messenger account

I was not able to log out of my messenger account. Tssss.

(15) @NAME pa followback [emoticon] please
@NAME M-PREFIX follow back EMO please
bitaw ni. schoolmate biya ta.
EMP this. schoolmates supposed IP-INCL
hehe! #teknoy
hehe! #teknoy

@NAME Follow me back. :) Please do this. We are supposed to be classmates. Hehe! #teknoy :)

5.9 Changes in Lexical Category

Words that functioned as verbs in code-switched Cebuano tweets sometimes belonged to other lexical categories in English. For example, nouns and adjectives were often utilized as verbs in Cebuano. Unlike English, which modifies lexical items through morphological and phonological changes in order to create a verb, Cebuano does not. In other words, English words embedded in Cebuano frames were mostly inserted in stem form. This phenomenon is demonstrated in (16) and (17) with Nov. 1 and beauty rest behaving as verbs or verbal nouns within the Cebuano ML yet originating as nouns in
English. Currently, there are no studies that investigate this phenomenon in Cebuano and English CS.

(16) ddto mi mag Nov. 1 sa bukidnon (:)
there 1P I-PREFIX November 1st at mountains EMO

We are going to go to the mountains on November 1st. :(

(17) @NAME yeah sure mas nindot if di
@NAME yeah sure more nice if NEG

sabutan pag beauty rest nalang oy kay
understood N-PREFIX beauty rest just INTJ because

di man ka maka jog lovelots muwah
NEG EMP 2S C-PREFIX jog lots of love muwah

@NAME yeah, sure. It'd be nicer if you didn't know. Just get some beauty rest man! Because you won't be able to jog. Lots of love. Muwah.

6. DISCUSSION

These findings show that the members of the Cebuano speaking community are diverging from each other with regards to their engagement in CS on Twitter. The data show that there is a strong relationship between CS and province, population density, and tweet length. The Northern Group engages in far less CS than the Southern Group. Regarding CS and population density, urban areas engage in more CS than rural areas. Lastly, these data show that there is a relationship between CS and tweet length among the tweets that exhibited CS. The following section further discusses reasons for these relationships given the evidence supported by the data.
6.1 Code-Switching and Province

The findings showed that the four provinces code-switched at significantly different rates from one another (Table 5.1). Furthermore, two groups emerged from the data, showing that the Southern Group code-switched more than the Northern Group. I propose two reasons to account for the disparity in CS frequency among these provinces and between these groups. The first reason may be that the Cebuano speaking community in Cebu exerts a strong influence over the surrounding area economically, historically, religiously, and socially. Cebu’s influence is partly driven by a resilient pro-Cebuano sentiment that prevails in the eponymous city and province. The second reason is that, Misamis Oriental and Davao del Sur are under pressure to dissociate themselves from Cebuano speakers in the Northern Group, who originally settled the islands about a century ago, nearly completely displacing the local indigenous ethnic and linguistic groups on Mindanao. Additionally, the formation of these local identities is driven by competition on the national stage. For example, the southern provinces are in direct competition developmentally and economically with Cebu, which may be driving the divergence in language patterns\textsuperscript{31}.

6.1.1 The Northern Group

Lower CS frequency in the Northern Group can be attributed to the Cebu and Cebu City’s strong influence on the surrounding area. With nicknames such as “The Queen City of the South” and “The First Capital”, Cebu City’s present and historical

\textsuperscript{31} Related to John Baugh (2003)’s findings on linguistic profiling and housing discrimination. Cagayan de Oro, the largest city in Misamis Oriental ranks #2 on the most competitive cities with regards to development (source: \url{http://www.competitive.org.ph/cmcindex/}).
significance is palpable. Cebu City and province of Cebu are historically tied to the spread of the Cebuano language as the lingua franca for the region, and known for its resistance to Tagalization (Yraola-Westfall, 1991). The spread of Cebuano was so effective that extinction of smaller minority languages in these regions have been attributed to the spread of the language (Thompson, 2003).

In addition to the strong resistance to Tagalization, there is also a strong pro-Cebuano sentiment in Cebu, which is evident by the formation of groups such as the *Lubas sa Dagang Bisaya* (LuDaBi) in 1959 and the Cebuano Studies Center in 1975, whose goals are to maintain and promote the Cebuano language. These characteristics are part and parcel of the construction of the Cebuano identity, which holds prestige in the surrounding area. Therefore, surrounding areas strive to be more like Cebu City, which is evident in the way that Negros Oriental engages in CS less than the Southern Group. Negros Oriental may be associating so strongly with Cebuano identity that it engages in CS far less than even Cebu itself (Map 6.1.1).

Negros Oriental is grouped with Cebu and exhibits less CS than Cebu. I argue that its smaller population and location in between Cebu and Negros Oriental, a non-Cebuano speaking province, may actually be driving its population to exhibit more pure Cebuano language on Twitter if Cebuano is used. Lower CS frequencies in the province could be attributed to smaller close-knit communities on the island, which has a low population density. These close-knit communities are typically more resistant to linguistic innovations, such as CS. In short, Negros Oriental’s sparse population and proximity to

---

32 The group’s goals are to protect, maintain, and promote the language.
other linguistic and ethnic groups may be the impetus for the province’s tendency to express less CS than Cebu and the Southern Group.

![Map 6.1.1](image)

**Map 6.1.1** The Northern Group. Shown: Cebu City (left) and Dumaguete City (right). 

\([CS+] = \text{Blue}; [CS-] = \text{Red}\)

### 6.1.2 The Southern Group

Higher CS frequency in the Southern Group can be attributed to the group falling outside of Cebu’s sphere of influence, and active motivation to compete directly with Cebu. Misamis Oriental and Davao del Sur are situated farther away from Cebu than Negros Oriental. These two cities experience different economic, historical, religious, and social influences than the Northern Group.

There are many minority ethnic groups still present on Mindanao who are descendants of the indigenous peoples on the island before the arrival of Cebuanos. A couple of these indigenous peoples include the Moros and the Lumads, who were in the majority on
Mindanao before resettlement of the island by Cebuanos in 1913 (Dai & Tai, 1974). These groups are now in the minority. Additionally, the island is home to several autonomous states and provinces who are known for their strong separatist movements, and even terrorist connections (NCC, 2014; U.S. Department of State, 2014).

Provinces in the southwestern part of Mindanao are home to non-state armed groups who use force to voice separatist and autonomous sentiments. Some of these groups are listed as terrorist organizations by the Philippines, United States, and the Geneva Academy of International Humanitarian Law and Human Rights (Banlaoi, 2012; RULAC, 2012; U.S. Department of State, 2014). These groups inhabit the western and southwestern parts of Mindanao and include groups such as the Moro National Liberation Front (MNLF), Moro Islamic Liberation Front (MILF), and Abu Sayyaf Group (ASG) (Banlaoi, 2012; RULAC, 2012; United States Institute for Peace, 2011; 2014; U.S. Department of State, 2014). The effect of these groups and the subsequent Arabic language influences from the groups, related to their Islamic religious affiliations, can be seen in the corpus from Mindanao. In the following excerpt, *kutob* (books) is utilized (18). *Kutob* is the plural form of *kitaab* (book) in Arabic. Normally, Cebuano uses *libro* to denote books. The plural is formed with the particle *mga* as in * mga libro* (books).

---

33 According to the Geneva Academy of International Humanitarian Law and Human Rights organization, this group was dormant for a while.
There are just 3 books, alright? Just take care of [them] if you are not going to answer me, @NAME.

The dynamic between Cebuano settlers and the minority groups in Mindanao may be the reason for language changes in Cebuano, which further affects the CS frequency in the Southern Group. The Cebuano speech community may be engaging in more CS in order to appear less Cebuano linguistically and ethnically (Map 6.1.2). This phenomenon is similar to the way that Tagalog and English speakers move away from purely Tagalog language and code-switch in order to appear less tied to the Tagalog ethnicity, which may have a negative nationalistic connotation (Thompson, 2003; Bautista, 2004). Therefore, the impetus to diverge from the Cebuano identity, which the Northern Group converges toward, is driven by a need to assimilate with preexisting groups on the island.

Although the difference in CS frequencies between the groups may differ by less than 10%, I propose that CS in spoken Cebuano could be much greater than this figure shows, since English is predominantly the default language used. Additionally, there is a salient increased use of Tagalog words in otherwise Cebuano tweets in the Southern provinces. In these instances, Tagalog words take on structural morpheme roles rather than solely providing content morpheme words, which are typical of cultural loan words.
In summary, differences in CS among the four regions are motivated by the need to associate or dissociate with the Cebuano identity. In the Northern Group, Cebuano identity enjoys local prestige. Association with that identity may allow speakers of Cebuano more economic success because of Cebu’s strong influence in the area. On the other hand, in the southern provinces, Cebuano identity may construct an unfavorable image for the individual since Cebuano communities displaced indigenous groups on Mindanao, many of which are proponents of secession and autonomy from the Philippines. In other words, antagonism towards the Philippines, who allowed Cebuano resettlement on Mindanao, may be driving Cebuano speakers to dissociate and diverge from the Cebuano identity that is embraced in the Northern Group through CS.

The next section explores reasons to explain the relationship between CS and population density. That is, whether or not urban or rural areas employ CS differently from each other. Unlike the relationship between CS and province, the results of the chi-square analysis found a non-statistically significant yet strong relationship between CS and population density.
6.2 Code-Switching and Density

Results of the chi-square analysis comparing CS patterns and population density were not statistically significant, but still showed a very strong association ($\chi^2 = 3.3085$, df = 1, p = .068). The data showed that individuals living in urban areas were more likely to code-switch than those living in rural areas. I reason that these differences in CS frequency between these two areas can be attributed to: social network strengths, and the subsequent acceptance of the linguistic innovation and standardization of CS.

The disparity between rural and urban CS frequencies can be attributed to the different kinds of networks that each area fosters. Rural areas foster more close-knit networks, whereas urban areas foster loose-knit networks. Based on this reason, CS is a linguistic innovation that is the standard variant of Cebuano utilized on Twitter. The data reflect

Map 6.1.2 The Southern Group. Shown: Davao City (left) and Cagayan de Oro (right). ([CS+] = Blue; [CS –] = Red)
this in that urban areas employ CS at a slightly higher rate (27.00%) than rural areas, which employ CS at a lower rate (21.69%).

Increased opportunities for language contact with English also play a role in the standardization and employment of CS in urban areas. Urban areas house many powerful public and private institutions that strongly influence the type of language by interlocutors interacting with them. For example, universities, government organizations, international business, and tourist areas are typically concentrated in urban areas, where English is spoken as the primary language of communication. English is preferred over Tagalog in Cebuano speaking areas in situations that call for either language. English use is particularly evident in the websites of such organizations. Even the official municipal websites for cities in these four provinces are in English rather than Cebuano or Filipino.

In contrast, rural areas contain fewer of these types of institutions, and are less exposed to English and less affected by language contact. Moreover, close-knit ties associated with rural areas prevent the adoption of CS as a standardized form. Therefore, they are less likely to adopt CS as a linguistic innovation or employ CS as a standardized variety of Cebuano.

In summary, urban areas’ exposure to English through more diverse contact and institutional language policies allow maintained language contact between Cebuano and English, which is conducive to CS. Furthermore, individuals propagate code-switched

---

34 These languages are typically English and, less frequently, Filipino or Cebuano.
language as the standardized form of the language evident through the CS rate in urban areas approximating the overall mean frequency. Rural areas, on the other hand are more under the influence of close-knit ties, which resist standardization and innovation of language.

6.3 Code-switching and Tweet Length

Lastly, the analysis shows that there is a statistically significant relationship and strong positive association between tweet length and CS. Longer tweets contain more CS elements than shorter tweets. The data do not provide causal evidence demonstrating one variable effecting the other. However, because CS is a function on language rather than a function on the length of the utterance, written or spoken, I argue that longer tweets facilitate CS. I provide two reasons to explain this relationship: that longer tweets have more room for words thereby providing support for CS to occur, and that longer tweets provide more context for the reader to recognize a CS element with an ML and EL or provide the author ample room to express CS in writing.

With regards to longer tweets allowing more opportunities for CS to occur, certain structural forms involving CS required minimum tweet lengths. I present *pa more* as an example of these limitations. The collocation *pa more* is a two word code-switched unit used across all provinces. Syntactically, this collocation generally calls for a predicate in the first position of the tweet. The predicate may be a word or phrase, but requires the individual to construct a tweet of at least 3 words.  

35 Hashtags (#) and personal references (@name) were not included.
Similarly, I proposed that longer tweets allow more context for the reader to recognize CS structures in the tweets. Due to stylistic differences and restrictions in the genre, tweets in the corpus did not typically follow conversational practices such as turn-taking. Therefore, there was insufficient discursive context to determine whether or not monolingual tweets were actually part of a string of code-switched tweets. In a conversational mode, the ML could be determined from the number of phrases uttered in language A or B (Myers-Scotton, 1993).

These complications affect both of the reasons I proposed explaining the relationship between CS and tweet length. Complications arise when Jejemon are involved in the production of a tweet. Jejemon are a result of lexical compression that results from the combination of multiple words. In the corpus, a large number of users incorporated many Jejemon and combined abbreviations into their tweets, which caused the word count to decrease. Because of this compression, tweet lengths could have generally been longer than accounted for, which further bolsters the argument that tweet length and CS are associated.

6.4 Limitations

In assessing CS on Twitter, there were several limitations stemming from the genre in which the language data appeared. The nature of the Twitter platform forces users to produce language within a very limited space. Therefore, CS occurring at intrasentential levels was difficult to detect and CS occurring at conversational levels were impossible
given the method in which the data were collected. Furthermore, the genre is not conducive to long discourse and therefore, discursive CS features were not captured.

In addition to the limitations set by the genre of the language data, stylistic choices made by users, such as *Jejemon*, added a level of complexity and difficulty in the analysis of CS in the corpus. This complexity resulted in many tokens encountered in the corpus with insufficient context whereby I could decipher, interpret, or translate the content. *Jejemon* are inherently difficult to decipher, but potentially contain useful information with regards to CS.

Non-word tokens also posed as limiting factors in that, at times, their presence outnumbered the presence of language specific tokens. Their presence would separate unitary sentences at critical points where CS could occur, which is especially true with hashtags. Occasionally, users implemented CS as part of sentences and phrases in their hashtags. Because I uniformly excluded hashtags, their presence may have interfered with word counts and CS detection. This interference also posed problematic in analyzing CS in this corpus. Also, there were a handful of times when tweets consisted entirely of tokens that were not representative of language production of the user. This included hashtags, @mentions, hyper-links, and emoticons.

Lastly, because certain language structures and lexical items in Cebuano and Tagalog shared much similitude with each other, short tweets were more difficult to classify if at all possible.
6.5 Suggestions for future research

Throughout the process, the process of seeking answers to address my research questions produced many more questions than this thesis originally set out to answer. Thus, there are several topics for research I find are suitable for further investigation. These topics relate to the genre of CMC and are unique to Cebuano and English CS since this particular subfield in the field of CS is understudied. These topics deal with CS as it effects language change. Specifically, these research suggestions focus on studying longitudinal changes, exploring more syncretized forms, reexamining CS grammars, and investigated Jejemon as a forge for lexical change and a source of neologisms.

The Internet is a rapidly changing environment that introduces new types of interactions and novel ways of making contact since the advent of social media. This thesis uncovered changes in CS patterns in the Philippines in social media, but captured only a single moment in time. A longitudinal study of language change with regards to CS in Cebuano and English may be able to provide further insight as to whether or not the divergence in the speech community is rapid, slow, or moving in a different direction.

As a result of CS and language change, the syncretized form "pa more" was found to have a skewed distribution across the provinces in the Philippines. I suggest that a further exploration of syncretized forms would provide insight to the extent of language change that language contact may induce. Because social media is especially available for public access, informal, and likely to produce language simulating oral speech, a study on
syncretized forms on CS in Cebuano and English in CMC may illuminate patterns of language change.

In conjunction to exploring syncretized forms, I suggest a further investigation of *Jejemon*. Findings thereof may lay the groundwork for CS researchers to further investigate the phenomenon of expressing phonological changes in oral and, subsequently, written language. That is, individual’s intentional application of multiple phonological systems evident through spelling changes could illuminate the relationship between oral language and informal written language.

Lastly, the ready availability of written language as produced by native speakers is ripe for further investigations into CS grammar. Because of the productive nature of affixes in Cebuano, individuals extensively use them when embedding English language elements into their messages, oftentimes separating them from English stems. Because of this seemingly unintentional application of the ML over the EL, grammar studies can benefit from natively produced CS. Moreover, because the language is written, transcription errors may be reduced by extracting language directly from the written form as produced by native speakers.

I leave these suggestions for future researchers to investigate. The results of these further forays into CS can potentially improve our understanding of the definition of language through better understanding language change as it occurs in real time.
7. CONCLUSION

On a superficial level, this study exposes that the Cebuano speech community utilize CS differently. Moreover, the study reveals that there is a divergence in the Cebuano speaking community in the Central and Southern Philippines. The study also further reinforces other research in CS with regards to identity, CMC, CS grammar, and approximation of CMC to oral speech. Furthermore, this study exemplifies the benefits of employing mixed methods in sociolinguistic work as well as highlights the power of employing programming and scripting languages in collecting quality CMC data directly from speakers instantaneously.

This investigation was successful in addressing the research questions I pose in the introduction. With regards to the first research question, which examines the heterogeneity or homogeneity of CS frequency across the four Cebuano speaking provinces, each province code-switches at significantly different rates from one another. Moreover, two groups emerge from the data, which exhibit CS tendencies alluding to the degree of their adoption of Cebuano identity. The Northern Group of Cebu and Negros Oriental are significantly more conservative in CS possibly due to strong pro-Cebuano sentiments and the tendency to associate with Cebuano identity. In contrast, the Southern Group of Misamis Oriental and Davao del Sur are significantly less conservative in CS possibly due to aversion to strong pro-Cebuano sentiment and tendency to dissociate from Cebuano identity. Within the Northern Group, Negros Oriental is considerably more conservative and employs CS less frequently than Cebu, the epicenter of Cebuano identity. I posit that this conservatism could be driven by socioeconomic reasons due to
the province’s proximity to Cebu and the effect that language can have one’s image as more Cebuano, which can positively or negatively affect the types of economic opportunities they are able to attain.

The second research question examines homogeneity of CS frequency in different population densities, urban areas code-switch more frequently than rural areas, which are considerably more conservative in adopting CS as a linguistic innovation. This is expected, and relates to Milroy & Muysken (1995)'s argument that close-knit networks, which are typically found in rural settings, are less prone to linguistic innovations. Moreover, the findings pertinent to the second question support the proposition that CS is a standard variant of Cebuano spoken among these areas, reasoning that loose-knit networks, which are typically found in urban settings, are more prone to adopting linguistic innovations, such as CS.

Lastly, the third research question inquires about the relationship between tweet length and CS frequency, longer tweets are more likely to exhibit CS than shorter tweets. This finding suggests that CS structures in written language are more salient with more textual context. Although both short and long tweets exhibited CS, the CS structure *pa more*, a syncretized CS phrase, were prevalent in shorter tweets. Longer tweets provide more opportunities and physical space for CS to emerge.

In addressing these research questions, my thesis, stating that CS patterns on CMC in Cebuano speaking communities differ depending on the relationship between the
individual and the Cebuano identity, is supported. Additionally, it directly provides evidence against the Erwin-Billones (2012)’s doubts that CS would be present in print in Mindanao, since Cebuano and English CS is clearly a standard variety on the island. These findings are clear in showing a divergence among individuals within the Cebuano speaking community. In addition to addressing my research questions, the data reinforce other CS research as it deals with identity, CMC, CS grammar, and approximation of CMC to oral speech.

CS is related to identity whether it be static (Messing, 2007; Sankoff & Poplack, 1981; Gross, 2006) or fluid (Weston, 2013; Mishra & Rahman, 2013; Thompson, 2003; Bautista, 2004). As I illustrated in the discussion section of this thesis, there is a strong Cebuano identity tied to conservative usage of the language, similar to strong Tagalog identity tied to conservative Tagalog usage (Bautista, 2004; Thompson, 2003). Individuals in the Cebuano speaking community adopted or avoided this identity by varying degrees through the expression of CS in CMC. By applying theoretical frameworks connecting standard and vernacular language, and identity by Milroy & Muysken (1995), Gumperz (1982), and Bullock & Toribio (2009), identification through CS reinforces research conducted by Weston (2013), Eppler (2010), Messing (2007), and Bullock & Toribio (2009). Additionally, in conjunction with demographic data provided by the Philippine Statistics Authority, the high presence of stylized forms of written language speak to Tagliamonte & Denis (2008)’s and Crystal (2008)'s propositions that standardized written language is exhibited more as the individual gets older. Lastly, with respect to the genre of Twitter, this study echoes Marley (2011) and Delamotte &
Desoutter (2011)'s studies that both discuss the increased presence of CS in informal CMC as well as the verisimilitude of written and oral speech in informal CMC.

What we learn from the results of this investigation forces us to rethink the definition and boundaries of language. In particular, CS brings us to think about the fluidity of language and its relationship to our fluid identities. Moreover, the phenomenon has us question whether or not languages themselves can be thought of as discrete entities as opposed to entities enclosed by blurred borders. Considering the dynamic relationship between CS and community in the Philippines, we find that languages and the identities they index are in continual flux as speech communities form, expand, diverge, and reform anew.
REFERENCES

N=2014085599&site=ehost-live&scope=site;


Asia’s Finest Forum (2010). What’s up with Tagalog hating on Cebu’s development?


Psychological, Computational and Theoretical Perspectives. Cambridge: Cambridge University Press, 190 – 204


Mander, J. (2014). As the Internet turns 25, China has 2.5 times more users than US. Retrieved from http://www.globalwebindex.net/blog/internet-turns-25


http://www.csun.edu/~lan56728/majorlanguages.htm


Philippine Statistics Authority (2013). “Cebu’s Population had Increased by Nearly Half A Million (Results from the 2010 Census of Population and Housing)”. Retrieved from [http://www.psa.gov.ph/content/cebu’s-population-had-increased-nearly-half-million-results-2010-census-population-and](http://www.psa.gov.ph/content/cebu’s-population-had-increased-nearly-half-million-results-2010-census-population-and)


http://search.proquest.com/docview/1460991126?accountid=11091


http://forum.wordreference.com/showthread.php?t=2187589

Yahoo! Answers (2014). How many dialects do the PHILIPPINES have? Retrieved from

https://answers.yahoo.com/question/index?qid=20080217053727AAA0gtV


http://search.proquest.com/docview/85641442?accountid=11091


http://books.google.com/books/about/Tagalization_of_Spanish_and_English.html?id=54yfnQEACAAJ