Cognizer Verbs and the First-Person Pronoun Yo in Contact: A Corpus Study

Proceedings of the Third Linguistics Conference at the University of Georgia
Stacey M. Jacobson
University of Alabama

Abstract. Recent research has investigated the optionality of first-person singular subject pronoun usage in pro-drop languages, where the morphology of the language encodes enough information to allow speakers to choose to include the pronoun (overt subject usage) or omit it (null subject usage), referencing the idea that the subject pronoun still exists in the syntax of the sentence, regardless of its expression. Previous studies have shown significant differences in overt and null subject usage in Peninsular Spanish (Posio 2011) and between Peninsula Spanish and Portuguese (Posio 2013). However, few studies have utilized bilingual speakers, nor have multiple Spanish-speaking countries been investigated. The present study analyzed data from two corpora (Spanish in Texas and PRESEEA) to investigate the use of these first-person singular subjects in a bilingual community in the US compared to Spanish-speaking countries. Four Cognizer verbs were analyzed (Posio 2011): creer [to believe], pensar [to think], entender [to understand], and recordar [to remember]. Significant differences between overt and null subjects were found within and between the two corpora, and between some of the verbs investigated. The bilingual community was also significantly different from the nearest available Spanish-speaking city. Effects of contact with English and differences between Spanish-speaking countries are discussed.

0. Introduction

Much investigation has been conducted on the syntax, semantics, and pragmatics of subject pronoun usage in Romance languages, and particularly those that allow null subjects (Kučerová 2014; Mayol 2010; Posio 2011, 2013; Travis & Torres Cacoullos 2012). Some studies have even attempted to discern exactly what causes this kind of variation. Amaral and Schwenter (2005), for example, group their findings on the optionality of subject pronouns by words or phrases that cause variability in subject expression, mainly adverbial phrases. Similarly, Posio (2011, 2013) attempt to use the semantic category of the verb itself to determine overt and null subject usage. However, there is less investigation of this phenomenon in relation to language contact; few studies have investigated pro-drop in bilingual communities (Torres Cacoullos & Travis 2011; Giancaspro 2015; Shin & Montes-Alcalá 2014). Of these studies, the manner in which they have investigated null and overt subjects is also relatively unexplored; these studies are often limited to the presence or absence of the subject pronoun itself (Torres Cacoullos & Travis 2011), attempting to discern where grammatical switches can occur (Giancaspro 2015), or to investigating the influence of the languages on each other (Torres Cacoullos & Travis 2011; Shin & Montes-Alcalá 2014). The present study explores the idea of null versus overt subjects by investigating four Cognizer verbs (creer [to believe], pensar [to think], entender [to understand], and recordar [to remember]) and their influence on the necessity of subject pronouns, as classified by Enríquez (1984) (as cited in Posio 2011). Moreover, the phenomenon of language contact is explored through the comparison of null and overt subject usage in a bilingual community in comparison to data from Spanish-speaking countries.
1. Literature Review

Much investigation has been conducted in recent years focusing on the use of null and overt subject pronouns in languages in which pronoun use is encoded morphologically and thus subject pronoun usage is not obligatory. Romance languages are typically investigated in these studies, as most languages in this language family have retained enough verbal inflection to permit the optionality of these pronouns. Amaral and Schwenter (2005) investigated the role of semantics in subject pronoun usage in Spanish and European Portuguese through the analysis of adverbial phrases that have information about who they refer to encoded within them, such as por mi parte, which actually contains a first-person pronoun, or honestamente, which they argue is semantically first person because no one but the speaker can know what he or she thinks for 100 percent certainty. They found that previous research focused on these pronouns in contrastive contexts, i.e., when a switch of reference occurred, but that this is not the only instance in which this variation can occur, as demonstrated by the aforementioned examples.

More recent studies have expanded this idea from adverbs to the semantic meaning of the verbs themselves. Posio (2011) investigated different semantic classifications of verbs (as proposed by Enríquez 1984) and their tendencies in overt or null subject usage by using a corpus of Peninsular Spanish conversational data as well as audio recordings from television or radio (Corpus de Referencia del Español Actual). Though Posio (2011) utilized a corpus, only 100 data points were analyzed for the frequency of null or overt subject; this is a limitation in that with so few tokens (for a corpus study) it might not reflect what is actually occurring in the general population of speakers. Posio (2011) also considers the idea of contrast mentioned in Amaral and Schwenter (2005) in that his findings confirm what Amaral and Schwenter (2005) found with adverbs and similar phrases: when there is a contrast or change in subject, it is marked with a subject pronoun.

Subsequently, Posio (2013) compared Peninsular Spanish with European Portuguese, the latter using more overt pronouns than the former. He found that the frequency of overt subjects between the two languages was significantly different, with Spanish having a lower frequency. Additionally, Cognizer verbs contributed to the largest portion of the data in Spanish, whereas Communicator verbs were the most frequent in European Portuguese. The most frequent of these was yo creo, a finding that was consistent with that of Travis and Torres Cacoullos (2012). Posio (2013) found that Cognizer verbs in Peninsular Spanish demonstrated a tendency toward overt subject usage. He also found a significant difference in overt versus null subject pronoun usage when there was a switch in referent; this was more prevalent in European Portuguese, but statistically significant in Peninsular Spanish as well. This prevalence of Cognizer verbs and the precedence of their preference of overt subjects (according to Posio 2013) informed the selection of this category of verbs for the present study.

Little research has been conducted on this phenomenon in terms of investigating language contact generally. However, recent studies have investigated this phenomenon within certain bilingual communities. Utilizing data from the New Mexico Spanish-English Bilingual Corpus, Torres Cacoullos and Travis (2011) found that the use of null and overt subjects was similar to varieties of Spanish which did not have contact with English, thus negating the idea of convergence (where one language adapts to the characteristics of another language). Interestingly, they also looked at how code-switching affected their data and found that when a subject was previously expressed, especially the English first person singular subject pronoun I, the speakers were more likely to include the overt pronoun in Spanish. In contrast to their first conclusion, this analysis of the data does demonstrate some convergence, as when the two
languages were used together, there were more instances of yo, mirroring English where the subject pronoun is obligatory. Another study that investigated the phenomenon of convergence using the New Mexico corpus was conducted by Benevento and Dietrich (2015), who found that, though the expectation would be convergence, the use of yo after the verb, in constructions such as digo yo, fell within the established range of use by monolingual speakers; thus, English was not affecting Spanish word order in this community in which English and Spanish are in constant contact. The expectation is that when there is contact, the more flexible language will become less flexible; because English has much less inflectional morphology than Spanish, speakers’ use of Spanish would change to mirror English more. These studies have great implications for the current study in that in a bilingual community one might expect to find convergence between the two languages (as one potential outcome of contact).

2. Research Questions & Hypotheses
The present study aims to investigate whether there is a difference in the use of null subjects in Spanish in a bilingual setting. In particular, it is of interest to this study to determine what occurs when Spanish, which is a pro-drop language, is in contact with a language, such as English, where overt subjects (including pronouns) are obligatory. Additionally, this study also aims to discover whether there are differences in subject pronoun usage within the bilingual community as well as the Spanish-speaking countries.

The null hypothesis (H₀) for this study is that there is no difference between null and overt subject pronoun usage pertaining to Cognizer verbs between these two communities. Alternatively, there are three hypotheses being tested in this study. First, (H₁) there will be a significant difference between the uses of pronouns with these verbs between the two groups when yo immediately precedes the verb; more specifically, there will be convergence in the bilingual community, in that Spanish will become more English-like and show more usage of overt subjects. Second, (H₂) there will be a significant difference no matter the position of the pronoun; and third, (H₃) within each corpus there will be a significant difference between null and overt subject usage prior to Cognizer verbs.

3. Methodology
3.1 Materials
Two corpora have been selected due to the populations their data were collected from. The Spanish in Texas (SiT) corpus provides information about the use of Spanish and English in bilingual communities throughout Texas (Bullock & Toribio 2010-2014). It is considered a bilingual corpus, as the speakers often switch from English to Spanish as they spoke. They were not prompted to do this; thus, the inclusion of code-switching demonstrates the manner in which both English and Spanish are used in Texas. The PRESEEA (Proyecto para el estudio sociolingüístico del español de España y de América) corpus is a sociolinguistic corpus comprised of interviews with native Spanish-speakers in many Spanish-speaking countries (PRESEEA 2014-). In addition to being analyzed in its own right for null and overt subject pronoun usage in Spanish-speaking countries, this corpus will also be used as a comparison tool by providing a baseline by which language change can be measured. Data on null and overt subject pronoun usage in Spanish-speaking countries will help to determine how and where subject pronoun usage is changing in communities whose languages are in contact with English. The countries whose cities were included in the corpus at the time of data retrieval and analysis were Spain, Colombia, Venezuela, Cuba, Mexico, Chile, Perú, and Uruguay.
3.2 Procedure
This study focused on comparisons between two corpora in order to discern differences in the inclusion or omission of the first person singular subject *yo* in reference to a specific classification of verbs, classified as verbs of cognition, or simply Cognizers. This classification of verbs is defined as involving a human, typically as the subject (Verdaguer 2010:666), can be classed as ‘Mental Process’ according to the ADESSE classification system (García-Miguel & Vaamonde 2010:135), often encompass semantic roles such as Thinker, Believer, Knower, and Presumer (Van Valin 1999:374), and include verbs such as “saber ‘know’, creer ‘believe’, pensar ‘think’, recordar ‘remember’, entender ‘understand’, olivdar ‘forget’, etc” among a total of 69 verbs classified as cognition in ADESSE (García-Miguel & Comesaña 2004:371). In the present study, three of the verbs that are focused on in this study are in the classification of Cognizer verbs as reported in Enríquez (1984), as verbs that have the semantic meaning of speaker opinion or judgment: *creer* [to believe], *pensar* [to think], and *entender* [to understand] (as cited in Posio 2011). The fourth verb (*recordar* [to remember]) maintains the same semantic classification, but has not been investigated much in previous research.

After collection, a concordance analysis was conducted on the data from each group. This concordance collected the frequency of *yo* usage two words prior to the verb (1), one word prior to the verb (2), and one word after the verb (3), to account for the most frequent locations of *yo*.

(1) yo no creo
I *don’t believe*

(2) yo creo que sí
Ibelieve that *yes*

(3) es lo que creo yo
is it that believe I

All other instances were considered as null because they have no instance of *yo* before or after the verb. Though infrequent, if instances were found of *yo* three words prior to the verb, as in *yo no lo creo* [I don’t believe it], for example, they were included in the analysis. After completing the concordance analysis, the null and overt subject pronouns were tabulated and Chi-squared tests were performed to determine if there are significant differences between the following pairings:

(4) Within the same corpus by specific verb (e.g., Spanish in Texas *yo creo* compared to Spanish in Texas *creo*).

(5) Across corpora by specific verb (e.g., Spanish in Texas (*yo*) *creo* compared to PRESEEA (*yo*) *creo*).

(6) Within the same corpus by Cognizer category (e.g., Spanish in Texas *yo* {all verbs} and (null) {all verbs} compared to expected values).

(7) Across corpora by Cognizer category (e.g., all Spanish in Texas *yo* and (null) compared to all PRESEEA *yo* and (null)).

This analysis will help to determine where there are differences in pronoun usage as well as whether the corpora as a whole are different from each other.
4. Results and Discussion

4.1 Across Corpora

Overall, the results indicate that there is a significant difference in the way that overt and null pronouns are used when Spanish is in contact with English. Specifically, in the Spanish in Texas corpus, when considering only instances where, if included, the subject pronoun immediately precedes the verb, there are 646 instances of overt subject usage and 871 instances of null subject, whereas in the PRESEEA corpus, there are 1420 instances of overt subject usage and 1575 instances of null subject when considering all four Cognizer verbs together. While both corpora demonstrate a tendency to favor null subjects, there is a more pronounced difference in the Spanish in Texas corpus. This suggests that contact with English is playing a role in making the Spanish spoken in this bilingual community different from that spoken in predominantly Spanish-speaking communities.

Chi-squared Goodness of Fit tests were performed in order to determine whether there was a significant difference between overt and null subject usage in general, while Chi-squared tests were utilized to determine whether the two corpora were significantly different from each other, both by specific verb and by overall verb semantic category (Cognizer). The results of each of these tests can be seen in Tables 1 and 2 below; these tables also include the results for each individual corpus, which will be discussed in later sections. Table 1 includes the results of the Chi-Squared calculations for null and overt subject usage only considering when yo occurred one word prior to the verb, as in yo creo, but not tokens such as yo no creo.

<table>
<thead>
<tr>
<th>Verb</th>
<th>PRESEEA</th>
<th>SiT</th>
<th>SiT v. PRESEEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>(yo) creo</td>
<td>*</td>
<td>*</td>
<td>**</td>
</tr>
<tr>
<td>(yo) pienso</td>
<td>nsd</td>
<td>nsd</td>
<td>nsd</td>
</tr>
<tr>
<td>(yo) entiendo</td>
<td>***</td>
<td>***</td>
<td>nsd</td>
</tr>
<tr>
<td>(yo) recuerdo</td>
<td>***</td>
<td>***</td>
<td>*</td>
</tr>
<tr>
<td>(yo) Cognizer</td>
<td>*</td>
<td>***</td>
<td>**</td>
</tr>
</tbody>
</table>

* p ≤ 0.05   ** p ≤ 0.01   *** p ≤ 0.001

As can be seen in Table 1, overall there is a significant difference between the two corpora in the Cognizer category \[X^2(1,N=4512) = 9.46, p=0.002\], demonstrating that overt and null subject pronoun usage differs between the two corpora. More specifically, the differences in the usage of yo with the verbs creer and recordar led to this statistically significant difference, and both are significantly different between the two corpora and within each corpus.

This difference is even more pronounced when all instances of yo, in any position pre- or post-verbal, found in both corpora are included in the analysis. Tokens included in this analysis would be yo no creo, yo no lo creo, and creo yo, for example. In comparison with the aforementioned raw data (which only considered instances of yo immediately preceding the verb), in total there are 718 instances of yo in the Spanish in Texas corpus, and 1597 instances of yo in the PRESEEA corpus (the null data remained the same: 871 and 1575, respectively). When considering all instances of yo in the data, the difference between the bilingual community (represented in the Spanish in Texas corpus) and the predominantly Spanish-speaking communities (represented in PRESEEA) is more pronounced; while there was a small increase in the number of tokens in the Spanish in Texas data, the frequency of the use of yo in the PRESEEA data overtook the frequency of the lack of (or null) pronoun. This reinforces the
conclusion drawn from the data that only included yo immediately preceding the verb: the predominantly Spanish-speaking communities use yo differently than the bilingual community.

The results of the Chi-Squared calculations for both corpora individually and in comparison with each other can be seen in Table 2; in this case, all positions of yo (two positions pre-verbal, one position pre-verbal, and one position post-verbal, see examples 1, 2, and 3) were considered in the tabulation and subsequent statistical analysis. The inclusion of all positions of yo creates a stronger picture of the use of overt or null subjects throughout the class of Cognizer verbs included in the study.

Table 2. All positions of yo

<table>
<thead>
<tr>
<th>Verb</th>
<th>PRESEEA</th>
<th>SiT</th>
<th>SiT v. PRESEEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>(yo) creo</td>
<td>***</td>
<td>nsd</td>
<td>**</td>
</tr>
<tr>
<td>(yo) pienso</td>
<td>nsd</td>
<td>nsd</td>
<td>nsd</td>
</tr>
<tr>
<td>(yo) entiendo</td>
<td>***</td>
<td>***</td>
<td>nsd</td>
</tr>
<tr>
<td>(yo) recuerdo</td>
<td>***</td>
<td>***</td>
<td>**</td>
</tr>
<tr>
<td>(yo) Cognizer</td>
<td>nsd</td>
<td>**</td>
<td>***</td>
</tr>
</tbody>
</table>

* p ≤ 0.05   ** p ≤ 0.01   *** p ≤ 0.001

After adding in the other positions of yo, there is an even higher level of significance between the two corpora in the Cognizer category \([X^2(1,N=4761)=11.29, \ p=0.0008]\). Interestingly, when considering the Cognizer category in only the PRESEEA corpus, there is no significant difference between the use of null and overt subjects, which contributes to the higher level of significance between the two corpora. It also signifies that in the Cognizer category, the Spanish spoken by the bilingual community uses these pronouns in significantly different ways than the predominantly Spanish-speaking communities, demonstrating a clear effect of contact with the English language.

While there is a significant difference in the Cognizer category overall, not every verb within the category showed a significant difference between null and overt pronoun usage. As was found when only considering yo immediately preceding the verb, the two verbs with significant differences between the two corpora were creer and recordar. As has been found in previous studies, creo had the most tokens out of the four verbs (Posio 2011; 2013), which could contribute to this level of significance, especially because in the PRESEEA corpus alone, creo had a very high level of significance.

4.2 Within Spanish in Texas

There is a significant difference in the manner in which null and overt pronouns are used in the bilingual communities interviewed to form the Spanish in Texas corpus. Figure 1 below represents the differences in number of tokens of yo found in the corpora across the four Cognizer verbs. The first column only contains instances where yo immediately preceded the verb, whereas the second column represents all instances of yo found in the corpus.
As can be seen in Figure 1, there is a large difference in the number of instances of *yo* regardless of the number of positions of *yo* tabulated. Tables 3 and 4 break these totals down by verb, ending with the category total; while most of the verbs show significant differences, it is important to note that the differences in *yo* usage depicted in Figure 1 are significantly different, regardless of the inclusion or exclusion of all positions and tokens of *yo*.

<table>
<thead>
<tr>
<th>Table 3. Yo one position prior to verb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Verb</td>
</tr>
<tr>
<td>(yo) creo</td>
</tr>
<tr>
<td>(yo) pienso</td>
</tr>
<tr>
<td>(yo) entiendo</td>
</tr>
<tr>
<td>(yo) recuerdo</td>
</tr>
<tr>
<td>(yo) Cognizer</td>
</tr>
</tbody>
</table>

As can be seen in Table 3, the Cognizer category showed highly significantly different usage of subject pronouns \[X^2(1,N=3034)=16.78, p<0.001\]. However, the difference between the number of *yo* and null tokens did not demonstrate convergence, but rather divergence, as there were 646 instances of overt subject versus 871 instances of null subject. According to previous research, when languages are in contact, there is a higher likelihood of finding convergence rather than divergence (Torres Cacoullos & Travis 2011; Benevento & Dietrich 2015), but that is not the case here. Instead, the disparity between the numbers of tokens in each category demonstrates divergence between English and Spanish in the bilingual community in Texas. Rather than adapt the more flexible language (Spanish) to English, these speakers are making a distinction between the two languages by their use of these pronouns; because they have no choice but to include the subject pronoun in English, they are marking the distinction between the two languages by dropping the pronoun more often in Spanish, where subject pronoun inclusion is not obligatory.

This same effect can be seen when considering all instances of *yo*, as in Table 4, but with a lower level of significance than was found when considering only tokens of *yo* that immediately preceded the verb.
When all positions of *yo* are considered, the relationship between the uses of the two pronouns is less (but still very) significant; this is caused by the number of null tokens remaining the same while the number of *yo* tokens increased. Interestingly, when all positions of *yo* are taken into account, the verb with the largest number of tokens overall, *creo*, does not show a significant difference, whereas when only instances of immediately pre-verbal *yo* were considered, there was a significant difference. One possible explanation for this is simply that the number of null tokens did not increase. However, another explanation could be that the divergence truly is found in the restricted pre-verbal *yo* position, and not in instances that combine possible subject pronoun usage with negation or pronominalization, for example. This is an aspect that should be explored in future research.

To further ensure that the divergence seen in the data is truly a result of contact with English, the data from the Spanish in Texas corpus were compared with data from Monterrey, Mexico, the geographically closest city available in the PRESEEA corpus. The results of this comparison were highly significant \[X^2(1,N=1778)=18.95, p<0.001\]. Thus, it is possible to claim that the divergence seen in the Spanish in Texas data is due to contact with English.

### 4.3 Within PRESEEA

Within the PRESEEA corpus itself, there were also significant differences in the manner in which null and overt subjects were used, demonstrating that this phenomenon is not limited to language contact, but rather that it exists within Spanish-speaking countries as well. Table 5 depicts the interactions between the overt and null subjects for each of the four Cognizer verbs, and then for the category as a whole.

<table>
<thead>
<tr>
<th>Verb</th>
<th>PRESEEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>(yo) creo</td>
<td>*</td>
</tr>
<tr>
<td>(yo) pienso</td>
<td>nsd</td>
</tr>
<tr>
<td>(yo) entiendo</td>
<td>***</td>
</tr>
<tr>
<td>(yo) recuerdo</td>
<td>***</td>
</tr>
<tr>
<td>(yo) Cognizer</td>
<td>*</td>
</tr>
</tbody>
</table>

In keeping with previous research (such as Posio 2011, 2013 which found significant differences between null and overt pronoun usage in Peninsular Spanish), there is a significant difference in the category of Cognizer verb in the use of these pronouns within predominantly Spanish-speaking communities as well. However, it is more interesting to note the changes in degree of
significance when all the instances of *yo* were added in to the analysis, which can be seen in Table 6.

**Table 6. All positions of *yo***

<table>
<thead>
<tr>
<th>Verb</th>
<th>PRESEEA</th>
</tr>
</thead>
<tbody>
<tr>
<td>(yo) creo</td>
<td>***</td>
</tr>
<tr>
<td>(yo) pienso</td>
<td>nsd</td>
</tr>
<tr>
<td>(yo) entiendo</td>
<td>***</td>
</tr>
<tr>
<td>(yo) recuerdo</td>
<td>***</td>
</tr>
<tr>
<td>(yo) Cognizer</td>
<td>nsd</td>
</tr>
</tbody>
</table>

* * p ≤ 0.05  ** * p ≤ 0.01  *** p ≤ 0.001

After this addition, instead of being significant only at the 0.05 level, the verb *creer* is significant at the 0.001 level. It is possible that this difference could be caused by the inclusion of tokens such as *yo no creo*, where the subject pronoun cannot immediately precede the verb due to the addition of negation, pronominalization, or both. If this is the case, then another instance of divergence may exist, as the opposite trend was found in the Spanish in Texas data.

It is interesting to note that while as a whole the null and overt subject pronoun usage in PRESEEA corpus was significantly different than the use of these pronouns in the Spanish in Texas corpus, all the cities included in the corpus do not act the same in regards to these pronouns. Figures 2 and 3 depict the null and overt subject usage in the PRESEEA corpus broken down by city; Figure 2 only contains instances of *yo* immediately preceding the verb, while Figure 3 contains all instances of *yo* in the corpus pertaining to Cognizer verbs.

![Figure 2. Yo usage in the PRESEEA corpus – immediately pre-verbal yo only](image-url)
As can be seen in these two figures, the inclusion of all locations of *yo* did not cause any inversion; however, in some cases, such as Valencia, the inclusion of all the instances of *yo* caused what was a large difference between the null and overt pronoun to become much smaller. More importantly, what these two figures demonstrate is that even in predominantly Spanish-speaking communities, there are differences in how these pronouns are used. Caracas and Medellin, for example, show almost exact opposite trends, while in Santiago, Chile there is a high level of usage of both pronouns. Thus, further investigation is needed in order to uncover the exact differences in pronoun use in the Spanish-speaking communities as well.

5. Discussion

Overall, this study has found that there is a significant difference in the use of overt and null subject pronouns between predominantly Spanish-speaking communities and a bilingual community in Texas when followed by four verbs from the Cognizer category of verbs. This held to be true when considering the four verbs from the category as a whole, regardless of whether *yo* immediately preceded the verb or if it were two positions pre-verbal, immediately preceding, or one position post-verbal (though individually, not all the verbs showed statistically different uses of null and overt subject pronouns). Moreover, the use of null and overt subject pronouns preceding the four verbs from this category was significantly different in the bilingual community, but not in the predominantly Spanish-speaking community, indicating a possible influence of language contact. This further appeared to be the case when comparing the bilingual community to the data from the nearest available city (Monterrey, Mexico); the use of null and overt subject pronouns differed significantly between these two groups.

Additionally, within each corpus, there often was a significant difference in the inclusion or omission of *yo*, but not in every instance. When considering the Spanish in Texas data, the overall use of overt versus null subjects was significantly different for the group of four verbs as a whole, regardless of the placement of the pronoun; however, this was not the case for all the verbs within the category, as some never showed significant differences (*pensar* [to think]), and others varied in significance based on the position of the pronoun (*creer* [to think] was significant when *yo* immediately preceded the verb, but was not when all positions were
When considering the PRESEEA data, the inclusion or omission of the subject pronoun was significantly different for all four verbs together when *yo* immediately preceded the verb, but was not significantly different when all positions were considered. As was found in the other corpus, *pensar* never showed significant differences in pronoun inclusion or omission, but the remaining three verbs did, regardless of the position of the pronoun.

These findings demonstrate that the treatment of overt and null subject pronouns appears to be susceptible to language contact, though not in the direction that was originally hypothesized. Convergence was hypothesized for the data from the bilingual community, as it was assumed that the presence of English, which requires overt subjects, would cause them to be represented more in Spanish. However, the data represent a divergence between Spanish and English; it is possible that the lower frequency of the use of *yo* represents a differentiation between English and Spanish, reflected in the more flexible syntax of Spanish.

5.1 Limitations and Future Research

Though the overall conclusion of this study is that there is a significant difference between the manner in which the subject pronoun *yo* is used when comparing bilingual and predominantly Spanish-speaking communities, there are some limitations to this study. First of all, only one bilingual community was included in the analysis; future research should attempt to include multiple communities from around the United States, represented in corpora such as the Salinas corpus and the New Mexico corpus (Brown 2017; Torres Cacoullos & Travis in preparation), in order to investigate whether this is a trend that is specific to Texas, or if it exists across the country. Moreover, these studies should also be replicated with other languages that are spoken in both monolingual and bilingual communities, such as French and English in Quebec, to determine whether this is a trend that exists when pro-drop languages come into contact with languages in which pronouns are obligatory.

A second limitation of this study is that only one semantic category of verbs was investigated. Future research should also investigate this trend in the other semantic categories proposed by Enríquez (1984) (as cited in Posio 2011) to determine what other categories demonstrate differences when in contact with English, and if these categories demonstrate the same divergence seen in the Cognizer category in this study. Similarly, future research should also investigate if this phenomenon occurs with other pronouns. Posio (2011) conducted some analysis of the use of *tú* [2nd person singular] in Peninsular Spanish, but other than this, the majority of the focus of research has been on *yo*. This, however, is explainable because the other pronouns can be problematic. For instance, *tú* is used differently depending on the country, while *él, ella*, and *usted* must be overtly stated to avoid ambiguity (as must *ellos, ellas*, and *ustedes*). However, if these issues were taken into account, primarily regarding *tú*, future studies could investigate these differences in countries that all use this pronoun in the same way.

Other aspects of interest for future research would be to investigate the differences in null and overt subject usage within the predominantly Spanish-speaking communities themselves. Figures 2 and 3, which depict the PRESEEA corpus tokens divided according to city, depict clearly that there are differences between cities and regions regarding this phenomenon. Finally, research should also investigate the effects of negation on this phenomenon. As was previously mentioned, the results from the Spanish in Texas corpus demonstrate a change of significance when the remaining *yo* tokens were added to the analysis; it would be interesting to note how many of them contained negation. This would also be interesting to investigate because it represents another area in which divergence could be seen, as the negation in English is also pre-
verbal (though with the added auxiliary verb). Despite these limitations and areas for future research, the present study contributes significantly to the understanding of null and overt subject pronoun usage, particularly in the area of language contact.

6. References


Torres Cacoullos, Rena & Travis, Catherine E. (In preparation). New Mexico Spanish-English Bilingual (NMSEB) corpus. The Pennsylvania State University. [http://nmcode-switching.la.psu.edu]

