Language contact in the US Southeast

The case of Spanish subject expression in an emerging bilingual community in Georgia

Philip P. Limerick
University of Georgia

This study examines the use of subject pronouns among Spanish speakers in the Southeastern US and explores the incipient stages of language contact through a case study of speakers in Roswell, Georgia, an emergent (recently developing) variety that thus far has rarely been studied in the literature. Sociolinguistic interviews were conducted in Roswell (Wilson 2013) and transcribed to allow for analysis of pronouns and factors that may influence subject expression (e.g. person/number) as well as social variables (e.g. length of residency). Results indicate an overall pronoun rate of 21%, similar to that of Mainland newcomers in New York (Otheguy, Zentella, and Livert 2007). However, results from the multivariate analysis suggest that pronoun usage in Roswell diverges from these communities, with differential effects observed for factors such as Coreferentiality Index (subject continuity). This analysis of subject expression reveals an intermediate stage of language shift in this particular community.

Keywords: subject pronouns, Spanish in the United States, Georgia

1. Introduction

The expression of subject pronouns (SPs) in Spanish is one of the most extensively studied features of the language (Silva-Corvalán 1982; Bentivoglio 1987; Cameron 1993; Flores-Ferrán 2004; Travis 2007; Torres Cacoullos and Travis 2010; Otheguy and Zentella 2012; Carvalho, Orozco, and Shin 2015). Whereas overt SPs are nearly obligatory in English, Spanish is a null-subject language, in which speakers can omit the SP, as in (1), or express it, as in (2).

(1) \( \emptyset \) Tengo que salir.
        ‘(I) have to leave.’
(2)  *Yo tengo que salir.*  
‘I have to leave.’

There are various linguistic and social factors that have been shown to influence such variation, and while many of these factors are consistent among dialects of Spanish, the differing social situations and grammars associated with each dialect reflect unique patterns concerning the explicit or implicit expression of SPs. The goal of this paper is to examine pronoun usage among Spanish speakers in the Southeastern US in order to explore the incipient stages of language contact through a case study of speakers in Roswell, Georgia (see Figure 1), an exurb of Atlanta.

Figure 1. Georgia. Map by Bert Sperling

Roswell reflects the patterns of Hispanic immigration in the Southeastern US in general, which, unlike other regions of the US which have contained stable Hispanic communities for decades or longer, is only recently experiencing the emergence of durable Spanish-speaking communities. Between 1990 and 2010 more than a million Hispanics from Mexico, El Salvador, and other Latin American countries immigrated to Virginia, North Carolina, South Carolina, and Georgia. In fact, the Hispanic populations of North Carolina and Georgia had the highest percentage of growth and currently have the largest percentages of monolingual Spanish speakers (Wolfram, Kohn, and Callahan-Price 2011). Due to such lack of an established presence, Spanish in the Southeastern US has received very little attention compared to the Southwest and Northeast.

However, there is a growing interest in Southeast Spanish as seen by the work of Smith (2006) and Montes-Alcalá and Sweetnich (2014), who investigate Georgia Spanish. However, very few studies to my knowledge have addressed language contact and change through the lens of Spanish subject expression in the Southeast (McKnight 2013 [Raleigh, NC] and Wilson 2014 [Roswell, GA] are exceptions). Therefore, a primary goal of the present research is to contribute to the
sociolinguistic literature and to compare subject expression among Spanish speakers in Georgia to that of speakers in other regions.

Given that the language contact situation in the Southeast is less established than that of other regions, this study offers the opportunity to examine the process of linguistic change at an incipient stage. In order to begin to explore Southeast Spanish, this paper will examine one community in particular, Roswell, GA, in order to further understand processes of language contact by investigating the influence of linguistic and social variables on SP variation and by comparing pronoun use with monolingual varieties and more established contact varieties. Reflecting immigration to the Southeast in general, the Hispanic population in Roswell (see Table 1) grew from 10.6% of the city’s population in 2000 to 16.6% in 2010 (U.S. Bureau of the Census 2010), experiencing a 75% increase.

Table 1. Roswell demographics

<table>
<thead>
<tr>
<th>Population</th>
<th>2000</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Population</td>
<td>79,334</td>
<td>88,346</td>
</tr>
<tr>
<td>White</td>
<td>64,666</td>
<td>66,010</td>
</tr>
<tr>
<td>Black or African American</td>
<td>6,773</td>
<td>10,373</td>
</tr>
<tr>
<td>Asian</td>
<td>2,964</td>
<td>3,565</td>
</tr>
<tr>
<td>Some Other Race</td>
<td>3,237</td>
<td>5,846</td>
</tr>
<tr>
<td>Two or More Races</td>
<td>1,511</td>
<td>2,241</td>
</tr>
<tr>
<td>American Indian or Alaska Native</td>
<td>160</td>
<td>261</td>
</tr>
<tr>
<td>Hawaiian Native/other Pacific Islander</td>
<td>23</td>
<td>50</td>
</tr>
<tr>
<td>Hispanic or Latino (of any race)</td>
<td>8,421</td>
<td>14,699</td>
</tr>
</tbody>
</table>

Source: U.S. Bureau of the Census, 2010 Census of Population

For the present study, I intend to answer the following questions:

a. How do the overall pronoun rates of Roswell speakers compare to more established varieties of US Spanish and to monolingual varieties of Spanish?
b. What linguistic variables have the greatest effect on SP usage?
c. How do social variables, such as age, gender, length of residency (LOR), and age of arrival (AOA) influence pronoun use?
d. Is language change observable in this community? If so, is English contact a significant factor contributing to this change?

To answer these questions, I carry out a quantitative analysis using sociolinguistic interviews that were conducted in Roswell (Wilson 2013). I examine factors such as person/number of the verb, same/switch reference, verb tense, and LOR in the US, among others, and their influence on subject expression in this Spanish variety.
In this paper, I first review overall pronoun rates and specific factors that influence pronoun usage among different Spanish varieties and summarize previous studies of Spanish subject expression. I then outline the methodology for the present study in section three and discuss the results in section four, including the overall pronoun rate and the conditioning factors of subject expression in Roswell. Finally, I discuss some general conclusions and considerations for future research.

2. Subject expression in Spanish

2.1 Pronoun rates and influencing factors

I use the term *overt* to indicate the presence of an SP and *null* when it is absent. According to Bosque and Demonte (1999), the primary function of using overt SPs in Spanish is to emphasize the referent or to express contrast among referents. The null pronoun is neutral, that is, not emphatic. Therefore, in example (2) above, the presence of the first-person singular pronoun *yo* ‘I’ would serve to emphasize the person speaking or to show a contrast with other interlocutors. However, when there is no particular emphasis desired, the *yo* can be omitted, while still leaving the sentence grammatical, as seen in (1). Additionally, overt SPs are used in contexts of potential ambiguity, for example, when the verb is in the third-person:

(3) *Yo trabajaba en el hotel.*

‘I worked at the hotel.’

In (3), the SP could have also been *él* ‘he’, *ella* ‘she’, or *usted* ‘you (formal)’ according to the verbal inflection. The overt *yo*, however, is necessary to disambiguate the referent, unless there is previous context from which the referent could be identified.

Below I present the reported overall overt pronoun rates for monolingual varieties (see Table 2) as well as Spanish-English contact varieties (see Table 3). It is important to note the cross-dialectal variability in rates among monolingual Spanish as well as between monolingual and contact varieties.

Regarding monolingual Spanish, varieties from Mexico and Spain tend to have the lowest pronoun rates, South American varieties such as Argentina exhibit mid-range rates, and Caribbean speakers have the highest rates out of all Spanish dialects. Contact varieties sometimes show higher rates than monolingual varieties, which is commonly attributed to English contact, as overt SPs are nearly obligatory in English but not in Spanish. Therefore, increased exposure to English has been argued to engender higher usage of overt pronouns (overt pro) in Spanish (Lapidus and Otheguy 2005; Otheguy, Zentella, and Livert 2007). However, the necessary evidence to conclude that language change is contact-induced or is occurring at
all is debated among scholars. Poplack and Levey (2010) posit that changes in SP frequencies do not constitute evidence for change, whether it be contact-induced or otherwise. They affirm that in order to consider contact-induced change, there must be differences in constraint hierarchies, that is, the ranking of factors probabilistically shown to favor or disfavor SP occurrence. Torres Cacoullos and Travis (2010, 189) are in agreement by stating that “reliance on overall rates of use to determine contact-induced change is problematic because it is well known that regional dialects vary enormously” and also suggest that divergences in grammatical patterning across dialects are necessary to substantiate contact-induced change. Other scholars attribute changes in SP usage to a process of simplification on part of the bilingual, whereby they lighten the cognitive load by simplifying certain morphosyntactic or discourse constraints (e.g. Sorace 2004). According to Prada Pérez (2009), simplification is different from convergence, the latter being a process in which two contact languages become more alike. In any case, speakers may show a weakening of pragmatic constraints on SP usage. For instance, various scholars have argued that Spanish-English bilinguals show a weakened sensitivity

<table>
<thead>
<tr>
<th>Variety</th>
<th>PN Rate</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yucatan, Mexico</td>
<td>16%</td>
<td>Michnowicz 2015</td>
</tr>
<tr>
<td>Valladolid, Mexico</td>
<td>21%</td>
<td>Solomon 1999</td>
</tr>
<tr>
<td>Mexico City</td>
<td>22%</td>
<td>Lastra and Martin Butragueño 2015</td>
</tr>
<tr>
<td>Puente Genil, Spain</td>
<td>24%</td>
<td>Ranson 1991</td>
</tr>
<tr>
<td>New York Newcomers (Mexican, S. American, Caribbean)</td>
<td>30%</td>
<td>Otheguy, Zentella, and Livert 2007</td>
</tr>
<tr>
<td>Barranquilla, Colombia</td>
<td>36%</td>
<td>Orozco and Guy 2008</td>
</tr>
<tr>
<td>Buenos Aires, Argentina</td>
<td>36%</td>
<td>Barrenechea and Alonso 1977</td>
</tr>
<tr>
<td>Santiago, Dominican Republic</td>
<td>39%</td>
<td>Olloqui de Montenegro 1987</td>
</tr>
<tr>
<td>San Juan, Puerto Rico</td>
<td>45%</td>
<td>Cameron 1993</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variety</th>
<th>PN Rate</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>North Carolina Mexicans</td>
<td>17%</td>
<td>McKnight 2013</td>
</tr>
<tr>
<td>New Jersey Mexicans</td>
<td>24%</td>
<td>Flores-Ferrán 2007</td>
</tr>
<tr>
<td>New York born/raised (Mexican, S. American, Caribbean)</td>
<td>38%</td>
<td>Otheguy, Zentella, and Livert 2007</td>
</tr>
<tr>
<td>Los Angeles Mexicans</td>
<td>43%</td>
<td>Silva-Corvalán 1982</td>
</tr>
<tr>
<td>New York Dominicans</td>
<td>48%</td>
<td>Flores and Toro 2000</td>
</tr>
</tbody>
</table>
to coreferrationality index (CI), that is, same vs. switch reference, a factor discussed in greater detail below (e.g. Silva-Corvalán 1994; Shin and Otheguy 2009; Otheguy and Zentella 2012).

The specific factors shown to influence subject expression across most dialects of Spanish are person/number of the verb (Silva-Corvalán 1994), same vs. switch reference (Bayley and Pease-Alvarez 1996), clause type where the verb appears (Morales 1997), tense-mood-aspect of the verb (TMA, Cameron 1994), lexical content of the verb (Travis 2007), specificity of the referent (Cameron 1992), reflexivity of the verb (Otheguy, Zentella, and Livert 2007), priming (Travis 2007), as well as social factors including age and gender (Carvalho and Child 2011). However, there are additional influencing variables in contact varieties including length of residency (LOR), age of arrival (AOA), and degree of use, proficiency, and exposure to English (Otheguy, Zentella, and Livert 2007). For example, Otheguy, Zentella, and Livert (2007) found that Spanish speakers living in NYC for longer periods of time (born/raised there) had higher pronoun rates (38%) than newcomers to the city (30%), suggesting a significant influence of English on NYC Spanish as well as a dialect contact influence from high-pronoun Caribbean varieties. Moreover, Shin and Otheguy (2013) employed a broader analysis of the role of social factors on increased pronoun rates in NYC, namely social class and gender, finding that more affluent communities showed significant increases of overt pro while less affluent communities were more resistant to such change due to loose social networks in the former group and more dense networks in the latter. Further, the study revealed that women lead this change, which is consistent with previous research on gender and language change (Eckert and McConnell-Ginet 2003; Romaine 2003). Specifically, women showed higher overt pronoun rates than men, which the authors attributed to women’s more extensive contact with second generation Latinos, including their children, who exhibit higher pronoun rates.

While previous research has indicated that contact varieties show some differences regarding SP expression in comparison to monolingual varieties, the same linguistic factors have been found to constrain SP use cross-dialectally, for instance, switch reference in Los Angeles (Silva-Corvalán 1982) person/number in NYC (Otheguy, Zentella, and Livert 2007), clause type in Raleigh, NC (McKnight 2013), and lexical frequency in Roswell, GA (Wilson 2014). Next, I discuss in greater detail each factor that is relevant to the present study.

1. Several factors not discussed here such as specificity, reflexivity, priming, and lexical frequency also greatly influence SP variation (see Cameron 1992; Otheguy, Zentella, and Livert 2007; Travis 2005; Erker and Guy 2012 for further discussion).
2.2 Linguistic variables

2.2.1 Person/number

In most Spanish varieties, both person and number of the verb highly influence SP variation. Specifically, first-person singular verbs tend to favor overt pro most frequently, with the exception of Madrid Spanish, where lower frequencies occur (Cifuentes 1980; Enríquez 1984; Cameron 1992; Silva-Corvalán 1994; Flores-Ferrán 2002). In fact, both Barrenechea and Alonso (1977) and Silva-Corvalán (1994) found that all singular forms are more likely to appear with overt pro compared to plural forms, which show a general disfavoring of overt pro (Enríquez 1984; Cameron 1992; Flores-Ferrán 2002). Otheguy, Zentella, and Livert (2007) reveal regional differences, finding that Caribbean newcomers to NYC preferred overt pro with second-person singular verbs compared to third-person singular verbs. Conversely, South American newcomers preferred overt pro with third-person singular verbs.

2.2.2 Coreferentiality index

Coreferentiality index (CI), which considers same vs. switch reference from one subject to another, has also shown a strong influence on the manifestation of subject expression cross-dialectically (Bentivoglio 1987; Cameron 1994; Bayley and Pease-Alvarez 1997; Travis 2005). Where there is a switch in subject referent, the SP is often overt, as seen in (4); when there is no switch, null SPs are preferred, as in (5).

(4) …también sé que ellos tienen eventos allí mismo en la iglesia [F37Mex]²
   ‘…I also know that they have events right there at the church’

(5) …después que me vine para acá ∅ empecé limpiando, casas, cuidando niños [F32Mex]
   ‘…after I came here I started cleaning, houses, taking care of children’

As noted above, certain contact varieties have shown a weakening of this constraint (Flores-Ferrán 2002; Shin and Otheguy 2009; Michnowicz 2015). For example, Flores-Ferrán (2002) found that NYC-born Spanish speakers exhibit an increased use of overt pro in same reference contexts, demonstrating their loss of sensitivity to null pronoun usage in such contexts. Similarly, such non-canonical patterns are found in the present study, as illustrated in (6) and (7):

(6) …una amiga mía se casó con un americano y ella está muy feliz [M23Mex]
   ‘…a friend of mine married an American and she is very happy’

---

2. Speaker codes indicate gender, age, and nationality.
(7) … yo te puedo decir que hace un año y medio yo dejé ya completamente a la construcción
‘…I can tell you that a year and a half ago I totally quit construction’

2.2.3 Clause type
Another factor conditioning subject expression is the type of clause in which the SP appears. For example, Morales (1997) found that overt SPs are more likely to occur in subordinate clauses than they are in main clauses. In addition, Otheguy and Zentella (2012) found that main clauses favored overt pro while coordinate clauses favored nulls for NYC Spanish.

2.2.4 Tense-mood-aspect
The tense-mood-aspect (TMA) of a verb has also been shown to condition SPs. Certain TMAs favor overt pro while others favor nulls. For instance, Silva Corvalán (1982) found that imperfects and conditionals favor overt pronouns while overt pro appearing with presents and preterits is less likely, as was also found by other researchers (Cameron 1994; Travis 2007). It has been proposed that imperfects and conditionals favor overt pro due to their potential ambiguity in the first and third persons. Overt SPs would serve to disambiguate the referents of such forms (Hochberg 1986). However, other studies have found no such correlation (Enríquez 1984; Bentivoglio 1987; Ranson 1991).

2.2.5 Lexical content
Several researchers have found that the lexical content of verbs can determine SP variation (Bentivoglio 1987; Silva-Corvalán 1994; Travis 2007). Psychological/mental verbs, verbs of communication, and copulas tend to favor overt pro, with psychological verbs showing the highest probability (e.g. creer, ‘to believe/think’). Contrastively, motion verbs generally disfavor overt pro (Bentivoglio 1987; Silva-Corvalán 1994; Travis 2007). It has been hypothesized that psychological verbs favor overt pro because they express the speaker’s point of view and because of their implied contrastive function (Silva-Corvalán 1994).

2.3 Social variables

2.3.1 Gender
Some authors have found that females and males differ in their SP frequencies. In their study of Uruguayan Spanish in contact with Portuguese, Carvalho and Child (2011) found that females favored overt pro whereas males favored nulls. Similarly, Shin and Otheguy (2013) observed a higher rate of overt pro among
women speakers in NYC, which they attribute to women’s extensive contact with US born bilinguals. Bayley and Pease-Alvarez (1996) hypothesized that women use narrative discourse more often than men and thus more SPs. Moreover, Alfaraz (2015) suggests that this gender pattern, particularly in the Dominican Republic, is linked to prestige.

2.3.2 Age
Age is another social factor that affects SP usage. Carvalho and Child (2011) found a favoring of nulls among younger speakers in Uruguay (ages 16–29) while older speakers favored overt pro, a trend they hypothesize is influenced by the Portuguese substratum (23). Similarly, Orozco and Guy (2008) observed that older Colombians favored overt pro while adolescents favored nulls, which they explain as a possible change in progress attributed to greater access to education by young people compared to prior generations. They also hypothesize that these results are due to influence of the Colombian highlands dialect. On the contrary, Flores-Ferrán (2002) found that older speakers (50+) exhibited lower rates of overt SPs than younger speakers and attributes this to the tendency of older speakers to prefer a more conservative use of language following prescriptive grammar norms. In the US, younger speakers may be using more overt pro due to English influence. Otheguy, Zentella, and Livert (2007), however, did not find age to be a significant factor for SP usage.

2.3.3 Investigating English contact: Length of residency and age of arrival
Two important factors to consider regarding potential influence of English on Spanish are LOR in the US, suggesting greater exposure to English, and AOA to the US, indicating speakers’ susceptibility to change. Drawing from Veltman (2000), Otheguy, Zentella, and Livert (2007, 779) state that older arrivals who speak Spanish more frequently and fluently should show greater resistance to English influence. In addition to finding that longer LORs in NYC correlated with higher frequencies of overt pro, the authors found that younger arrivals, being more susceptible to English influence, exhibited higher pronoun rates than older arrivals, who tend to use less English and whose Spanish is more proficient, thereby making them less susceptible to contact influence.

Through a quantitative analysis of SP usage using conversational data from Mexican speakers, I show that subject expression in immigrant Spanish in Georgia responds to the same factors governing pronoun usage in other varieties such as person/number, clause type, and TMA, among others (Otheguy and Zentella 2012; Michnowicz 2015). I also show that Spanish in this region exhibits a lower pronoun rate than more established varieties of US Spanish and is similar to monolingual Mexican varieties. However, divergences are shown in the ranking
of variables on pronoun use, specifically with regard to CI. Finally, I argue that the similarity in pronoun rate with monolingual Mexican varieties and the insignificance of social factors provide no evidence for English influence on SP use in Roswell, but that the CI patterns observed potentially reflect a weakened sensitivity to discourse-pragmatic constraints.

3. Methodology

3.1 Speakers

Using a variationist sociolinguistic approach, I examine subject expression among 10 Spanish speakers born in various regions of Mexico, who are residents of Roswell, GA and whose average LOR in the US is 15 years. The sample includes five females and five males ranging in age from 23 to 54 and whose AOA to the US ranges from 13 to 38. Additionally, their education levels range from primary school to university, and the majority work in hotel housekeeping (see Table 4).

Sociolinguistic interviews were conducted in Roswell (Wilson 2013) and later transcribed to allow for analysis of pronoun usage and the variables found in previous literature to influence such usage. The five linguistic variables in this study are person/number, CI, clause type, TMA, and lexical content. The four social variables include gender, age, LOR, and AOA. The coding of these variables will be described and illustrated below.

3.2 The envelope of variation

Approximately 2,312 finite clauses were identified in the transcripts, of which 1,307 were excluded from the present study, leaving 1,005 tokens for analysis. These exclusions were made because they were outside the envelope of variation (Otheguy, Zentella, and Livert 2007). In the present study, the envelope of variation includes environments in which variation between overt and null SPs is possible. Any context of null SP expression in which an overt SP could also occur belongs inside the envelope of variation and is included. Likewise, the appearance of an overt SP in a context in which a null pronoun could alternatively occur would fit inside the envelope of variation. However, contexts in which either the null or overt pronoun is obligatory were eliminated since they do not generally exhibit variation and consist of the following contexts: verbs within subject headed relative clauses (La mujer que vino ayer ‘The woman that came yesterday’), clauses with full noun phrases (Las chicas estudiaron mucho ‘The girls studied a lot’), verbs, such as haber or ser, when used existentially (Hay muchas personas ‘There are/
were many people’; ∅ Es que no salgo mucho ‘It’s just that I don’t go out much’), verbs such as hacer when making reference to chronological periods (∅ Hace cinco años ‘Five years ago’), verbs with inanimate referents (∅ es bonita, making reference to a city, ‘It’s nice’), and set phrases where an overt or null SP is categorical (¿Qué sé yo? ‘I don’t know’ [literally ‘What do I know?’]; ¿me entiendes ∅? ‘Do you know what I mean?’ [literally ‘Do you understand me?’]). Spanish speakers would rarely (if ever) alternate between an overt and null in these cases.3

Below is a passage from the present data in which I illustrate the verbs included, which are underlined, and those which I excluded, which are bolded:

(8) Yo creo que donde (9) yo soy, lo más agradezco (10) es que este país me (11) hice a [sic] madurar. Entonces de allí, (12) empecé a trabajar, a trabajar en esta compañía de construcción. Posteriormente por lo mismo la, la habilidad que me (13) dio Dios de aprender tan rápido que, este, después de tres años que (14) trabajé para esta compañía, pues (15) di las gracias.  

[38 Mex]

3. Additionally, in contact situations one might expect to see overt pro used where they aren’t permitted in Spanish, but are in English, such as with weather expressions (i.e. ∅ Está lloviendo, ‘it’s raining’). However, I have found no such examples in my data. Moreover, I strongly suspect that this phenomenon does not occur since, to the best of my knowledge, it has not been observed in previous research on Spanish-English contact varieties.
'I think that where I am, I am grateful the most is that this country made me mature. So from there, I started to work, to work at this construction company. Later, for that reason, the, the ability that God gave me to learn so quickly that, umm, after three years of working for this company, well I was thankful.'

For the analysis, I first extracted all finite verbs within the envelope of variation and then coded whether the verb appeared with a null or overt SP. Next, I coded for the five linguistic variables and the four social variables. To determine the significance of each variable as well as the constraints that condition SP use, I carried out a mixed-effects multivariate analysis using Rbrul (Johnson 2009).

3.3 Coding of variables

3.3.1 Person/number
The present study includes five person/number categories: first-person singular ('yo 'I'), second-person singular ('tú 'you' [informal] and 'usted 'you'[formal]), third-person singular ('él/ella 'he/she), first-person plural ('nosotros/nosotras 'we' [masc./fem.]), and third-person plural ('ellos/ellas 'they' [masc./fem]).

3.3.2 Coreferentiality index (CI)
Adopting a more extensive method from previous SP studies (Silva-Corvalán 1994; Orozco and Guy 2008), I have coded this variable considering not only subjects, but also the coreferentiality with preceding objects (direct/indirect object, object of preposition, etc.) within switch reference contexts. Various studies that have examined this factor found a somewhat neutral effect, with neither a favoring nor disfavoring in these contexts (Orozco and Guy 2008; Otheguy and Zentella 2012; Lastra and Martín Butragüeño 2015). Moreover, following Otheguy and Zentella (2012), I took the interviewer’s speech into account regarding coreferentiality and considered the previous reference made by the interviewer when looking at the respondent’s first token. For example, I coded same reference for any case in which the interviewer’s final reference was the same as the respondent’s first reference, as seen in (16).

(16) Interviewer: ¿Me puedes hablar un poquito sobre cualquier educación que has tenido en tu vida?
   ‘Can you talk to me about any education you have had in your life?’
   Respondent: Solamente fui a la primaria.
   ‘I only went to primary school.’ [F37Mex]

Since there is no switch in reference between the respondent’s self-reference fui and the interviewer’s reference has, this could influence the use of a null pronoun by the respondent.
3.3.3 Clause type
The clause categories I coded for are Main, Subordinate, and Coordinate in order to determine which of these types of clauses might have an effect on pronoun usage. A fourth category, Indeterminate Clause, was included for cases in which the structure of the surrounding discourse could not be determined, thus making it impossible to identify the clause type in which the verb appeared (see Otheguy and Zentella 2012).

3.3.4 TMA
Nine TMA categories were included: present indicative, preterit, imperfect, perfect – which includes present perfect and pluperfect –, present subjunctive, imperfect subjunctive, synthetic future, periphrastic future, and conditional.

3.3.5 Lexical content
I employed four semantic categories: psychological, motion, communication, and copulas. Additionally, I coded verbs that did not fit into any of these categories as “other”, adopting Travis’s (2007) method. (17) and (18) below illustrate some of these verb types in the data.

(17) …yo pienso que, para mí, yo he escuchado mucho, que les gusta Roswell, que a mí me gusta, está bonito aquí.
    ‘I think that, for me, I’ve heard a lot, that they like Roswell, that, I like it, it’s nice here.’  [F51Mex] (Coded as a psychological verb)

(18) No puedo contar mucho, pero….
    ‘I can’t tell much, but….’  [M23Mex] (Coded as Other)

3.4 Social variables
In the present study, gender was coded as M or F, and age was coded with two levels: older (40+) and younger (< 40). Both LOR and AOA were also coded with two levels: LOR (16+ years vs. < 16 years), AOA (26+ vs. < 26).

3.5 Predictions
Based on the findings of previous studies regarding correlations between independent variables and Spanish subject expression, I predict the following concerning SP use in Roswell:

a. Singular verb forms, especially first-person forms, will favor overt SPs while plural forms will disfavor overt pro.
b. Switch reference contexts will favor overt pro while same reference contexts will favor nulls. However, these contextual effects may be weaker than those observed in monolingual varieties since I have noted an increase of overt pro in same reference contexts in the data. Additionally, contexts in which there is a switch in subject but no switch with the preceding object referent will show a neutral effect.

c. Verbs within subordinate clauses will prefer overt pro compared to verbs in main and coordinate clauses.

d. Imperfect and conditional verb forms will favor overt pro while present and preterit forms disfavor them.

e. Psychological verbs, copulas, and verbs of communication will favor overt pro, with psychological verbs showing the highest probability, while motion verbs will disfavor overt pro.

f. Speakers with longer LORs will have higher frequencies of overt SPs than those with shorter LORs. Also, those with lower AOAs will exhibit higher overt frequencies than older arrivals. Due to inconsistent findings on age and gender and given that most SP studies do not find such factors to be significant, I do not expect either variable to be a conditioning factor of SP use.

4. Results

4.1 Overall pronoun rate

As seen in Table 5, out of 1,005 verbs, 214 occurred with an overt SP and 791 occurred with a null SP, indicating an overall overt pronoun rate of 21% for Mexicans in Roswell.

<table>
<thead>
<tr>
<th>N Verbs</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overt SPs</td>
<td>214</td>
</tr>
<tr>
<td>Null SPs</td>
<td>791</td>
</tr>
<tr>
<td>Total</td>
<td>1,005</td>
</tr>
</tbody>
</table>

Comparing this result to other studies (see Table 6), the pronoun rate in the Roswell data is only marginally higher than that of Mexican newcomers in NYC (19%) and Mexicans in North Carolina (17%), marginally lower than that of Mexican newcomers in New Jersey (24), virtually identical to that of monolingual Mexicans in Mexico City (22%), and substantially lower than
the rate of Mexicans in Los Angeles (43%). Therefore, in terms of overall pronoun rate, Mexican Spanish in Roswell is more like monolingual and incipient contact varieties and less like established contact varieties.

Table 6. Comparison of pronouns rates between Roswell Spanish and other varieties

<table>
<thead>
<tr>
<th>Variety</th>
<th>PN Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>San Juan, Puerto Rico (Cameron 1993)</td>
<td>45%</td>
</tr>
<tr>
<td>Los Angeles Mexicans (Silva-Corvalán 1982)</td>
<td>43%</td>
</tr>
<tr>
<td>NYC born/raised (Mexican, S. American, Caribbean; Otheguy, Zentella, and Livert 2007)</td>
<td>38%</td>
</tr>
<tr>
<td>Barranquilla, Colombia (Orozco and Guy 2008)</td>
<td>36%</td>
</tr>
<tr>
<td>NYC Newcomers (Mexican, S. American, Caribbean; Otheguy, Zentella, and Livert 2007)</td>
<td>30%</td>
</tr>
<tr>
<td>Puente Genil, Spain (Ranson 1991)</td>
<td>24%</td>
</tr>
<tr>
<td>New Jersey Mexicans (Flores-Ferrán 2007)</td>
<td>24%</td>
</tr>
<tr>
<td>Mexico City (Lastra and Martín Butragueño 2015)</td>
<td>22%</td>
</tr>
<tr>
<td>Roswell Mexicans (present study)</td>
<td>21%</td>
</tr>
<tr>
<td>NYC Mexican Newcomers (Shin 2012)</td>
<td>19%</td>
</tr>
<tr>
<td>North Carolina Mexicans (McKnight 2013)</td>
<td>17%</td>
</tr>
<tr>
<td>Yucatan, Mexico (Spanish monolinguals, Michnowicz 2015)</td>
<td>16%</td>
</tr>
</tbody>
</table>

4.2 Conditioning factors of SP expression

For the Rbrul analysis (one-level), I included all aforementioned independent linguistic and social variables. With regard to TMA, I excluded all uses of the synthetic future (e.g. comeré), conditional, and imperfect subjunctive due to the very low number of tokens for these categories and recoded the very low number of periphrastic futures (e.g. voy a comer), adding them to the present indicative category. I also modified the Lexical Content variable due to the small size of the data set in order to avoid an overly complex statistical model. Following Torres Cacoullos and Travis (2010), I simplified the verb types to a binary categorization: Psychological and All other verbs. Regarding CI, I recoded the cases of object coreference as simple switch reference due to a small token count. I also combined the main and subordinate levels for Clause Type due to collinearity/overlap of these categories.4 Additionally, I included the speaker as a random effect.

4. The % overt did not descend in order with the factor weights for subordinate and main clauses; the former showed a factor weight of .66 with 24% overt while the latter had a factor weight of .62 with 25% overt. See Tagliamonte (2012) for further discussion of collinearity.
All linguistic variables were selected as significant in the multivariate analysis. Regarding the social variables of gender, age, LOR and AOA, none were selected as significant to subject expression. The resulting variable hierarchy is presented below (see Table 7), showing the ranking of significant factors (by range) as compared with other studies. Person/number, which is ranked first, has the most powerful effect on subject expression while CI, which is ranked fifth, has the weakest effect. When comparing this ranking with that of Mainland Newcomers in NYC, two main differences are noted: Clause type is ranked much higher in Roswell than in NYC and CI is ranked much lower. This reduced effect for CI is also seen when compared to Mexico City and Yucatan speakers (Lastra and Martín Butragueño 2015; Michnowicz 2015) and will be further discussed below. Despite these differences, there are many similarities across studies: Person/number ranks first for all studies, and TMA has similar rankings. Further, Lexical Content ranks low in the two studies that employed this factor. Table 8 shows the constraint hierarchy for SP usage.

Table 7. Variable hierarchy as compared to other varieties

<table>
<thead>
<tr>
<th>NYC Mainland Newcomers (Otheguy and Zentella 2012)</th>
<th>Mexico City speakers (Lastra and Martín Butragueño 2015)</th>
<th>Yucatan speakers (Michnowicz 2015)</th>
<th>Roswell speakers (present study)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Person/Number</td>
<td>1. Person/Number</td>
<td>1. Person/Number</td>
<td>1. Person/Number**</td>
</tr>
<tr>
<td>2. Coreferentiality Index</td>
<td>2. Coreferentiality Index</td>
<td>2. TMA</td>
<td>2. Clause Type**</td>
</tr>
<tr>
<td>3. TMA</td>
<td>3. Coreferentiality Index</td>
<td>3. Coreferentiality Index</td>
<td>3. TMA*</td>
</tr>
<tr>
<td>4. Lexical Content</td>
<td></td>
<td></td>
<td>4. Lexical Content**</td>
</tr>
<tr>
<td>5. Clause Type</td>
<td></td>
<td></td>
<td>5. Coreferentiality Index*</td>
</tr>
</tbody>
</table>

* p < 0.05
** p < 0.01

Table 8. Hierarchy of constraints

<table>
<thead>
<tr>
<th>Factor Group</th>
<th>Probability Weight</th>
<th>% Overt</th>
<th>N</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Person/Number</td>
<td>.75</td>
<td>30</td>
<td>80</td>
<td>6.61e–10</td>
</tr>
<tr>
<td>3s</td>
<td>.65</td>
<td>29</td>
<td>497</td>
<td></td>
</tr>
<tr>
<td>1s</td>
<td>.51</td>
<td>18</td>
<td>74</td>
<td></td>
</tr>
<tr>
<td>2s</td>
<td>.46</td>
<td>13</td>
<td>222</td>
<td></td>
</tr>
<tr>
<td>3pl</td>
<td>.17</td>
<td>4</td>
<td>126</td>
<td></td>
</tr>
</tbody>
</table>

RANGE 58
Table 8. (continued)

<table>
<thead>
<tr>
<th>Factor Group</th>
<th>Probability Weight</th>
<th>% Overt</th>
<th>N</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clause Type</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Main/Subordinate</td>
<td>.68</td>
<td>25</td>
<td>749</td>
<td>0.00028</td>
</tr>
<tr>
<td>Coordinate</td>
<td>.48</td>
<td>12</td>
<td>222</td>
<td></td>
</tr>
<tr>
<td>Indeterminate</td>
<td>.33</td>
<td>7</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td><strong>RANGE 35</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TMA</strong></td>
<td></td>
<td></td>
<td></td>
<td>0.035</td>
</tr>
<tr>
<td>Imperfect</td>
<td>.62</td>
<td>36</td>
<td>83</td>
<td></td>
</tr>
<tr>
<td>Perfect</td>
<td>.53</td>
<td>26</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td>Present</td>
<td>.47</td>
<td>19</td>
<td>700</td>
<td></td>
</tr>
<tr>
<td>Preterit</td>
<td>.38</td>
<td>19</td>
<td>123</td>
<td></td>
</tr>
<tr>
<td><strong>RANGE 23</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Lexical Content</strong></td>
<td></td>
<td></td>
<td></td>
<td>0.00597</td>
</tr>
<tr>
<td>Psychological</td>
<td>.57</td>
<td>34</td>
<td>163</td>
<td></td>
</tr>
<tr>
<td>All Other Verbs</td>
<td>.43</td>
<td>19</td>
<td>836</td>
<td></td>
</tr>
<tr>
<td><strong>RANGE 14</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Coreferentiality Index</strong></td>
<td></td>
<td></td>
<td></td>
<td>0.0108</td>
</tr>
<tr>
<td>Switch reference</td>
<td>.56</td>
<td>25</td>
<td>511</td>
<td></td>
</tr>
<tr>
<td>Same reference</td>
<td>.45</td>
<td>18</td>
<td>488</td>
<td></td>
</tr>
<tr>
<td><strong>RANGE 11</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Speaker (random) Std. Dev. .42

4.2.1 Person/number

As predicted, the singular verb forms show a favoring of overt SPs whereas the plural forms disfavor overt pro. Third-person singular forms favor overt pro the most with a probability weight of .75 and first-person plural forms highly disfavor overt pro with a weight of .17. In fact, only 4% of first-person plural forms in the data occur with an overt SP. This general finding regarding singular and plural forms is consistent with previous studies (Cameron 1992; Silva-Corvalán 1994; Flores-Ferrán 2002; Otheguy and Zentella 2012). Moreover, it is interesting to note that third-person singular forms outweigh first-person singular forms in overt pro (probability weights of .75 and .65, respectively), which is the inverse of what most studies have found for Mexican Spanish (e.g. Shin 2012; Michnowicz 2015). However, this finding is inconclusive due to the low number of 3s tokens \((n = 80)\) and the virtually identical overt percentages for 3s (30%) and 1s (29%).
4.2.2 **Clause type**
Verbs that appear in main/subordinate clauses were found to favor overt SPs with a probability weight of .68. Contrastively, coordinate clause contexts slightly disfavored overt pro with a weight of .48. Further, verbs that appeared in indeterminate clause types disfavored overt pro. The observation that subordinate clauses favor overt SPs is in line with Morales (1997). Additionally, the finding that main clauses favor overt pro while coordinate clauses disfavor overt pro mirrors what Otheguy and Zentella (2012) observed overall for Spanish in NYC.

4.2.3 **TMA**
In line with previous studies, the verb’s TMA had a significant effect on pronoun expression in the Roswell data (Cameron 1994; Bayley and Pease-Alvarez 1997; Travis 2007). Imperfects most highly favored overt pro (probability weight = .62) and preterits least favored overt pro (probability weight = .38). As mentioned above, this is potentially due to the ambiguity of the first and third-person singular imperfect forms. Another explanation is given by Silva-Corvalán (2001), who discusses imperfects and preterits in relation to discourse functions. Particularly, more overt pro are used with imperfects due to their backgrounded nature, and fewer overt pro are used with preterits since they tend to foreground events. Finally, the perfect and present verb forms in these data showed little effect with weights of .53 and .47, respectively.

4.2.4 **Lexical content**
Regarding lexical content, Roswell speakers follow the trend of favoring overt pro with psychological verbs (probability weight = .57), an observation that Silva-Corvalán (1994) attributes to the fact that psychological verbs tend to express the speaker’s point of view and have an implied contrastive function. In contrast, the other verb types showed a disfavoring of overt pro with a weight of .43. These findings are consistent with previous studies such as Torres Cacoullos and Travis (2010) and Travis and Torres Cacoullos (2012).

4.2.5 **CI**
As reported in all prior studies, same reference contexts favor null SPs and disfavor overt pro (e.g. Silva-Corvalán 1982; Enríquez 1984; Cameron 1994; Travis 2005). Likewise, Roswell Spanish follows this pattern: where there is no change in subject, the use of an overt SP is disfavored with a weight of .45. When there is a change in subject, there is a tendency to produce overt pro (probability weight = .56). However, there is a comparatively weaker effect for this factor as a whole as demonstrated by its low position on the variable hierarchy (5th) (e.g. 2nd on variable hierarchy in monolingual Mexican Spanish [Lastra and Martín Butragueño
2015]). This difference potentially reflects a weakened sensitivity to these particular pragmatic constraints on SP use in the community. Specifically, as other authors have found, there is an increased use of overt pro in same reference contexts (e.g. 18% overt among bilinguals in present study vs. 11% overt for monolingual Mexicans in Yucatan (Michnowicz 2015, 112). This result lends support to previous research and could be taken as evidence of bilingual simplification (e.g. Shin and Otheguy 2009; Michnowicz 2015).

As shown above, the general findings concerning the effect of linguistic variables are in line with previous SP research, but some differences are also apparent. The next section discusses English contact and the possibility of language change within the Hispanic community of Roswell.

5. Discussion

Is English contact a major factor influencing Roswell Spanish? Considering the lack of increase in the overall pronoun rate (21%) and the fact that neither LOR nor AOA had a significant impact on overt pronoun frequencies, there is little or no evidence for the argument of English contact in the use of SPs in the present data. However, a language contact hypothesis should not be completely discarded. Although the overall pronoun rate of Roswell Mexicans did not indicate any considerable divergence from monolingual or incipient contact varieties, a comparison of the effects of certain variables suggests that Spanish in Roswell may be undergoing change. Specifically, the reduced importance observed for CI when compared to monolingual Mexican varieties could be due to a process of bilingual simplification among Roswell speakers whereby they are utilizing a higher proportion of overt SPs in same reference contexts. This finding is consistent with more established contact varieties of Spanish that have shown a decreased sensitivity to CI such as NYC-born speakers (Flores-Ferrán 2002; Shin and Otheguy 2009) as well as Maya-Spanish bilinguals in Yucatan, whose overt rates for same reference doubled compared to monolinguals (Michnowicz 2015). In this regard, speakers in Roswell, despite being at a relatively initial stage of language contact, could be moving toward more established communities’ norms regarding coreferentiality. However, this explanation (as well as the findings in general) is tentative, and the results should be interpreted with caution due to the small sample size, and also to the challenges of comparison across studies that often differ in their statistical methods. Additionally, a direct comparison of coreferentiality effects between first generation immigrants and second generation speakers in the Southeast is warranted.
6. Conclusion

In this paper, I have attempted to describe the behavior of SPs in Southeast Spanish by examining Spanish in Roswell, Georgia, a region that, despite its large and increasing Latino population, lacks the linguistic attention that has been given to regions with more established Spanish-English contact situations in the US. Through a variationist sociolinguistic analysis of the factors contributing to the expression or non-expression of an SP in this particular Spanish variety as well as a comparison of pronoun behavior with other varieties of Spanish, I have been able to draw some conclusions. First, I was able to determine the general pronoun rate among the 10 speakers in the sample and to locate, to a certain degree, the Roswell data with respect to the rates of other US contact varieties as well as monolingual Spanish varieties. I have come to the conclusion that Mexican Spanish in Roswell, at least in terms of overall pronoun rate, is much more similar to monolingual and incipient contact varieties of Mexican Spanish than it is to more established contact varieties. With an overall rate of 21% overt pronouns in Roswell, this is very similar to the rates observed in New York (Otheguy, Zentella, and Livert 2007), New Jersey (Flores-Ferrán 2007), and North Carolina (McKnight 2013), and is virtually identical to the rate observed in Mexico City (Lastra and Martín Butragueño 2015). However, this rate is substantially lower than that observed in Los Angeles, which is a much more established contact variety and thus, may be influenced by English to a greater degree.

Second, I have found that the language internal factors having the strongest effect on pronoun usage in Roswell are the person/number of the verb, the clause type in which the verb appears, and the TMA of the verb. These factors are followed by lexical content of the verb and CI, which have weaker effects on subject expression but are nevertheless statistically significant. On the other hand, the language external factors under analysis – gender, age, LOR in the US, and AOA to the US – were not found to be significant. This lack of effect for LOR and AOA goes against a language contact hypothesis.

However, there are notable discourse-pragmatic differences of Roswell Mexicans when compared to monolingual and incipient contact varieties. The relative strength of CI is weaker from that of other varieties such as mainland newcomers in NYC and monolingual speakers in both Yucatan and Mexico City. This was interpreted, tentatively, as a possible case of bilingual simplification by Roswell speakers; furthermore, the general weakened sensitivity to this factor supports previous research (e.g. Shin and Otheguy 2009) and suggests that language change may be occurring among these Spanish-speakers despite them being part of more recent and less established situations of language contact in the Southeast. But, is this change attributable to English contact?
As part of my broader research goal, I have considered potential influence from English on the Spanish of Roswell. While it is certainly possible that English is a force shaping Roswell Spanish (specifically concerning CI), no conclusive evidence was found of such contact-induced language change given the similarity in overall pronoun rate to monolingual and incipient Mexican varieties as well as the insignificance of social factors, namely LOR and AOA.

While the results of this study reveal, to some degree, a number of important findings about Spanish in Roswell and give us a glimpse into incipient contact varieties in the Southeast more generally, the small sample size of only 10 speakers makes it difficult to generalize such findings as being representative of the speech community as a whole. Therefore, future research should employ a larger sample of speakers as well as compare first and second generation Latinos in the same community. In addition, it would be useful for future studies to consider a wider range of social factors, for example to take into account the speakers’ social class and level of education as well as specific information regarding the speakers’ English proficiency as well as amount of exposure to and use of English. Such a wider scope would be particularly beneficial for gaining more insight into the factor of English contact, and more generally, would lead to a more complete description of Southeast Spanish with regard to subject pronoun expression.

Acknowledgements

I would like to thank Chad Howe and one anonymous reviewer for their invaluable feedback as well as Anna Wilson for the use of her data.

References


Flores-Ferrán, Nydia. 2004. “Spanish subject personal pronoun use in New York City Puerto Ricans: Can we rest the case of English contact?” Language Variation and Change 16: 49–73. doi: 10.1017/S0954394504161048


Sociolinguistics, ed. by Lotfi Sayahi and Maurice Westmoreland, 67–75. Somerville, MA: Cascadilla Proceedings Project


Philip P. Limerick


Author’s address

Philip P. Limerick
University of Georgia
Department of Romance Languages
Athens, GA 30602
USA
nztm@uga.edu

Biographical notes

Philip P. Limerick is a PhD student in Hispanic Linguistics in the Department of Romance Languages at the University of Georgia. His primary research interests are sociolinguistics, language variation and change, language contact, and pragmatics. His current projects involve the analysis of the preterit/present perfect distinction as well as person-reference and identity in Spanish in Georgia. He has a forthcoming chapter entitled ‘Variable clitic placement in U.S. Spanish,’ which will appear in the book series Issues in Hispanic and Lusophone Linguistics.

Publication history

Date received: 1 February 2015
Revision requested: 7 June 2015
Date accepted: 10 June 2015