

## Subject Pronoun Expression in Mexican Spanish: *¿Qué pasa en Xalapa?*

Rafael Orozco\*

**Abstract.** This study is the first variationist analysis of subject personal pronoun expression (SPE) in the Spanish of Xalapa, Mexico. The overall pronominal rate (25%)—the highest such rate found in Mexican Spanish so far—also constitutes one of the highest in a mainland Spanish variety. Six predictors—four internal and two external—significantly condition SPE. The internal conditioning—congruent with what occurs elsewhere—reveals grammatical number and person of the subject as the strongest predictor. It also shows that verb class has tendencies similar to those found in other communities. However, further analysis uncovers that lexical frequency provides more definite answers regarding how verbs condition SPE, as within the copulative verb class category *ser* ‘be’ favors overt subjects but *estar* ‘be’ favors null subjects. Moreover, the unusually robust effect of age sets Xalapa Spanish apart from most other varieties. Interestingly, the pronominal rate among teenagers (11%)—below the lowest overall pronominal rate anywhere—is consistent with what occurs in other Spanish varieties such as Colombian, European, Dominican, and Mexican. These findings call for further research on the effects of verb semantics and age on SPE.

**Keywords.** Subject pronoun expression (SPE); sociolinguistics; language variation; Latin American Spanish; Mexico;

**1. Introduction.** The variable presence and absence of subject personal pronouns (e.g. *ellos cantan* alternating with *cantan* to mean ‘they sing’) constitutes a morphosyntactic feature that Spanish inherited from Latin. Variationist subject pronoun expression (SPE) studies were pioneered by Barrenechea & Alonso (1973), Bentivoglio (1980), and Morales (1980) who explored the Spanish spoken in Buenos Aires, Argentina; Caracas, Venezuela; and San Juan, Puerto Rico, respectively. Those groundbreaking studies led to numerous others that have explored the alternation between null and overt pronominal subjects in Caribbean Spanish (Alfaraz 2015; Bentivoglio 1987, Orozco 2015, Ortiz López 2009; among others), Mainland Latin American Spanish (Cerrón Palomino 2014, Lastra & Martín Butragueño 2015; Travis 2005a, 2005b), Peninsular Spanish (Cameron 1993, Enríquez 1984, Posio 2011, de Prada Pérez 2009, 2015), and Spanish in the United States (Bayley & Pease-Alvarez 1997; Cameron 1992, 1995, 1996, 1998; Cameron & Flores-Ferrán 2004; Flores-Ferrán 2002, 2004, 2007; Hurtado; 2001, Otheguy & Zentella 2007, 2012; Otheguy, Zentella & Livert 2007; Silva-Corvalán 1982, 1994, 1997; and others). These studies have determined that variable SPE displays marked regional differences in terms of overt pronoun rates, and that overall frequency of use differs dialectally. The highest overt pronoun expression rates are found in the Caribbean (~38%) while lower pronominal rates occur in Spain and mainland Latin American varieties such as those of Mexico and Ecuador (~21%).

\* This research was partially supported by *Awards To Louisiana Artists and Scholars (ATLAS)* under grant # LEQSF(2011-12)RD-ATL-05. I am grateful to my research assistants Maritza Nemogá and Monika Estrada Andino for their valuable help with transcription, token extraction, and coding. Most of all, I am greatly thankful to the residents of the city of Xalapa, Veracruz, Mexico who kindly provided the data for this study.  
Author: Rafael Orozco, Louisiana State University (rorozc1@lsu.edu).

Despite overt pronominal rate differences, the existing body of literature reflects relative uniformity in the tendencies exhibited by the predictors that condition SPE (Carvalho, Orozco, & Shin 2015:xiii). Although this linguistic variable has been extensively explored throughout the Hispanic World, it remains understudied in Mexican Spanish. Thus, besides adding to previous treatments of variable SPE in the Spanish of Mexico by Lastra & Martín Butragueño (2015), Michnowicz (2015), Shin (2012, 2015), and Shin & Erker (2015), this pilot study seeks to provide a data baseline for further research. Concurrently, this investigation intends to contribute to the emerging sociolinguistic exploration of the State of Veracruz initiated almost a decade ago by Brizuela Casimir (2008).

**2. Methodology.** This section describes the speech community and the data sample used in this study. It also states the research questions, and discusses the predictors included in the analysis.

**2.1. THE SPEECH COMMUNITY, THE CORPUS, AND THE DATA SAMPLE.** Xalapa (Figure 1) was founded by Spanish explorers in 1739, at the same location where four indigenous settlements dating back to 1313 had existed. This city is the capital of the Mexican state of Veracruz and has a population of approximately half a million people. Despite being located in a coastal state and roughly 80 miles from the Gulf of Mexico, the Spanish of Xalapa features the phonological characteristics of the Mexican inland regions.



Figure 1. Xalapa and the State of Veracruz, Mexico <sup>1</sup>

The data sample examined here was extracted from the *Corpus del Castellano Xalapeño* (CorCaXa). This corpus consists of 35 hours of sociolinguistic conversations with 30 (15 women and 15 men) socially stratified residents of the Xalapa metropolitan area. These consultants—born between 1936 and 1994—had spent most of their lives in the Xalapa metropolitan area when the data was collected. The orthographic transcription of this corpus consists of 161,500 words.

**2.2. RESEARCH QUESTIONS AND HYPOTHESIS.** This investigation contributes to the formation of a baseline of data on contemporary SPE in monolingual Spanish speaking communities. In exploring the constraints that affect the alternation of overt and null pronominal subjects in the Spanish of Xalapa, I seek to answer three main research questions. These questions are guided by the findings of numerous previous studies including most of those cited in the preceding paragraphs.

<sup>1</sup> Image from pickatrail.com (<http://pickatrail.com/>).

- a) *How are overt and null pronominal subjects distributed in the Spanish of Xalapa? How does this variety compare with other varieties of Spanish in terms of subject pronoun expression?*
- b) *Is the internal conditioning on subject pronoun expression in Xalapa Spanish similar to that throughout the Hispanic World? Do all verbs within the same syntactic/semantic category similarly condition SPE?*
- c) *How do age and gender condition SPE in Xalapa? How do their effects in this speech community compare to those in other communities?*

Concurrently, I seek to test the following hypothesis: *Different verbs within a single semantically-based category condition SPE differently*. This hypothesis has been informed by studies indicating that we lack conclusive information as to the effects of the verb on SPE (cf., Erker & Guy 2012; Orozco 2015; Orozco, Méndez Vallejo, & Vidal-Covas 2014; Orozco & Guy 2008; Posio 2011; *inter alia*). I additionally test a series of hypotheses that directly address each one of the predictors explored here and discussed below.

**2.3. PREDICTORS EXPLORED.** To answer the above research questions and test my hypotheses, I explored the effects of four internal and two external predictors. I based my choice of predictors on the findings of a multitude of previous SPE investigations (cf. Cameron 1992, 1993, 1995; Enríquez 1984; Flores-Ferrán 2002, 2004, 2007; Lastra & Martín Butragueño 2015; Orozco 2015; Otheguy & Zentella 2007, 2012; Otheguy, Zentella & Livert 2007; Torres Cacoullos & Travis 2010, 2011; Travis 2005a, 2005b, 2007; Travis & Torres Cacoullos 2012; among others). The four internal predictors analyzed here are (1) Verb class, (2) Prior subject's grammatical person and number, (3) Verbal tense, mood and aspect (TMA) of the verb, and (4) Grammatical person and number of the subject. The two external predictors analyzed are Gender and Age.

**2.4. THE ENVELOPE OF VARIATION AND THE ANALYSIS.** The envelope of variation used here adheres to the Principle of Accountability (Labov 1972:72). It also follows the comprehensive parameters defined by Otheguy & Zentella (2012:48 ff.), which are regarded as standard for SPE studies. I included in the data sample only those clauses with ascertainable animate pronominal subjects that contain a conjugated verb where the alternation between a null and an overt subject is clearly possible. Thus, all tokens constitute one of at least two possible different ways of saying the same thing. The data sample used in this study is comprised of 3,656 tokens. I coded all tokens in terms of the predictors discussed above and conducted a series of multivariate statistical regression analyses using Rbrul as my statistical tool.

In the sections that follow, as I walk the reader through my results, I begin by setting forth the distribution of the overt and null SPPs. My discussion of the internal conditioning on pronominal usage precedes that of external constraints. I subsequently, draw conclusions and formulate their implications.

**3. Distribution of variable pronominal subjects and predictor model.** The distribution of overt and null pronominal subjects is presented in Table 1. Xalapa's overall 24.8% overt pronominal rate is higher than Mexico City's 21.7% (Lastra & Martín Butragueño 2015) and Yucatan's 20% (Michnowicz 2015). In fact, it constitutes one of the highest overt pronominal rates found in a monolingual mainland speech community, as mainland pronominal rates are consistently below 28% (cf. Lastra & Martín Butragueño 2015; Michnowicz 2015; Orozco & Guy 2008; Otheguy & Zentella 2007, 2012). One reason for this relatively high pronominal rate

may be Xalapa’s geographical location in a state that is on the Gulf of Mexico and in close proximity to Caribbean varieties, which are known for their high pronominal rates.

Variant	N	%
Overt subjects ( <i>ellos cantan</i> ‘they sing’)	907	(24.8%)
Null subjects ( <i>Ø cantan</i> ‘[they] sing’)	2,749	(75.2%)
Total	3,656	(100%)

Table 1: Distribution of overt and null subjects

The results of the multivariate analysis reveal a complex model that includes linguistic and social forces with all six predictors explored (four internal and two external) reaching statistical significance (See Table 2). The order of selection shows person & number of the subject as the strongest predictor with a *p*-value of  $1.05^{-75}$ . Interestingly, the effect of age, which registers the second highest *p*-value, appears to be unusually high. In general, internal constraints have a greater conditioning effect on overt SPP occurrence based on their order of selection and *p*-values. Moreover, the constraint hierarchy found in Xalapa with (a) grammatical person and number of the subject and (b) prior subject’s grammatical person and number being the strongest internal predictors is largely consonant with findings around the Hispanic World including Barranquilla, Colombia (Orozco 2015); Los Angeles (Silva-Corvalán 1982, 1997); Madrid, (Enríquez 1984); Mexico City (Lastra & Martín Butragueño 2015); Puerto Rico (Cameron 1993, 1995); New York City (Otheguy & Zentella 2007, 2012), Rivera, Uruguay (Carvalho & Bessett 2015); and Yucatan, Mexico (Michnowicz 2015); *inter alia*. This finding corroborates that despite varying pronominal rates at the surface level, the grammar underlying SPE across varieties remains essentially the same (Cameron 1993; Michnowicz 2015; Travis 2007; Torres Cacoullos & Travis 2010). A detailed discussion of the effects of the predictors conditioning SPE in Xalapa follows.

Predictor	<i>p</i> -value
Person & number of the subject	$1.05^{-75}$
Age	$9.63^{-21}$
Prior subject’s grammatical person & number	$1.1^{-17}$
Verb class	$1.22^{-11}$
Verbal tense, mood & aspect (TMA)	$1.08^{-07}$
Gender	0.0162

Table 2: The statistical model for SPE in Xalapa

**4. Internal conditioning.** The internal conditioning on SPE reveals the effects of two predictors pertaining to the subject (grammatical person and number of the subject and prior subject’s grammatical person and number) and two verb-related predictors (Verb class and TMA). My discussion of the internal predictors that condition SPE in Xalapa Spanish follows. Individual predictors are discussed in descending order of strength, i.e., according to the *p*-values provided in Table 2. Thus, I will first deal with the subject-related predictors and subsequently with those pertaining to the verb.

**4.1. GRAMMATICAL PERSON AND NUMBER OF THE SUBJECT.** This predictor is known to most strongly condition SPE across the board. Singular pronouns consistently favor the occurrence of overt subjects and register higher pronominal rates than plural pronouns (Carvalho, Orozco, &

Shin 2015:xiv). Nevertheless, in Mexican Spanish, pronominal rates vary by grammatical person. In Mexico City, third person singular pronouns register the highest overt pronominal rate among all pronouns with 27% (Lastra & Martín Butragueño 2015:43). Conversely, in Yucatan, first person singular overt subjects are the most frequent with 28% whereas third person singular pronouns register 21% (Michnowicz 2015:109). Thus, among other things, I intended to determine whether the effect of grammatical person in Xalapa is more similar to either of these Mexican communities. The results are presented in Table 3.

Factor	Prob. <sup>2</sup>	% overt	N
3 <sup>rd</sup> singular ( <i>él</i> ‘he,’ <i>ella</i> ‘she’)	.71	39%	507
1 <sup>st</sup> singular ( <i>yo</i> , ‘I’)	.65	33%	1,557
2 <sup>nd</sup> singular ( <i>tú</i> , ‘you’)	.44	19%	308
All plural ( <i>nosotros</i> ‘we,’ <i>ustedes</i> ‘you,’ <i>ellas</i> ‘they,’ <i>ellos</i> ‘they’)	.21	9%	1,173
<i>Range</i> = 50		<i>p</i> -value = 1.05 <sup>-75</sup>	

Table 3: Effects of grammatical person & number

The effects of grammatical person and number of the subject show that the third person singular strongly favors overt pronominal subjects with a probability value of .71 (pronominal rate 39%). The first person singular also favors overt subjects (.65). On the other hand, the second person singular and all plural grammatical persons combined favor null subjects with probability values of .44 and .21, respectively. These findings are largely commensurate with the wide-ranging tendency (Abreu 2009, 2012; Bayley & Pease-Álvarez 1997; Carvalho & Child 2011; Erker & Guy 2012; Flores-Ferrán 2002, 2004, 2007, 2009; Orozco 2015, under review; Otheguy & Zentella 2007, 2012; Otheguy, Zentella, & Livert 2007; Posio 2011; Ortiz López 2011, among others) that singular SPPs occur more frequently as overt subjects than plural pronouns.

Despite the large-scale similarities with other speech communities in terms of grammatical number, some subtle differences within the singular persons exist. There is a series of differences between the first and the third persons. As shown in Table 3, and congruent with what occurs in Mexico City (Lastra & Martín Butragueño 2015:43), the third person has the highest proun rate (39%). This makes the effect of grammatical person in Xalapa more similar to that in Mexico City but less similar to that in Yucatan and Oaxaca, respectively, where the first person has the highest overt pronominal rate (Michnowicz 2015:109; Shin 2015:13). Further, the effect of the third person singular, by most strongly favoring overt subjects, sets Xalapa apart from other speech communities including NYC Mexicans (Shin & Erker 2015:179), Mexico City (Lastra & Martín Butragueño 2015:43), Barranquilla, Colombia (Orozco 2015:27), San Juan, Puerto Rico (Claes 2011:199), and Yucatan, Mexico (Michnowicz 2015:109), where first person singular pronouns favors most strongly the occurrence of overt pronominal subjects. Concurrently, second person singular pronouns, by promoting null subjects, behave differently from their first and third person singular counterparts. Although the favorable effect of the second person singular on null subjects in Xalapa is consonant with what occurs among Mexican speakers (Lastra & Martín Butragueño 2015:43; Shin & Erker 2015:179), it is different from what happens in other communities such as Barranquilla and the NYC Colombian enclave (Orozco 2015, under review) where

<sup>2</sup> In this and all subsequent tables reporting probability values, the closer individual values are to 1, the more they favor the occurrence of an overt subject. The closer a value is to zero, the more it favors a null subject.

the second person singular favors overt subjects. Thus, we might be in the presence of a feature that differentiates Mexican Spanish from other varieties. Further research is needed to better inform our knowledge of the effects of grammatical person and number and, specifically, to uncover the differences in the effects of the singular grammatical persons on SPE.

4.2. PRIOR SUBJECT'S PERSON AND NUMBER. My analysis of this predictor aims at learning more about the effects of priming on SPE. I initially coded my data using four factors: (1) same person and number, (2) same person different number, (3) different person same number, (4) different person and number. Preliminary results revealed similar tendencies for all cases of different grammatical person. Thus, I combined them into a single factor, and obtained the results presented in Table 4.

Factor	Prob.	% overt	N
Different grammatical person	.58	30%	1,632
Same person, different number	.55	21%	372
Same person & number	.37	20%	1,501
<i>Range = 21</i>			<i>p-value = 1.1<sup>-17</sup></i>

Table 4: Effects of prior subject's person & number

The effects of prior subject's person and number uncover that prior subjects with a different grammatical person, regardless of number, favor overt subjects with a probability of .58. Prior subjects with the same person show opposing tendencies according to grammatical number, as follows. Those with different number favor overt subjects with a statistical weight of .55 whereas prior subjects with the same person and number favor null subjects with a value of .37. Thus, continuity of a given subject manifested by a subject with the same person and number as its predecessor promotes null subjects while a change in person and/or number favors overt pronominal subjects. These findings confirm that the presence of one form or structure motivates the occurrence of subsequently higher frequencies of the same forms or structures (Bayley & Pease-Álvarez 1997; Cameron & Schwenter 2013:476; Flores-Ferrán 2002; Shin & Otheguy 2009:128; Travis 2005b, 2007; Torres Cacoullos & Travis 2015; among others). Moreover, these results further inform our collective knowledge of the effects of both switch reference and priming since a referential switch goes hand in hand with a change in grammatical subject.

4.3. VERB CLASS. As done by Orozco (2015), in exploring this predictor, I adapted the pioneering classification advanced by Bentivoglio (1980). By testing this predictor together with TMA, I sought further evidence that despite the effects of TMA being fairly uniform across the board, the effects of verb class, and, more specifically, lexical frequency on SPE are not uniform across varieties of Spanish (cf. Orozco et al. 2014).

The findings for verb class (presented in Table 5) reveal that copulative, perception, and motion verbs, respectively, promote overt pronominal subjects. Copulative and perception verbs exert similar favorable effects on overt subjects with respective probability weights of .63 (pronominal rate 35%) and .61 (pronominal rate 37%). Motion verbs register a rather modest favorable effect with a probability value of .53 and a pronominal rate of 22%. At the same time, both cognition and speech verbs promote null subjects with probability weights of .39 and .35, respectively. Verbs in the “other” category have a neutral effect with a statistical value of .49 and a pronominal rate of 21%. The favorable effects of copulative and perception verbs on overt pronominal subjects are consonant with previous findings (Enríquez 1984:240; Orozco 2015:24,

under review; Otheguy & Zentella 2012:164; Shin 2015; Silva-Corvalán 1994:162; Torres Cacoullos & Travis 2011: 250; Travis 2007:115; among others).

Factor	Prob.	% overt	N
Copulative	.63	35%	410
Perception	.61	37%	409
Motion	.53	22%	556
Other	.49	21%	1,583
Cognition	.39	23%	314
Speech	.35	19%	273
<i>Range</i> = 28		<i>p</i> -value = 1.22 <sup>-11</sup>	

Table 5: Effects of verb class

On the other hand, the effects of motion, speech, and cognition verbs, respectively, do not appear to be entirely consistent with what occurs in other speech communities. In Xalapa motion verbs (probability value .53) promote overt pronominal subjects whereas they have the opposite effect in Cali, Colombia (Travis 2005b:340, 2007:115), Barranquilla, Colombia (Orozco 2015:24), Lima, Peru (Cerrón-Palomino 2014:69), and New Mexico (Travis 2007:115). In the present analysis verbs of speech (probability value .35) favor null subjects while in Barranquilla, Cali, New Mexico, and among NYC Colombians (cf. Travis 2007; Orozco 2015, under review) these verbs have the opposite tendency by favoring overt SPPs. By the same token, verbs of cognition (probability value .39) promote null subjects in Xalapa but have a neutral effect in both Barranquilla and the New York Colombian community. Furthermore, in analyses where verbs of cognition and perception have been grouped as psychological verbs (cf. Travis 2005b, 2007; Orozco 2015:23), psychological verbs have been found to promote overt pronominal subjects. Concurrently, the conditioning effect of verb class on SPE appears to constitute another differentiating factor between different Mexican Speech communities. Although in Xalapa and among NYC Mexicans (Shin & Erker 2015:179) verb class conditions SPE, it does not in Mexico City (Lastra & Martín Butrageño 2015:43) nor Yucatan (Michnowicz 2015:109).

The apparent inconsistencies in the effects of most verb classes (motion, other, cognition, speech) when they are compared across different speech communities suggest that the effects of the verb on SPE may exceptionally lack the overarching uniformity evidenced by other internal predictors and factors (cf. Carvalho, Orozco, & Shin 2015 and references therein; Lastra & Martín Butrageño 2015:53). These findings along with those of Erker & Guy 2012; Orozco 2015, under review; Orozco Méndez Vallejo & Vidal Covas 2014; Posio 2011, 2015; Travis 2005b, 2007; among others imply that verb class, or for that matter another classification based on syntactic or semantic criteria, does not constitute the most accurate way to explore the effects of the verb on SPE. Moreover, our findings appear to suggest that the effect of the verb on SPE may be idiosyncratic. Further study of how the verb conditions SPE using criteria other than syntactic or semantic verb groupings shall provide more definite answers.

**4.4. TENSE, MOOD AND ASPECT.** Based on the results of previous investigations (cf. Erker & Guy 2012; Orozco 2015, under review; Otheguy & Zentella 2012; among others), and to facilitate comparisons with Mexico City Spanish (Lastra & Martín Butrageño 2015), I used the following four factors for this analysis: (1) Imperfect indicative, (2) present indicative, (3) preterite indicative, and (4) all other tenses. The findings for TMA are presented below in Table 6.

Factor	Prob.	% overt	N
Imperfect indicative	<b>.61</b>	29%	906
Present indicative	.48	25%	1,551
All other tenses	.48	24%	333
Preterite indicative	.43	20%	755
<i>Range = 18</i>		<i>p-value = 1.08<sup>-07</sup></i>	

Table 6: Effects of verbal tense, mood, and aspect

The results show that the imperfect indicative tense favors overt pronominal subjects with a .61 probability value. The preterite indicative has the opposite effect, i.e., favors null subjects, with a value of .43. Concurrently, the present indicative and all other tenses, respectively, have rather neutral effects with equal probability values of .48 (see Table 6). Interestingly the two tenses that express past-time actions or events constitute polar opposites, as the imperfect promotes overt subjects and the preterite indicative favors null subjects.

The favorable effect of the imperfect indicative on overt subjects concurs with findings in other communities in Mexico, specifically Oaxaca (Shin & Erker 2015:180) and Yucatan (Michnowicz 2015:109). Moreover, the tendencies registered by the different TMA factors are largely congruent with those in Mexico City (Lastra & Martín Butragueño 2015:43). These results imply that the effects of TMA are similar across Mexico. These tendencies are further commensurate with what occurs throughout the Hispanic World in monolingual speech communities (Abreu 2009, 2012; Cameron 1993; Claes 2011; Orozco 2015; Travis 2005b, 2007; among others) as well as in situation of contact with other languages (Bayley & Pease-Álvarez 1997; Erker & Guy 2012; Hochberg 1986; Otheguy & Zentella 2007, 2012; Shin & Erker 2015:180). Thus, our results support the premise that morphologically ambiguous verbal paradigms such as the imperfect tense promote more overt pronominal subjects than unambiguous forms. Furthermore, the effects of TMA provide additional evidence as to the consistency of the grammar across varieties of Spanish (cf. Carvalho et al. 2015).

4.5. ANOTHER TAKE AT THE EFFECTS OF THE VERB ON SPE USING LEXICAL FREQUENCY. The following analysis addresses the need for more definite answers as to how verbs condition SPE discussed above in Section 4.3. I explore the effects of lexical frequency by testing the effects of the ten most frequently occurring verbs in the sample as random effects factors.

Factor	Prob.	% overt	N
<i>ir</i> ‘go’	<b>.67</b>	36%	100
<i>ser</i> ‘be’	.59	43%	268
<i>poder</i> ‘be able to, can’	.58	32%	94
<i>creer</i> ‘believe’	.57	48%	102
<i>hacer</i> ‘make, do’	.47	18%	111
<i>decir</i> ‘say, tell’	.46	18%	188
<i>tener</i> ‘have’	.45	18%	311
<i>ver</i> ‘see’	.42	23%	103
<i>estar</i> ‘be’	.37	18%	131
<i>saber</i> ‘know’	.29	12%	156

Table 7: Effects of lexical frequency as a random effects predictor for the 10 most frequent verbs in the data sample

The lexical frequency analysis—whose results are presented in Table 7—reveals that four of the ten most frequent verbs in the data sample (*ir*, *ser*, *poder*, and *creer*) promote overt subjects. *Ir* ‘go’ is the strongest promoter of overt subjects with a probability value of .67 and an overt pronominal rate of 36%. Contrariwise, the remaining six verbs favor null subjects. A comparison of these findings with the results for verb class—discussed above in §4.3 and presented in Table 5—uncovers discrepancies between verbs that fall within three of the verb class categories, as follows. Firstly, among copulative verbs *ser* ‘be’ favors overt subjects with a probability value of .59 and the second highest pronominal rate (43%) whereas *estar* ‘be’ promotes null subjects with a weight of .37 and one of the lowest pronominal rates (18%). Secondly, among perception verbs *creer* ‘believe’ promotes overt subjects with a probability value of .57 while *ver* ‘see’ has the opposite effect with .42. Thirdly, within the “other verbs” category whereas *poder* ‘be able to, can’ favors overt subjects with a value of .58, *tener* ‘have’ promotes null subjects with a weight of .45. At the same time, our random effects analysis also uncovers discrepancies between the effects of verbs within the lexical content of the verb classification advanced by Enríquez (1984). For instance, within the external activity category, *ir* ‘go’ promotes overt pronominal subjects, but both *decir* ‘say, tell’ and *ver* ‘see’ favor null subjects.

These findings suggest that exploring the effects of the verb on SPE by using classifications based on syntactic, pragmatic, or semantic criteria such as verb class does not considerably inform our collective knowledge beyond what we already know, as such classifications may leave important differences uncovered. It appears that a random effects analysis using a lexical frequency approach can help us provide a more detailed account of how verbs condition SPE. Moreover, these results concur with recent findings (Posio 2011, 2015; Orozco 2015, under review; Orozco et al. 2014) in providing mounting evidence that despite four decades of research, we are yet to know the real effects of the verb on SPE. It appears that by using analyses that take into account the configuration of the corpus at hand, and, most importantly, lexical frequency, we will be able to obtain more conclusive answers as to the effects of the verb on SPE and, perhaps, other linguistic variables. Fortunately, such analyses can be done with the 21<sup>st</sup> century state-of-the art statistical tools that are now available to us.

**5. External conditioning.** While we possess a great deal of information as to the effects of internal predictors, information as to how external predictors constrain SPE is less abundant. With this analysis, I sought to probe the contention that social constraints do not significantly condition SPE in monolingual speech communities (cf. Alfaraz 2015:5; Bentivoglio 1987; Cameron 1992, 1993; Lastra & Martín Butragueño 2015:41).

**5.1. SPEAKER’S AGE.** With a *p*-value of 9.63<sup>21</sup>, age constitutes the second strongest predictor overall in the present analysis. This robust effect of age sets Xalapa apart from most other Hispanic speech communities, as it does not consistently condition SPE (Alfaraz 2015:11; Bentivoglio 1987; Michnowicz 2015). In the cases where age significantly constrains SPE, its effect is rather weak (e.g., Carvalho & Bessett 2015:151; Lastra & Martín Butragueño 2015:41; Orozco 2015).

In this analysis, I divided my consultants into five different groups according to their year of birth. The findings, presented in Table 8, do not reveal any discernible age patterns. Speakers born in the 1970s, i.e., those 30-39 years old most strongly favor overt SPPs with a statistical weight of .63. Those born in the 1980s as well as those born before 1950 also favor overt subjects, both groups with a probability value of .57. Speakers born in the 1950s and 1960s have a neutral effect with a value of .48. Teenagers strongly favor null subjects with a probability value of .28 and an overt pronominal rate of 11%.

Factor	Prob.	% overt	N
Born in the 1970s (30-39 years old)	.63	33%	535
Born in the 1980s (20-29 years old)	.57	29%	1,014
Born before 1950 (>60 years old)	.57	28%	778
Born in the 1950s and 1960s (40-59)	.48	23%	549
Born in the 1990s (teenagers)	.28	11%	669
<i>Range = 35</i>		<i>p-value = 9.63<sup>21</sup></i>	

Table 8: Effects of age

Given that Spanish is considered to be evolving toward becoming a semi-pro-drop language, with higher occurrences of overt pronominal subjects, speakers in their 30s seem to be leading the evolutionary trend. Conversely, teenagers by strongly favoring null subjects and registering a very low pronominal rate of 11% appear to be disfavoring the change toward increased pronoun occurrence. Interestingly, the lowest pronominal rate among the different age groups in this study registered by Xalapa's teenagers is consistent with findings elsewhere in Mexico. The youngest segment of Mexico City speakers registers the lowest pronominal rate (Lastra & Martín Butragueño 2015:49). Moreover, adolescents in Oaxaca also register an overt pronominal rate of 11% (Shin 2015:12). Similarly, lower overt pronominal rates have been reported for other Spanish varieties such as Peninsular (de Prada Perez 2015), Colombian (Orozco 2015, forthcoming), and Dominican Spanish (Alfaraz 2015). Thus, the general effect of age in Xalapa is commensurate with what occurs in other speech communities with younger individuals using fewer overt pronouns than their elders (Alfaraz 2015; Lastra & Martín Butragueño 2015; Orozco 2015; Orozco & Guy 2008).

Concurrently, over half a century of variationist research on the effects of age, recognized as the principal social correlate of language change (Chambers 2002:349) shows younger speakers consistently promoting linguistic innovations (Labov 2001:437). Thus, the trend exhibited by Xalapa's youth can be explained as an acquisitional feature of SPE that goes hand in hand with age-grading. As children acquire adult SPE usage patterns, their pronominal rates increase gradually (Shin 2015: 11; Shin & Erker 2015). Further research on the effects of age on SPE in Spanish and other languages shall shed more light on this fascinating issue.

5.2. GENDER. The findings for gender are presented in Table 9. They show a gender gap as Xalapa men favor overt pronominal subjects with a probability value of .53. Conversely, women promote null subjects with a probability of .47.

Factor	Prob.	% overt	N
Men	.53	25%	1,745
Women	.47	24%	1,800
<i>Range = 6</i>		<i>p-value = .0162</i>	

Table 9: Effects of gender

The significant effect of gender in this analysis sets Xalapa apart from other Mexican speech communities as gender does not constrain SPE in Mexico City (Lastra & Martín Butragueño 2015:43) nor Yucatan (Michnowicz 2015:109). In fact, gender is not reported to condition SPE in a number of other communities (cf. Carvalho & Bessett 2015:151; Travis 2007). At the same time, the effects of gender on SPE in Xalapa run contrary to those in other Spanish varieties including Dominican (Alfaraz 2015:11), Colombian in Barranquilla and New York City, (Orozco

2015:30, under review), and Peninsular (Prada Pérez 2015:126). It is apparent that gender does not constitute a strong SPE predictor, and it has been found to interact with age (Orozco 2015, under review) and other predictors (Shin & Otheguy 2013). Thus, future research into the effects of gender on SPE could benefit from exploring its effects in analyses that combine gender with other external constraints.

**6. Summary and conclusion.** The results of this exploratory study contribute to the formation of a baseline of data for further inquiry into Mexican Spanish. Answering our first research question, the overall pronominal rate for Xalapa Spanish (24.8%) constitutes a relatively high rate among mainland Spanish varieties. The answer to the second research question reveals that the internal conditioning on SPE, with grammatical person and number of the subject as the strongest predictor in Xalapa, is largely similar to that throughout the Hispanic World. Concurrently, a more detailed analysis of the effects of the verb supports the main hypothesis tested here by revealing that all verbs within a given class do not condition SPE similarly. Specifically, there are discrepancies between verbs in the copulative, perception, and “other” categories. These discrepancies are best illustrated by the copulative verbs with *ser* ‘be’ favoring overt subjects and *estar* ‘be’ having the opposite effect.

The effects of social constraints provide the answer to my third research question. In spite of reports that social constraints do not significantly condition SPE in monolingual Spanish varieties (Bayley & Pease-Alvarez 1997; Bentivoglio 1987; Cameron 1992, 1993; Flores-Ferrán 2002; Martínez-Sanz 2011), this study concurs with Ávila-Jiménez (1995), Alfaraz (2015), Lastra & Martín Butragueño (2015), and Orozco (2015) in finding that external constraints indeed condition pronominal expression. In fact, the effect of age—the second strongest predictor overall—is quite robust, as it conditions SPE more strongly than verb class, TMA, and prior subject’s person and number, respectively. Interestingly, teenagers, with an overt pronominal rate of 11%, strongly promote null subjects. This tendency, although perhaps surprising at first sight, is congruent with findings in other speech communities. Additionally, it corroborates findings indicating that the pronominal rates of children and adolescents increase gradually as they acquire adult SPE usage patterns (Shin 2015:11, Shin & Erker 2015).

Notwithstanding the wide-ranging similarities in terms of the constraint hierarchy and the effects of internal factors suggesting that the grammar underlying SPE across varieties of Spanish is essentially the same (Cameron 1993; Carvalho et al. 2015 and references therein), our findings show some subtle differences with other speech communities in the effects of verb class and grammatical person and number of the subject. For instance, the third person singular has both the strongest favoring effect on overt SPPs and the highest pronominal rate (39%), setting Xalapa apart from other speech communities where first person singular pronouns have the strongest effect on the occurrence of overt pronominal subjects (cf. Claes 2011; Lastra & Martín Butragueño 2015; Orozco 2015; among others). Furthermore, results suggest that the effects of the verb on SPE (a) may lack the overarching uniformity shown by other internal predictors and factors, and (b) may be idiosyncratic to each speech community.

In general, this study contributes to enrich our knowledge of SPE. It also helps to open new research paths as it highlights shortcomings in how we have been exploring the effects of the verb on SPE, and perhaps other linguistic variables. Moreover, our results show differences between the effects of internal constraints in Xalapa and other communities. Further study shall provide more definite information regarding the nature of these differences as well as to the effects of age, which appear to be unusually strong in this speech community.

## References

Abreu, Laurel. 2009. Spanish subject personal pronoun use by monolinguals, bilinguals and second language learners. Gainesville, FL: University of Florida dissertation.

Abreu, Laurel. 2012. Subject Pronoun Expression and Priming Effects among Bilingual Speakers of Puerto Rican Spanish. In Kimberly Geeslin & Manuel Díaz-Campos (eds.), *Selected Proceedings of the 14th Hispanic Linguistics Symposium*. 1-8. Somerville, MA: Cascadilla Proceedings Project.

Alfaraz, Gabriela. 2015. Variation of Overt and Null Subject Pronouns in the Spanish of Santo Domingo. In Ana M. Carvalho, Rafael Orozco & Naomi Lapidus Shin (eds.), *Subject Pronoun Expression in Spanish: A cross-dialectal perspective*. 3-17. Washington DC: Georgetown University Press.

Ávila-Jiménez, Bárbara I. 1995. A sociolinguistic analysis of a change in progress: Pronominal overtness in Puerto Rican Spanish. *Cornell Working Papers in Linguistics* 13. 25-47.

Barrenechea, Ana María & Alicia Alonso. 1973. Los pronombres personales sujetos en el español de Buenos Aires. In K. Karl-Hermann & K. Rühl (eds.), *Studia Iberica: Festschrift für Hans Flasche*. 75-91. Bern; München: Francke.

Bayley, Robert & Lucinda Pease-Alvarez. 1997. Null Pronoun variation in Mexican-descent children's narrative discourse. *Language Variation and Change* 9. 349-371.

Bentivoglio, Paola. 1980. *Why canto and not yo canto? The problem of first-person subject pronoun in spoken Venezuelan Spanish*. Los Angeles: University of California Los Angeles M.A. thesis.

Bentivoglio, Paola. 1987. *Los sujetos pronominales de primera persona en el habla de Caracas*. Caracas: Universidad Central de Venezuela, Consejo de Desarrollo Científico y Humanístico.

Brizuela Casimir, Hernán. 2008. *Variables estigmatizadas y actitudes hacia la lengua: Un estudio de caso en Xalapa, Veracruz*. Xalapa, Mexico: Universidad Veracruzana M.A. thesis.

Cameron, Richard. 1992. *Pronominal and null subject variation in Spanish: Constraints, dialects, and functional compensation*. Philadelphia, PA: University of Pennsylvania dissertation.

Cameron, Richard. 1993. Ambiguous agreement, functional compensation, and non-specific *tú* in the Spanish of San Juan, Puerto Rico and Madrid, Spain. *Language Variation and Change* 5. 305-334.

Cameron, Richard. 1995. The scope and limits of switch reference as a constraint on pronominal subject expression. *Hispanic Linguistics* 6/7. 1-27.

Cameron, Richard. 1996. A community-based test of a linguistic hypothesis. *Language in Society* 25(1). 61-111.

Cameron, Richard. 1998. A variable syntax of speech, gesture, and sound effect: Direct quotations in Spanish. *Language Variation and Change* 10(1). 43-83.

Cameron, Richard & Nydia Flores-Ferrán. 2004. Preservation of subject expression across regional dialects of Spanish. *Spanish in Context*. 1(1). 41-65.

Cameron, Richard & Scott Schwenter. 2013. Pragmatics and Variationist sociolinguistics. In Robert Bayley, Richard Cameron & Ceil Lucas (eds.), *The Oxford Handbook of Sociolinguistics*. 464-483. Oxford and New York: Oxford University Press.

Carvalho, Ana M. & Michael Child. 2011. Subject Pronoun Expression in a Variety of Spanish in Contact with Portuguese. In Jim Michnowicz & Robin Dodsworth (eds.), *Selected Proceedings of the 5th Workshop on Spanish Sociolinguistics*. 14-25. Somerville, MA: Cascadilla Proceedings Project.

Carvalho, Ana M. & Ryan M. Bessett. 2015. Subject Pronoun Expression in Spanish in Contact with Portuguese. In Ana Maria Carvalho, Rafael Orozco & Naomi Lapidus Shin (eds.), *Subject Pronoun Expression in Spanish: A cross-dialectal perspective*. 145-167. Washington DC: Georgetown University Press.

Carvalho, Ana M., Rafael Orozco, & Naomi Lapidus Shin. 2015. Introduction. In Ana Maria Carvalho, Rafael Orozco & Naomi Lapidus Shin (eds.), *Subject Pronoun Expression in Spanish: A cross-dialectal perspective*. xiii-xxvi. Washington DC: Georgetown University Press.

Cerrón-Palomino, Álvaro. 2014. Ser más *pro* o menos *pro*: variación en la expresión de sujeto pronominal en el castellano limeño. *Lingüística* 30(1). 61-83.

Chambers, J.K. 2002. Patterns of Variation Including Change. In J.K. Chambers, Peter Trudgill, & Natalie Schilling-Estes (eds.), *The handbook of language variation and change*. 349–372. Oxford: Blackwell.

Claes, Jeroen. 2011. ¿Constituyen las Antillas y el Caribe continental una sola zona dialectal? Datos de la variable expresión del sujeto pronominal en San Juan de Puerto Rico y Barranquilla, Colombia. *Spanish in Context* 8(2). 191-212.

Enríquez, Emilia V. 1984. *El pronombre personal sujeto en la lengua española hablada en Madrid*. Madrid: Consejo Superior de Investigaciones Científicas.

Erker, Daniel. 2005. *Functional compensation for morphological ambiguity in New York Spanish*. New York City: City University of New York M.A. thesis.

Erker, Daniel & Gregory R. Guy. 2012. The role of lexical frequency in syntactic variability: Variable subject personal pronoun expression in Spanish. *Language* 88(3). 526-557.

Flores-Ferrán, Nydia. 2002. *A sociolinguistic perspective on the use of subject personal pronouns in Spanish narratives of Puerto Ricans in New York City*. Munich: Lincom-Europa.

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(1). 49-73.

Flores-Ferrán, Nydia. 2007. Los Mexicanos in New Jersey: Pronominal expression and ethnolinguistic aspects. In Jonathan Holmquist, Augusto Lorenzino, & Lotfi Sayahi (eds.), *Selected Proceedings of the Third Workshop on Spanish Sociolinguistics*. 85-91. Somerville, MA: Cascadilla Proceedings Project.

Flores-Ferrán, Nydia. 2009. Are you referring to me? The variable use of *UNO* and *YO* in oral discourse. *Journal of Pragmatics* 41. 1810-1824.

Guy, Gregory R. 1990. The sociolinguistic types of language change. *Diachronica* 7. 47-67.

Hurtado, Luz Marcela. 2001. *La variable expresión del sujeto en el español de los colombianos y colombo-americanos residentes en el condado de Miami-Dade*. Gainesville, FL: University of Florida dissertation.

Hurtado, Luz Marcela. 2005. El uso de tú, usted y uno en el español de los colombianos y colomboamericanos. In: Luis Ortiz López & Manuel Lacorte (eds.), *Contactos y contextos lingüísticos: el español en los Estados Unidos y en contacto con otras lenguas*. 185-200. Madrid/Frankfurt: Iberoamericana/Vervuert.

Labov, William. 1972. *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

Labov, William. 2001. *Principles of Linguistic Change: Social factors*. Malden: Blackwell.

Lastra, Yolanda & Pedro Martín Butragueño. 2015. Subject Pronoun Expression in Oral Mexican Spanish. In Ana Maria Carvalho, Rafael Orozco & Naomi Lapidus Shin (eds.), *Subject Pronoun Expression in Spanish: A cross-dialectal perspective*. 41-59. Washington DC:

Georgetown University Press.

Martínez-Sanz, Cristina. 2011. *Null and overt subjects in a variable system: The case of Dominican Spanish*. Ottawa, Canada: University of Ottawa dissertation.

Michnowicz, Jim. 2015. Subject Pronoun Expression in Contact with Maya in Yucatan Spanish. In Ana M. Carvalho, Rafael Orozco, & Naomi Lapidus Shin (eds.), *Subject Pronoun Expression in Spanish: A Cross-dialectal Perspective*. 103-122. Washington, DC: Georgetown University Press.

Morales, Amparo. 1980. La expresión de sujeto pronominal de primera persona en el español de Puerto Rico. *Boletín de la Academia Puertorriqueña de la Lengua Española* 8(2). 91-102.

Orozco, Rafael. 2015. Pronominal variation in Costeño Spanish. In Ana M. Carvalho, Rafael Orozco, & Naomi Lapidus Shin (eds.), *Subject Pronoun Expression in Spanish: A Cross-dialectal Perspective*. 17-37 Washington, DC: Georgetown University Press.

Orozco, Rafael. Under review. El castellano colombiano en la ciudad de Nueva York: Uso variable de sujetos pronominales. Submitted to *Studies in Lusophone and Hispanic Linguistics*.

Orozco, Rafael & Gregory Guy. 2008. El uso variable de los pronombres sujetos: ¿Qué pasa en la costa Caribe colombiana? In Maurice Westmoreland & Juan Antonio Thomas (eds.), *Selected proceedings of the Fourth Workshop on Spanish Sociolinguistics*. 70-80. Somerville, MA: Cascadilla Proceedings Project.

Orozco, Rafael, Catalina Méndez Vallejo, & Lee-Ann Vidal-Covas. 2014. Los efectos condicionantes del verbo en el uso variable de los pronombres personales de sujeto. In *Actas del XVII Congreso Internacional de la Asociación de Lingüística y Filología de la América Latina: João Pessoa, Paraíba, Brazil*. 2342-2358.

Ortiz López, Luis. 2009. Pronombres de sujeto en el español (L2 vs L1) del Caribe. In Manel Lácora & Jennifer Leeman (eds.), *El español en Estados Unidos y otros contextos de contactos: sociolingüística, ideología y pedagogía*. 85-110. Frankfurt Main/ Madrid: Verlag/Iberoamericana.

Ortiz López, Luis A. 2011. Spanish in contact with Haitian Creole. In Manuel Díaz-Campos (ed.), *The handbook of Spanish sociolinguistics*. 418-445. Malden, MA: Wiley-Blackwell.

Otheguy, Ricardo & Ana Celia Zentella. 2007. Apuntes preliminares sobre el contacto lingüístico y dialectal en el uso pronominal del español en Nueva York. In Kim Potowski and Richard Cameron (eds.), *Spanish in Contact: policy, social and linguistic inquiries*. 275-295. Amsterdam/ Philadelphia: John Benjamins.

Otheguy, Ricardo & Ana Celia Zentella. 2012. *Spanish in New York: Language contact, dialectal leveling, and structural continuity*. Oxford: Oxford University Press.

Otheguy, Ricardo, Ana Celia Zentella & David Livert. 2007. Language contact in Spanish in New York toward the formation of a speech community. *Language* 83(4). 770-802.

Posio, Pekka. 2011. Spanish subject pronoun usage and verb semantics revisited: First and second person singular subject pronouns and focusing of attention in spoken Peninsular Spanish. *Journal of Pragmatics* 43(3). 777-798.

Prada-Pérez, Ana de. 2009. *Subject expression in Minorcan Spanish: Consequences of contact with Catalan*. State College, PA: Pennsylvania State University dissertation.

Prada-Pérez, Ana de. 2015. First Person Singular Subject Pronoun Expression in Spanish in Contact with Catalan. In Ana Maria Carvalho, Rafael Orozco & Naomi Lapidus Shin (eds.), *Subject Pronoun Expression in Spanish: A cross-dialectal perspective*. 123-144. Washington DC: Georgetown University Press.

Shin, Naomi L. 2012. Variable use of Spanish subject pronouns by monolingual children in Mexico. In Kimberly Geeslin & Manuel Díaz-Campos (eds.), *Selected Proceedings of the 14th Hispanic Linguistics Symposium*. 130-141. Somerville, MA: Cascadilla Proceedings Project.

Shin, Naomi L. 2015. Acquiring constraints on morphosyntactic variation: children's Spanish subject pronoun expression. *Journal of Child Language, First View*. Available on CJO 2015 doi: [10.1017/S0305000915000380](https://doi.org/10.1017/S0305000915000380).

Shin, Naomi & Daniel Erker. 2015. The emergence of structured variability in morphosyntax: Childhood acquisition of Spanish subject pronouns. In Ana Maria Carvalho, Rafael Orozco & Naomi Lapidus Shin (eds.), *Subject Pronoun Expression in Spanish: A cross-dialectal perspective*. 169-190. Washington DC: Georgetown University Press.

Shin Naomi & Ricardo Otheguy. 2009. Shifting sensitivity to continuity of reference: Subject pronoun use in Spanish in New York City. In Manel Lacorte & Jennifer Leeman (eds.), *Español en Estados Unidos y otros contextos de contactos: Sociolinguística, ideología, y pedagogía*. 111-136. Frankfurt/Madrid: Iberoamericana/Vervuert.

Shin Naomi & Ricardo Otheguy. 2013. Social class and gender impacting change in bilingual settings: Spanish subject pronoun use in New York. *Language in Society* 42(4). 429-452.

Silva-Corvalán, Carmen. 1982. Subject Expression and Placement in Spoken Mexican-American Spanish. In Jon Amastae & Lucía Elías-Olivares (eds.), *Spanish in the United States: Sociolinguistic Aspects*. 93-120. New York: Cambridge University Press.

Silva-Corvalán, Carmen. 1994. *Language contact and change: Spanish in Los Angeles*. New York: Oxford University Press.

Silva-Corvalán, Carmen. 1997. Avances en el estudio de la variación sintáctica: La expresión del sujeto. *Cuadernos del Sur* 27. 35-49.

Torres Cacoullos, Rena & Catherine E. Travis. 2010. Variable *yo* expression in New Mexico: English influence? In Susana Rivera Mills & Daniel Villa (eds.), *Spanish of the U.S. Southwest: A Language in Transition*. 185-206. Madrid: Iberoamericana/Vervuert.

Torres Cacoullos, Rena & Catherine E. Travis. 2011. Testing convergence via code-switching: priming and the structure of variable subject expression. *International Journal of Bilingualism* 15(3). 241-267.

Torres Cacoullos, Rena & Catherine E. Travis. 2015. Foundations for the study of subject pronoun expression in Spanish in contact with English. In Ana Maria Carvalho, Rafael Orozco & Naomi Lapidus Shin (eds.), *Subject Pronoun Expression in Spanish: A cross-dialectal perspective*. 81-100. Washington DC: Georgetown University Press.

Travis, Catherine E. 2005a. *Discourse markers in Colombian Spanish: A study in polysemy* (Cognitive Linguistics Research). Berlin/New York: Mouton de Gruyter.

Travis, Catherine E. 2005b. The *yo-yo* effect: Priming in subject expression in Colombian Spanish. In Randall Gess & Edward J. Rubin (eds.), *Theoretical and experimental approaches to Romance linguistics: Selected papers from the 34th Linguistic Symposium on Romance Languages, 2004*. 329-349. Amsterdam / Philadelphia: John Benjamins.

Travis, Catherine E. 2007. Genre effects on subject expression in Spanish: Priming in narrative and conversation. *Language Variation and Change* 19(2). 101-135.

Travis, Catherine E. & Rena Torres Cacoullos. 2012. What do subject pronouns do in discourse? Cognitive, mechanical and constructional factors in variation. *Cognitive Linguistics* 23(4). 711-748.