

Variable subject pronoun expression revisited: This is what the *Paisas* do

Rafael Orozco & Luz Marcela Hurtado*

Abstract. We explore subject pronoun expression (SPE) in Medellín, Colombia using 4,623 tokens to test eight predictors. The 28% overall pronominal rate found is significantly higher than those in other mainland communities. Grammatical person exerts the greatest conditioning effect, with *uno* ‘one’ strongly favoring overt subjects. Findings for verb class reveal that speech and cognitive verbs promote overt subjects. However, our in-depth analysis unveils opposing tendencies between different pronominal subject + verb collocations for the same verb. E.g., whereas *(yo) soy* ‘I am’ strongly favors overt subjects, *(ellos) son* ‘they are’ favors null subjects. These findings suggest that analyses focusing on infinitives do not constitute the most accurate way to explore verb effects on SPE. Moreover, the effect of age reveals a low pronominal rate among the youngest speakers, a finding that appears to have cognitive and acquisitional implications, as younger speakers would be expected to have higher pronominal rates. In general, this study contributes to expand our knowledge of SPE. Further, the findings regarding age and the lexical effect of the verb open promising research paths.

Keywords: subject pronoun expression (SPE); sociolinguistics; language variation; Latin American Spanish; Colombian Spanish; lexical effects

1. Introduction. This paper expands on recent research on the variable alternation between null and overt pronominal subjects in Spanish (Hurtado & Ortega-Santos 2019; Orozco 2016, 2018a, 2018b; Orozco & Hurtado 2020; inter alia). *Pronombrismo*, the study of subject pronoun expression (SPE), was pioneered by Barrenechea & Alonso (1973) with their analysis of the Spanish of Buenos Aires, Argentina. That seminal investigation was followed by a multitude of studies throughout the Hispanic World and beyond. Five decades of SPE scholarship have provided such important contributions to sociolinguistics that this research strand has been considered a show-case sociolinguistic variable (Bayley et al. 2012:50).

The vast body of SPE literature provided a solid foundation for its study in Colombian Spanish, where it was first explored by Hurtado (2001), who studied SPE among Colombians in Miami-Dade County, FL. Subsequent research analyzed SPE among Colombians in Bogotá and Florida, respectively (Hurtado 2005a, 2005b), in the city of Cali (Travis 2005, 2007; Travis & Torres Cacoullos 2012), and among Mainlander Colombians in NYC (Otheguy & Zentella 2012; Otheguy et al. 2007). That research was followed by studies of the Caribbean city of Barranquilla (Orozco & Guy 2008; Orozco 2015, 2018a; Hurtado & Ortega-Santos 2019) and among Caribbean Colombians in NYC (Orozco 2018a, 2018b). Other work explores SPE in Cartagena,

* This research was partially supported by a Louisiana State University 2018 ASPIRE Summer Internship Award granted to Rafael Orozco and Noelle Primeaux. We thank Marianne Dieck and María Claudia González Rátiva for providing us generous access to the PRESEEA-Medellín corpus as well as the speakers who provided the data. This research benefited from valuable contributions by the following LSU students who helped with token extraction: Alex Thomas, Alex Yandell, Bailly Daberkow, Cecelia Morise, Cecilia Vazquez, Joelle Semplonius, and Sabrina Valenti. We are especially grateful to Noelle Primeaux for her valuable help with data extraction and coding. We are thankful to the audience at the 95th LSA annual meeting. We take responsibility for all remaining infelicities. Authors: Rafael Orozco, Louisiana State University (rorozc1@lsu.edu) & Luz Marcela Hurtado, Central Michigan University (hurtal1m@cmich.edu).

Barranquilla, Valledupar, and San Basilio de Palenque, respectively (Pérez & Gomes 2019; De la Rosa 2020). Yet, SPE remains understudied in Andean Spanish despite being the variety spoken by most Colombians. With this study of SPE in Medellín, we seek to answer some questions opened by Orozco & Hurtado (2020). Among other things, this analysis explores the role of *uno* ‘one’ as a subject pronoun and the effect of kinesics as a verb classification. Moreover, employing an analytical approach motivated by recent findings revealing opposing tendencies between verbs in the same lexical category (Orozco 2016, 2018a, 2018b; Orozco & Hurtado 2020) and with the multiprong goal of gaining a more detailed understanding of how verbs condition SPE and further informing linguistic theory, we explore the lexical effect of the verb.

2. Background. As *pronombrista* research evolved, most studies followed the model laid out by Barrenechea & Alonso (1973) including all subject pronouns. Others focused exclusively on the first-person pronouns, as done by Bentivoglio (1980) and Morales (1980), who respectively analyzed SPE in Caracas, Venezuela and San Juan, Puerto Rico. Regardless of whether only the first person or all pronouns have been explored, *pronombrista* research has determined that variable SPE displays dialectal or regional overt pronominal rate differences. The lowest overall overt pronominal rates have been found in Spain (21%, Enríquez 1984; Cameron 1993) and the continental Latin American varieties such as those of Lima, Peru (16.8%, Cerrón-Palomino 2014) and Mexico City (21.8%, Lastra & Martín Butragueño 2015), with an average of 24%. Concurrently, the highest overall overt pronominal expression rates have been found in the Caribbean, ranging from 33% (Cuban newcomers to New York City, Otheguy & Zentella 2012) to 45% (San Juan, Puerto Rico, Cameron 1993), with an average of 38% (Orozco & Hurtado 2020:2).

Despite known overt pronominal rate differences between speech communities in different dialect regions, five decades of SPE research have found much uniformity regarding both the predictors probabilistically conditioning pronominal expression and the tendencies exhibited by their individual factors (Carvalho, Orozco & Shin 2015:xiii). The notable similarity of effects found supports the notion that structured linguistic variation reflects an intrinsic part of our grammatical knowledge: usage patterns are deeply embedded in our knowledge of grammar. SPE is mainly conditioned by grammatical person and number of the subject; priming; coreference; tense, mood, and aspect (TMA) morphology; clause type; verbal reflexivity; and lexical semantics or verb type. Thus, overt pronominal subjects occur more frequently with singular subjects, with verb tenses that have ambiguous person morphology such as the imperfect, after a change in referent, immediately after an overt subject, and in main clauses. Overt subjects are disfavored when a reflexive pronoun is used with the same verb. Previous research exploring verb semantics found that psychological and cognitive verbs (e.g., *recordar* ‘remember,’ *saber* ‘know,’ *creer* ‘believe’) favor overt pronominal subjects, but external activity verbs (e.g., *salir* ‘leave,’ *trabajar* ‘work,’ *ir* ‘go’) promote null subjects (Bentivoglio 1980, 1987; Enríquez 1984; Hurtado 2005b; Travis 2007; Orozco & Guy 2008; Abreu 2009; Posio 2011; among others). High pronominal expression with psychological verbs has been explained as an indicator of the speaker’s stance towards the utterance (Travis 2007:117). This conceptualization is also supported by the fact that singular pronouns (especially *yo* ‘I’ and *uno* ‘one’) register higher overt pronominal rates and probabilistic weights across Spanish varieties (Orozco & Hurtado 2020:3). The rates of overt *uno* ‘one’ are the highest across different varieties of Spanish (Cameron 1993; Hurtado 2005a; Otheguy, Zentella & Livert 2007). Hurtado (2005a) found the strong effect of non-specific referents on SPE among Colombians living in Miami, Florida. Specifically, probabilistic weights showed that impersonal uses of *tú* ‘you (sing.),’ *usted* ‘you (sing.),’ and *uno* ‘one’ favored overt pronoun expression (.55, .63 and .85, respectively). Pronominal rates

indicated that that effect advanced in the same direction, and the pronoun with the highest rates of expression was *uno*.

3. Methodology. This section describes the speech community and the dataset analyzed. It also states the research questions, the hypothesis, and discusses the predictors explored.

3.1. THE SPEECH COMMUNITY, THE CORPUS, AND THE DATASET. Medellín, with a population of 2,372,330, 59, is Colombia's second largest metropolitan area and one of its main industrial centers. This city is located in the department of Antioquia, where two dialectal varieties, the Andean and the Coastal converge. Medellín's Spanish is part of Western Andean Colombian Spanish (Montes Giraldo 1982). The *Paisa* speech community, as the people of this region are called, is characterized, among other things, by a tripartite second person singular address system where *tú*, *usted*, and *vos*, all meaning 'you (singular),' alternate (Millán 2014:92). The data sample examined here was culled from the *Proyecto para el Estudio Sociolingüístico del Español de España y de América (PRESEEA) Medellín Corpus* collected between 2007 and 2010 (González-Rátiva 2008). We used 39 of the 119 interviews in the corpus, which correspond to 20 women and 19 men whose ages ranged from 15 to 85 years old at the time they were interviewed. All consultants were born in Medellín or in the surrounding region.

3.2. RESEARCH QUESTIONS AND HYPOTHESES. With this *pronombrista* study of Medellín Spanish, we aim to answer the following three main research questions.

- (a) *How does Medellín Spanish compare with other varieties of Spanish in terms of subject pronoun expression?*
- (b) *How does the internal conditioning on SPE including the lexical effect of the verb in Medellín Spanish compare to what is found throughout the Hispanic World?*
- (c) *How do age and gender condition SPE in Medellín, and how do their effects in this speech community compare to those in other communities?*

Concurrently, we seek to probe the following main hypothesis: *Despite an internal conditioning concurrent with what is found across the board, different pronominal subject + verb collocations corresponding to a single verb condition SPE differently.* Our research questions and hypothesis are guided by the findings of many previous studies cited throughout this paper and by findings suggesting that we lack conclusive information as to the effects of the verb on SPE (Erker & Guy 2012; Orozco 2015; Orozco & Guy 2008; Posio 2011). We also test hypotheses directly related to each one of the predictors we probe, which are discussed below.

3.3. PREDICTORS EXPLORED. To answer our research questions and test our hypotheses, we probed the effects of six linguistic and two social predictors chosen based on many prior findings (Enriquez 1984; Cameron 1992, 1993, 1995; Hurtado 2001, 2005a; Hurtado & Ortega-Santos 2019; Travis 2005, 2007; Otheguy et al. 2007; Otheguy & Zentella 2012; Posio (2011); Travis & Torres Cacoullos 2012; Lastra & Martín Butragueño 2015; Orozco 2015, 2018a; Torres Cacoullos & Travis 2011, 2018; inter alios). The predictors analyzed here are discussed below.

3.3.1. LINGUISTIC PREDICTORS. The six internal predictors analyzed here and their factors are (1) Grammatical person and number of the subject: 1st person singular, 2nd person singular, 3rd person singular, *uno*, plurals tested as a single factor; (2) Discourse type: narrative, opinion, hypothetical situation, description, routine; (3) Coreferentiality: same person and number as the preceding subject, different person and number, same person different number, same number different person; (4) Verb tense, mood and aspect (TMA): present indicative, imperfect indicative,

preterit indicative, and other tenses; (5) Semantic verb class: verbs of activity, motion, communication, physical perception, cognition, and stative; and (6) Lexical effect of the verb tested by means of pronominal subject + verb collocations.

3.3.2. SOCIAL PREDICTORS. Our analysis probes the effects of gender and age, which condition SPE in Colombia and other Hispanic communities (Orozco 2016, 2018a; Alfaraz 2015; Shin & Erker 2015; among others). We follow the parameters described in Orozco & Hurtado (2020:6).

3.4. THE ENVELOPE OF VARIATION AND THE ANALYSIS. The envelope of variation used here follows the Principle of Accountability and the comprehensive parameters defined by Barrenechea & Alonso (1973), and refined by Otheguy & Zentella (2012:48 ff.), and the PRESEEA Project (Silva-Corvalán & Enrique-Arias 2017:173), which are regarded as standard for *pronombrista* studies. Our analysis includes only those clauses with ascertainable animate pronominal subjects containing a conjugated verb where the alternation between an overt and a null pronominal subject is clearly possible. Thus, all tokens constitute one of at least two possible different ways of saying the same thing. Our data sample consists of 4623 tokens coded in terms of the predictors discussed above. We tested our data by means of a series of multivariate statistical regression analyses intended to probe hypotheses pertaining to each of the predictors we explored.

In what follows, we present the distribution of overt and null pronominal subjects, discuss the predictors that condition SPE, draw conclusions and formulate their implications.

4. Distribution of variable pronominal subjects and predictor model. The distribution of overt and null pronominal subjects (Table 1) shows Medellín's overall overt pronominal rate 28% to be significantly lower ($X^2 = 33.6$; $p < .001$) than Barranquilla's 34% (Orozco 2015). But it is the highest overt pronominal rate found in a mainland speech community, where pronominal rates average 24% (Orozco & Guy 2008; Otheguy & Zentella 2012; Lastra & Martín Butragueño 2015; Michnowicz 2015). One reason for Medellín's relatively high overt pronominal rate may be its geographical location in a department with a Caribbean coastal region. Thus, Paisa speech is closely in contact with Caribbean Spanish varieties, whose pronominal rates are high.

Variant	N	%
Overt subjects (<i>ellos cantan</i> 'they sing')	1,294	(28%)
Null subjects (<i>Ø cantan</i> '[they] sing')	3,329	(72%)
Total	4,623	(100%)

Table 1. Distribution of overt and null pronominal subjects in Medellín

The results of several rounds of multivariate tests including all predictors but the lexical effect of the verb uncovered a complex model that includes linguistic and social forces with six predictors (one social and five linguistic) reaching statistical significance (See Table 2).

Predictor	Range
Grammatical person & number of the subject	67
Discourse type	36
Coreferentiality	29
Verb tense, mood & aspect (TMA)	21
Semantic verb class (Kinesis)	18
Age	15

Table 2. The SPE model for Medellín

The order of selection shows grammatical person and number of the subject as the strongest predictor with a range of 67. In general, internal constraints have a greater conditioning effect on SPE based on their order of selection, which was established according to range values. The greater internal conditioning found in Medellín concurs with what obtains across the board in both mon-olingual and bilingual speech communities (Carvalho et al. 2015).

Furthermore, the constraint hierarchy found in Medellín with grammatical person and number of the subject being, by far, the strongest linguistic predictor is consonant with findings in other Colombian communities including Barranquilla, New York City Colombians (Orozco 2018a), Cartagena, and Valledupar (Pérez & Gomes 2019). It also concurs with what happens around the Spanish-speaking World including Madrid (Enríquez 1984); Mexico City (Lastra & Martín Butragueño 2015); Los Angeles (Silva-Corvalán 1982); NYC (Otheguy & Zentella 2012); Puerto Rico (Cameron 1993, 1995); Rivera, Uruguay (Carvalho & Bessett 2015); and Yucatan, Mexico (Michnowicz 2015), among other speech communities. The similarity between the SPE conditioning in Medellín and the rest of the world corroborates that the grammar underlying SPE across varieties remains essentially the same despite varying pronominal rates at the surface level (Cameron 1993; Travis 2007; Torres Cacoullos & Travis 2011, Michnowicz 2015).

5. Linguistic conditioning. The internal conditioning on SPE in Medellín reflects the effects of two subject-related predictors (grammatical person and number of the subject and coreferentiality), one predictor pertaining to the whole clause (discourse type), and two verb-related predictors (TMA and semantic verb class). Our presentation of the results pertaining to internal predictors follows the order in which they appear on Table 2. Thus, we first deal with grammatical person and number of the subject and, subsequently, with the remaining predictors.

5.1. GRAMMATICAL PERSON AND NUMBER OF THE SUBJECT. Our findings, presented in Table 3, show that Medellín adheres to the general tendency for grammatical person and number across varieties of Spanish with singular pronouns favoring overt subjects and plural pronouns disfavoring them. Our findings also reflect an idiosyncratic feature of Medellín speech, as coreferent *uno* ‘one’ is the factor that most strongly promotes overt pronominal subjects with a statistical weight of .83. This dominance of overt coreferential *uno* over its null coreferential counterpart concurs with findings in Barranquilla, Colombia (Hurtado & Ortega-Santos 2019:51).

Factor	Prob. ¹	% overt	N	% Data
<i>Uno</i> ‘one’	.83	60%	192/320	6.9%
2 nd singular (<i>tú, usted, vos</i> ‘you’)	.52	33%	67/204	4.4%
1 st singular (<i>yo</i> , ‘I’)	.51	32%	816/2,559	55.4%
3 rd singular (<i>él</i> ‘he,’ <i>ella</i> ‘she’)	.49	27%	105/394	8.5%
All plural (<i>nosotros</i> ‘we,’ <i>ustedes</i> ‘you,’ <i>ellas</i> ‘they,’ <i>ellos</i> ‘they’)	.16	10%	114/1,146	24.8%
Range = 67		p-value = <.001		

Table 3. Effects of grammatical person & number in Paisa speech

The overt pronominal rates for the different grammatical persons corroborate the fact that overt pronominal subjects occur much more frequently with singular pronouns than with their plural counterparts, which despite constituting one fourth of the data, only register a 10% overt

¹ 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.

pronominal rate (Bentivoglio 1987:36; Bayley & Pease-Alvarez 1997; Flores-Ferrán 2002, 2004, 2007b; Abreu 2009; Carvalho & Child 2011; Erker & Guy 2012; Otheguy & Zentella 2012; Otheguy et al. 2007; Posio 2011; Prada Pérez 2009; Ortiz López 2011; among others) It also shows a higher incidence of first-person subjects (55% of the data).

5.2. DISCOURSE TYPE. As Table 4 shows, discourse type strongly conditions overt pronominal expression, with narrative, opinion, and hypothetical situations as the factors favoring explicit subject pronouns with probability values of .61, .59, and .55, respectively. Descriptive statements, with a probability value of .53 barely favor overt pronominal subjects. At the same time, routine statements disfavor overt subjects with a probability value of .25. Our findings clearly indicate the link between pronominal expression and speaker's stance and experiences. This concurs with findings by Lastra & Martín Butragueño (2015) who explored discourse type in Mexico City, where they found a favoring effect of argumentation (.66), which they connected to positioning points of view and highlighting opinions.

Factor	Prob.	% overt	N	% Data
Narrative	.61	29%	587/2,033	44.0%
Opinion	.59	33%	433/1,326	28.7%
Hypothetical Situation	.55	28%	159/560	12.1%
Description	.53	25%	89/355	7.7%
Routine	.25	7%	26/349	7.5%
<i>Range = 36</i>		<i>p-value = <.001</i>		

Table 4. Effects of discourse type in Medellín

The results for discourse type provide baseline information for future pronombrista studies in Colombian Spanish given that this predictor has not been found to condition SPE in other Colombian speech communities.

5.3. COREFERENTIALITY. The findings presented in Table 5 show that a complete change of reference favors subsequent overt subjects with a probabilistic weight of .62. A partial change of reference also favors overt subjects (.56). Conversely, a prior subject with the same person and number favors null subjects with a low probabilistic weight (.33). The tendencies found in Medellín concur with findings in Barranquilla, Colombia and among Costeño Colombian speakers in New York City (Orozco 2018a:104).

Factor	Prob.	% overt	N	% Data
Different grammatical person & number	.62	32%	173/548	11.9%
Partially different person or number	.56	38%	675/1,839	39.8%
Same person & number	.33	20%	446/2,236	48.4%
<i>Range = 29</i>		<i>p-value = <.001</i>		

Table 5. Effects of prior subject's person & number (Coreferentiality)

Our results are also congruent with findings in Cali, Colombia (Torres Cacoullós & Travis 2019:671), where null subjects are favored by previous unexpressed subjects (.63) but disfavored by preceding overt pronouns (.37). Thus, the effects of coreferentiality appear to be consistently similar across different varieties of Spanish including Madrid, the speech of Puerto Ricans in San Juan and NYC (Cameron & Flores-Ferrán 2004), Mexico City (Lastra and Martín Butragueño 2015), Yucatán, Mexico (Michnowicz 2015), and Spain (Prada Pérez 2015), among others.

5.4. VERB TENSE, MOOD, AND ASPECT. Results, presented in Table 6, show the imperfect tense favoring overt pronominal expression (.61); the present with a probability value of .50 has a neutral effect and so does the preterit (.49). In contrast, all other tenses, acting as a single factor, favor null pronouns (.40). These tendencies concur with those in the *Costeño* Colombian variety of Barranquilla (imperfect .61, present .52, preterit .47, others .40) and in New York (imperfect .58, present .56, preterit .47, others .38) found by Orozco (2018a:109). Our results also concur with findings in other Spanish varieties (Enríquez 1984; Bentivoglio 1987; Travis 2007; Carvalho & Bessett 2015; Lastra & Martín Butragueño 2015; inter alios).

Factor	Prob.	% overt	N	% Data
Imperfect indicative	.61	37%	154/422	9.1%
Present indicative	.50	29%	829/2,832	61.3%
Preterit indicative	.49	26%	146/563	12.2%
All other tenses	.40	21%	165/806	17.4%
<i>Range = 21</i>		<i>p-value = <.001</i>		

Table 6. Effects of verb tense, mood, and aspect in Medellín

These tendencies also support the premise that the imperfect tense is low in transitivity. Thus, the focus of attention remains on the subject, promoting overt pronominal expression.

5.5. SEMANTIC VERB CLASS (KINESIS). The results, presented in Table 7, reveal that verbs of communication (.59) and cognition (.57) promote overt pronominal subjects. Stative and activity verbs register neutral effects whereas verbs of emotion and volition (.47), motion (.47) and physical perception (.41) promote null subjects. Our findings appear to concur with previous findings reporting that overt pronominal subjects are favored by verbs of cognition (Carvalho et al. 2015:xv) and speech (Orozco 2018a:112).

Factor	Prob.	% overt	N	% Data
Communication	.59	37%	118/317	6.9%
Cognition	.57	41%	307/747	16.2%
Stative	.51	26%	293/1,127	24.4%
Activity	.49	22%	257/1,145	24.8%
Emotion & volition	.47	28%	108/385	8.3%
Motion	.47	22%	140/640	13.8%
Physical Perception	.41	27%	71/262	5.7%
<i>Range = 18</i>		<i>p-value = <.001</i>		

Table 7. Effects of verb class on SPE in Medellín

At the same time, being aware of recent findings reporting opposing tendencies among verbs within a single lexical category (Orozco 2015, 2016, 2018a; Orozco & Hurtado 2020), we also probed the lexical effects of the verb, whose findings are presented in the next section.

5.6. LEXICAL EFFECT OF THE VERB. As shown in Table 8, we explored the lexical effect of the verb by testing how pronominal subject + verb collocations condition SPE. Overt subjects are most strongly favored by (*yo*) *creo* ‘I think, I believe’ with a probability value of .877. This finding is consonant with what occurs in Cali, Colombia (Travis & Torres Cacoullos 2012:739), and with this collocation consistently promoting overt subjects across the board. In general, singular collocations promote SPE over plural forms. These results support the idea of the influence of

speaker stance, as first person singular *yo* (*creo* ‘I think, I believe,’ *pienso* ‘I think,’ *tengo*, ‘I have,’ and *digo* ‘I say’) and 3-person singular verbs linked with the use of *uno* (*es* ‘she, he, one is,’ *tiene* ‘she, he, one has,’ *tenía* ‘she, he, one had,’ *ve* ‘she, he, one sees’) favor overt subjects.

Factor	Prob.	% overt	N	% Data
<i>Creo</i> ‘I believe, I think’	.877	73.0%	108/148	3.2%
<i>Sabe</i> ‘she/he/one knows’	.876	81.0%	17/21	0.5%
<i>Soy</i> ‘I am’	.807	60.0%	39/65	1.4%
<i>Vivo</i> ‘I live’	.785	63.0%	17/27	0.6%
<i>Tenía</i> ‘I/she/he/one had’	.770	51.6%	16/31	0.67%
<i>Pienso</i> ‘I think’	.757	54.5%	30/55	1.19%
<i>Digo</i> ‘I say’	.754	52.3%	46/88	1.90%
<i>Estaba</i> I/she/he/one was’	.747	51.9%	14/27	0.58%
<i>Ve</i> ‘she/he/one sees’	.739	53.6%	15/28	0.61%
<i>Dije</i> ‘I said’	.731	52.2%	12/23	0.50%
<i>Puede</i> ‘she/he/one can’	.721	45.0%	9/20	0.43%
<i>Es</i> ‘she/he/one is’	.700	37.4%	34/91	1.97%
<i>Tiene</i> ‘she/he/one has’	.690	41.2%	14/34	0.74%
<i>Está</i> ‘she/he/one is’	.689	40.0%	8/20	0.43%
<i>Era</i> ‘she/he/one was’	.669	36.7%	11/30	0.65%
<i>Hago</i> ‘I do’	.591	25.6%	10/39	0.84%
<i>Conozco</i> ‘I know’	.583	30.8%	12/39	0.84%
<i>Conoci</i> ‘I met’	.542	32.0%	8/25	0.54%
<i>Levanto</i> ‘I stand up’	.538	26.1%	6/23	0.50%
<i>Voy</i> ‘I go’	.538	25.0%	7/28	0.61%
<i>Tengo</i> ‘I have’	.530	25.0%	34/136	2.94%
<i>Veo</i> ‘I see’	.521	27.8%	15/54	1.17%
<i>Somos</i> ‘we are’	.500	23.5%	8/34	0.74%
<i>Salgo</i> ‘I leave’	.495	20.0%	5/25	0.54%
<i>Siento</i> ‘I feel’	.492	22.2%	8/36	0.78%
<i>Me voy</i> ‘I leave’	.468	13.6%	3/22	0.48%
<i>Sé</i> ‘I know’	.468	22.4%	34/152	3.29%
<i>Estoy</i> ‘I am’	.421	18.0%	11/61	1.32%
<i>Imagino</i> ‘I imagine’	.384	13.2%	5/38	0.82%
<i>Eramos</i> ‘we were’	.382	14.3%	3/21	0.45%
<i>Tenemos</i> ‘we have’	.350	12.7%	8/63	1.36%
<i>Son</i> ‘they are’	.323	10.3%	3/29	0.63%
<i>Vea</i> ‘she/he/one sees’	.300	7.4%	2/27	0.58%
<i>Estamos</i> ‘we are’	.267	5.9%	2/34	0.74%
<i>Hacemos</i> ‘we do’	.250	3.6%	1/28	0.61%
<i>Vamos</i> ‘we go’	.227	0.0%	0/23	0.50%

Table 8. Effects of lexical frequency according to pronominal subject + verb collocations in Medellín, Colombia

In contrast, plural conjugations (*son* ‘they are,’ *estamos* ‘we are,’ *hacemos* ‘we do,’ *vamos* ‘we go’) overwhelmingly have null subjects, thus disfavoring overt pronominal subjects. In terms of the relationship between semantic class and lexical frequency, we observe a discrepancy in the verbs of cognition, the second most influential factor. *Creo* ‘I think, I believe,’ *sabe* ‘she, he, one knows,’ *pienso* ‘I think,’ *conozco* ‘I know,’ and *conoci* ‘I knew’ favor SPE whereas

imagino ‘I imagine’ and *sé* ‘I know,’ disfavor it. In the case of *imagino*, the disfavoring effect is linked to the use of the reflexive pronoun *me*.

Our findings regarding the effect of pronominal subject + verb collocations lead to the conclusion that these paradigms constitute two kinds of prefabricated units in Spanish: *yo creo* vs. all others. (*Yo*) *creo* ‘I think, I believe’ appears to have been reanalyzed cognitively and consistently promotes overt subjects because it has become grammaticalized as a discourse formula. In contrast, all other collocations remain grammatically productive. This analysis of the effect of collocations opens the door to similar analyses in other speech communities.

6. Social conditioning. Our study explores two social predictors: speaker’s age and gender. Our findings reveal that while age significantly constrains SPE in Medellín Spanish, speaker’s gender does not. The results for gender (Table 9) show both women and men with identical pronominal rates (28%) and similar probability values. The lack of significance of gender in Medellín differs from what occurs in other Colombian speech communities and elsewhere in the Hispanic World given that a sizable body of research finds overt pronominal subjects being favored by women (e.g., Bayley and Pease-Alvarez 1996; Carvalho & Child 2011; Otheguy & Zentella 2012; Shin & Otheguy 2013; Alfaraz 2015; Orozco 2015, 2018b). Concurrently, the lack of significance of gender in Medellín concurs with findings from Caracas (Bentivoglio 1980, 1987), the Uruguay-Brazil border region (Carvalho & Bessett 2015), Mexico City (Lastra & Martín Butragueño 2015), and Yucatán, Mexico (Michnowicz 2015), among other speech communities, where women and men display similar SPE tendencies. These results suggest the existence of different gender effects for SPE in different speech communities, disproving our hypothesis that gender would condition SPE in Medellín with women favoring overt pronominal subjects.

Factor	Prob.	% Overt	N	% Data
<i>Speaker’s Gender</i>				
Women	[.50]	28%	660/2,357	51.0%
Men	[.50]	28%	634/2,266	49.0%
Range	0			
<i>Speaker’s Age</i>				
Over 55	.58	33%	549/1,687	36.5%
30 to 54	.49	27%	487/1,839	39.8%
15 to 29	.43	24%	258/1,097	23.7%
Range	15			

Table 9. Social conditioning on SPE in Medellín, Colombia.

Guided by a preliminary distributional analysis, we determined the best configuration fit for our data by dividing speakers into three age groups: (a) 15–29, (b) 30–54, (c) over 55 years old. Findings (Table 9) reveal an age grading effect, as pronominal rates proportionally increase with age. The youngest speakers (<30 years old) favor null subjects (.43) with a pronominal rate of 24%; middle-aged speakers (30 to 54 years old) have a neutral effect (.49) with a pronominal rate of 27% ; and our oldest speakers (>55 years old), with the highest pronominal rate (33%), favor overt pronominal subjects (.58). These results support our hypothesis that older speakers promote overt pronominal subjects whereas our youngest speakers favor null subjects. Our results also concur with findings in other speech communities including Oaxaca, Mexico (Shin & Erker 2015); Mexico City (Lastra & Martín Butragueño 2015); Barranquilla, Colombia (Orozco & Guy 2008; Orozco 2015); and Spain (Prada Pérez 2015); among others.

7. Discussion. We have addressed three research questions and a main hypothesis. The answer to our first research question (*How does Medellín Spanish compare with other varieties of Spanish in terms of subject pronoun expression?*) reveals that Medellín's overt pronominal rate of 28% is both the lowest found in Colombia and the highest in a highland Hispanic community. This pronominal rate concurs with dialectal classifications placing the *Paisa* variety within the Andean dialect region (Montes 1982). It also reflects Medellín's proximity to the Caribbean coast and the influx of *Costeño* speakers taking place in recent years (Orozco & Hurtado 2020:22).

Answering our second research question (*How does the internal conditioning on SPE including the lexical effect of the verb in Medellín Spanish compare to what is found throughout the Hispanic World?*), we find an internal conditioning congruent with what occurs across the board, as attested in the rich body of SPE literature (Carvalho et al. 2015). SPE is significantly conditioned by grammatical person and number of the subject (the strongest predictor), discourse type, coreferentiality, TMA, and semantic verb type. Thus, the internal conditioning provides further evidence of a subjacent grammar for all Spanish varieties despite significant pronominal rate differences (Carvalho et al. 2015; Torres Cacoullos & Travis 2019; Travis 2005, 2007). The effects of grammatical person reveal that *uno* 'one' strongly favors overt pronominal subjects, a tendency already attested among Colombians living in Miami (Hurtado 2001, 2005a). Although Cameron (1993) suggested a favoring effect for nonspecific pronominal uses in SPE, the prevailing referential interpretation in Medellín's use of *uno* is its connection with the speaker. In fact, 224 of the 320 occurrences of overt and null coreferent *uno* were cases with the experiences of the I-speaker being the focus of attention and *yo* the previous clause's subject. In contrast, only 96 were more general references that excluded the speaker. This referential interpretation of *uno*, as analyzed in Bogotá (Hurtado 2015) and Barranquilla (Hurtado & Gutiérrez-Rivas 2016), can be interpreted as a discursive marker of the speaker's stance and as subjectification of discourse. This effect seems to reflect the subjectivity that *uno* confers in discourse genres such as narration and argumentation (Company & Pozas 2009:1207). Discourse-type tendencies (narration, opinion, and hypothetical situations), imperfective tenses (low in transitivity, Hopper & Thompson 1980), and semantic verb class (verbs that indicate the speaker's knowledge and opinion), reflect the role of subjectivity and the focus of attention on subject pronoun expression.

The answer to our third research question (*How do age and gender condition SPE in Medellín, and how do their effects in this speech community compare to those in other communities?*) shows age as the only social predictor significantly conditioning SPE. While our youngest speakers (15 to 29 years old) promote null subjects, the oldest speakers (over 55 years old) favor overt pronominal subjects. That is, we find an age grading effect, as pronominal rates progressively increase with age. Besides concurring with findings in Barranquilla (Orozco 2015), Santo Domingo (Alfaraz 2015), and Mexico City (Lastra & Martín Butragueño 2015), among other communities, our results corroborate findings showing that children's and adolescents' pronominal rates increase gradually as they acquire adult SPE usage patterns (Shin 2015:11, Shin & Erker 2015). The youngest speakers' disfavoring effect on overt subjects seems to have language acquisition and cognitive implications since Spanish appears to be changing toward higher pronominal rates. Our results, including the intriguing fact that the youngest speakers are not being the promoters of this change, are explained by the finding that, in monolingual speech communities, pronominal rates increase developmentally with age (Shin 2015; Shin & Erker 2015; Orozco 2016; among others). Thus, native Spanish speakers appear to reach adult pronominal usage in their 20s rather than as young children or teenagers. However, the lack of significance of gender is incongruent with what happens in many other communities, where women promote overt

pronominal subjects including Barranquilla, Colombia (Orozco 2015, 2018b); Santo Domingo, Dominican Republic (Alfaraz 2015); New York City (Otheguy & Zentella 2012; Shin & Otheguy 2013; Orozco 2018a, 2018b); and Spain (Prada Pérez 2015). Concomitantly, gender conditions other linguistic variables such as the expression of nominal possession in other Colombian communities (Orozco 2018a; Orozco & Hurtado 2020:21). Thus, our findings contribute to provide mounting evidence suggesting that the effect of gender varies with respect to different speech communities and linguistic variables.

Our analysis of the lexical effects of the verb goes beyond previous research (Otheguy & Zentella 2012; Orozco 2015; Torres Cacoullos & Travis 2018; among others). It also helps respond to recent findings showing verb effects not to depend on lexical frequency, i.e., the most frequent verbs' conditioning effect on SPE is not different from that of the less frequent ones (Posio 2011, 2015; Orozco 2016, 2018a, 2018b). This analysis also contributes to validate our main hypothesis (*Despite an internal conditioning concurrent with what is found across the board, different pronominal subject + verb collocations corresponding to a single verb condition SPE differently.*). Results show different tendencies for different collocations corresponding to a single verb. For instance, *soy* 'I am' strongly promotes overt subjects, *somos* 'we are' has a neutral tendency, and *son* 'they are' strongly promotes null subjects. Additionally, there appears to be a difference between singular and plural verb forms. In fact, as reflected on Table 8, collocations with singular subjects consistently promote overt subjects, but those with plural subjects have the opposite effect. Furthermore, it appears that the lexical effect of the verb measured by means of specific pronominal subject + verb collocations or prefabs (Bybee 2010; Bybee & Eddington 2006; Bybee & Torres Cacoullos 2008; Croft & Cruse 2004; Goldberg 2006) can help us provide a more detailed account of how verbs condition SPE. Thus, these results set the verb apart from all other internal language variation and change predictors. Our results, besides helping to account for the effects of pronominal subject + verb collocations on SPE, have substantial implications regarding the relationship between language variation and change and grammaticalization theory (Croft & Cruse 2004; Goldberg 2006; Bybee 2010). As Travis & Torres Cacoullos indicate (2012: 739), one of these implications relates to the role of such collocations as *yo creo* 'I believe, I think,' behaving as a single grammaticalized unit rather than as an analyzable pronoun + verb combination. The grammaticalization undergone by collocations would, subsequently, impact their variation patterns.

8. Conclusion. This study contributes to enrich our knowledge of SPE in Colombian Spanish. The internal conditioning in Medellín largely concurs with what obtains cross-linguistically, providing further evidence of the stability of the Spanish grammar as well as of the overall internal conditioning on SPE. This fact suggests the existence of a degree of universality regarding pronominal expression. The results of our analysis of pronominal subject + verb collocations responds to recent findings suggesting that despite four decades of research, we are yet to know how verbs condition SPE and language variation. Concurrently, our detailed analysis of the effects of the verb supports the main hypothesis tested here by revealing discrepancies between collocations corresponding to a single verb. This suggests that the analysis of collocations provides more definitive results than either analyses of verbs divided into semantic or syntactic categories (Bentivoglio 1980, 1987; Enríquez 1984; Otheguy & Zentella 2012) or using infinitives (Orozco 2016, 2018a). One important theoretical implication of our investigation is that it informs linguistic theory, as it reveals limitations in the analytical usefulness of morphosyntactically or semantically based lexical categories and their applicability to account for the lexical effects on language variation and change. The social conditioning shows the effects of age

grading, as pronominal rates increase proportionally with age. This constitutes an acquisitional trend, so far attested in Spanish, that may occur in other pro drop languages.

Moreover, this investigation helps to open new research paths, as it highlights shortcomings in how we, variationist sociolinguists, have been exploring lexical effects on language variation for decades. The analysis of collocations contributes to highlight the role of grammaticalization in language variation and change. Furthermore, our research contributes to show that we stand to benefit from integrating the analysis of syntactic and pragmatic predictors to increase the explanatory power of the forces that condition SPE. Among other things, such an analytical approach would provide us with a more comprehensive understanding of such phenomena as the effects of the verb, competition for the focus of attention, volition, and intention. This study widens our collective analytical scope and enhances the explanatory power of our findings. Our analysis opens unprecedented inquiry avenues, as it further demonstrates that the current state of affairs regarding lexical effects on SPE and, perhaps, other linguistic variables, merits further research.

References

- Abreu, Laurel. 2009. *Spanish subject personal pronoun use by monolinguals, bilinguals and second language learners*. Gainesville, FL: University of Florida dissertation.
- 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: Georgetown University Press.
- Bayley, Robert & Lucinda Pease-Alvarez. 1996. Null and expressed subject pronoun variation in Mexican-descent children's Spanish. In Jennifer Arnold, Renee Blake & Brad Davidson (eds.), *Sociolinguistic variation: Data, theory, and analysis*, 85–99. Stanford: Center for the Study of Language and Information.
- Bayley, Robert & Lucinda Pease-Alvarez. 1997. Null pronoun variation in Mexican-descent children's narrative discourse. *Language Variation and Change* 9. 349–71.
- Bayley, Robert, Norma L. Cárdenas, Belinda Treviño Schouten & Carlos Martín Vélez Salas. 2012. Spanish dialect contact in San Antonio, Texas: An exploratory study. In Kimberly Geeslin & Manuel Díaz-Campos (eds.), *Selected proceedings of the 14th Hispanic Linguistics Symposium*, 48–60. Somerville, MA: Cascadilla Proceedings Project.
- 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.
- Bentivoglio, Paola. 1980. *Why canto and not yo canto? The Problem of first-person subject pronoun in spoken Venezuelan Spanish*. Los Angeles, CA: Univ. of California MA 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.
- Bybee, Joan & David Eddington. 2006. A usage-based approach to Spanish verbs of 'becoming'. *Language* 82. 323–55. <https://doi.org/10.1353/lan.2006.0081>.
- Bybee, Joan. 2010. *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Bybee, Joan & Rena Torres Cacoullos. 2008. Phonological and grammatical variation in exemplar models. *Studies in Hispanic and Lusophone Linguistics* 1. 399–413.
- Cameron, Richard. 1992. *Pronominal and null subject variation in Spanish: Constraints, dialects, and functional compensation*. Philadelphia, PA: Univ. of Pennsylvania dissertation.

- Cameron, Richard. 1993. Ambiguous agreement, functional compensation, and non-specific *tú* in the Spanish of San Juan, PR and Madrid. *Language Variation and Change* 5. 305–34.
- Cameron, Richard. 1995. The scope and limits of switch reference as a constraint on pronominal subject expression. *Hispanic Linguistics* 6. 1–27.
- Cameron, Richard & Nydia Flores-Ferrán. 2004. Perseveration of subject expression across regional dialects of Spanish. *Spanish in Context* 1. 41–65.
- Carvalho, Ana M. & Ryan M. Bessett. 2015. Subject Pronoun Expression in Spanish in Contact with Portuguese. In Ana M. Carvalho, Rafael Orozco & Naomi Lapidus Shin (eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective*, 145–67. Washington: Georgetown 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., Rafael Orozco, & Naomi Lapidus Shin (eds.). 2015. *Subject pronoun expression in Spanish: A cross-dialectal perspective*. Washington: Georgetown University Press.
- Cerrón-Palomino, Álvaro. 2014. Ser más pro o menos pro: Variación en la expresión de sujeto pronominal en el castellano limeño. *Linguística* 30. 61–83.
- Company, Concepción & Julia Pozas. 2009. Los indefinidos compuestos y los pronombres genéricos-impersonales *omme* y *uno*. In Concepción Company (dir.), *Sintaxis histórica de la lengua española. Segunda parte: La frase nominal*. 1073–222. Mexico City: Fondo Cultura.
- Croft, William & Alan Cruse. 2004. *Cognitive linguistics*. Cambridge: Cambridge Univ. Press.
- De la Rosa, Johan. 2020. *Subject pronoun expression in Spanish-Palenquero bilinguals: Contact and language acquisition*. University Park, PA: Pennsylvania State University dissertation.
- Enríquez, Emilia. 1984. *El pronombre personal sujeto en la lengua española hablada en Madrid*. Madrid: Consejo Superior de Investigaciones Científicas.
- Erker, Daniel & Gregory R. Guy. 2012. The role of lexical frequency in syntactic variability: Variable subject personal pronoun expression in Spanish. *Language* 88. 526–57.
<https://doi.org/10.1353/lan.2012.0050>.
- 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. 49–73.
- Flores-Ferrán, Nydia. 2007a. Los Mexicanos in New Jersey: Pronominal expression and ethno-linguistic 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. 2007b. A bend in the road: Subject personal pronoun expression in Spanish after 30 years of sociolinguistic research. *Language and Linguistics Compass* 1. 624–52.
- Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- González-Rátiva, María Claudia (Coord.). 2008. *Corpus Sociolingüístico de Medellín [Electronic Portal]*. Medellín: Facultad de Comunicaciones, Universidad de Antioquia, Colombia.: <http://comunicaciones.udea.edu.co/corpuslinguistico/> (accessed on 22 August 2017).
- 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. 2005a. El uso de tú, usted y uno en el español de los colombianos y colombianoamericanos. In Luis Ortiz López & Manel Lacorte (eds.), *Contactos y contextos lingüísticos: El español en los Estados Unidos y en contacto con otras lenguas*. 185–200. Madrid/ Frankfurt: Iberoamericana/Vervuert.
- Hurtado, Luz Marcela. 2005b. Condicionamientos sintáctico-semánticos en la expresión del sujeto en el español colombiano. *Hispania* 88. 335–48.
- Hurtado, Luz Marcela. 2015. El predominio de *uno* en Bogotá, Colombia: ¿Estrategia para desfocalizar o para subjetivizar? *Anuario de Letras* (3)2. 105–151.
- Hurtado, Luz Marcela & Carolina Gutiérrez-Rivas. 2016. La versatilidad del pronombre *uno* para expresar posicionamiento frente a lo enunciado en el español de Barranquilla (Colombia). *Forma y Función* 29(1). 37–60.
- Hurtado, Luz Marcela & Ivan Ortega-Santos. 2019. On the use of *uno* in Colombian Spanish: The role of transitivity. *Studies in Hispanic and Lusophone Linguistics* 12. 35–64. <https://doi.org/10.1515/shll-2019-2001>.
- Hopper, Paul & Sandra Thompson. 1980. Transitivity in grammar and discourse. *Language* 56. 252–99. <https://doi.org/10.2307/413757>.
- Lastra, Yolanda & Pedro Martín Butragueño. 2015. Subject pronoun expression in oral Mexican Spanish. In Ana M. Carvalho, Rafael Orozco & Naomi Lapidus Shin (eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective*, 41–59. Washington: Georgetown University Press.
- 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*. 103–22. Washington: Georgetown U Press.
- Millán, Mónica. 2014. “*Vos sos paisa*”: A study of address forms in Medellín, Colombia. In Rafael Orozco (ed.), *New directions in Hispanic linguistics*, 92–111. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Montes Giraldo, José Joaquín. 1982. El español de Colombia. *Thesaurus* 37. 23–92.
- 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. 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: Georgetown University Press.
- Orozco, Rafael. 2016. Subject pronoun expression in Mexican Spanish: ¿*Qué pasa en Xalapa*? *Proceedings of the Linguistic Society of America* 1(7). 1–15. <https://doi.org/10.3765/plsa.v1i0.3703>.
- Orozco, Rafael. 2018a. *Spanish in Colombia and New York City: Language contact meets dialectal convergence*. Amsterdam/Philadelphia: John Benjamins.
- Orozco, Rafael. 2018b. El castellano colombiano en la ciudad de Nueva York: Uso variable de sujetos pronominales. *Studies in Lusophone and Hispanic Linguistics* 11(1). 89–129. <https://doi.org/10.1515/shll-2018-0004>.
- 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: Cascadilla Proceedings Project.
- Orozco, Rafael & Luz Marcela Hurtado. 2020. A variationist study of subject pronoun expression in Medellín, Colombia. *Languages* 6(1). 5. <https://doi.org/10.3390/languages6010005>.

- Ortiz López, Luis A. 2011. Spanish in contact with Haitian Creole. In Manuel Díaz-Campos (ed.), *The handbook of Spanish sociolinguistics*, 418–45. Malden: Wiley-Blackwell.
- 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 and dialect contact in Spanish in New York: Toward the formation of a speech community. *Language* 83. 770–802. <https://doi.org/10.1353/lan.2008.0019>.
- Pérez, Alder L. & Roberto Gomes. 2019. A expressão do pronome pessoal sujeito no espanhol falado no Caribe colombiano. *Revista Estudos Linguísticos* 48(1). 404–424, <https://doi.org/10.21165/el.v48i1.2148>.
- 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. 777–98. <https://doi.org/10.1016/j.pragma.2010.10.012>.
- de Prada Pérez, Ana. 2009. *Subject expression in Minorcan Spanish: Consequences of contact with Catalan*. State College, PA: Pennsylvania State University dissertation.
- de Prada Pérez, Ana. 2015. First person singular subject pronoun expression in Spanish in contact with Catalan. In Ana M. Carvalho, Rafael Orozco & Naomi Lapidus Shin (eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective*, 123–44. Washington: Georgetown University Press.
- Shin, Naomi. 2015. Bilingual language acquisition: Spanish and English in the first six years. *Heritage Language Journal* 12. 314.
- 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. 429–52.
- Shin, Naomi & Daniel Erker. 2015. The emergence of structured variability in morphosyntax: Childhood acquisition of Spanish subject pronouns. In Ana M. Carvalho, Rafael Orozco & Naomi Lapidus Shin (eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective*, 169–90. Washington: Georgetown University Press.
- 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 & Andrés Enrique-Arias. 2017. *Sociolingüística y pragmática del español* (2nd edition). Washington: Georgetown University Press.
- Torres Cacoullos, Rena & Catherine Travis. 2011. Testing convergence via code-switching: priming and the structure of variable subject expression. *International Journal of Bilingualism* 15. 241–67. <https://doi.org/10.1177/1367006910371025>.
- Torres Cacoullos, Rena & Catherine Travis. 2018. *Bilingualism in the community. Code-switching and grammars in contact*. Cambridge: Cambridge University Press.
- Torres Cacoullos & Catherine Travis. 2019. Variationist typology: Shared probabilistic constraints across (non-)null subject languages. *Linguistics* 57. 653–92.
- Travis, Catherine. 2005. 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*, 329–49. Amsterdam and Philadelphia: John Benjamins.
- Travis, Catherine. 2007. Genre effects on subject expression in Spanish: Priming in narrative and conversation. *Language Variation and Change* 19. 101–35.
- Travis, Catherine & Rena Torres Cacoullos. 2012. What do subject pronouns do in discourse? Cognitive, mechanical and constructional factors in variation. *Cognitive Linguistics* 23. 711–48.