2022. *Proc Ling Soc Amer* 7(1). 5286. https://doi.org/10.3765/plsa.v7i1.5286.

The effect of the verb on pronominal expression: A reanalysis

Rafael Orozco & Johnny Orozco*

Abstract. The verb is a known constraint on subject pronoun expression (SPE) that has been explored using several different predictor/factor configurations. This paper expands on recent investigations of how verbs condition SPE by analyzing 17,500 tokens from five locales: Barranquilla, Cali, Medellín, New York, and Xalapa. We hypothesized that verb groups do not constitute natural functional SPE constraining categories. We analyze the effect of the verb with multivariate regressions testing —as random effects factors—verbs and pronominal subject +verb collocations. Results uncover —within each corpus— significant opposite tendencies between (a) verbs in the same category and (b) finite forms of a single verb. Findings support our hypothesis revealing that pronominal subject + verb collocations provide a more definitive account of how verbs condition SPE. Findings also suggest that pragmatic/ cognitive dynamics govern verb effects on SPE and provide mounting evidence that despite four decades of research, we have not found the real effects of the verb on SPE. This research expands our analytical scope and improves the accountability of findings on SPE by offering new perspectives on the lexical effect of the verb.

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

1. Introduction. The variable alternation between null and overt pronominal subjects —*yo bailo* alternating with *bailo*, both meaning 'I dance'— constitutes the main focus of the study of subject pronoun expression (SPE). As discussed in the specialized literature (Otheguy & Zentella 2012; Carvalho, Orozco & Shin 2015; Orozco & Hurtado 2021, inter alia), the study of this morphosyntactic characteristic of pro-drop languages started in the 1970s and is now one of the most productive research strands in variationist sociolinguistics. Besides studying most varieties of Spanish, SPE scholarship has explored Portuguese, English, Chinese, Fang (spoken in Equatorial Guinea), Persian, and several other languages (Orozco forthcoming). In general terms, SPE studies have established pronominal expression tendencies which reflect marked regional differences with respect to absolute overt subject pronoun rates. The highest such rates —averaging 38% for Spanish— are found in Brazil, where it peaks at 75%, Chile, the Caribbean, and communities where Spanish is in contact with non-pro-drop languages such as English. Lower pronominal rates —generally below 30% for Spanish— are characteristic of Portugal, Spain, Latin American continental varieties, and communities where Spanish is in contact with other pro-drop languages, as in Ecuador, Peru, and Equatorial Guinea, respectively (Orozco, forthcoming).

SPE research has found statistically significant differences between pronominal rates found in different regions of the Hispanic World. Notwithstanding, there is great uniformity regarding both the linguistic predictors that probabilistically condition SPE and the trends exhibited by

*We are immensely grateful to the speakers who provided the data analyzed. We thank the generous access provided to the PRESEEA corpora by Marianne Dieck & María Claudia González (Medellín) and Róbinson Grajales & Margareth Marmolejo (Cali). We thankfully acknowledge Luz Marcela Hurtado's contribution. We are grateful for the contributions of research assistants Maritza Nemogá, Monika Estrada, Bailey Nunez, Noelle Primeaux, Alex Yandell, Alex Thomas, Baily Daberkow, Cecelia Morise, Cecilia Vazquez, Joelle Semplonius, and Sabrina Valenti. We are thankful to the audience at the 96th LSA annual meeting. We take responsibility for all remaining infelicities.
Authors: Rafael Orozco (rorozc1@lsu.edu) & Johnny Orozco, Louisiana State University (rorozc4@lsu.edu).

their corresponding factors (Carvalho et al. 2015: xiii). The notable conditioning effect similarities found support the theoretical postulate that structured linguistic variation reflects an intrinsic part of our grammatical knowledge: the patterns of use of a linguistic variable are deeply rooted at the cognitive level in our grammatical competence. SPE is mainly conditioned by grammatical person and number of the subject. Other robust predictors are coreference, priming, TMA, verb reflexivity, clause type, and lexical semantics or verb type. Generally, singular subjects favor overt subjects while plural pronouns promote null subjects. Overt pronominal subjects also occur more frequently after a change of referent, immediately after an overt subject, with verb tenses of ambiguous morphology such as the imperfect, and in main or independent clauses.

This study expands on prior research exploring SPE in five speech communities: the Colombian cities of Barranquilla, Cali, and Medellín; Xalapa, Mexico; and the metropolitan New York City Colombian enclave. We answer questions generated by recent research analyzing the effect of the verb on SPE (Orozco 2018a, 2018b, forthcoming; Orozco & Hurtado 2021; inter alia) following parameters in force for four decades (Bentivoglio 1980; Enríquez 1984) whose findings reveal opposite tendencies between verbs within the same lexical category We explore the lexical effect of the verb through an analytical approach conceived with the multipronged goal of obtaining detailed information on how verbs condition SPE and of contributing to linguistic theory with empirical findings that test the applicability of lexical groupings in variationist studies.

2. Background. Lexical categories or classes (adjectives, adverbs, pronouns, nouns, verbs) are fundamental linguistic concepts that have been used since the dawn of the study of language. Thus, the morphosyntactic theoretical concept of lexical class and directly related concepts have a deep-rooted and irrefutable tradition, as they have been fruitfully applied for centuries in morphological, syntactic, semantic analyses, etc. This seems to have facilitated on the one hand subdivisions of the different lexical classes in, e.g., different classes of adjectives, nouns, verbs, etc. However, more complex theoretical concepts such as phrase, clause, sentence, etc. have been formulated based on the concept of lexical class all these with their corresponding application beyond their immediate analytical environment. The concept of lexical class was incorporated into variationist studies practically since they began, in the late twentieth century. Despite constituting an apparently simple theoretical conceptualization, the concept of lexical class and its derivatives have profound ramifications for sociolinguistic theory since, until now, concepts such as phoneme, syllable, pronoun, verb, phrase, sentence, etc., originating in the basic branches of linguistics, have been used productively in variationist scholarship. Pronouns and verbs constitute the most relevant lexical classes in SPE studies. Specifically, within the lexical class pronoun, subject personal pronouns and the study of their variable use are the raison d'être of pronombrismo. Classifications of various types have been used to analyze how verbs condition SPE. But concepts that fit within the structuralist postulate (Chomsky 1965; Saussure 1966) according to which the linguist must essentially be concerned with the ideal speaker-listener in a homogeneous speech community are, in principle, incompatible with the postulates of variationist sociolinguistic theory (see Silva-Corvalán & Enrique-Arias 2017:2). Thus, this paper tests the applicability of verbal lexical classifications as functional categories in the analysis of SPE.

2.1. ANALYSIS OF VERB EFFECTS ON SPE. The study of the effect of the verb on SPE —initiated by Bentivoglio (1980)— has been approached from various perspectives with primarily semantic orientations and has been explored in countless investigations. The different approaches to study the effects of the verb on SPE assume that certain groups of verbs motivate the use of overt pronominal subjects; thus, suggesting that other types of verbs favor the occurrence of null subjects. The most commonly used verb classifications are described below.

2.1.1. VERB TYPE. Introduced by Bentivoglio (1980:47), this was the first classification used to explore the effect of the verb on SPE. Based on semantic criteria verbs are divided as follows.

1. Cognitive or mental activity*: aprender* 'learn,' *creer* 'believe,' *imaginar* 'imagine,' etc.
2. Perception: *oler* 'smell,' *oír* 'hear,' *ver* 'see,' *sentir* 'feel,' etc.
3. Speech: *afirmar* 'state,' *decir* 'say, tell,' *comentar* 'comment,' *hablar* 'speak,' etc.
4. Desiderative and manipulative: *desear* 'wish, desire,' *querer* 'want,' *pedir* 'request,' etc.
5. Motion: *andar* 'wander, walk' *caminar* 'walk,' *correr* 'run,' *ir* 'go,' etc.
6. Other verbs: all those that do not correspond to the above categories, e.g., *tener* 'have,' *dar* 'give,' *poder* 'be able to, can,' *vivir* 'live,' etc.

2.1.2 LEXICALCONTENT OF THE VERB. This classification was first used in Enríquez's (1984) study of SPE in Madrid. According to their lexical content, verbs are divided into four groups:

1. *External activity verbs* suppose some activity, whether physical, social, or behavioral, either in movement or rest: *comprar* 'buy,' *decir* 'tell,' *ir* 'go,' *hacer* 'do, make,' etc.
2. *Mental activity* verbs demand a psychic activity: *recordar* 'remember,' *desear* 'wish,' etc.
3. *Stative verbs* are outside of any dynamic process, being alien to all activity developed by the subject, both physical and mental: *crecer* 'grow,' *estar* 'be,' *ser* 'be,' *tener* 'have,' etc.
4. *Estimative verbs* assume a judgment or opinion: *admirar* 'admire,' *considerar* 'consider,' etc.

2.1.3. PYCHOLOGICAL VS. OTHER VERBS. Based on findings obtained with the two previous classifications, Torres Cacoullos & Travis (2010) combine them and establish this binary classification which probes the effect of psychological verbs against all other categories.

2.1.4. LEXICAL FREQUENCY. The analysis of the lexical frequency effect on SPE advanced by Erker & Guy (2012) provides an alternative to the above classifications. It also responds to calls for more definitive information about how verbs condition SPE (cf. Orozco & Guy 2008:77).

2.1.5. KINESIS OR VERB CLASS. This approach introduced by Orozco & Hurtado (2021) divides verbs according to a semantic scale consisting of seven categories: 1) activity, 2) communication, 3) emotive and volitional, 4) cognition, 5) movement, 6) physical perception, and 7) static.

Numerous research findings since the 1980s (Bentivoglio 1980; Enríquez 1984; Carvalho, Orozco & Shin 2015; Otheguy & Zentella 2012; Torres Cacoullos & Travis 2010; among others) concur that mental, cognitive, or psychological activity verbs (*pensar* 'think,' *creer* 'believe,' *acordarse* 'remember,' etc.) favor overt subjects while verbs of external activity or motion (*ir* 'go,' *llegar* 'arrive,' *venir* 'come,' etc.) promote null subjects. However, the use of categories based on semantic or lexical criteria to study how verbs condition SPE has been rendered problematic (Posio 2015:59; Orozco 2018b:112; among others). Thus, recent studies (Orozco & Hurtado 2021, Orozco forthcoming, inter alia) have called for alternatives or proposed them.

3. Methodology. In this section, we address the corpora and data analyzed. We also state our research questions and hypothesis. Additionally, we describe our research procedure.

3.1. CORPORA AND DATA. The data analyzed here were extracted from socially stratified corpora made from semi-directed conversations with residents of five Spanish-speaking speech communities: Barranquilla, Cali, Medellín, Xalapa, and the New York City Colombian enclave. In total, the data were provided by 203 speakers. The data from Barranquilla —a Colombian city on the Caribbean—comes from Corpus del Castellano Barranquillero, collected between 1997 and 1999, fully described in Orozco (2018a:20). Our second corpus is from Cali, a city in the Pacific region of Colombia. The Project for the Sociolinguistic Study of Spanish in Spain and America

(PRESEEA)-Cali Corpus (Grajales & Marmolejo 2019) was collected between 2016 and 2019. Our third data source is the PRESEEA-Medellín Corpus (González Rátiva 2008), compiled between 2007 and 2010. Medellín has a unique sociolinguistic situation among Colombian metropolises, being located in a region where Caribbean and Andean speech converge. Corpus del Castellano Xalapeño, collected between 2009 and 2010, is our fourth corpus. It is described in Orozco (2016). Xalapa is located on a mountain range, roughly 80 miles from the Gulf of Mexico, and the Spanish spoken there is representative of inland Mexican Spanish. The Corpus of Colombian Spanish in New York was collected between 2000 and 2002. The NYC Colombian community simultaneously coexists with English and NYC Spanish, which is essentially a variety of Caribbean Spanish. Orozco (2018a) describes the speech community and the corpus.

3.2. RESEARCH QUESTIONS AND HYPOTHESES. This study seeks to fill gaps in SPE research by exploring in detail the lexical effect of the verb on the alternation between overt and null pronominal subjects with the main objective of probing the analytical efficacy of the traditional lexical classifications. This research is guided by the following main research questions.

(a) *Is there a clear conditioning effect of lexical frequency on subject pronoun expression?*

(b) *Is pronominal expression similarly conditioned by all verbs within a given lexical category?*

(c) *What do we learn from the analysis of the effect of subject pronoun + verb collocations on SPE?*

We also test the following main hypothesis: *Verb groups (semantic or syntactic) do not constitute natural functional categories in variable subject pronoun expression.*

Our research questions and hypothesis are informed by the findings of many prior studies cited through this paper, particularly those that point to the lack of conclusive results on how the verb conditions SPE (Erker & Guy 2012; Orozco 2016, 2018a, 2018b; Orozco & Guy 2008; Orozco & Hurtado 2021; Orozco forthcoming, Posio 2015; among others).

3.3. THE ENVELOPE OF VARIATION AND THE ANALYTICAL PROCEDURE. The envelope of variation used in this investigation conforms to the comprehensive parameters practiced as a norm in pronombrista, i.e., SPE studies (cf. Bentivoglio 1980 and Otheguy & Zentella 2012:48 inter alia). The envelope of variation is fully described in Orozco (2018a:99) and in Orozco & Hurtado (2021:10). Thus, our analyses include only those clauses with ascertainable animate pronominal subjects that contain a finite or conjugated verb where the alternation between an overt pronominal subject and a null one is clearly possible. That is, all the tokens selected for analysis constitute one of at least two different possible ways of saying the same thing.

This study expands on previous research that explores the effects of linguistic, cognitive, and social predictors in each of our five speech communities (Orozco 2016, 2018a, 2018b; Orozco & Hurtado 2021, among others). We analyzed 17,077 tokens (3009 from Barranquilla, 2679 from Cali, 4623 from Medellín, 3006 from NYC, and 3760 from Xalapa) coded in terms of ten internal and cognitive predictors operating at different morphosyntactic and discursive levels and are divided in three groups. (a) predictors pertaining to the whole clause: discourse style and clause type; (b) subject-related predictors: switch reference, person and number of the subject, and priming; (c) verb-related predictors: TMA form of the verb, preceding TMA, verb regularity, verb type, and lexical content of the verb. These predictors are fully described in Orozco (2018a: 97). As with our hypothesis and research questions, our predictors were selected based on the findings of five decades of pronombrista studies (cf. Carvalho et al. 2015; Enríquez 1984; Orozco 2018a, 2018b; Otheguy & Zentella 2012; Torres Cacoullos & Travis 2010; inter alia).

The data were also coded in terms of five social predictors fully described in Orozco (2018a:125 ff.): conversation conditions, education, socioeconomic status, age, and gender as well as others that only apply to the NYC Colombian community. We proceeded with a series of multipronged multivariate mixed-effects regression analyses to first determine the overall conditioning effects for each corpus and, subsequently, explore in detail the effects of the verb. We first corroborated that the verb, as it has traditionally been explored (using lexical groupings based on semantic criteria), conditions SPE in each community. Then, we analyzed the lexical effect of the verb by means of additional multivariate analyses using the infinitive as a random predictor. We also analyzed as a random predictor the effects of pronominal subject plus verb collocations.

4. Results. We begin by presenting the distribution of overt and null pronominal subjects. Then, we present the results for lexical content of the verb followed by those for the lexical effects of the verb, first according to infinitives, and then according to subject plus verb collocations.

4.1. BASELINE TENDENCIES. The distribution of overt and null subjects is presented in Table 1. The pronominal rates, i.e., the overt subject percentages (in bold), increase progressively from left to right, corroborating pan-Hispanic SPE parameters and regional differences. The pronominal rates in Xalapa, Medellín, and Cali, being below 30%, concur with what is established for continental speech communities. At the same time, the pronominal rates in Barranquilla and New York, being higher than 30%, reflect the most frequent occurrence of overt pronominal subjects in the Caribbean and in communities where Spanish is in contact with English. (cf. Otheguy & Zentella 2012; Carvalho et al. 2015; Orozco 2018a; among others).

Variant	Xalapa	Medellín	Cali	Barranquilla	New York
Overt subjects	**25.1%**	**27.9%**	**29.3%**	**34.3%**	**43.3%**
(***yo*** *canto* 'I sing')	943/3760	1294/4623	784/2,679	1031/3009	1303/3006
Null subjects	74.9%	72.1%	70.7%	65.7%	56.7%
(***Ø*** *canto* '[I] sing')	2817/3760	3329/4623	1895/2,679	1978/3009	1703/3006

Table 1. Distribution of overt and null pronominal subjects

Despite statistically significant pronominal rate differences between our continental speech communities (Cali, Medellín, and Xalapa) and our lowlands communities (Barranquilla and NYC), our preliminary analyses also revealed a high degree of uniformity in terms of linguistic conditioning. The internal conditioning found reveals that in every one of our speech communities, grammatical person and number of the subject exerts the strongest conditioning effect on SPE. The consistent predictor hierarchies as well as the consistent conditioning effects of switch reference, TMA, and verb semantics found in our five speech communities are consonant with what prevails throughout the Hispanic World and beyond in both monolingual and bilingual speech communities (cf. Bouchard 2018; Carvalho et al. 2015, Orozco & Hurtado 2021; Otheguy & Zentella 2012; Silva-Corvalán & Enrique-Arias 2017, among others).

Furthermore, the similarity between the conditioning effects in our five speech communities and the rest of the world supports the pronombrista theoretical postulate of uniform linguistic conditioning, which is corroborated by the findings widely reported in the pronombrista literature cited in this paper. Those findings reveal a high degree of agreement regarding both the predictors that condition SPE and the tendencies of the factors within each predictor (Carvalho et al. 2015, Orozco & Hurtado 2021:2, inter alios). The uniformity in linguistic conditioning also corroborates that the underlying SPE grammar in all varieties of Spanish remains essentially the same despite significant pronominal rate differences at the surface level (Erker & Guy 2012;

Carvalho et al. 2015, Torres Cacoullos & Travis 2019, Silva Corvalán & Enrique-Arias 2017, among others). Although details pertaining to how linguistic, cognitive, and social predictors condition SPE are beyond the scope of this paper, they are addressed in the foundational investigations for this research: Orozco (2018a) for Barranquilla and NYC; Orozco & Hurtado (2021) for Medellín; Orozco, Marmolejo & Grajales (2021) for Cali; and Orozco (2016) for Xalapa.

4.2. EFFECTS OF VERB GROUPINGS ON SPE: LEXICAL CONTENT OF THE VERB. Our analysis of how verbs condition pronominal subjects probes the perspective practiced in pronombrismo since 1980 —conceived based on traditional theoretical parameters (cf. Chomsky 1965, Saussure 1966)— which presumes the existence of verb groups that promote the occurrence of overt pronominal subjects and others that promote null subjects. To test the validity of that premise, in our more detailed analysis, we probed predictors based on verb groupings used to explore the effects of the verb in pronombrista studies, namely: type of verb (Bentivoglio 1980), lexical content of the verb (Enríquez 1984), kinesis (Hurtado & Ortega-Santos 2019), and transitivity (Orozco & Hurtado 2021). Although our detailed analyses employ four different verb classifications, here we focus on the effects of lexical content of the verb, which are also representative of the other classifications.

To explore the effect of lexical content, we initially divided verbs into the four categories used by Enríquez (1984): stative, mental activity, estimative, and external activity. Preliminary results revealed similar trends for mental activity and estimative verbs. Thus, as Erker & Guy (2012), among others, did, we amalgamated these two categories under the label *Mental Activity*. The results in Table 2 show a great deal of congruence among the different speech communities. Mental activity verbs favor overt subjects in Xalapa (.53), Medellín (.56), Cali (.54) and New York (.55) but have a neutral effect in Barranquilla (50). Stative verbs fluctuate, moderately favoring overt subjects in Xalapa (.53) and Barranquilla (.55) while registering neutral effects in the other three communities with probabilistic weights of .52, .51, and .48, respectively. At the same time, external activity verbs favor null subjects in all five communities with probabilistic weights between .43 and .47. In sum, mental activity verbs and stative verbs fluctuate from one speech community to another between neutral tendencies and tendencies that favor overt pronominal subjects while external activity verbs consistently favor null subjects in all our five speech communities. These trends are consistent with the findings of Enriquez (1984: 240) and Erker & Guy (2012: 541), and many others over four decades of study of the effects of the verb on SPE.

Lexical content	Xalapa		Medellín		Cali		Barranquilla		New York	
	Prob.	%	Prob.	%	Prob.	%	Prob.	%	Prob.	%
Mental activity	.53	33%	.56	36%	.54	41%	.50	46%	.55	60%
Stative verbs	.53	27%	.52	32%	.51	36%	.55	41%	.48	41%
External activity	.45	22%	.47	25%	.43	27%	.45	29%	.47	38%
Range	*8*		*9*		*11*		*10*		*8*	

Table 2. Effect of Lexical Content of the Verb

Besides the trends exhibited by the statistical probability values, the overt pronominal percentage rates allow us to identify a common trend: mental activity verbs have the highest overt pronominal rates —New York's (60%) being the highest— whereas external activity verbs have the lowest rates, fluctuating between 22% in Xalapa and 38% in New York City. Concurrently, stative verbs generally have lower pronominal rates than mental activity verbs but higher than

external activity verbs. Broadly speaking, the trends registered by lexical content of the verb corroborate those found both in other varieties of Spanish (Carvalho, Orozco, & Shin, 2015) and in Portuguese (Bouchard 2018).

4.3. LEXICAL EFFECT OF THE VERB: THE INFINITIVE. The tendencies registered by lexical content of the verb, one of the classifications most frequently used in pronombrista studies, serve to corroborate that verb semantics conditions SPE in our speech communities. Next, aiming to provide alternatives to lexical groupings, we explore how the verb conditions SPE in terms of the lexical effect of the infinitive.

Verb	Prob.	% Overt	N	% data
Creer 'believe'	**.856**	75.4%	43/57	1.9%
Recordar 'remember'	.793	71.4%	15/21	0.7%
Pensar 'think'	.781	63.6%	14/22	0.7%
Estudiar 'study'	.737	63.6%	14/22	0.7%
Saber 'know'	.705	50.5%	49/97	3.2%
Ser 'be'	.704	50.9%	81/159	5.3%
Ver 'see'	.680	45.3%	24/53	1.8%
Decir 'tell'	.647	42.4%	87/205	6.8%
Llegar 'arrive'	.635	40.5%	17/42	1.4%
Quedar 'stay'	.621	40.0%	16/40	1.3%
Vivir 'live'	.606	40.0%	22/55	1.8%
Estar 'be'	.606	42.2%	57/135	4.5%
Ir 'go'	.589	38.4%	33/86	2.9%
Tener 'have'	.579	36.8%	89/242	8.0%
Dejar 'leave'	.576	34.1%	15/44	1.5%
Trabajar 'work'	.576	38.2%	13/34	1.1%
Querer 'want'	.541	34.0%	17/50	1.7%
Irse 'leave'	.530	31.5%	17/54	1.8%
Venir 'come'	.526	30.5%	18/59	2.0%
Salir 'exit, go out'	.521	28.3%	13/46	1.5%
Poder 'be able to, can'	.513	31.4%	11/35	1.2%
Hacer 'do'	.509	28.4%	25/88	2.9%
Pasar 'pass'	.508	27.6%	8/29	1.0%
Acordarse 'remember'	.477	25.0%	18/72	2.4%
Comprar 'buy'	.467	23.8%	5/21	0.7%
Poner 'put'	.415	21.1%	15/71	2.4%
Coger 'take, grab'	.386	17.1%	7/41	1.4%
Dar 'give'	.355	16.7%	8/48	1.6%
Llevar 'carry'	.298	9.1%	3/33	1.1%
Jugar 'play'	.264	7.1%	2/28	0.9%

Table 3. lexical effect of the verb in Barranquilla (pronominal rate: 34.3%)

The findings in Table 3 show the lexical effect of the verb in Barranquilla. These tendencies are representative of what obtains in our other four speech communities. Case in point, *tener* 'have' (.579) emerges as the most frequent verb in our five corpora, corroborating that it is the

most frequent verb in pronominal contexts. In Barranquilla, it registers 242 occurrences which represent 8% of the data. In Medellín *tener* accounts for 9.2% of the data, in NYC for 9.5%, and in Xalapa 11% (Orozco forthcoming).

In addition, in Barranquilla, *creer* 'believe' (.856), *recordar* 'remember' (.793), and *pensar* 'think' (.781) stand out as the verbs that most strongly favor overt subjects. However, *dar* 'give' (.355), *llevar* 'carry' (.298), and *jugar* 'play' (.264) emerge respectively as the verbs that most strongly favor null subjects. None of these six verbs is one of the most frequent. These effects are consistent with those observed in Xalapa and Medellín. Moreover, by corroborating that the most frequent verbs do not exert a consistent effect either favorable or unfavorable towards overt or null pronominal subjects, we have answered our first research question

The findings in Tables 3 and 4 also reveal that not all verbs grouped within the same semantic category exhibit identical tendencies either favoring the occurrence of overt subjects or promoting the occurrence of null subjects. The results presented in Table 4 report five of the largest statistically significant differences between the effects of verbs grouped within a single lexical category. The differences between the remaining pairs of verbs found in Barranquilla correspond to the external activity (lexical content), motion (type of verb and kinesis), other (type of verb), and transitive (transitivity) categories. The two largest statistically significant differences found in Barranquilla for verbs in the same category are the following:

- *Creer* 'believe' (.856) registers the greatest favorable effect on overt subjects among transitive verbs while *coger* 'take' (.386) strongly favors null subjects.
- Among cognitive verbs (type of verb and kinesis) or mental activity (lexical content), *recordar* 'remember' (.793) favors overt subjects while *acordarse* 'remember' (.477) favors null subjects.

Verb	Prob.	% Overt	N	X^2	*P*	Lexical category
Creer 'believe'	.856	75.4%	43/57	30.2	<.001	Transitive
Coger 'take'	.386	17.1%	7/41			
Recordar 'remember'	.793	71.4%	15/21	13.4	<.001	Cognitive Mental activity
Acordarse 'remember'	.477	25.0%	18/73			
Ver 'see'	.680	45.3%	24/53	10.5	.001	External activity
Jugar 'play'	.264	7.1%	2/28			
Ir 'go'	.589	38.4%	33/86	9.6	.002	Motion External activity
Llevar 'take'	.298	8.4%	3/33			
Tener 'have'	.579	36.8%	89/242	6.4	.011	Other Transitive
Dar 'give'	.355	16.7%	8/48			

Table 4. Same category discrepancies in the effects of verbs on SPE: Barranquilla

Moreover, in the other four speech communities, we find verbs with significant opposite tendencies in each of the lexical classifications that we used to explore the effects of verb groups on SPE. By showing significant differences between verbs within a single lexical category, we answered our second research question. Thus, we can say that *pronominal expression is not similarly conditioned by all verbs within a given lexical category* in any of the corpora analyzed. At the same time, it is evident that by exploring the lexical effect of the verb in terms of infinitives, we continue to generalize by grouping several forms of the same verb under a single designation.

4.4. LEXICAL EFFECT OF THE VERB: SUBJECT PRONOUN + VERB COLLOCATIONS. We now address our third research question. In so doing, we sought to deepen our analysis by going beyond the infinitive and avoiding generalizations. Thus, we probed the effects of subject + verb collocations as random effects factors on SPE. The results of this analysis are provided in Table 5.

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

Table 5. Effects of lexical frequency according to pronominal subject + verb collocations in Medellín, Colombia

Our pronominal subject + verb collocations analysis reveals *(yo) creo* 'I think, I believe' as the collocation that most strongly promotes overt pronominal subjects with a probability value of .877 and an astronomical pronominal rate of 73%. This concurs with findings in Cali (Travis & Torres Cacoullos 2012:739). This also corroborates that *(yo) creo* 'I think, I believe' consistently exerts a strong favorable effect on overt pronominal subjects in many other speech communities

(Travis & Torres Cacoullos 2012; Torres Cacoullos & Travis 2018). Table 5 also shows that, despite few exceptions, singular collocations promote overt pronominal subjects whereas plural collocations favor null subjects. In fact, plural inflectional forms overwhelmingly have null subjects, thus disfavoring overt pronominal subjects. For instance, *estamos* 'we are,' *hacemos* 'we do,' and *vamos* 'we go' —registering the lowest probability values and pronominal rates— are located at the very bottom of Table 5.

Moreover, an examination of the relationship between semantic class and lexical frequency reveals a discrepancy within verbs of cognition, traditionally considered to promote overt subjects (Carvalho et al. 2015; Otheguy & Zentella 2012; Travis & Torres Cacoullos, among others), as follows. *Creo* 'I think, I believe,' *sabe* 'she, he, one knows,' *pienso* 'I think,' *conozco* 'I know,' and *conocí* 'I knew,' respectively favor overt subject pronouns. Contrariwise, *imagino* '[I] imagine' and *sé* '[I] know,' favor null subjects. In terms of lexical frequency, the results in Table 5 also uncover that the most frequent subject + verb collocations, i.e., those on that table, do not seem to exert a specific frequency effect on SPE (cf. Orozco 2018a, 2018b). That is, the more frequent collocations do not exert tendencies consistently different or opposite to those of the less frequent ones. Therefore, we now possess mounting evidence that lexical frequency alone cannot account for the effect of the verb on SPE.

C ollocation	**Prob.**	**% Overt**	**N**	X^2	***P***	**Verb**
Soy '[I] am'	.807	60.0%	39/65	18.04	<.001	*Ser* 'be'
Son '[They] are'	.323	10.3%	3/29			
Soy '[I] am'	.807	60.0%	39/65	3.60	.05	*Ser* 'be'
Era '[I] was'	.669	36.7%	11/30			
Tenía '[I] had'	.770	51.6%	16/31	14.6	<.001	*Tener* 'have'
Tenemos '[We] have'	.350	12.7%	8/63			
Tenía '[I] had'	.770	51.6%	16/31	7.3	.006	*Tener* 'have'
Tengo '[I] have'	.530	25.0%	34/136			

Table 6. Same verb discrepancies in the effects of verbs on SPE: Barranquilla

We also find, analogous to the effects of infinitives, discrepancies between collocations corresponding to a single verb. These discrepancies are representative of what happens across the board, and they occur in all our five corpora. Concurrently, as Tables 5 and 6 show, the findings regarding the effects of pronominal subject + verb collocations also reveal opposing tendencies between inflectional forms of a single verb within each speaker cohort. Table 6 shows opposing tendencies between finite forms of *tener* 'have' and *ser* 'be' in Barranquilla.

5. Discussion. This paper has addressed three research questions and a main hypothesis with the main goal of exploring the complex conditioning effect of the verb on SPE beyond what has been done so far. We analyzed five socially stratified corpora (Barranquilla, Cali, Medellín, Xalapa, and the NYC Colombian enclave). The pronominal rates found, between 25.1% in Xalapa and 43.3% in NYC, concur with well-known pan-Hispanic pronombrista parameters. The pronominal rates in Xalapa, Medellín, and Cali — below 30%— fit within those characteristic of continental speech communities. Concurrently, the pronominal rates in Barranquilla and New York are typical of both Caribbean communities and those in contact with English, where pronominal rates higher than 32%, prevail (Orozco 2018a:100). Despite significant pronominal rate

differences among our five cohorts, the pronombrista theoretical postulate of the universality of linguistic conditioning (Carvalho et al. 2015) holds, with grammatical person and number of the subject as the predictor that most strongly conditions SPE. The remaining predictors that significantly condition SPE in the speech communities studied here concur with the in the rest of the Hispanic World. Thus, evidence is provided for the stability of Spanish grammar, i.e., of an underlying grammar for all varieties of Spanish despite of significant pronominal rate differences (Carvalho et al. 2015; Torres Cacoullos & Travis 2019; Travis 2005, 2007).

To analyze the effects of verbal groupings on SPE, we explored three semantic predictors (verb type, lexical content of the verb, and kinesis) and one syntactic (transitivity). Multivariate analysis results –essentially what has been done since 1980– find that the four classifications explored condition SPE, corroborating the conditioning effect of the verb on SPE. However, this part of the present research —as recent studies indicate (Orozco 2016, 2018a, 2018b; Orozco & Hurtado 2021, among others)— only proves what we already know without increasing our knowledge of the effects of the verb on SPE. Thus, we explore the lexical effect of the verb by examining how infinitives condition the alternation between null and overt subjects. This analysis serves multiple functions; on the one hand, it serves to probe the effect of lexical frequency on SPE; on the other hand, it tests the entrenched pronombrista premise (see e.g., Carvalho, Orozco, & Shin 2015: xv) of the existence of lexical groups that favor the occurrence of overt subjects and other groups of verbs that promote null subjects. The results of the analyses of the lexical effect of the verb in the five corpora analyzed (presented in Tables 3, 4, 5, and 6) reveal that the effect of the verb on pronominal expression is not governed by lexical frequency. The lack of a consistent effect either favorable or unfavorable towards overt pronominal subjects by the most frequent verbs is illustrated in Tables 3 and 5. Table 3 contains the most frequent infinitives in Barranquilla whereas Table 5 contains the most frequent pronominal verb + verb collocations in Medellín. Both of those tables include verbs that favor overt subjects, others with a neutral effect, and verbs that favor null subjects. Thus, the different tendencies found among the most frequent verbs constitutes evidence that in none of the five corpora analyzed the most frequent verbs in pronominal contexts register a consistent favorable effect toward overt or null subjects. That is, the conditioning effect of the most frequent verbs on pronominal expression is not radically different from that of the less frequent ones. This finding corroborates what was found in New York Spanish by Erker & Guy (2012:527) and answers our first research question (*Is there a clear conditioning effect of lexical frequency on subject pronoun expression?*).

When comparing the lexical effects results with those of the verb classes, we find discrepancies in the effects of verbs grouped into different categories for each of the four classifications analyzed. Table 4 shows the five greatest statistical discrepancies for infinitives in Barranquilla, which are representative of what occurs in each of the corpora analyzed. The greatest discrepancy is found in Medellín, among the transitive verbs *creer* 'believe' and *poner* 'put' ($X^2 = 49.9$; $p = <.001$). *Creer* 'believe' strongly favors overt subjects with a statistical weight of .882 and a pronominal rate of 72.3%; on the contrary, *poner* 'put' strongly favors null subjects with a weight of .328 (pronominal rate 7.9%). The second greatest opposition between verbs within the same lexical category occurs in New York between *pensar 'think'* and *acordarse 'remember'* ($X^{2=}$ 30.8; $p = <.001$). These verbs are classified either as cognitive (verb type, kinesis) or as mental activity verbs (lexical content). *Pensar* 'think' strongly favors overt subjects (statistical weight .896, pronominal rate 87.7%) while *acordarse* 'remember' promotes null subjects (.386; 22.7%). By probing the reliability of the various verb classes with the analyses of the lexical effect of the verb, we find serious shortcomings. The verb groupings hide statistically significant

differences between verbs grouped in the same class. These findings answer our second research question (*Is pronominal expression similarly conditioned by all verbs within a given lexical category?*). They also support our main hypothesis (*Verb groups (semantic or syntactic) do not constitute natural functional categories in variable subject pronoun expression.*), as we find different tendencies for verbs within the same lexical classification.).

By discovering opposite tendencies between verbs traditionally grouped in the same categories based on the verbal classifications of Bentivoglio (1980) and Enríquez (1984) and others that emerged later, the analysis of the lexical effect also discovers that lexical groupings based on semantic or syntactic criteria do not constitute natural functional classes for the analysis of the variable pronominal expression. Consequently, this research demonstrates the existence of limitations in the application of lexical categories to pronombrismo. That is, although lexical categories are functional in semantic and morphosyntactic contexts, they are not completely applicable in SPE contexts. Therefore, this empirical analysis generates theoretical questions about the applicability of lexical groupings in variationist research exploring other linguistic variables.

Our findings regarding the effect of pronominal subject + verb collocations suggest that these constructions constitute two kinds of prefabricated units in Spanish: (*yo) creo* 'I think, I believe' vs. all others. *(Yo) creo* stands out as a cognitively reanalyzed unit grammaticalized as a discourse formula (cf. Torres Cacoullos & Walker 2009a). Thus, it consistently favors overt pronominal subjects cross-dialectally (cf. Travis & Torres Cacoullos 2012) and exemplifies an instance of *autonomy* (Bybee 2003, 2006; Hopper & Traugott 2003: 127). Yet, all other pronominal subject + verb collocations are idiosyncratic and remain grammatically productive. Thus, although the effect of the verb on SPE was considered a resolved issue with cognitive-psych verbs promoting overt subjects (Carvalho et at. 2015:xv, Orozco 2015, inter alios), we now have evidence that the semantically-based classifications used to explore the effect of the verb on SPE for the last four decades (Bentivoglio 1980, Enríquez 1984) fail to uncover important differences not only between verbs in a given category but between inflectional forms of the same verb. Further, our four-way comparison sheds light on the effects of lexical idiosyncrasy. It suggests that lexical idiosyncrasy is the norm and seems to more reliably account for the effects of the verb on SPE and perhaps other linguistic variables.

Moreover, by exploring the verb in terms of pronominal subject + verb collocations (cf. Bybee 2010; Bybee & Eddington 2006; Bybee & Torres Cacoullos 2009; Croft & Cruse 2004; Travis & Torres Cacoullos 2012, among others), our analysis is more detailed than previous analyses and uncovers important facts, including numerous instances of opposing tendencies regarding finite forms of a single verb. These findings suggest that exploring the effects of the verb on SPE by using collocations informs our collective knowledge beyond what we already know. Thus, it appears that by using analyses that probe the effects of pronominal subject + verb collocations (cf. Bybee & Eddington 2006; Bybee & Torres Cacoullos 2009; Goldberg 2006; Torres Cacoullos & Walker 2009b; Travis & Torres Cacoullos 2012, among others), we can obtain more conclusive answers as to how lexical effects condition language variation and change. Our lexical effects analysis helps respond to recent research that shows the lexical effect of the verb a) not to depend on lexical frequency; that is, the most frequent verbs do not behave differently from the less frequent ones; and b) to exert different conditioning effects in different speech communities (Posio 2011, 2015; Orozco 2018a, 2018b; Orozco et al. 2014). That is, recent findings provide mounting evidence that despite four decades of research, we are yet to know the real effects of the verb on SPE. It appears that the lexical effect of the verb measured by means of specific pronominal subject + verb collocations or prefabs (cf. Bybee 2010; Bybee &

Eddington 2006; Bybee & Torres Cacoullos 2009; Croft & Cruse 2004; Goldberg 2006) can help us provide a more detailed account of how verbs condition SPE.

6. Conclusion. Five decades of SPE research contributed remarkably to variationist sociolinguistics. The pronombrista research tradition has informed us about the congruence of internal conditioning in SPE, which extends beyond the Hispanic world (Bouchard 2018; Carvalho, Orozco & Shin 2015; Jia & Bayley 2002; Padilla 2020; Torres Cacoullos & Travis 2018; among others). The present investigation seeks to offer new perspectives regarding the study of the effect of the verb on SPE by going beyond previous research (Cerrón-Palomino 2014; Orozco 2018a; Otheguy & Zentella 2012; Torres Cacoullos & Travis 2010, 2018; among others) and allows us to conclude that SPE is not directly governed by lexical frequency and that verbs do not constitute natural functional classes as they relate to variable subject pronoun expression.

Our results prove that, despite four decades of investigating the effects of the verb on SPE, we still do not know in detail what its effects are. Our analysis answers our first research question by revealing opposite tendencies between verbs classified within a single category. For example, in Barranquilla, *recordar* 'remember,' (pronominal rate 71%) strongly favors overt subjects while *acordarse 'remember,'* (pronominal rate 25%) has the opposite effect. This shows that the analysis of verbs divided into lexical categories (Bentivoglio 1980; Enríquez 1984; Orozco & Guy 2008; Otheguy & Zentella 2012; among others) obscures important differences between verbs grouped in the same category. The answers to our research questions tell us that a) there is not a clear lexical frequency effect on SPE; b) SPE is not similarly conditioned by all verbs within a given lexical, and c) the analysis of the effect of subject pronoun + verb collocations on SPE provides more definitive results that prior analyses. Thus, our research validates our main hypothesis, as we prove that verb groupings, either semantic or syntactic do not constitute natural functional categories as far as variable subject pronoun expression is concerned.

This study helps open promising research trails, since it contributes to show that, leaving lexical groupings behind when exploring the effects of the verb, we can increase the explanatory power of SPE research. Our results, also help to explain the effects of the verb on SPE with substantial implications regarding the relationship between variation and linguistic theory. Despite concluding that it is time to go beyond verbal classifications —verb type (Bentivoglio 1980) and verbal lexical content (Enríquez 1984) have been the most copiously used in pronombrista studies— it is important to recognize their merits. These verb classifications emerged during an incipient period of variationist sociolinguistics and contributed to the advance of pronombrismo when VARBRUL was the quantitative tool of choice in the late twentieth century. But now that we have more sophisticated quantitative tools, we can reevaluate what has been done previously and propose new approaches to investigate, as we have done in this work, the lexical effect of the verb and other lexical effects within variationist studies. Among other things, this work makes us reconsider the role of lexical classifications and informs us that what may make sense in syntactic analytical reasoning, does not infallibly match what can occur in the brains of speakers or in variable contexts. An important theoretical impact of our conclusions is related to other linguistic variables. Given the nature of linguistic variation (Chambers 2009; Labov 1994; Tagliamonte 2006, 2012), and that one case of linguistic variation and/or change generally goes hand in hand with others, it is to be expected that limitations will be discovered in the lexical categories used to explore other cases of linguistic variation. Possibly, our results have implications in the study of linguistic variables in which the effects of lexical groupings such as the expression of futurity have been explored (Blas Arroyo 2018, Orozco 2018a; Sedano 1994) and the expression of the nominal possession (Freeman 2019; Orozco 2018a)

Thus, our results set the verb apart from all other internal language variation and change predictors. They provide evidence that the differences in how verbs condition language variation in our corpora may be triggered by lexical idiosyncrasy. That is, the lexical effects on a given linguistic variable differ among speech communities due to the intrinsic idiosyncratic characteristics of every community's lexicon. These findings, by helping to account for the idiosyncratic effects of pronominal subject + verb collocations on SPE, have important implications in terms of the relationship between grammaticalization theory and language variation (cf. Bybee 2010; Croft & Cruse 2004; Goldberg 2006). This pilot study widens our collective analytical scope and enhances the explanatory power of our findings. This study further demonstrates that the current state of affairs both in the communities explored and regarding the effects of the verb merits further investigation, as it opens unprecedented inquiry avenues for research endeavors such as that proposed here. These opposing tendencies reveal a clear idiosyncratic effect that lends support to the lexical idiosyncrasy theory, which we seek to further develop with the research proposed here.

This study contributes to closing a chapter and opening new research avenues in SPE and variationist sociolinguistics. The results expand our collective knowledge of the phenomena that condition variable SPE in Spanish. By analyzing the effect of the verb through lexical groupings, pronombrismo has reached a barely superficial knowledge of this important predictor. These results provide additional information regarding the effects of the verb on pronominal expression and reveal the need to study it in more detail both in pronombrismo and in the analysis of other linguistic variables. In addition, with the advent of more sophisticated quantitative tools than those used by previous generations of variationist researchers, which allow for more detailed analyses, we are on the threshold of being able to address the questions opened by this research.

References

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 MA thesis.

Bouchard, Marie-Eve. 2018. Subject pronoun expression in Santomean Portuguese. *Journal of Portuguese Linguistics* 17(5) 1–29. https://doi.org/10.5334/jpl.191.

Carvalho, Ana M., Rafael Orozco, & Naomi Lapidus Shin (eds). 2015. *Subject pronoun expression in Spanish: A cross-dialectal perspective*. Washington: Georgetown University Press.

Chomsky, Noam. 1986. *Knowledge of language*. New York: Praeger.

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.

González-Rátiva, María Claudia (coord.). 2008. *Corpus sociolingüístico de Medellín.* Medellín: Universidad de Antioquia. http://comunicaciones.udea.edu.co/corpuslinguistico/.

Grajales, Róbinson & Margareth Marmolejo (coords.). 2019. *Corpus lingüístico de Cali* Cali: Universidad del Valle. https://preseea.linguas.net/Equipos/Cali.aspx/.

Hurtado, Luz Marcela & Rafael Orozco. Forthcoming. La variable expresión de sujetos pronominales al estilo paisa. In Néstor Ruíz (ed.), *Perspectivas actuales de la investigación en lingüística*, Bogotá, Colombia: Publicaciones del Instituto Caro y Cuervo.

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.
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. Forthcoming. El efecto léxico en la variación lingüística: Expresión de sujetos pronominales. In Néstor Ruíz (ed.), *Perspectivas actuales de la investigación en lingüística*, Bogotá, Colombia: Publicaciones del Instituto Caro y Cuervo.
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 & Luz Marcela Hurtado. 2021. A variationist study of subject pronoun expression in Medellín, Colombia. *Languages* 6(1). 5. https://doi.org/10.3390/languages6010005.
Orozco, Rafael, Margareth Marmolejo & Róbinson Grajales. 2021. La variable expresión de sujetos pronominales en Cali, Colombia. Paper presented at *50th Annual Meeting of the Linguistic Association of the Southwest (LASSO)*, Las Vegas, NV.
Otheguy, Ricardo & Ana Celia Zentella. 2012. *Spanish in New York: Language contact, dialectal leveling, and structural continuity.* Oxford: Oxford University Press.
Posio, Pekka. 2015. Subject pronoun usage in formulaic sequences: Evidence from Peninsular Spanish. In Ana Carvalho, Rafael Orozco & Naomi Shin (eds.), *Subject pronoun expression in Spanish: A cross-dialectal perspective*, 61–80. Washington: Georgetown University Press.
Saussure, Ferndinand de. 1966. *Course in general linguistics.* English translation by Wade Baskin [1959]. New York: McGraw-Hill.
Silva-Corvalán, Carmen & Andrés Enrique-Arias. 2017. *Sociolingüística y pragmática del español* (2nd edition). Washington: Georgetown University Press.
Torres Cacoullos, Rena. & Catherine Travis. 2010. Variable *yo* expression in New Mexico: English influence? In Susana Rivera-Mills & Daniel Villa (eds.), *Spanish of the US Southwest: A language in transition*, 185–206. Madrid: Iberoamericana.
Torres Cacoullos, Rena & Catherine E. Travis. 2018. *Bilingualism in the community: Codeswitching and grammars in contact.* Cambridge, UK.: 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.