

# Using Sociolinguistic Analyses of Discourse-Level Features to Expand Research on L2 Variation in Forms of Spanish Subject Expression

Kimberly L. Geeslin and Aarnes Gudmestad

Indiana University and Virginia Polytechnic Institute and State University

## 1. Introduction

The ability to vary one's speech according to the interlocutor, setting and other social variables (i.e., sociolinguistic competence) is essential for native-like use of a second language (L2) (Canale & Swain, 1980). Our understanding of native speaker (NS) sociolinguistic variation continues to grow, and areas such as the study of subject expression in Spanish are quite well developed (e.g., Bayley & Pease-Álvarez, 1997; Bentivoglio, 1993; Cameron, 1995; Cameron & Flores-Ferrán, 2004; Comajoan 2006; Davidson, 1996; Flores-Ferrán, 2007; Lapidus & Otheguy, 2005; Otheguy & Zentella, 2007). In contrast, L2 research on variable structures, or those that have the potential to vary, has gained attention more recently and is far outnumbered by studies that compare the language of non-native speakers (NNSs) to prescriptive rules or researchers' judgments. Thus, the primary goal of the current study is to foster cross-disciplinary dialogue between research in sociolinguistics and research in second language acquisition (SLA) while furthering our knowledge of the forms used to express grammatical subjects in Spanish by NSs and NNSs. Although the need for research on variation in L2s is clear, the pursuit of such investigations raises several methodological issues, such as the group to whom learners should be compared, the identification of an individual token of analysis and the methods through which comparisons across groups should be quantified. The exploration of these methodological issues constitutes one of the primary goals of our research program and each will be addressed as they relate to the current study. Finally, in order to advance our understanding of subject expression in Spanish, which has previously focused largely on syntactic and morphological variables, we extend this line of inquiry to examine the level of discourse by analyzing the relationship of the forms of subject expression to two variables: referent cohesiveness and perseveration.

## 2. Subject Expression in Spanish

Spanish allows the subject of finite verbs to be expressed through a variety of forms, including a lexical noun phrase (NP), a personal pronoun, a null pronoun, and several less frequent forms such as a demonstrative pronoun (e.g., *ésta* 'that one'), an interrogative pronoun (e.g., *quien* 'who') or an indefinite pronoun (e.g., *alguien* 'someone'). The contrast between the first three is illustrated in example (1) below.

- (1) a. *Martina patina frecuentemente.* 'Martina skates often.'  
b. *Ella patina frecuentemente.* 'She skates often.'  
c. *∅ patina frecuentemente.* '[She] skates often.'

Prescriptive accounts of which option is preferred generally focus on the contrast between (1b) and (1c), and claim that overt pronouns are required when necessary for contrast, emphasis or disambiguation and are prohibited when such conditions are not met. Additionally, some syntactic structures are said to prohibit overt subject pronouns, such as those expressing atmospheric conditions

as in (2), and in certain syntactic structures, such as the subject of a verb in a subject-headed relative clause, as in (3).

- (2) *¿Llovió bastante ayer.*  
 ‘[It] rained a lot yesterday.’
- (3) *Silvia es la mujer que ¿corre todos los días conmigo.*  
 ‘Silvia is the women who [she] runs every day with me.’

There is little argument that the examples in (2) and (3) are ungrammatical when an overt subject pronoun is present, but there is evidence of NS variation in most syntactic structures. Thus, we turn immediately to the review of sociolinguistic research on this phenomenon in order to demonstrate some of the factors that have been shown to influence this variation.

### 3. Previous Research on Subject Expression by NSs

There is a wealth of research on subject expression in Spanish, focusing nearly exclusively on animate personal subject pronouns. In other words, the contrast between null and overt subject pronouns that refers to animate referents. In this body of work tokens are defined as potential contexts for variation, meaning that those syntactic structures and lexical items that either require or prohibit overt pronouns according to prescriptive accounts are excluded from the analysis. Research has examined subject expression in the Spanish of Colombia, Cuba, the Dominican Republic, Mexico, Puerto Rico, Spain, and Venezuela (Bentivoglio, 1987; Cameron, 1993, 1994, 1995; Miyajima, 2000), as well as in the contact varieties of speakers from each of these countries residing in the United States (Bayley & Pease-Álvarez, 1997; Comajoan, 2006; Flores-Ferrán & Toro 2000, Lapidus & Otheguy, 2005; Otheguy & Zentella, 2007; Silva-Corvalán, 1982). General findings have shown that the choice between overt and null pronominal forms with animate subjects can be conditioned by variables such as the person, number, tense, mood and aspect (TMA) of the verb, the specificity of the referent, the position of the subject, the presence of verbal negation and of object pronouns, as well as by discourse-level factors such as switch reference, perseveration and discourse connectedness. Serrano (1996) argues that an analysis that includes the level of discourse is essential and, because these very factors have not yet been explored for L2 speakers, they will constitute the primary focus of the remainder of this review.

The most commonly examined discourse-level variable in sociolinguistic research on the contrast between null and overt pronominal subject forms is switch reference. This variable distinguishes contexts where a given pronoun (overt or null) refers to the same referent as the subject of the preceding tensed verb form from those where the referent is different from one tensed verb to the next. Research on Spanish speakers of Mexican descent in the United States as well as those from Puerto Rico and Spain has shown that overt subject pronouns are more frequent in contexts of switch reference than in contexts of co-reference (Bayley & Pease-Álvarez, 1997; Cameron 1994, 1995). In other words, when the referent differs from one tensed verb to the next, the subject of the second verb is likely to be expressed overtly. Both Flores-Ferrán (2002) and Bayley and Pease-Álvarez (1997) suggest that the switch reference variable might be further developed to include additional detail about the way in which switch reference influences subject form expression. We follow the model established by Bayley and Pease-Álvarez who, in their study of oral and written narratives produced by Mexican-American school children between the ages of 8 and 12, provide a comparison between the switch reference variable with only the categories switch reference and same reference, and one that further divides the categories of this variable. Specifically, they follow work on Portuguese by Paredes Silva (1993) on written Brazilian Portuguese and code their data for a variable called discourse cohesiveness, which includes five categories: (1) same subject, same verbal TMA, (2) same subject, different verbal TMA, (3) 1 or more intervening clauses, (4) different subject and first mention was another function (e.g., a direct object), and (5) new topic or new narrative section (based on Labovian narrative structure). Thus, there are two categories for same reference (1 and 2) and there are three for switch reference (3, 4 and 5) so that greater detail is provided about the ways in which

referents may contrast. Their findings show that the overt pronouns were used most often in the last category and least often in the first. Bayley and Pease-Álvarez provide a revealing comparison of the discourse cohesiveness variable with the traditional switch reference variable by entering each into separate predictive models (along with several other linguistic variables that were held constant) and show that the more detailed categories in the discourse cohesiveness variable lend greater predictive power to the model. Based on these findings, the current study will employ a multi-category variable to address the relationship between previous mention of the referent and the forms of subject expression.

A second discourse-level variable that has received attention in the studies on NS variation is perseveration (Cameron, 1994; Cameron & Flores-Ferrán, 2004; Flores-Ferrán, 2005). This variable describes the degree of continuity in the form of the previous trigger in order to compare two distinct hypotheses. The first predicts that an overt subject pronoun is likely to be followed by a null subject pronoun (the Functional Hypothesis), whereas the second predicts that an overt subject pronoun is actually more likely to be followed by an overt subject pronoun and a null subject pronoun is most likely to be followed by a null subject pronoun. In fact, support has been found for the second hypothesis and there is no evidence in favor of the Functional Hypothesis (Cameron, 1994; Cameron & Flores-Ferrán, 2004; Flores-Ferrán, 2005). In addition, Cameron (1994) showed that even though frequencies of use of the overt subject pronouns differ from one dialect to another, in this case from San Juan to Madrid, the factor weights for both groups are similar. In other words, the relationship between this discourse-related variable and the forms of subject expression is the same, even when the rates of use of a given form are different. In the current study, we will therefore examine both the frequency of use of forms of subject expression as well as the way the distribution of forms of subject expression changes across categories of the perseveration variable.

#### **4. Previous Research on Variation L2 Spanish**

For several decades there has been steady interest in the ways in which the language of L2 speakers varies, and issues such as the role of free variation in language acquisition, the relationship between variation and a variety of linguistic theories, and the increasing detail of methods of analysis have been widely discussed (Geeslin & Gudmestad, 2010; Ellis, 1999; Preston, 2000; Tarone, 2007). One key distinction in the field is between Type I variation, which is developmental and includes among those forms that vary at least one non-native-like form, and Type II variation, which is the variation between two or more native-like forms (Mougeon & Dewaele, 2004). Just as is the case for NSs we know that linguistic, social, situational and attitudinal factors influence the use of a given form for NNSs as well (e.g., Bayley & Preston, 1996; Young, 1991). Although most early work focused on L2 speakers of either English or French, the development of situationally-appropriate use of several linguistic forms in L2 Spanish has grown in the last decade. To date, research has been published that explores the frequency of use and predictors of that use for the copula contrast, mood distinction and future-time expression (Geeslin, 2003; Geeslin & Gudmestad, 2008; Geeslin & Guijarro-Fuentes, 2006; Gudmestad, forthcoming; Gudmestad & Geeslin, in press). The current study builds on our early work on subject expression with the goal of continuing to expand the body of research on the SLA of potentially-variable structures.

The contrast between null and overt subject pronouns has, in fact, received attention in the investigations of L2 Spanish. Many earlier studies were concerned with the Null Subject Parameter (Jaeggli & Safir, 1989) and whether or not new settings to this parameter can be acquired (e.g., Al-Kasey & Pérez-Leroux 1998; Bini, 1993; Emberson, 1987; Galvan, 1999; Isabelli, 2004; Liceras, 1989; Liceras, Maxwell, Laguardia, Fernández, Fernández, & Díaz, 1997; Phinney, 1987; White, 1985). Recent research has also examined the discourse-pragmatic features that play a role in the L2 acquisition of subject expression (e.g., LaFond, 2002; Lafond, Hayes & Bhatt, 2000; Montrul & Rodríguez Louro, 2006; Rothman, 2007). To our knowledge, the only previous work on variation in the forms of subject expression in L2 Spanish is our own. Thus, we briefly review this research in order to demonstrate the unique contribution of the current study and to contextualize the present analysis of discourse-level features.

In Geeslin and Gudmestad (2010), we analyzed interview data collected from 16 NSs and 16 NNSs to ascertain the distribution of forms used in a variety of potentially-variable structures, including subject expression. There are several methodological choices that separate this work from both the research in sociolinguistics and on L2s. Firstly, our definition of a token was any subject of any tensed verb, rather than only those contexts where both null and overt subject pronouns were possible and the referent was animate. The rationale for this change is that we expect that a priori definitions of obligatory contexts will not apply to learners and may not apply to NSs either. Moreover, we hope to examine the forms of subject expression in all contexts, rather than just a subset. A second distinction is that our dependent variable is not binary (i.e., null vs. overt), but rather it includes lexical NPs, and the additional types of pronouns described in section 2. We broaden the scope of our research in this way in order to include an account of the relationship of all forms of subject expression because at the very core of all SLA research is the premise that interlanguage is a system (Selinker, 1972) and, thus, understanding language development means exploring how each of the forms that fulfill a given function are related to one another. We argue that this is especially appropriate with learners who may also vary between use of lexical NPs and other types of pronouns since the ability to know when pronouns are interpretable is linked to understanding discourse constraints in a given language. Finally, we make inferences about the differences between NSs and NNSs using a comparison group that differs from some previous studies in that our NSs are also bilingual (i.e., know English). Although we recognize that this means that the Spanish spoken by this group may be influenced by knowledge of English, we argue that this is, in fact, the only appropriate starting point for comparison. This is because, if there is convergence on this structure in Spanish/English bilinguals, our NNSs will never arrive at the state of being monolingual Spanish speakers. Instead, the appropriate goal for them is to achieve the same level of competence as an educated bilingual with knowledge of the same languages. Moreover, our NNSs have contact with many different varieties of Spanish, both domestically and through time spent abroad, and this variety of experience is found even within the individual. Thus, it would be quite arbitrary to select a given NS population as the basis for comparison without regard to the sociolinguistic variation in language use known to exist. We certainly expect that future research may compare NNSs to a particular NS group when their language contact experiences are limited to a single variety of Spanish and, thus, our argument is that ours is the best starting point, not the only item on the research agenda. Following these methodological decisions, in Geeslin and Gudmestad (2010) we found that for the dependent variable subject expression there were significant differences in the distribution of forms of subject expression between groups. A more detailed analysis further showed that these differences were not limited to a single form or pair of forms (e.g., only null subjects or only lexical NPs), but rather they extended across the range of categories of the dependent variable. Our subsequent research on this topic has the central goal of exploring more fully these differences.

Using the existing research on NS variation in subject expression as a guide, we began the task of exploring constructs that had explained portions of the variation in NS use. In Geeslin and Gudmestad (2008), we examined the relationship of the forms of subject expression in this same data set to the person and number of the verb and to the specificity of the referent, both of which were significant predictors of NSs use in earlier studies (e.g., Bayley & Pease-Álvarez, 1997). Noting the continual reference to the role of ambiguity in predicting NS use of subject pronouns and the disagreement in the literature as to how such a construct was operationalized, in Gudmestad and Geeslin (2010) we conducted an exploration of the relationship of subject forms to the potential ambiguity of the verb form and the potential ambiguity of the discourse context (i.e., switch reference). In both of those studies we found that by looking at the frequency of use of each form of subject expression within the categories of each of these linguistic variables, important insights could be gleaned about how NSs differed from NNSs. Nevertheless, we recognized that the variable switch reference, with only the categories same reference and switch reference, was likely to exclude relevant details about the role of continuity of the referent. Moreover, aside from that variable, our previous research has neglected discourse-level factors, despite clear evidence of their importance in the research on NSs to date. Thus, the goal of the current study is to improve our treatment of the variable switch reference and extend this body of research to include the additional factor of perseveration in order to glean a more complete understanding of the relationship of these discourse-level factors to the forms of subject

expression and the ways in which these factors can identify differences and similarities between NSs and NNSs.

## 5. The Current Study

In order to meet the aforementioned goals, the current study was guided by the following research question and sub-questions:

1. What is the relationship between discourse-level factors and forms of subject expression?
  - a. What is the relationship of referent cohesiveness to forms of subject expression for NNSs and for NSs?
  - b. What is the relationship of perseveration to forms of subject expression for NNSs and for NSs?

### 5.1. *Participants*

The participants in the current study included two groups, 16 NNSs of Spanish and 16 NSs of Spanish. The NNSs of Spanish all spoke English as a first language, ranged in age from 22 to 35 (mean = 26.4) and had received between 3 and 17 years of formal study (mean = 8.6 years). The group included 10 males and 6 females and scored between 16 and 25 points on a 25-point proficiency test (to be described in section 5.2). The participants in this group had spent between 3 months and 11 years abroad (mean = 18.1), although when the participant who spent 11 years abroad was removed from the calculation the range was between 3 and 27 months (mean = 9.7 months). This participant group was selected because previous research on variation in L2s has shown that only at relatively advanced levels of proficiency are native-like manners of variation acquired and, despite the wide range of proficiency levels found on our own independent measure, these learners were all highly motivated, highly functional (enough to take graduate classes in Spanish and to teach lower division language classes) L2 learners and all of them had daily contact with NSs of Spanish.

Our NS group was also residing in the US at the time of the study but came from a wide range of countries of origin: Argentina, Chile, Colombia, Ecuador, Mexico, Spain, United States (Puerto Rico & the Southwest) and Uruguay. The group was evenly divided between males and females and ranged in age from 24 to 37 (mean = 29.8). On the same independent measure of proficiency completed by the NNSs, this group scored between 22 and 25 points (mean = 23.6). All of the speakers in this group were highly educated and came to the US for academic purposes, although not all were studying Spanish. As mentioned earlier, this group was selected deliberately for their comparability to our own NNSs in terms of social class, language contact (several varieties of Spanish and English), and academic achievement. We acknowledge that the language spoken by these speakers may differ from monolinguals in each of their individual countries of origin as a result of their contact both with English and with other varieties of Spanish, but that this contact is preferable to comparing our NNSs with a NS group from an arbitrarily selected monolingual population that may not reflect a valid target for their own acquisition.

### 5.2. *Elicitation tasks*

Each of the participants (NSs and NNSs) completed the same three activities. First, all participants completed a detailed background questionnaire from which the previous description was taken. Secondly, we administered a 25-item proficiency exam that covered a range of grammatical structures generally taught in language courses. The format of the exam was multiple-choice and each item was contextualized as part of a story (see the Appendix for a sample item). As reported earlier, not all NSs scored 100 percent on this test so it is likely that some of these structures are indeed variable. The final task completed by all participants was a sociolinguistic interview conducted by two researcher assistants, one male and one female, one from Latin America and one from Spain. The list of interview questions was the same for each participant although speakers were allowed to take the content of the interview in the direction of their interests. Interviewers maintained the conversation but tried to limit their contributions to the content to the degree possible. Sample topics covered in the interview include

plans for the future, opinion questions and recounting of recent and past experiences (see the Appendix for sample questions). Each interview lasted approximately 30 minutes and was digitally recorded in a language laboratory.

### 5.3. Coding Scheme and Analysis

The interview data were transcribed and each subject position of a tensed verb was identified. The dependent variable was the form of the subject and included the categories, null pronoun, lexical NP, personal pronoun (e.g., *ella* ‘she’), demonstrative pronoun (e.g., *éste* ‘this one’), indefinite pronoun (e.g., *alguien* ‘someone’), and interrogative pronoun (e.g., *quién* ‘who’). The data were also coded for two independent discourse-level linguistic variables. The first was referent cohesiveness, which refers to the distance and function of the previous mention of the referent and represents an expansion in the number of categories from the variable discourse cohesiveness previously investigated (e.g., Bayley & Pease-Álvarez, 1997). The second was perseveration, which describes the continuity of form across mentions of the referent. The categories for each of these variables are listed in Table 1.

Table 1. Summary of Categories for Referent Cohesiveness and Perseveration

Variable	Categories
Referent Cohesiveness	(1) the subject of <b>the</b> preceding V, same TMA
	(2) the subject of <b>the</b> preceding V, different TMA
	(3) the subject of <b>a</b> previous V, but 1 or more clauses intervene
	(4) another function of <b>the</b> preceding V (DO, IO, possessive)
	(5) another function of <b>a</b> preceding V, 1 or more clauses intervene
	(6) not in the preceding 10 clauses, first mention, not identifiable
	(7) subject of <b>the</b> preceding V, different P/N, same TMA
	(8) subject of <b>the</b> preceding V, different P/N, different TMA
	(9) subject of <b>a</b> previous V, different P/N, but 1 or more clauses intervene
Perseveration	The previous mention of the referent is a lexical NP
	The previous mention of the referent is a pronoun (person or ‘other’)
	The previous mention of the referent is null
	N/A, first mention or not identifiable (not included in the analysis)

Whereas the categories we employed for the perseveration variable are consistent with previous research, a comparison of the categories for the variable discourse cohesiveness with those used by Bayley and Pease-Álvarez (1997) reveals that we have made several modifications. Our new categories (4) and (5) (theirs was a single category) allow us to account for change in function and also whether or not there is intervening material. Category (6) has been modified to generalize outside of narrative structure. The addition of categories (7), (8), and (9) was necessary to account for mutually-constructed discourse such as cases where a referent was mentioned by one speaker and then referred to again by another speaker, using a different person and number (e.g., switching from *tú* ‘you’ to *yo* ‘I’). These last categories are necessary for interview data but not for oral or written narratives such as those analyzed by Bayley and Pease-Álvarez. Thus, the spirit of our coding is very much like that used by Bayley and Pease-Álvarez, even though our data required several modifications to the coding scheme.

The data were coded by one of the researchers according to the classifications in Table 1. In our previous research we have employed a system of a single coder followed by a discussion between both researchers in cases that required a subjective decision for application. In the current study, no such cases existed because the coding scheme is applied objectively (e.g., there is no dispute as to whether one verb has a different TMA than another). This means that the only source of variation from one coder to another is random human error, most of which can be caught through spot checking and through data analysis. Thus, inter-rater reliability with this type of coding scheme is by nature quite high.

The initial analysis performed on this data set included cross-tabulations for speaker groups and the dependent variable, subject expression, showing the distribution of each form (i.e., frequency of use) for each speaker group. We will begin the presentation of the results with that same cross-tabulation. From that starting point we analyzed the data for the discourse-level variables by conducting cross-tabulations for subject expression with the variables referent cohesiveness and perseveration for each speaker group. In other words, four separate cross-tabulations were conducted. In addition, when appropriate a chi-square test was run on each separate cross-tabulation to see whether the forms of subject expression differed significantly across the categories of each independent variable for each speaker group. Thus, the goal of this analysis is to see whether contexts with certain discourse constraints demonstrate a different distribution of forms of subject expression than other contexts. The results from each of these tests will be presented in the sections that follow.

## 6. Results

We begin the report of our results with a presentation of the frequency of use of each of the forms of subject expression for both the NNSs and the NSs.<sup>1</sup> These frequencies are summarized in Table 2. The category ‘other’ pronouns is shown as an aggregate figure and also with details below for each of the types of pronouns in this group.

Table 2. Distribution of Forms of Subject Expression for NNSs and NSs

Subject form	NNSs		NSs	
	#	%	#	%
Null	3792	70.2%	5150	67.6%
Lexical NP	770	14.3%	1042	13.7%
Personal pronoun	589	10.9%	793	10.4%
‘Other’ pronouns (includes 3 categories below)	249	4.6%	637	8.4%
<i>Demonstrative pronoun (éste)</i>	166	3.1%	366	4.8%
<i>Indefinite pronoun (alguien)</i>	61	1.1%	229	3.0%
<i>Interrogative pronoun (quién)</i>	22	0.4%	42	0.6%
Total	5400	100	7622	100

Table 2 shows that both NNSs and NSs produced the same range of forms in the subject position of tensed verbs. Likewise, the most frequent form produced by both groups was the null subject and the least frequent form was the interrogative pronoun. Although the use of personal pronouns is similar for both groups, we note that use of the null subject and the lexical NP were slightly higher for NNSs than for NSs and that the use of demonstrative, indefinite and interrogative pronouns (ie., ‘other’ pronouns) was lower for NNSs than for NSs. Given the relative infrequency of each of the types of pronouns included in the umbrella term ‘other’ pronoun, the remainder of the analysis deals with these forms in aggregate terms. This is especially useful for the purpose of statistical testing where extremely small cells make the tests less reliable. A chi-square test showed that for both the six-category dependent variable and the four-category dependent variable there are significant differences in the distribution of these forms between the NNSs and the NSs (six category:  $X^2=79.10$ ,  $df=5$ ,  $p<0.001$ , Cramer’s  $V=0.08$  and four-category:  $X^2=73.897$ ,  $df=3$ ,  $p<0.001$ , Cramer’s  $V=0.076$ ).<sup>2</sup> After

<sup>1</sup> We base our decision to present group results in our analysis on recent research that has shown that despite individual variation the constraints on group behavior are similar to those that govern individual language use (Bayley & Langman, 2004; Regan, 2004). Although this decision is justified in the literature and makes sense from a practical standpoint, we do not wish to imply that individual variation is of no interest. On the contrary, we hope to examine individual language use in greater detail once we have gained a better understanding of the group trends as a whole.

<sup>2</sup> Later in the paper we will refer to the fact that NNSs use more null subjects than NSs. Although we do not wish to imply that this is the sole source of variation between groups, an additional chi-square test with all of the categories collapsed into “null” and “all other forms” shows that there is a significant difference between groups ( $X^2=189.31$ ,  $df=1$ , Cramer’s  $V=0.126$ ,  $p<0.001$ , no small cells).

the initial distribution analysis, we excluded *vosotros*, *usted* and *ustedes* (all second-person subjects) forms because the participants used them infrequently in the interview. In the remainder of this paper we will explore how referent cohesiveness and perseveration are related to the distribution of the dependent variable for both of these groups.

### 6.1. Results for Referent Cohesiveness

In order to examine the changes in distribution of the dependent variable across the categories of the reference cohesiveness variable (summarized in Table 1), a cross-tabulation was conducted, as was a chi-square test to see whether these distributions varied significantly across categories of the independent variable. The results of these analyses are summarized in Table 3, where category (1) shows the greatest referent cohesiveness (the subject of **the** preceding V, same TMA), category (6) shows the least (not in the preceding 10 clauses, first mention, not identifiable) and categories (7) through (9) refer to co-constructed discourse, with (7) being the most cohesive (subject of **the** preceding V, different P/N, same TMA) and (9) being the least cohesive (subject of **a** previous V, different P/N, but 1 or more clauses intervene). The note at the bottom of Table 3 provides the details for the chi-square test, which demonstrated significant differences in the distribution of forms of subject expression across the categories of the referent cohesiveness variable.

Table 3. Distribution of Subject Expression According to Referent Cohesiveness for NNSs

Referent Cohesiveness	Subject-expression form								Total
	Null		Personal pronoun		Lexical NP		Other pronoun		
	#	%	#	%	#	%	#	%	
1	1065	87.9	111	9.2	17	1.4	18	1.5	1221
2	364	84.8	62	14.5	3	0.7	0	0	429
3	787	72.7	211	19.5	63	5.8	21	1.9	1082
4	265	58	57	12.5	38	8.3	97	21.2	457
5	131	59.8	32	14.6	33	15.1	23	10.5	219
6	871	54.4	30	1.9	610	38.1	90	5.6	1601
7	81	75	27	25	0	0	0	0	108
8	55	77.5	16	22.5	0	0	0	0	71
9	112	75.2	36	24.2	1	0.7	0	0	149

Note.  $\chi^2 = 1700.537$ ,  $df = 24$ , Cramer's  $V = 0.326$ ,  $p < 0.001$ , 1 cell has expected count less than 5

In examining the results for referent cohesiveness, one notes that for NNSs the null subject is the most frequently used form in all contexts. In categories (1) through (3) where there is relative cohesion in the discourse (the referent is the subject of a preceding verb, whether or not there are intervening forms), we can see that personal pronouns are the second most frequent form used (after null subjects) and that there is a larger proportion of use of these overt pronouns when the TMA is different between verbs or when there are intervening clauses. It can also be seen that lexical NPs make up 5.8 percent of use in category (3) but are otherwise infrequent. Finally, in all of these contexts, the use of 'other' pronouns falls below two percent. In categories (4) through (6), where there is lesser cohesiveness (there is a change in function or no previous mention), null subjects continue to be the most frequent form, but they constitute a lower proportion of use (i.e., below 60 percent) than in other categories. Lexical NPs and 'other' pronouns increase in relative frequency of use in these categories and lexical NPs are used more than personal and other pronouns in categories (5) and (6), where their rates of use are 15.1 and 38.1 percent, respectively. Finally, 'other' pronouns are used more than personal pronouns and lexical NPs in category (4), where use is at 21.2 percent. In other words, the role of lexical NPs shows increasing prominence in these categories even though null subjects remain frequent. Finally, in the categories of the referent cohesiveness variable that refer to co-constructed discourse, NNSs show an increased relative frequency of use of personal pronouns (at least 22 percent) and, with the exception of a single token, lexical NPs and 'other' pronouns were not used.

Turning our attention to the distribution of subject forms across the categories of the referent cohesiveness variable for NSs, summarized in Table 4, it is also the case that the null subject is the most frequent form in all categories of this variable. In categories (1) through (3), where there is some degree of referent cohesiveness, personal pronouns are the second most frequent forms and constitute a larger proportion of the forms produced when the TMA is different or when additional clauses intervene. Furthermore, the use of lexical NPs and of ‘other’ pronouns is slightly higher in these categories than it was for the NNS group. In categories (4) through (6), where there is a lesser degree of cohesiveness, the null subject constitutes a lower proportion of use than in other categories of this variable (i.e., below 61 percent), and lexical NPs and ‘other’ pronouns show a higher relative frequency of use. Lexical NPs are more frequent than personal pronouns and ‘other’ pronouns (38.1 percent) in category (6). In categories (4) and (5), ‘other’ pronouns are used more than personal pronouns or lexical NPs, with rates of 21.2 and 27.9 percent, respectively. Thus, one key difference between the NSs and the NNSs is in category (5) where NNSs used more lexical NPs than personal and ‘other’ pronouns and NSs used more ‘other’ pronouns than personal pronouns and lexical NPs. Finally, in categories (7) through (9) where discourse was co-constructed and the referent changed in person and/or number, personal pronouns are seen to increase in proportion of use relative to null subjects, constituting at least 16 percent of use in each category. Lexical NPs and ‘other’ pronouns were not used in category (7) and were used infrequently in the other two categories. This constitutes another difference between the NNSs and the NSs, where the former showed stronger representation of personal pronouns in categories (7) and (8).

Table 4. Distribution of Subject Expression According to Referent Cohesiveness for NSs

Referent Cohesiveness	Subject-expression form								Total
	Null		Personal pronoun		Lexical NP		Other pronoun		
	#	%	#	%	#	%	#	%	
1	1377	84.3	145	8.9	49	3	62	3.8	1633
2	618	83.2	100	13.5	8	1.1	17	2.3	743
3	1124	73	278	18.1	58	3.8	80	5.2	1540
4	361	57.3	52	8.3	41	6.5	176	27.9	630
5	254	60.9	50	12	37	8.9	76	18.2	417
6	991	46.8	79	3.7	828	39.1	221	10.4	2119
7	87	80.6	21	19.4	0	0	0	0	108
8	56	78.9	12	16.9	3	4.2	0	0	71
9	69	72.6	24	25.3	1	1.1	1	1.1	95

Note.  $\chi^2 = 2294.654$ ,  $df = 24$ , Cramer's  $V = 0.322$ ,  $p < 0.001$ , no small cells

To summarize the results for the referent cohesiveness variable, there are several tendencies that are accurate for both groups, despite differing frequencies of use of a given form. Firstly, we see that both groups show highest use of the null subject in contexts of same reference (categories (1) and (2)). Secondly, when the person or number of the referent changes, a personal pronoun is preferred over ‘other’ pronouns and lexical NPs. Finally, when the referent has not been mentioned in the previous 10 clauses, the lexical NP moves in on use of the null subjects and surpasses the use of all other types of pronouns.

## 6.2. Results for Perseveration

As with the referent cohesiveness variable, a cross-tabulation and chi-square tests were run for the perseveration variable and forms of subject expression for both groups. The results of this procedure for the NNSs are summarized in Table 5 and the note below that table reports the results for the chi-square test, demonstrating that the apparent differences in the distribution of forms of subject expression across the categories of this variable are indeed significant.

Table 5. Distribution of Subject Expression Forms According to Perseveration for NNSs.

Previous mention of referent	Subject-expression form								Total
	Null		Personal pronoun		Lexical NP		Other pronoun		
	#	%	#	%	#	%	#	%	
Lexical NP	156	72.2	14	6.5	42	19.4	4	1.9	216
Pronoun	333	60	202	36.4	4	0.7	16	2.9	555
Null	4224	92.3	290	6.3	41	0.9	20	0.4	4575

Note.  $X^2 = 1022.135$ ,  $df = 6$ , Cramer's  $V = 0.309$ ,  $p < 0.001$ , 3 small cells

Table 5 shows that for the NNSs, when the previous mention of the referent is a lexical NP, null subjects constituted the largest portion of the referents (72.2 percent), followed by lexical NPs (19.4 percent), personal pronouns (6.5 percent) and 'other' pronouns (1.9 percent). In the case where the preceding referent was a null subject, the rates of use of null subjects rose to nearly exclusive use (92.3 percent), personal pronouns were used at a similar rate (6.3 percent) and use of lexical NPs and 'other' pronouns was less than one percent. Finally, in contexts where the preceding referent was a pronoun, the use of personal pronouns rose considerably from the other two conditions (36.4 percent) and null subject use fell to 60 percent. Use of 'other' pronouns and of lexical NPs was below 3 percent. These results can be compared to those for the NSs, presented in Table 6, as can the statistical results shown in the note below the table, which are also significant for NSs.

Table 6. Distribution of Subject Expression Forms According to Perseveration for NSs

Previous mention of referent	Subject-expression form								Total
	Null		Personal pronoun		Lexical NP		Other pronoun		
	#	%	#	%	#	%	#	%	
Lexical NP	161	65.7	9	3.7	63	25.7	12	4.9	245
Pronoun	482	60.3	211	26.4	7	0.9	100	12.5	800
Null	84.8	2910	424	12.4	49	1.4	50	1.5	3433

Note.  $X^2 = 903.743$ ,  $df = 6$ , Cramer's  $V = 0.318$ ,  $p < 0.001$ , no small cells

Table 6 demonstrates that for the NS participant group, when the previous mention of the referent is a lexical NP, null subjects constituted the largest portion of the referents (65.7 percent), followed by lexical NPs (25.7%), 'other' pronouns (4.9 percent) and then by personal pronouns (3.7 percent). The order of these last two categories is the opposite for our two participant groups. When the preceding referent is a null subject, null subjects increase to 84.8 percent of all forms produced, as sharp but smaller increase than was found with NNSs. Rates of use for personal pronouns also rise to 12.4 percent and use of lexical NPs and 'other' pronouns falls below 2 percent. Despite differences in magnitude, the overall trend found for this condition is the same for both NSs and NNSs. Finally, in contexts where the preceding referent is a pronoun, rates of use of personal pronouns and 'other' pronouns rise to 26.4 and 12.5 percent, respectively, whereas use of null subjects falls to 60.3 percent. Rates of use of lexical NPs in this context are below one percent. These results show that the general patterns for the NNSs and for the NSs are similar. The two most notable differences are that in contexts where the preceding referent is a null subject, the NNSs show nearly exclusively use of the null subject pronoun whereas NSs show a sharp increase in use of null subjects but the magnitude is smaller than for NNSs. In the context where the preceding referent is a pronoun, the forms used by NNSs largely include null subjects and personal pronouns. In contrast the NSs use 'other' pronouns with preceding referents that are pronouns to a greater extent. The implications of these findings will be discussed in the section that follows.

## 7. Discussion

The specific research goal of the current study was to extend our previous research on subject expression to the level of discourse by examining the relationship of the two most important

(according to previous sociolinguistic research) discourse-level variables to the range and frequency of forms used to express the subject function in Spanish for both NSs and NNSs. Our first research question explored the relationship of a variable called referent cohesiveness to the forms of subject expression. This variable is a more developed version of the switch reference variable found in numerous studies on the contrast between null and overt personal subject pronouns (e.g., Bayley & Pease-Álvarez, 1997; Bentivoglio, 1987; Cameron, 1993, 1994, 1995; Flores-Ferrán, 2004), which has been shown to have greater predictive power of the form used to express the subject in those contexts (Bayley & Pease-Álvarez, 1997). Our results show that increased distance from the original mention of the referent and changes in function of that mention are related to the form chosen to express the subject. Despite the fact that null subjects were the most frequent form used by both groups in all contexts, their rates of use decreased for both groups in favor of overt pronouns and/or lexical NPs as distance increased or functions changed.

Our second research question explored the relationship between the variable perseveration and the forms used in the subject position of tensed verbs. We found that for both NSs and NNSs, our speakers always used null subjects more than any other form, but that the rates of use fluctuated sharply according to the form of the previous referent. We found that null subjects were followed by a higher frequency of null subjects and overt forms were followed by a higher frequency of overt forms. Thus, although no form ever surpassed the use of the null subject, the direction of these changes across categories is consistent with the results for perseveration in other sociolinguistics studies (e.g., Cameron, 1994).

These findings have implications for both SLA research and for sociolinguistics. Firstly, there is considerable research in the field of SLA in the area of ultimate attainment (e.g., Birdsong, 1992; Montrul & Slabakova, 2003; Papp, 2000; White & Genesee, 1996) because one of the fundamental differences between learners and NSs is the rate of acquisition and the end state of that process such that L2 learners tend to need more time and rarely reach the same level of competence as NSs. Research on variation in an L2 has identified several potentially-variable structures that are acquired late (if at all), possibly because variable input is more difficult to incorporate into a learner's grammar than cases where a given function is fulfilled by only one form, or by two forms that follow categorical rules of contrast (Geeslin & Gudmestad, 2010; Gudmestad, forthcoming). In the context of this issue, our research shows that subject expression by our NNSs does indeed show a relationship to the same discourse-level variables as NSs and that their overall rates of use of the forms of subject expression vary in the same direction as NSs across the categories of these variables. Moreover, when subject forms are ranked in order of frequency of use, these rankings are identical for both groups and this is an additional indication that the NS and NNS grammars are similar. Nevertheless, there are important differences revealed in the current study. For example, NNSs use more null subjects and fewer 'other' pronouns than NSs.<sup>3</sup> We stop short of an explanation of this result because we believe there are several potential reasons for this difference. It may be the case that the type of discourse produced by these two groups differs such that speech by NNSs is inherently simpler and therefore allows greater use of null subjects. It may also be the case, however, that our learners produce syntactically comparable language but that NNSs allow a greater level of ambiguity in their speech than NSs. This second explanation is consistent with other variable structures in L2 Spanish where once a second form is recognized as permissible it tends to be integrated into the grammar gradually but reaches a stage of overgeneralization such that rates of use actually exceed those of NSs. For example, Geeslin (2003) shows that the copular verb *estar* 'to be' is the second form to be acquired by learners but reaches frequencies beyond those of NSs before more native-like levels of use are reached. These hypotheses show that the current study has served to refine the questions we might ask about differences between NSs and NNSs on this structure while at the same time contributing to our knowledge of the differences between these two groups. Along these lines, it may also be worth noting that the current

---

<sup>3</sup> We do note that the direct comparison in the frequency of use of any single form shows relatively similar rates of use between groups. Nevertheless, our large sample size allows us to show that these differences are significant (see note 1) and we anticipate that these results will be fortified by similar findings in future studies.

study suggests that in terms of methodological design, a monolingual Spanish-speaking group is not necessary in order to identify similarities and differences between NSs and advanced NNSs.

The implication for sociolinguistic research is methodological. The results of our NS group do not represent a mere replication of existing studies. Instead, we have shown that the known influencing factors of the contrast between null and overt pronouns with animate referents are also relevant to other subject contexts. Although our motivation for examining the full range of forms in a functionally-defined context (i.e, subject of a tensed verb) stems from a desire to see how learner language as a system operates, the way we have defined a token and the inclusion of all of the forms that fulfill this function are also consistent with recent developments in sociolinguistic research (e.g., Schwenter & Torres-Cacoulllos, 2008), which is beginning to show greater reference to the range of forms that fulfill a given function, rather than examining two or more forms that convey the same meaning. Thus, our study is an example of the importance of effective dialogue between fields and we hope to foster this avenue of communication.

## 8. Conclusions and Future Directions

The current study has shown that the in-depth examination of linguistic variables and their relationship to the range and frequency of use of forms used to fulfill a given function can provide essential information about language use by NSs and by NNSs, as well as the similarities and differences between those two groups. We now know that when expressing the subject of a tensed verb NNSs respond to discourse-level variables such as referent cohesiveness and perseveration in similar ways to NSs of Spanish. The analysis of the distribution of forms of subject expression across the categories of another variable allows us to pinpoint subtle differences between groups (e.g., NSs use ‘other’ pronouns when the preceding referent is a pronoun but NNSs do not). Just as Bayley and Pease-Alvarez (1997) demonstrated that the effect of discourse cohesiveness generalizes from written Portuguese to oral and spoken Spanish, we have shown that this variable can be further refined and applied to a broader range of tokens in order to improve our knowledge of NS use as well as that of NNSs. Likewise, the categories of the variable perseveration can be applied to a more comprehensive range of tokens as a means to the same end.

The current study also points to several future projects that would enhance our current knowledgebase in this area. For example, given that NNSs use more null subjects than NSs despite the fact that their overall response to a variety of linguistic variables is quite similar, we need to know more about why this might be. It is possible that these differences are the result of differing types of sentence or discourse structures produced by the two groups or that there is a greater degree of ambiguity and lack of cohesion in non-native discourse. Thus, one future step in research on subject expression is to examine the types of sentence and discourse structures in which these forms appear, in order to explore these hypotheses. Secondly, our research to date has focused on very advanced learners because previous research has shown that the ability to vary one’s speech in native-like ways is acquired rather late in the process of acquisition. Nevertheless, it will be necessary to examine the developmental path of these structures and how a learner comes to respond to linguistic variables in ways that are similar to NSs. It may be the case that some linguistic predictors are universally influential whereas others must be acquired, but only longitudinal or cross-sectional research with learners from a variety of first-language backgrounds can answer this question. Another important avenue for research on this topic lies in the investigation of multi-level predictive models. Such models are capable of accounting for dependent variables that include several categories as well as linguistic and extra-linguistic factors, which may apply at the level of the speaker or at the level of the token. Such a research design would also allow us to address the issue of individual variation, because these models allow for the inclusion of a “participant” factor in addition to the group and linguistic factors. Finally, as mentioned in our description of our participants, it is likely that there are subtle differences between bilinguals and monolinguals, as well as across monolingual varieties of Spanish. Thus, other NS populations should also be investigated using this newer methodology to increase our understanding of these differences.

## Appendix: Sample translated items from the grammar test and the sociolinguistic interview

I think it is very interesting \_\_\_\_\_ (I talk / to talk / talking) about people's eating habits. As for me, \_\_\_\_\_ (I am (*ser*) / I am (*estar*) / I have) a vegetarian.

What plans do you have for the weekend?

Where do you see yourself in five years?

Who is the best language teacher, a native or non-native speaker?

## References

- Al-Kasey, Tamara & Pérez-Leroux, Ana Teresa. (1998). Second language acquisition of Spanish null subjects. In Suzanne Flynn, Gita Martohardijono, & Wayne O'Neil (Eds.), *The generative study of second language acquisition* (pp.161-185). Hillsdale, N.J.: Lawrence Erlbaum.
- Bayley, Robert & Pease-Álvarez, Lucinda. (1997). Null pronoun variation in Mexican-descent children's narrative discourse. *Language Variation and Change*, 9, 349-371.
- Bayley, Robert & Langman, Juliette. (2004). Comparing Variation in the Group and the Individual: Evidence from Second Language Acquisition. *IRAL*, 42, 4, 303-319.
- Bayley, Robert & Preston, Dennis. R. (Eds.) (1996). *Second language acquisition and linguistic variation*. Amsterdam: John Benjamins.
- Bentivoglio, Paola. (1987). *Los sujetos pronominales de primera persona en el habla de Caracas*. Caracas: Universidad Central de Venezuela.
- Bentivoglio, Paola. (1993). Full NPs in spoken Spanish: A discourse profile. In William Ashby, Marianne Mithun, Giorgio Perissinotto & Eduardo Raposo (Eds.), *Linguistic perspectives on Romance languages. Selected papers from the 21<sup>st</sup> LSRL* (pp. 211-224). Amsterdam: John Benjamins.
- Bini, Milena. (1993). La adquisición del italiano: Más allá de las propiedades sintácticas del parámetro *pro-drop*. In Juana.M. Liceras (Ed.), *La lingüística y el análisis de los sistemas no nativos* (pp. 126-193). Ottawa: Dovehouse.
- Birdsong, David. (1992). Ultimate attainment in second language acquisition. *Language*, 68, 706-55.
- Cameron, Richard. (1993). Ambiguous agreement, functional compensation, and non-specific *tú* in the Spanish of San Juan, Puerto Rico and Madrid, Spain. *Language Variation and Change*, 5, 305-334.
- Cameron, Richard. (1994). Switch reference, verb class and priming in a variable syntax. In Katharine Beals (ed.), *Papers from the 30<sup>th</sup> regional meeting of the Chicago Linguistics Society: Volume 2: The parasession on variation in linguistic theory* (pp. 27-45). Chicago Linguistics Society.
- Cameron, Richard. (1995). The scope and limits of switch reference as a constraint on pronominal subject expression. *Hispanic Linguistics* 6/7, 1-28.
- Cameron, Richard. & Flores-Ferrán, Nydia. (2004). Perseveration of subject expression across regional dialects of Spanish. *Spanish in Context*, 1, 41-65.
- Canale, Michael & Swain, Merrill. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1-47
- Comajoan, Llorenç. (2005). Continuity and episodic structure in Spanish subject reference. In J. Clancy Clements & Jiyoung Yoon (eds.), *Functional approaches to Spanish syntax: Lexical semantics, discourse and transitivity* (pp. 53-79). London: Palgrave.
- Davidson, Brad. (1996). Pragmatic weight and Spanish subject pronouns: The pragmatic and discourse uses of *tú* and *yo* in spoken Madrid Spanish. *Journal of Pragmatics*, 26(4), 543-565.
- Ellis, Rod. (1999). Item versus system learning: Explaining free variation. *Applied Linguistics*, 20 (4), 460-480.
- Emberson, Jan E. (1987). The pro drop parameter in the acquisition of Spanish as a second language (Doctoral dissertation, University of Texas at Austin, 1987). *Dissertation Abstracts International*, 47, 12.
- Flores-Ferrán, Nydia. (2002). Subject personal pronouns in Spanish narratives of Puerto Ricans in NYC: A sociolinguistic Perspective. Munchen: Lincom Europa.
- Flores-Ferrán, Nydia. (2004). Spanish subject personal pronoun use in New Cork City Puerto Ricans: Can we rest the case of English contact? *Language Variation and Change*, 16, 49-73.
- Flores-Ferrán, Nydia. (2005). La expresión del pronombre personal sujeto en narrativas orales de puertorriqueños de Nueva Cork. In Luis Ortiz López & Manel Lacorte (Eds.), *Contactos y contextos lingüísticos: el español en los EEUU y en contacto con otras lenguas*, (pp. 119-129). Madrid: Iberoamericana.
- Flores-Ferrán, Nydia. (2007). Los Mexicanos in New Jersey: Pronominal expression and ethnolinguistic aspects. In Jonathan Holmquist, Augusto Lorenzino & Lofti Sayahi (Eds.) *Selected proceedings of the Third*

- Workshop on Spanish Sociolinguistics* (pp. 85-91). Somerville, MA: Cascadilla Proceedings Project.
- Flores-Ferrán, Nydia & Toro, Jeannette. (2000). The persistence of dialect features under conditions of contact and leveling. *Southwest Journal of Linguistics*, 19(2), 31-42.
- Galvan, Javier A. (1999). The distribution of lexical and null subjects in the acquisition of L2 Spanish by adult speakers of English (Doctoral dissertation, University of Southern California, 1999). *Dissertation Abstracts International*, 60, 5.
- Geeslin, Kimberly L. (2003). A comparison of copula choice in advanced and native Spanish. *Language Learning*, 53, 4, 703-764.
- Geeslin, Kimberly L. & Gudmestad, Aarnes. (2010). An exploration of the range and frequency of occurrence of forms in potentially-variable structures in second language Spanish. *Studies in Second Language Acquisition*, 32(3), 433-463.
- Geeslin, Kimberly L. & Gudmestad, Aarnes. (2008). Comparing interview and written elicitation task in native and non-native data: Do speakers do what we think they do? In Joyce Bruhn de Garavito & Elena Valenzuela (Eds.), *Selected proceedings of the 10th Hispanic Linguistics Symposium*, (pp. 64-77). Somerville, MA: Cascadilla Proceedings Project.
- Geeslin, Kimberly L. & Guijarro-Fuentes, Pedro. (2006). The second language acquisition of variable structures in Spanish by Portuguese speakers. *Language Learning*, 56(1), 53-107.
- Gudmestad, Aarnes. (forthcoming). Acquiring a variable structure: An interlanguage analysis of second-language mood use in Spanish. *Language Learning*.
- Gudmestad, Aarnes & Geeslin, Kimberly L. (in press). Assessing the use of multiple forms in variable contexts: The relationship between linguistic factors and future-time reference in Spanish. *Studies in Hispanic and Lusophone Linguistics*, 4(1).
- Gudmestad, Aarnes & Geeslin, Kimberly L. (2010). Exploring the roles of redundancy and ambiguity in variable subject expression: A comparison of native and non-native speakers. In Claudia Borgonovo, Manuel Español-Echevarría & Philippe Prévost (Eds.), *Selected proceedings of the 12th Hispanic Linguistics Symposium* (pp. 270-283). Somerville, MA: Cascadilla Proceedings Project.
- Isabelli, Casilde. (2004). The acquisition of the null subject parameter properties in SLA: Some effects of positive evidence in a naturalistic learning context. *Hispania*, 87(1), 150-162.
- Jaeggli, Oswaldo & Safir, Kenneth. (1989). *The Null Subject Parameter*. Dordrecht: Kluwer Academic Press.
- Labov, William. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- LaFond, Larry. (2002). The pro-drop parameter in second language acquisition revisited: A developmental account (Doctoral dissertation, University of South Carolina, 2002), *Dissertation Abstracts International*, 62, 7.
- Lafond, Larry; Hayes, Rachel; & Bhat, Rakesh. (2000). Constraint demotion and null-subjects in Spanish L2 acquisition. In Joaquim Camps & Caroline Wiltshire (Eds.), *Romance syntax, semantics and L2 acquisition: Selected papers from the 30th Linguistic Symposium on Romance Languages* (pp. 121-135). Amsterdam: John Benjamins.
- Lapidus, Naomi & Otheguy, Ricardo. (2005b). Overt nonspecific *ellos* in the Spanish of New York. *Spanish in Context*, 2, 157-176.
- Licerias, Juana M. (1989). On some properties of the “pro-drop” parameter: Looking for missing subjects in non-native Spanish. In Susan Gass & Jacqueline Schachter (Eds.), *Linguistic Perspectives on Second Language Acquisition* (pp.109-133). Cambridge: CUP.
- Licerias, Juana M.; Maxwell, Denyse; Laguardia, Biana; Fernández, Zara; Fernández, Raquel; & Díaz, Lourdes. (1997). A longitudinal study of Spanish non-native grammars: Beyond parameters. In Pérez-Leroux, Ana & William Glass (Eds.), *Contemporary perspectives on the acquisition of Spanish Volume 1: Developing grammars* (99-132).
- Montrul, Silvina & Rodriguez Louro, Celeste. (2006). Beyond the syntax of the Null Subject Parameter: A look at the discourse-pragmatic distribution of null and overt subjects by L2 learners of Spanish. In Torrens, Vincent and Linda Escobar (Eds.), *The Acquisition of syntax in Romance languages* (pp. 401-418). Amsterdam: John Benjamins
- Miyajima, Atsuko. (2000). Spanish subject pronoun expression and verb semantics. *Sophia Linguistica*, 46-47, 73-88.
- Mougeon, Raymond & Dewaele, Jean-Marc. (2004). Preface to variation in the interlanguage of advanced second language learners. *IRAL*, 42(4), 295-301.
- Otheguy, Ricardo & Zentella, Ana Cecilia. (2007). Apuntes preliminares sobre el contacto lingüístico y dialectal en el uso pronominal del español en Nueva York. In Kim Potowski & Richard Cameron (Eds.), *Spanish in Contact: Policy, Social and Linguistic Inquiries* (pp. 275-298). Amsterdam: John Benjamins.
- Papp, Szilvia. (2000). Stable and developmental optionality in native and non-native Hungarian grammar. *Second Language Research*, 16, 173-200.

- Paredes Silva, Vera Lúcia. (1993). Subject omission and functional compensation: Evidence from written Brazilian Portuguese. *Language Variation and Change*, 5, 35-50.
- Phinney, Marianne. (1987). The pro-drop parameter in second language acquisition. In Thomas Roeper & Edwin Williams (Eds.) *Parameter Setting* (pp. 221-238). Dordrecht: Reidel.
- Preston, Dennis. (2000). Three kinds of sociolinguistics and SLA: A psycholinguistic perspective. In Bonnie Swierzbins, Frank Morris, Michael E. Anderson, Carol A. Klee, & Elaine Tarone (Eds.), *Social and Cognitive Factors in SLA* (pp. 3-30). Somerville, MA: Cascadilla.
- Regan, Vera. (2004). The relationship between the group and the individual and the acquisition of native speaker variation patterns: A preliminary study. *IRAL*, 42, 4, 335-348.
- Rothman, Jason. (2007). Pragmatic solutions for syntactic problems: Understanding some L2 syntactic errors in terms of discourse-pragmatic deficits. In Baauw, Drijckoningen, & Pinto (Eds). *Romance Languages and Linguistic Theory 2005: Selected papers from 'Going Romance'* (pp. 299-320). Amsterdam: John Benjamins.
- Serrano, María José. (1996). El análisis del discurso en variación sintáctica. *Hispanic Linguistics*, 8, 154-177.
- Schwenter, Scott & Torres Cacoullos, Rena. (2008). Defaults and interdeterminacy in temporal grammaticalization: The 'perfect' road to perfective. *Language Variation and Change*, 20, 1-39.
- Silva-Corvalán, Carmen. (1982). Subject expression and placement in Mexican-American Spanish. In Jon Amastae & Lucía Elías-Olivares (Eds.), *Spanish in the United States: Sociolinguistic aspects* (pp. 93-120). New York: Cambridge University Press.
- Tarone, Elaine. (2007). Sociolinguistic approaches to second language acquisition-1997-2007. *Modern Language Journal*, 95(5), 837-848.
- White, Lydia. (1985). The "pro-drop" parameter in adult second language acquisition. *Language Learning*, 35(1), 47-62.
- White, Lydia & Genesee, Fred. (1996). How native is near-native? The issue of age and ultimate attainment in the acquisition of a second language. *Second Language Research*, 12, 233-265.
- Young, Richard. (1991). *Variation in interlanguage morphology*. New York: P. Lang.

# Selected Proceedings of the 2009 Second Language Research Forum: Diverse Contributions to SLA

edited by Luke Plonsky  
and Maren Schierloh

Cascadilla Proceedings Project    Somerville, MA    2011

## Copyright information

Selected Proceedings of the 2009 Second Language Research Forum:  
Diverse Contributions to SLA  
© 2011 Cascadilla Proceedings Project, Somerville, MA. All rights reserved

ISBN 978-1-57473-444-7 library binding

A copyright notice for each paper is located at the bottom of the first page of the paper.  
Reprints for course packs can be authorized by Cascadilla Proceedings Project.

## Ordering information

Orders for the library binding edition are handled by Cascadilla Press.  
To place an order, go to [www.lingref.com](http://www.lingref.com) or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA  
phone: 1-617-776-2370, fax: 1-617-776-2271, [sales@cascadilla.com](mailto:sales@cascadilla.com)

## Web access and citation information

This entire proceedings can also be viewed on the web at [www.lingref.com](http://www.lingref.com). Each paper has a unique document # which can be added to citations to facilitate access. The document # should not replace the full citation.

This paper can be cited as:

Geeslin, Kimberly L. and Aarnes Gudmestad. 2011. Using Sociolinguistic Analyses of Discourse-Level Features to Expand Research on L2 Variation in Forms of Spanish Subject Expression. In *Selected Proceedings of the 2009 Second Language Research Forum*, ed. Luke Plonsky and Maren Schierloh, 16-30. Somerville, MA: Cascadilla Proceedings Project. [www.lingref.com](http://www.lingref.com), document #2522.