

Variable Subject Expression in Second-language Spanish: A Comparison of Native and Non-native Speakers

Kimberly L. Geeslin and Aarnes Gudmestad

Indiana University and Virginia Polytechnic Institute and State University

Research on the second language acquisition (SLA) of sociolinguistic variation in Spanish is in emergent stages. While the process of learning to vary language use like a native speaker (NS) has been examined more thoroughly in French (e.g., Mougeon, Rehner, & Nadasdi, 2004) and English (e.g., Adamson & Regan, 1991), studies on second language (L2) Spanish have been limited to mood distinction, copula choice, and task-based variation (Geeslin, 2003; Geeslin & Gudmestad, in press; Geeslin & Guijarro-Fuentes, 2006; Gudmestad, 2006). Findings in Spanish indicate that there are important differences between even very advanced non-native speakers (NNSs) and NSs of Spanish and that the contrasts between NSs and NNSs can be described by the frequency of selection of a variant and those linguistic and social factors that condition its use. Nevertheless, these conclusions must remain tentative given the limited number of grammatical structures studied. One important goal of the current study is to expand our knowledge of variation in L2 Spanish and to continue to refine a research program for the study of the subtle differences between NSs and NNSs of Spanish.

Subject expression in Spanish, the focus of the current investigation, can take a variety of forms (e.g., a lexical noun phrase, a pronoun, or non-overt subject) and has been the focus of extensive research on NSs (e.g., Bayley & Pease-Álvarez, 1997; Cameron & Flores-Ferrán, 2004; Flores-Ferrán, 2004; Silva-Corvalán, 1982). In contrast, SLA research on subject expression has focused on the formal theoretical implications of L2 data (e.g., Al-Kasey & Pérez-Leroux, 1998; Isabelli, 2004; LaFond, Hayes, & Bhat, 2000; LaFond, 2002) rather than the sociolinguistic factors conditioning this variable use. Thus, a second goal of our study is to bridge this gap with an analysis that incorporates insights from research on the SLA of variable structures and on NS variable use of subject expression by identifying the forms learners use to express subjects and linguistic variables conditioning this use.

Our research is based on two important assumptions, both of which highlight the importance of using insights from SLA and sociolinguistics in our design. Firstly, it has been assumed that among NSs there are certain contexts that allow variation between null and overt subject pronouns and some contexts that do not allow such variation. However, it would be unwise to assume that NNSs only show variation in contexts where NSs do. In fact, there is extensive research on the acquisition of pronouns that shows that L2 learners often repeat a full noun phrase when a NS would have used a pronoun. Thus, pronouns may appear in contexts where they should not and lexical noun phrases may occur where pronouns would in NS speech (Malovrh, 2006; VanPatten & Cadierno, 1993). Thus, our first research assumption is that variability among NNSs may occur outside traditionally-defined contexts of variation (i.e., in contexts where NSs are assumed not to vary their speech), so our analysis must include all contexts where a verbal subject occurs rather than some subset of these contexts. Our second, related research assumption is that because interlanguage should be viewed as a system, it is important to consider all of the forms produced by language learners in relationship to the linguistic variables we examine, rather than only the pronominalized forms. These assumptions will be better understood once the relevant research has been reviewed, but the practical consequence is that our analysis contains a much wider range of tokens (all subjects of finite verbs) and our dependent variable is not binary (null vs. overt), but rather it includes all of the forms learners produce in these contexts. The theoretical foundations upon which this research rests come from both SLA and sociolinguistics. We agree with the many approaches to SLA that state that acquiring a language includes acquiring the ability to vary language use according to both social and linguistic factors (Adamson & Regan, 1991; Canale & Swain, 1980; Preston, 2000). The incorporation of variation into any analysis of SLA

necessarily entails the examination of the factors that govern language use and the ways in which the strength of influence of such factors varies across levels of proficiency and between NNSs and NSs.

In keeping with this approach, our data come from sociolinguistic interviews conducted with NSs and graduate-level NNSs of Spanish who belong to the same speech community. Each token was coded for the person and number of the subject and for specificity. The focus on these two variables is due to the fact that person and number are consistently important for NSs and the effect of this variable is relatively consistent across varieties (Bayley & Pease-Álvarez, 1997; Bentivoglio, 1987; Cameron, 1993, 1994; Flores-Ferrán, 2002, 2004) and because specificity has been key in the exploration of dialectal differences and interacts with person and number in important ways. The relationship between the dependent variable and each independent variable was examined statistically, as was the intersection of the two. We begin with an overview of subject expression in Spanish and research within the fields of sociolinguistics and SLA. We continue with details of the current study and provide both general results for subject expression as well as a careful account of the effects of person and number and specificity on subject expression across groups. We conclude with an assessment of the importance of our findings and directions for future research.

1. Background

1.1 Subject expression in Spanish

Perhaps one of the best known characteristics of Spanish, especially in contrast with English, is that it allows the subject of an inflected verb to be variably explicit or to be understood as a function of context and verbal inflection without being overtly expressed. As shown in example (1) below, when introducing a referent for the first time, as in (a), the subject may be expressed as a full noun phrase. Subsequent mentions of the same referent may be null, as in (b) and (c), or may take the form of an overt subject personal pronoun, as in (d), when contrast is implied or when the referent is ambiguous.

- (1) *(a) Juan es mi amigo de Venezuela. (b) ∅ Tiene tres hijos y (c) ∅ trabaja con Maria. Julia dice que (d) él es muy inteligente.*
 ‘(a) John is my friend from Venezuela. (b) [he] has 3 kids and (c) [he] works with Mary. Julie says that (d) he is very intelligent.’

The null subject, or that which is not overtly expressed, is often expected in cases where a referent is clear to the listener and where no contrast is implied and is obligatory with certain verbs, such as those expressing atmospheric conditions as in (2), and in certain syntactic structures, such as the subject of a verb in an 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 woman who [she] runs every day with me.’

1.2 Sociolinguistic research on subject expression

There is an extensive body of sociolinguistic research on the use of overt subjects (e.g., Bentivoglio, 1987, 1988, 1993; Comajoan, 2006; Davidson, 1996; Miyajima, 2000; Serrano, 1996). Generally speaking, these studies have focused exclusively on subject personal pronouns (SPPs) (i.e., the [+human] subject of a tensed verb) in contexts where variability is possible (Otheguy & Zentella, 2007).¹ Silva-Corvalán (1982) found that overt subject pronouns were most likely to occur with new information, with verb forms that are ambiguous (i.e., those that may correspond to more than one person and number), in contexts where a contrast is expressed, and when new topics of discussion are established. Building on Silva-Corvalán’s work, Bayley and Pease-Álvarez (1997) found that degree of

¹ We note, as do Otheguy and Zentella (2007), that ‘variable contexts’ have yet to be exhaustively detailed or empirically demonstrated. We cast a wider net in terms of contexts analyzed for precisely this reason.

cohesiveness of the narrative, switch reference, the person and number of the verb, the surface ambiguity of the verb, and verb type (based on tense and aspect) were all significant predictors of overt subject pronoun use. The person and number of the verb yielded the strongest effect. The importance of the person and number of the verb was also confirmed by Flores-Ferrán (2004)

Most recent research has confirmed the effectiveness of the variables found to predict the presence of an overt subject pronoun and have added new variables to this list. For example, Cameron and Flores-Ferrán (2004) found that an overt subject is more likely to follow a preceding overt subject and a null subject is more likely to follow a null subject. This process, called perseveration, seems to hold for this and other phenomena (see also Cameron, 1993, 1994; 1995; Flores-Ferrán, 2005; Morales, 1997; Poplack, 1981). One of the most relevant findings of these studies, for the purpose of the current investigation, is the role played by the variable specificity, which refers to whether or not the referent is a clearly identifiable entity or an unspecified group or individual. It was shown that apparent dialectal differences in the frequency of use of overt subject pronouns were actually due to differences in the use of forms such as the non-specific *tú* ‘you (singular)’ and that rates of overt pronoun use in other contexts were more consistent across groups (Cameron, 1993; Flores-Ferrán, 2002, 2004; Lapidus & Otheguy 2005).

The aforementioned sociolinguistic research has several implications. Firstly, it is clear that subject expression is variable across linguistic contexts, across dialectal groups, and most likely even within individuals. Secondly, variation in subject expression is conditioned by linguistic and social variables, including the person and number of the verb and the specificity of the referent. We have opted to focus on these two variables because the former is the strongest factor in several models and because the latter was the key to interpreting important dialectal differences. While we believe that these are but two of the many relevant factors for subject expression, we find them to be an appropriate starting place given their importance across studies. In contrast with the sociolinguistic studies cited here, however, we will focus on the subject position of all finite verbs, rather than predetermined ‘contexts of variation’ because in an L2 context variability may be more widespread.

1.3 Research on the second language acquisition of subject expression

Research on the SLA of null subject expression in Spanish has been approached from a different viewpoint than that in the field of sociolinguistics. During the Principles and Parameters era of Generative Theory (Chomsky, 1981), it was believed that several properties of a language clustered together such that the acquisition of one ought to lead to the acquisition of the others. The Null Subject Parameter included the knowledge that Spanish allowed null subjects, had no expletive subjects (e.g., ‘it’ as in ‘it rains’), allowed subject-verb inversion and required a ‘that’-trace (as in *¿qué dijiste que hizo?* ‘what did you say (that) he did?’) (Jaeggli & Safir, 1989). Thus, numerous studies emerged that examined the degree to which English-speaking learners were able to reset this parameter (from disallowing null subjects to allowing them) and the degree to which the allowance of null subjects also led to grammatical use of the other three properties (e.g., Al-Kasey & Pérez-Leroux 1998; Bini, 1993; Emberson, 1987; Galvan, 1999; Isabelli, 2004; LaFond, 2002; Lafond, Hayes, & Bhat, 2000; Liceras, 1989; Liceras, Maxwell, Laguardia, Fernández, Fernández, & Díaz, 1997; Phinney, 1987; White, 1985). What these studies have in common is that the most important evidence for their research goals is the existence (or lack thereof) of null subject pronouns. In other words, the focus is on whether or not a learner’s grammar allows null subjects rather than the frequency with which such forms are used or the factors that condition their use.

To our knowledge, the only study that addresses subject expression in L2 Spanish as a variable structure is our own large-scale analysis of several variable structures (Gudmestad & Geeslin, 2007, April). In that study, which is based on the same data as the current study, we compared the overall frequency of use of several different grammatical structures by advanced NNSs and NSs of Spanish and found significant differences in the frequency of use of the various forms of subject expression between groups, but we provided no additional details about the linguistic factors that influence the use of those forms. Thus, we begin the presentation of the results below, with a summary of what was found in our earlier study and subsequently build on those results by providing greater detail of subject expression in particular.

Finally, the current study was also informed by two important insights from the wider body of L2 research. The first, based on studies of the SLA of variation in other languages or of other forms, is

that an important aspect of native-like language use is the ability to vary one's speech according to both social and linguistic factors (e.g., Adamson & Regan, 1991; Mougeon, Rehner, & Nadasdi, 2004). In fact, it appears that such variable contexts are one of the areas that demonstrate the most complex differences between NSs and NNSs, even when those NNSs have achieved a very high level of proficiency in other areas of the grammar (Geeslin, 2003; Geeslin & Gudmestad, in press; Geeslin & Guijarro-Fuentes, 2006; Gudmestad, 2006). Under this approach, understanding these differences is the key to understanding language development. The second insight, which comes from the very foundation of contemporary thought within the field of SLA (Selinker, 1972), is that interlanguage is a system. This means, among other things, that an analysis of a single component of the learner grammar would miss important connections between other related elements. In sum, one must look at what learners do, not what they do wrong in certain pre-designated contexts. In light of our current goals, this means that unlike in sociolinguistic research, where the focus of study is the contrast between null and overt pronouns, an L2 researcher must look at all of the forms produced in subject position in order to fully describe the learner's L2 system. Thus, while in sociolinguistics it may be possible to conduct an analysis based on a binary dependent variable (null vs. overt), this is only possible in an L2 analysis if these are the only two forms that appear in the learner data, a finding we have already shown to be false for subject expression (Gudmestad & Geeslin, 2007, April).

2. The current study

Given the goals of the current study and the need for extensive expansion of the database on L2 subject expression in Spanish, the current study was guided by the following research questions:

1. How can the subject expression of English-speaking learners of Spanish be described?
 - a. What forms do L2 learners produce in the subject position of finite verbs in Spanish?
 - b. What is the relationship of the person and number of the verb to the forms produced by these learners?
 - c. What is the relationship of the specificity of the referent to the forms produced by these learners?
2. How do these characteristics compare to the forms produced by native speakers in the same interview task?

2.1 Participants

The first group of participants included 16 graduate-level English-speaking learners of Spanish, all of whom were language instructors, capable of teaching and socializing effectively in Spanish. There were 10 males and 6 females, ranging in age from 22 to 35 (mean = 26.4). The speakers had spent between 3 months and 11 years in a Spanish-speaking country (mean = 18.1 months)² and had studied Spanish between 3 and 17 years (mean = 8.6 years). On a 25-item discrete-point grammar test, described in greater detail below, the scores for this group ranged from 16-25 points (mean = 20.8).

The second participant group included 16 NSs of Spanish belonging to the same speech community. Their countries of origin included Argentina, Chile, Colombia, Ecuador, Mexico, Spain, United States (Puerto Rico and the Southwest), and Uruguay and participants were evenly divided between genders. Participants in this group ranged in age from 24 to 37 (mean = 29.8) and scored between 22 and 25 points on the same grammar test as given to the NNSs (mean = 23.6). By including our own NS group we are able to account for the differences across groups found in previous sociolinguistic research and we can address speech norms specific to this speech community. Although the speakers hail from several countries of origin, this diversity reflects the range of input to which our NNS group was exposed, both through study abroad and in the United States. We also note that our NSs completed the same tasks as the NNSs and, thus, all speakers had the same opportunities for production, even though this does not mean that each speaker produced the same tokens.

² One speaker had been abroad for 11 years and when this participant is removed from the calculation the group range was 3 to 27 months abroad and the mean was 9.7 months.

2.2 Data elicitation

Each participant completed three tasks. The first was a background questionnaire from which the previous descriptions of the participants were taken. The second task was a discrete-point multiple-choice grammar test, examining knowledge of a range of characteristics of Spanish grammar. One notes that the NS group did not score perfectly and that the range of scores for the NNSs is more different from the NSs than the mean score for all participants in those groups. In other words, some but not all of our NNSs performed like NSs on this measure. Sample items are provided in Appendix A. The final task was a semi-structured digitally-recorded sociolinguistic interview conducted by two NSs of Spanish (one male and one female) from different countries of origin. Each 30-minute interview was guided by the same list of topics, including questions regarding their plans for the future, their opinions about certain issues, and recent and past experiences. Sample questions are available in Appendix B.

2.3 Coding and analysis

Because it is not clear where variation will occur with L2 learners, the current study examined the subject of every finite verb produced in the interview task. Thus, our dependent variable was the form used to express the subject and included lexical noun phrases, overt subject pronouns, null subject pronouns, demonstrative pronouns (e.g., *éste* ‘this one’), interrogative pronouns (e.g., *quién* ‘who’), and indefinite pronouns (e.g., *alguien* ‘someone’). Although we recognize that there are many independent linguistic variables that are likely to contribute to the subtle differences between NSs and NNSs, we focused on two: the grammatical person and number of the finite verb and the specificity of the referent. The former includes the following categories: first person singular, first person plural, second person singular, second person plural, third person singular, and third person plural. Within these last two categories a further distinction was made between formal address referents (*usted* or *ustedes*) and all other third person forms. The specificity variable determines whether the human referents in the sample can be identified by name or not and includes four categories: specific, non-specific, group, and not applicable, which applies to all non-human referents. The first category refers to humans that can be named individually, whereas the second, non-specific, refers to humans for which no specific referent is readily apparent (e.g., *alguien, uno* ‘one’, etc.). The category ‘group’ is an intermediate category that accounts for referents for which a group is named but in which the members of the group itself are not specified to the degree that they can be individually named. Although these categories are slightly different from some of those used in sociolinguistic research, in order to have more clearly operationalizable definitions across tokens, we capture the important distinction between a non-specific referent and a specific referent and this is precisely where the dialectal differences in previous studies were found. All coding was conducted by one of the researchers because the variables and their categories were defined objectively, but any unclear cases were discussed until agreement was reached.

We began our analysis by using a cross-tabulation to identify the distribution of forms produced by the NS and NNS groups. Then, a chi-square test determined whether the apparent differences between the groups were significant. Next, we assessed the relationship between the two independent variables included in the current study and the forms used to express subjects by each group. We conducted cross-tabulations and chi-square analyses for each variable for each participant group. Finally, we explored the intersection of specificity with person and number by cross-tabulating for specificity within each category of the person and number variable. This final analysis allows us to determine whether the effect of specificity is different depending on the person and number of the referent.

3. Results

We begin here with a presentation of the overall distribution of the forms produced in the subject position of all finite verbs by each group, as was reported in Gudmestad and Geeslin (2007, April). It will be recalled that the categories of the dependent variable include all forms produced. Both the number of each type of form produced by each group and the percentage of the total tokens that this

represents are presented in Table 1. In the case where a context was found to be categorically null for both groups of speakers, the data were excluded from the analysis. These contexts include clauses with *hay* 'there is/there are', atmospheric expressions, subject-headed relative clauses and the impersonal/passive *se*.

Table 1. Distribution of subject expression forms

Forms	NNSs		NSs	
	#	%	#	%
Null subject	3792	70.2	5150	67.6
Lexical NP	770	14.3	1042	13.7
Subject pronoun	589	10.9	793	10.4
Demonstrative pronoun	166	3.1	366	4.8
Indefinite pronoun	61	1.1	229	3.0
Interrogative pronoun	22	0.4	42	0.6
Total	5400	100	7622	100

Table 1 shows that the most common forms produced in the subject position of finite verbs were null subjects, subject pronouns, and lexical noun phrases. Additionally, several instances of demonstrative pronouns, indefinite pronouns, and interrogative pronouns were produced. The proportion of tokens in each category is similar for NNSs and NSs, but a chi-square test revealed that the two groups were significantly different ($X^2=79.10$, $df=5$, $p<0.001$, Cramer's $V=0.08$). Thus, the remainder of this analysis will examine some of the subtle differences between groups.

Although several independent linguistic variables are likely to contribute to subject form use for both groups, our analysis focuses on two: the person and number of the verb and the specificity of the referent. Table 2 reveals the distribution of the number of forms produced in the subject position of finite verbs across the categories of the person and number variable for the NSs and the NNSs.

Table 2. Distribution of subject expression across person/number

Factors	NNSs		NSs	
	#	%	#	%
1 st , singular	2321	43.0	3022	39.6
1 st , plural	295	5.5	289	3.8
2 nd , singular (<i>tú</i>)	142	2.6	473	6.2
2 nd , plural (<i>vosotros</i>)	1	0.0	12	0.2
3 rd , singular	2230	41.3	3254	42.7
3 rd , plural	399	7.4	538	7.1
<i>Usted</i>	3	0.1	17	0.2
<i>Ustedes</i>	9	0.2	17	0.2
Total	5400	100	7622	100

The most frequently used forms were the first person singular and the third person singular, and this was true for both NSs and NNSs. Chi-square tests revealed that there is a significant difference in the distribution of these forms across categories of the person and number variable for both speaker groups (NNSs: $X^2=1538.33$, $df=35$, $p<0.001$, Cramer's $V=0.24$, 24 small cells and NSs: $X^2=2409.93$, $df=35$, $p<0.001$, Cramer's $V=0.25$, 18 small cells). Thus, what we know so far is that while NSs and NNSs differ significantly in their frequency of use of forms of subject expression, the person and number variable is important in understanding form use for both groups. In order to examine the effect of this variable more closely, and to make a qualitative comparison of NS and NNS use, each category of the person and number variable will be described in detail. Because so few tokens were produced in the second person plural form and the *usted* and *ustedes* forms, these forms will not be explored further.

3.1 Results for forms used within each category of the person and number variable

We begin with the first person singular referents, which are described in Table 3. Throughout the qualitative discussion, we identify those differences that are smaller than two percentage points as similar (and henceforth use the terms such as ‘similar’) and those that are larger than two percentage points as differing to some degree (and henceforth use terms such as ‘more’ and ‘fewer’ to represent these differences). The use of two percentage points, thus, merely defines our use of terms such as ‘similar’ and ‘different’, it does not affect the quantitative analysis, which included all data with no such distinctions made.

Table 3. Distribution of subject expression within ‘1st person singular’ referents

Forms	NNSs		NSs	
	#	%	#	%
Null	1840	79.3	2459	81.4
Subject pronoun	481	20.7	563	18.6
Total	2321	100	3022	100

Table 3 shows that only null subjects and subject pronouns (i.e., *yo* ‘I’) were produced in this context and that the distribution of these forms within this category of the person and number variable is fairly similar for both groups; NNSs used slightly more null subjects and slightly fewer subject pronouns (2.1% for each form). Table 4 shows the distribution of forms produced for first person plural referents.

Table 4. Distribution of subject expression within ‘1st person plural’ referents

Forms	NNSs		NSs	
	#	%	#	%
Null	275	93.2	257	88.9
Lexical NP	16	5.4	16	5.5
Subject pronoun	4	1.4	16	5.5
Total	295	100	289	100

In this category of the person and number variable, null subjects, subject pronouns and lexical noun phrases were produced by both groups. One notes that the NNSs used more null subjects and fewer subject pronouns (i.e., *nosotros* ‘we’) than the NSs. If the process of acquisition were merely one of allowing null subjects, one would expect the NNSs to have fewer null subjects, rather than more of them. In the case of our advanced NNSs, the question is more accurately described as learning the appropriate discourse constraints that trigger the use of null subjects, rather than allowing them in general. The distribution of forms produced in second person singular contexts is described in Table 5.

Table 5. Distribution of subject expression within ‘2nd person singular’ referents

Forms	NNSs		NSs	
	#	%	#	%
Null	128	90.1	386	81.6
Subject pronoun	14	9.9	87	18.4
Total	142	100	473	100

In second person singular contexts, like those for first person singular, only null subjects and subject pronouns (i.e., *tú* ‘you’) were produced by both groups. Like with the first person plural subjects, however, the NNSs used more null forms than the NSs. Table 6 shows the distribution of forms within the category of third person singular referents.

Table 6. Distribution of subject expression within ‘3rd person singular’ referents

Forms	NNSs		NSs	
	#	%	#	%
Null	1329	59.6	1694	52.1
Lexical NP	620	27.8	866	26.6
Subject pronoun	56	2.5	96	3.0
Demon. pronoun	145	6.5	332	10.2
Indef. pronoun	61	2.7	227	7.0
Interrog. pronoun	19	0.9	39	1.2
Total	2230	100	3254	100

In third person singular contexts, one notes a much wider range of forms produced, which is not surprising given the more limited options in the other contexts.³ In comparing the two groups, it can be seen that the NNSs used more null subjects and fewer demonstrative pronouns and indefinite pronouns than the NSs but that the use of lexical noun phrases and subject pronouns was relatively similar. Thus, the pattern of greater use of null forms is seen here, but a new difference emerges in that the NSs were using a wider range of forms to a greater degree, rather than simply more subject pronouns. The final category to be examined is the third person plural contexts, which are shown in Table 7.

Table 7. Distribution of subject expression within ‘3rd person plural’ referents

Forms	NNSs		NSs	
	#	%	#	%
Null	210	52.6	319	59.3
Lexical NP	134	33.6	159	29.6
Subject pronoun	31	7.8	21	3.9
Demon. pronoun	21	5.3	34	6.3
Indef. pronoun	0	0	2	0.4
Interrog. pronoun	3	0.8	3	0.6
Total	399	100	538	100

Table 7 shows that both groups used the same range of subject expression forms with third person plural contexts as they did with third person singular contexts (with the exception of the NNSs who did not use any indefinite pronouns). Moreover, the third person plural category is the only person where NNSs had appreciably fewer null subject pronouns than the NSs (the first person singular forms move in that direction but the difference is just 2.1%). It can also be seen that NNSs used more lexical noun phrases and subject pronouns than the NSs. Thus, this distribution is unlike that seen in the other categories of this variable.

In summary, we have shown that the distribution of forms was closest for the two participant groups for the first person forms, especially for first person singular. We also note that the NNSs used more null subject pronouns than the NSs in all forms except for first person singular (where the difference is 2.1%) and in the third person plural forms. The third person plural forms also showed the greatest difference between groups in the use of lexical noun phrases, where use among NNSs was higher. What the analysis has shown thus far is that the use of subject forms can vary across the person and number of the verb form for both NSs and NNSs and that the inclusion of all categories of the dependent variable (i.e., all forms produced, not just null and overt subject pronouns) is especially important in the third person forms. In other words, variability within and between groups is not likely to be limited only to null and overt subject pronoun forms.

3.2 Results for forms used across categories of the specificity variable

We begin our presentation of the results for the specificity variable in Table 8 by summarizing the number of forms produced by NNSs and NSs across the categories of the specificity variable.

³ Target-like agreement between a demonstrative, indefinite, or interrogative pronoun and a verb entails using a third person form. Both groups produced this native-like agreement, so first and second person forms were never used.

Table 8. Distribution of subject expression across specificity

Factors	NNSs		NSs	
	#	%	#	%
Specific	2932	54.3	3858	50.6
Group	195	3.6	144	1.9
Nonspecific	438	8.1	1085	14.2
Not applicable	1835	34.0	2535	33.3
Total	5400	100	7622	100

For NNSs and NSs approximately one-third of the referents were non-human and were excluded from the analysis and just over one half of the referents were specific, although the NNSs produced slightly more specific referents than the NSs. The NNSs produced fewer non-specific referents and slightly more group referents than the NSs. A chi-square test for each speaker group showed that there is a significant relationship between specificity and subject expression forms (NNSs: $X^2=1410.14$, $df=15$, $p<0.001$, Cramer's $V=0.30$, 4 small cells and NSs: $X^2=2763.55$, $df=15$, $p<0.001$, Cramer's $V=0.35$, 2 small cells). Thus, like the person and number variable, this variable has an important relationship with subject expression for both groups. The distribution of forms across these categories and a qualitative comparison of the NNSs and the NSs will follow, beginning in Table 9 with contexts where the referent was specific. This comparison focuses on the three most frequently used forms: null subjects, lexical noun phrases, and subject pronouns.

Table 9. Distribution of subject expression within 'specific' referents

Factors	NNSs		NSs	
	#	%	#	%
Null	2236	76.2	2996	77.7
Lexical NP	143	4.9	145	3.8
Subject pronoun	551	18.8	703	18.2
Total	2930	99.9	3844	99.7

Not only did specific referents make up the largest proportion of tokens for the NSs and the NNSs (about half), but the distribution of the forms produced was similar for NNSs and NSs. This result is especially striking given the differences shown for the other contexts. Table 10 shows the distribution of forms produced within the context of group referents.

Table 10. Distribution of subject expression within 'group' referents

Factors	NNSs		NSs	
	#	%	#	%
Null	159	81.5	107	74.3
Lexical NP	27	13.8	23	16.0
Subject pronoun	9	4.6	8	5.6
Total	195	99.9	138	95.9

In contexts where the referent belonged to the group category, the NNSs used more null subject pronouns and fewer lexical noun phrases than the NSs. The use of subject pronouns was similar (1% difference) for both groups. The final category of the specificity variable was the non-specific category and the distribution of forms in this category for NNSs and NSs is shown in Table 11. One notes that the NSs produced many more non-specific referents than the NNSs, and this may be a mere reflection of the degree to which NSs spoke more in general (NSs produced 7622 finite verb contexts whereas NNSs produced 5400).

Table 11. Distribution of subject expression within ‘non-specific’ referents

Factors	NNSs		NSs	
	#	%	#	%
Null	281	64.4	687	63.3
Lexical NP	83	19.0	130	12.0
Subject pronoun	22	5.0	76	7.0
Total	386	88.4	893	82.3

When the referent was non-specific, the NNSs used more lexical noun phrases than the NSs and the NSs used slightly more subject pronouns. As with the other categories of this variable, the distribution of subject expression forms is different for each group.

To summarize, as was the case for the person and number variable, the distribution of the forms of subject expression was significantly different across categories of the specificity variable for both NNSs and NSs. Likewise, there are notable qualitative differences between the NNSs and the NSs in the use of particular forms within some of the categories of this variable. For specific referents, the distribution is relatively similar for NNSs and NSs. However, when the referent was a group, the use of subject pronouns was similar for all participants, but the NNSs used more null subjects and fewer lexical noun phrases than the NSs. When the referent was non-specific, the use of null subject pronouns was similar but the NNSs used more lexical noun phrases and fewer subject pronouns. Taken together, these results show that there is an individual pattern of use within each of these three categories and differences between NNSs and NSs emerge when these patterns are compared.

3.3 Person and number with specificity

It was mentioned earlier that not only do person and number on the one hand and specificity on the other hand contribute to form use in subject position, but rather these two variables might also interact in such a way that the cross-section of the categories of the two variables provides additional details about subject expression for both participant groups. The remainder of this results section is dedicated to precisely this type of analysis. Aside from indirect quotes, the first person does not include non-specific referents and is not included. Thus, we focus this portion of the analysis on the cross-section of the specificity variable with second person singular and with third person singular and plural forms. We begin with the distribution of forms in contexts where the second person singular referent was specific, shown in Table 12.

Table 12. Distribution of subject expression within ‘specific *tú*’ referents

Forms	NSs		NNSs	
	#	%	#	%
Null	50	94.3	87	75.7
Subject pronoun	3	5.7	28	24.3
Total	53	100	115	100

In the second person singular contexts, the use of null subjects was higher for NNSs than for NSs and the use of subject pronouns was lower. This result further supports the idea that the task for learners at this high proficiency level is a matter of acquiring the appropriate constraints on subject expression. Table 13 describes the contexts where the second person singular referent was non-specific.

Table 13. Distribution of subject expression within ‘non-specific *tú*’ referents

Forms	NNSs		NSs	
	#	%	#	%
Null	78	87.6	299	83.5
Subject pronoun	11	12.4	59	16.5
Total	89	100	358	100

Table 13 shows that NNSs used slightly more null and slightly fewer subject pronouns than the NSs. The results for this category of the specificity variable are like those for specific second person singular referents. No group second person singular tokens were produced, as these categories are mutually exclusive. The analysis continues with Table 14, which shows the distribution of forms in the contexts where third person singular referents were specific.

Table 14. Distribution of subject expression within ‘specific 3rd person singular’ referents

Forms	NNSs		NSs	
	#	%	#	%
Null	131	48.0	203	49.9
Lexical NP	98	35.9	109	26.8
Subject pronoun	43	15.8	84	20.6
Demon. pronoun	0	0	11	2.7
Interrog. pronoun	1	0.4	0	0
Total	273	100	407	100

Table 14 shows that the NNSs used more lexical noun phrases and fewer demonstrative pronouns than the NSs. In contrast with the results for the second person singular, NNSs used slightly fewer null subjects and fewer subject pronouns for the third person singular specific referents. The distribution for third person singular group referents is shown in Table 15.

Table 15. Distribution of subject expression within ‘group 3rd person singular’ referents

Forms	NNSs		NSs	
	#	%	#	%
Null	2	33.3	8	47.1
Lexical NP	4	66.7	9	52.9
Total	6	100	17	100

As with specific third person singular contexts, the NNSs used fewer null subjects and more lexical noun phrases than the NSs in these contexts. The results for the use of null subjects in specific and group third person singular contexts differ from the others in that the overall trend in this data set is that NNSs used null subjects with greater frequency and lexical noun phrases with lower frequency than the NSs. We note, however, that this category represents very few tokens; 6 for the NNSs and 17 for the NSs. Finally, third person singular contexts for non-specific referents are provided in Table 16.

Table 16. Distribution of subject expression within ‘non-specific person singular’ referents

Forms	NNSs		NSs	
	#	%	#	%
Null	115	52.0	205	42.4
Lexical NP	55	24.9	87	18.0
Subject pronoun	7	3.2	6	1.2
Demon. pronoun	0	0	5	1.0
Indef. pronoun	40	18.1	175	36.2
Interrog. pronoun	4	1.8	6	1.2
Total	221	100	484	100

In contrast with the group and specific contexts, Table 16 shows that NNSs used more null subjects than NSs for third person singular non-specific referents. Unlike the case of specific referents, third person singular non-specific referents were referred to with a greater number of subject pronouns and a smaller number of indefinite pronouns by the NNSs than by the NSs. Finally, as with group referents, the non-specific third person singular referents are referred to with more lexical noun phrases by the NNSs than by the NSs. Once again, for a single category of the person and number variable, the variable specificity demonstrates different patterns of use for each type of referent. Moving to third person plural referents, the results for specific contexts are summarized in Table 17.

Table 17. Distribution of subject expression within ‘specific 3rd person plural’ referents

Forms	NNSs		NSs	
	#	%	#	%
Null	60	53.6	41	51.3
Lexical NP	34	30.4	26	32.5
Subject pronoun	17	15.2	10	12.5
Demon. pronoun	1	0.9	3	3.8
Total	112	100	80	100

Table 17 shows a similar distribution of null subject pronouns and lexical noun phrases in specific third person plural contexts for both NNSs and NSs. We note that NNSs used slightly more subject pronouns and fewer demonstrative pronouns and this result is like that found for non-specific referents in the third person singular context. Although the difference between the two speaker groups for each category of subject expression is greater than two percent, NNS and NS use is more similar within specific third person plural referents than any of the second and third person singular referents. The results for third person plural group contexts are shown in Table 18.

Table 18. Distribution of subject expression for ‘group 3rd person plural’ referents

Forms	NNSs		NSs	
	#	%	#	%
Null	43	60.6	72	76.6
Lexical NP	19	26.8	11	11.7
Subject pronoun	9	12.7	5	5.3
Demon. Pronoun	0	0	6	6.4
Total	71	100	94	100

In third person plural contexts with a group referent, NNSs used fewer null subjects and fewer demonstrative pronouns than the NSs. Like the third person singular specific contexts, the NNSs used more lexical noun phrases and more subject pronouns than the NSs. The last context analyzed in the current study was third person plural contexts with non-specific referents, shown in Table 19.

Table 19. Distribution of subject expression forms within ‘non-specific 3rd person plural’ referents

Forms	NNSs		NSs	
	#	%	#	%
Null	72	64.9	124	70.9
Lexical NP	27	24.3	39	22.3
Subject pronoun	4	3.6	6	3.4
Demon. pronoun	8	7.2	4	2.3
Indef. pronoun	0	0	2	1.1
Total	111	100	175	100

Like group referents in the third person plural contexts, NNSs used fewer null subjects and more lexical noun phrases in third person plural non-specific contexts than the NSs. Unlike those same contexts, however, the non-specific contexts showed greater use of demonstrative pronouns by the NNSs. Subject pronoun use in third person plural non-specific contexts was similar for both groups.

We briefly summarize these results before continuing. In second person singular contexts, NNSs used more null subjects and fewer subject pronouns than the NSs in non-specific contexts and this trend is the same but stronger for specific contexts. In third person singular contexts, the NNSs used fewer null subjects and more lexical noun phrases in specific and group contexts. Moreover, the NSs used the wider range of forms to a greater degree than the NNSs. Specific referents also appeared as subject pronouns and demonstrative pronouns, although in smaller numbers for the NNSs than the NSs, and there were a few instances of other forms as well. The non-specific contexts were different: there was greater use of null subjects and lower use of indefinite pronouns for NNSs. Despite these differences, the NNSs used more lexical noun phrases than the NSs, just as in the specific and the group third person singular contexts. Finally, in third person plural contexts, specific referents showed

a more similar distribution for NNSs and NSs than the other two categories. Group and non-specific contexts showed fewer null subjects and more lexical noun phrases for the NNSs than for the NSs. Additionally, group third person plural contexts showed more subject pronouns and fewer demonstratives for the NNSs and the non-specific contexts showed more demonstratives for the NNSs than for the NSs. The upshot of these findings is that not only do the person and number and the specificity variables provide important details about subject form use for NNSs and for NSs, but the intersection of these two variables also provides additional knowledge regarding language use.

4. Discussion

The current study was guided by two research questions. The first sought to describe subject expression by learners of Spanish, including the range of forms produced and the relationship between the subject forms used and the person and number and the specificity of the referent. The second question sought to compare those data with the same data for NSs in order to determine the degree of similarity between the two. Beginning with the range of forms produced, we showed that in the subject position of finite verbs, six different forms were produced: lexical noun phrases, null subject pronouns, overt subject pronouns and indefinite, interrogative, and demonstrative pronouns. Although the range of forms produced was the same for NNSs and for NSs, the frequency with which each form was produced was significantly different. Consequently, the person and number and the specificity variables were examined not only to see whether or not they had a relationship to use of the subject form but also as source of comparison in the use of subject forms between NNSs and NSs.

Looking at the person and number of the referent, we found that the frequency of occurrence of each of the forms of subject expression was significantly different across categories for both NNSs and for NSs. Likewise, important differences emerged between the forms used by NNSs and those used by NSs. Although the distribution of forms was similar for the first person forms, especially the first person singular, there was a general tendency on the part of the NNSs to use more null subject forms in all except for first person singular and third person plural contexts, where NSs used more null forms than the NNSs. The tendency to use more null forms on the part of NNSs demonstrates that the acquisition problem for English-speaking learners of Spanish is not merely one of allowing or disallowing null subjects (in which case a lower frequency of null subjects would be expected among learners), but rather acquisition involves learning to respond to the additional factors in the context that determine when use of null subjects is appropriate. That said, we want to be clear that we do not wish to imply that the use of these null subjects on the part of our learners is inappropriate. In fact, no evaluative assessment has been made. It is quite possible that NNSs simply produced different types of discourse in which more null subjects are possible. For example, it may be the case that NNSs manage fewer referents at one time and chances for misunderstanding are less frequent.

In addition to this finding, there are other important results from the analysis of the person and number variable that warrant discussion. Firstly, the person and number variable is relevant for both participant groups. This means that like previous sociolinguistic research, our two groups respond to this linguistic factor and, thus, our learners have a grammar that corresponds to other NS grammars in this way (even though use within those categories still differs). In other words, our learners recognize this as a valuable factor in subject expression, even if the distribution of the forms they use remains different across categories. Secondly, we have shown that the full range of forms produced must be examined in order to fully understand subject expression; this is especially true in the third person contexts where more forms were produced. While previous research has focused on the opposition between null and overt subject pronouns in contexts where 'variation was possible', we have shown that a wider range of forms provides a more detailed picture of language use among these populations.

The analysis based on the specificity of the referent helped to underscore several of these results. Firstly, the specificity variable is significantly related to subject expression for both groups and there are qualitative differences between the NNSs and the NSs. Unlike the person and number variable, however, fewer general tendencies were seen across categories. For example, NNSs used approximately the same number of null subject pronouns in contexts where the referent was specific and where the referent was non-specific, but a greater number for group referents. The rates of use of overt subject pronouns were similar for NNSs and NSs in specific and group contexts but NNSs used fewer subject pronouns than NSs in non-specific contexts. What emerges with the specificity variable

are independent patterns of use for each category of the variable. In relating this variable to the person number variable, it bears mentioning that the specificity variable does not include non-human referents (in order to be consistent with previous research) and thus, it is possible that the tendency for NNSs to use more null subject pronouns than the NSs is more prevalent in non-human contexts.

These complex differences across categories and between participant groups demonstrate the value in examining each of these variables independently. Although sociolinguistic research often lends itself to a regression analysis (e.g., VARBRUL), ours is a case where a binary dependent variable does not exist given the range of forms produced. Regression analyses also assume that there is no interaction between the variables in the statistical model and our final analysis, that of the examination of the forms produced within the cross-section of each of the categories of the person and number and the specificity variable, shows that there are different patterns of use when the two variables are considered together, and this would further complicate a predictive statistical analysis. For example, NNSs used more null subjects in non-specific second person singular contexts, and in third person singular non-specific contexts but fewer null subjects in third person plural non-specific contexts than NSs. In contrast, NNSs used more null subject pronouns in second person singular specific contexts, fewer null subjects in third person singular specific contexts, and a relatively similar distribution of null subject pronouns in third person plural specific contexts as compared to the NSs. These two examples show that there are distinct patterns of use within the cross-section of each of the categories of these two variables such that no category of either variable has a uniform effect across the categories of the other. Thus, both the individual examination of linguistic variables and their relationship to subject expression and the examination of how these two variables interact are important in understanding subject expression for NNSs and for NSs.

5. Conclusion and future directions

The primary contribution of this paper has been to provide a rich, quantitative and qualitative description of the forms used to express a subject with finite verbs. In fact, we have demonstrated that subject expression is tremendously complex and that the person and number of the referent, the specificity of the referent, and the interaction of these two properties are all important in providing a complete description of the forms used for subject expression. We have demonstrated both similarities and differences between our NNSs and the NSs, even though our participants have a relatively high level of proficiency. Any theory that recognizes variation of the target as an integral part of L2 competence benefits from increased knowledge of these issues.

There are two big-picture issues that remain unresolved and are worthy of future study. The first is a question of analysis: how do we examine the multiple influences of the many linguistic features likely to be implicated in subject expression in the absence of a binary dependent variable? It is possible that an appropriate start to this question is to look at smaller samples of the data. For example in looking at second person singular referents, only null and over subject pronouns appear in the data and a binary dependent variable was produced naturally by the participants. In cases where this is not possible, additional detailed analyses like the current one may provide an indication of which subsets of the data can logically be examined as parts of this whole so that a more complete picture can begin to emerge.

The second important issue is related to what has been called the ‘envelope of variation’ (e.g., Cameron, 1994), or those contexts where variation is possible. In the existing sociolinguistic research it is noted that these contexts are not well defined (Otheguy & Zentella, 2007) and we contend that for L2 learners the issue is even more difficult because there is no reason to believe that learners limit variation in their speech to contexts where NSs allow variation. In response to this, we have included every subject of every finite verb in our analysis, but it is likely that future research can help to identify those contexts that are variable for both groups and to provide a clearly operationalizable manner for identifying these contexts. Once this is accomplished, the existing research will be more comparable across studies and researchers will have better assurance that important facets of the variation are not being overlooked. What we have shown thus far, however, is that important differences exist between NSs and NNSs outside the categories of null and overt subject pronouns, suggesting that a broader scope is appropriate. We expect that future research of the type we have conducted here will also take us closer to answering this second research question.

Appendix A: Sample items for the grammar test

Creo que es muy interesante _____ de los hábitos alimenticios de la gente. Yo, por mi parte,
 a. hablo
 b. hablar
 c. hablando

_____ vegetariana. Cuando voy a eventos sociales, como por ejemplo fiestas, bodas o bailes,
 a. soy
 b. estoy
 c. tengo

espero que _____ comida vegetariana allí. Algunas personas dicen que _____ representa un
 a. hay
 b. haya
 c. sea

a. le
 b. los
 c. les

inconveniente proveer _____, pero yo creo que no _____ que ser así.

- | | |
|-------|----------|
| a. lo | a. tiene |
| b. la | b. tenga |
| c. le | c. tengo |

English translation

I think that it is very interesting _____ about people's eating habits. As for me,

- a. I talk
 b. to talk
 c. talking

_____ a vegetarian. When I go to social events, like for example parties, weddings, or
 a. I am (verb form that implies that this is the norm)

b. I am (verb form that implies a change)

c. I have

dances, I hope that _____ vegetarian food there. Some people say that it represents an

- a. there is
 b. there may be
 c. it may be

inconvenience _____ to provide _____, but I think that _____ to be that way.

- | | | |
|------------|---------------|--------------------|
| a. to him | a. it (masc.) | a. it doesn't have |
| b. them | b. it (fem.) | b. it may not have |
| c. to them | c. to it | c. I don't have |

Appendix B: Sample questions for the sociolinguistic interview

¿Qué planes tienes para el fin de semana?

¿Dónde te ves en cinco años?

¿Quién es el mejor instructor de idiomas, un hablante nativo o no-nativo?

¿Cuáles son las ideas o los principios que tú crees que deberían motivar o guiar el gobierno de los EEUU (o tu país de origen)?

English translation

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?

What are the ideas or principles that you believe should motivate or guide the United States' government (or your country of origin)?

References

- Adamson, H.D., & Regan, Vera. (1991). The acquisition of community speech norms by Asian immigrants learning English as a second language. *Studies in Second Language Acquisition*, 13, 1, 1-22.
- Al-Kasey, Tamara, & Pérez-Leroux, AnaTeresa. (1998). Second language acquisition of Spanish null subjects. In S. Flynn, G. Martohardjono, & W. 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.
- Bentivoglio, Paola. (1987). *Los sujetos pronominales de primera persona en el habla de Caracas*. Caracas: Universidad Central de Venezuela.
- Bentivoglio, Paola. (1988). La posición del sujeto en el español de Caracas: Un análisis de los factores lingüísticos y extralingüísticos. In R. Hammond & M. Resnick (Eds.), *Studies in Caribbean Spanish dialectology* (pp. 13-23). Washington, DC: Georgetown University Press.
- Bentivoglio, Paola. (1993). Full NPs in spoken Spanish: A discourse profile. In W. Ashby, M. Mithun, G. Perissinotto & E. Raposo (Eds.), *Linguistic perspectives on Romance languages: Selected papers from the 21st 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 J.M. Licerias (Ed.), *La lingüística y el análisis de los sistemas no nativos* (pp. 126-193). Ottawa: Dovehouse.
- 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 K. Beals (Ed.), *Papers from the 30th regional meeting of the Chicago Linguistics Society: Volume 2: The parasession on variation in linguistic theory* (pp. 27-45). Chicago: 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, N. (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.
- Chomsky, Noam. (1981). *Lectures on Government and Binding*. Dordrecht: Foris.
- Comajoan, Llorenç (2006). Continuity and episodic structure in Spanish subject reference. In J. C. Clements & J. 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.
- 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. (2007). Los Mexicanos in New Jersey: Pronominal expression and ethnolinguistic aspects. In J. Holmquist, et. al. (Eds.), *Selected proceedings of the Third Workshop on Spanish Sociolinguistics* (pp. 85-91). Somerville, MA: Cascadilla Proceedings Project.
- Flores-Ferrán, Nydia. (2005). La expresión del pronombre personal sujeto en narrativas orales de puertorriqueños de Nueva Cork. In L. Ortiz López & M. 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. (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.
- 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. (in press). Comparing interview and written elicitation task in native and non-native data: Do speakers do what we think they do? In J. Bruhn de Garavito (Ed.), *Selected proceedings of the Hispanic Linguistics Symposium*. 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. (2006). L2 variation and the Spanish subjunctive: Linguistic features predicting mood selection. In C. Klee & T. Face (Eds.), *Selected proceedings of the 7th Conference on the Acquisition of*

- Spanish and Portuguese as First and Second Languages* (pp. 170-184). Somerville, MA: Cascadilla Proceedings Project.
- Gudmestad, Aarnes, & Geeslin, Kimberly L. (2007, April). Becoming native-like: A corpus-based study of variable structures in Spanish. American Association for Applied Linguistics, Costa Mesa, California.
- 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, Osvaldo, & Safir, Kenneth. (1989). *The null subject parameter*. Dordrecht: Kluwer Academic Press.
- Lapidus, Naomi, & Otheguy, Ricardo. (2005). Overt nonspecific *ellos* in the Spanish of New York. *Spanish in Context*, 2, 157-176.
- 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 J. Camps, & C. Wiltshire (Eds.), *Romance syntax, semantics and L2 acquisition: Selected papers from the 30th linguistic symposium on romance languages* (pp. 121-135). Amsterdam: John Benjamins.
- Liceras, Juana M. (1989). On some properties of the "pro-drop" parameter: Looking for missing subjects in non-native Spanish. In S. Gass & J. Schachter (Eds.), *Linguistic Perspectives on Second Language Acquisition* (pp.109-133). Cambridge: CUP.
- Liceras, 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, A. & W. Glass (Eds.), *Contemporary perspectives on the acquisition of Spanish Volume 1: Developing grammars* (99-132).
- Malovrh, Paul, A. (2006). L2 sentence processing of Spanish OVS word order and direct object pronouns: An analysis of contextual constraints. In N. Sagarra & A.J. Toribio (Eds.), *Selected proceedings of the 9th Hispanic Linguistics Symposium* (pp. 169-179). Somerville, MA: Cascadilla Proceedings Project.
- Morales, Amparo. (1997). La hipótesis funcional y la aparición de sujeto no nominal: El español de Puerto Rico. *Hispania*, 80, 153-165.
- Mougeon, Raymond; Rehner, Katherine; & Nadasdi, Terry. (2004). The learning of spoken French variation by immersion students from Toronto, Canada. *Journal of Sociolinguistics*, 8(3), 408-432.
- Miyajima, Atsuko. (2000). Spanish subject pronoun expression and verb semantics. *Sophia linguistica*, 46-47, 73-88.
- Otheguy, Ricardo, & Zentella, Ana Celia. (2007). Apuntes preliminares sobre el contacto lingüístico y dialectal en el uso pronominal del español en Nueva York. In K. Potowski & R. Cameron (Eds.), *Spanish in contact: Policy, social and linguistic inquiries* (pp. 275-298). Amsterdam: John Benjamins.
- Phinney, Marianne. (1987). The pro-drop parameter in second language acquisition. In T. Roeper & E. Williams (Eds.) *Parameter Setting* (pp. 221-238). Dordrecht: Reidel.
- Poplack, Shana. (1981). Mortal phonemes as plural morphemes. In D. Sankoff & H. Cedergren (Eds.), *Variation Omnibus* (pp. 59-71). Edmonton: Linguistic Research Inc.
- Preston, Dennis. Three kinds of sociolinguistics and SLA: A psycholinguistic perspective. In B. Swierzbins et al. (Eds.), *Social and cognitive factors in SLA* (pp. 3-30). Somerville, MA: Cascadilla.
- Selinker, Larry, (1972). Interlanguage. *IRAL*, 10, 209-231.
- Serrano, María José. (1996). El análisis del discurso en variación sintáctica. *Hispanic Linguistics*, 8, 154-177.
- Silva-Corvalán, Carmen. (1982). Subject expression and placement in Mexican-American Spanish. In J. Amastae & L. Elias-Olivares (Eds.), *Spanish in the United States: Sociolinguistic aspects* (pp. 93-120). Cambridge: CUP.
- VanPatten, Bill, & Cadierno, Teresa. (1993). Explicit instruction and input processing. *Studies in Second Language Acquisition*, 15, 225-243.
- White, Lydia. (1985). The "pro-drop" parameter in adult second language acquisition. *Language Learning*, 35(1), 47-62.

Selected Proceedings of the 2007 Second Language Research Forum

edited by Melissa Bowles, Rebecca Foote,
Silvia Perpiñán, and Rakesh Bhatt

Cascadilla Proceedings Project Somerville, MA 2008

Copyright information

Selected Proceedings of the 2007 Second Language Research Forum
© 2008 Cascadilla Proceedings Project, Somerville, MA. All rights reserved

ISBN 978-1-57473-425-6 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 or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, e-mail: sales@cascadilla.com

Web access and citation information

This entire proceedings can also be viewed on the web at 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. 2008. Variable Subject Expression in Second-language Spanish: A Comparison of Native and Non-native Speakers. In *Selected Proceedings of the 2007 Second Language Research Forum*, ed. Melissa Bowles, Rebecca Foote, Silvia Perpiñán, and Rakesh Bhatt, 69-85. Somerville, MA: Cascadilla Proceedings Project.

or:

Geeslin, Kimberly L. and Aarnes Gudmestad. 2008. Variable Subject Expression in Second-language Spanish: A Comparison of Native and Non-native Speakers. In *Selected Proceedings of the 2007 Second Language Research Forum*, ed. Melissa Bowles, Rebecca Foote, Silvia Perpiñán, and Rakesh Bhatt, 69-85. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #1736.