

# The L2 Development of Subject Form Variation in Spanish: The Individual vs. the Group

Kimberly Geeslin, Bret Linford, Stephen Fafulas, Avizia Long, and Manuel Díaz-Campos

## 1. Introduction

For half a century countless studies have employed the methods of quantitative sociolinguistics in order to describe linguistic variation and change. This field of research is known as variationist linguistics (cf. Labov, 1963) and has shown that linguistic and extra-linguistic constraints probabilistically guide the variation between linguistic forms and structures found in human language. While this methodology has been applied extensively to native languages, only since the early nineties have variationist techniques been used to examine the use and acquisition of categorical forms (e.g., Young, 1991) as well as variable forms in second languages (e.g., Adamson & Regan, 1991). This line of research has shown that like native speakers, second language (L2) learners show systematic variation in their use of linguistic forms, and that use is constrained by linguistic and extra-linguistic factors (e.g., Geeslin, 2011; Geeslin & Gudmestad, 2011; Geeslin & Guijarro-Fuentes, 2006; Gudmestad, 2012; Killam, 2012). This type of variation is essential for effective communication across speech settings (Canale & Swain, 1980). Given the importance of this feature of first and second languages, one of the principle goals of the current study is to foster cross-disciplinary dialogue between second language acquisition (SLA) and sociolinguistics by expanding the growing database on the acquisition of variation in second languages.

Because research on L2 variation is relatively new, several issues remain under-examined. For example, one potential challenge in applying quantitative sociolinguistics to L2 data is the fact that sociolinguistic research assumes that patterns of linguistic variation of the individual reflect those of the group (e.g., Guy, 1980; Otheguy & Zentella, 2012). Despite the general agreement surrounding this assumption, there are reasons to believe that this claim is not so irrefutable in the context of second languages. Research has shown that there is tremendous individual variation in L2 learning outcomes and the influence of social, cognitive, and affective factors is widely attested in the field (Ellis, 2004; Dörnyei, 2006, 2009). Some studies even suggest that the path of acquisition varies across both individuals and individual contexts of learning (e.g., Tarone & Liu, 1995). It is quite possible, then, that each language learners' interlanguage system varies idiosyncratically in a way that is more deviant from the group norms than was initially thought. This makes it essential to examine both group and individual norms of variation so as not to obscure individual differences through the reporting of only group norms of use (Howard, 2012). To date, only two studies have explored this issue using a variationist approach. The first study, Bayley and Langman (2004), produced findings consistent with the claim that the patterns of the individual do reflect the patterns of the group. In contrast, the second study, Geeslin and Gudmestad (2012), showed a great deal of individual variation between the participants, even at the highest levels of proficiency. In the detailed review that follows, we will explore the connection to these differences and the distinction between Type I variation (between a target-like and a non-target-like form) and Type II variation (between two target-like forms) because prior to the present study this was seen to be a likely explanation for the contrasting findings. For the purpose of this introduction, however, it is sufficient to state that previous research would not lead us

---

\* Kimberly Geeslin, Indiana University. Bret Linford, Indiana University. Stephen Fafulas, East Carolina University. Avizia Long, Indiana University. Manuel Díaz-Campos, Indiana University. For questions, contact Kimberly Geeslin at kgeeslin@indiana.edu. We would like to thank the anonymous reviewers and the editors of this volume for their helpful comments on this paper.

to expect, without question, group patterns that are wholly representative of individuals within the same group. In fact, the current study seeks to pursue precisely this line of inquiry by taking a closer look at group and individual norms of subject form selection in L2 Spanish.

## 2. The research context

As mentioned in the preceding introduction, the current study represents an example of cross-disciplinary research that seeks to expand our knowledge of subject form selection in Spanish, as well as how language variation contributes to our knowledge of first and second languages. More specifically, this project will contribute to the still limited body of research that examines the degree to which individual variation reflects group norms and vice versa. In order to design such a study, it is necessary to incorporate insights from several different fields of research. Consequently, the review that follows begins with an overview of how subject expression works in Spanish, how sociolinguistic research has described first language (L1) variation, what is known about L2 subject expression thus far and, finally, what we know about the relationship between the individual and the group in quantitative variationist L2 research. This review concludes with a summary of the goals of the current project.

### 2.1. Subject expression in Spanish

Spanish allows the grammatical subject of finite verbs to be null, as in example (1) or overt<sup>1</sup> as in example (2). While some contexts in Spanish generally require categorical use of either a null or an overt subject form (e.g., a null subject is used with verbs referring to weather such as ‘llover’ (*to rain*) and with the existential verb ‘haber’), variationist research has shown that in many contexts there is variation between null and overt forms of the subject (Amaral & Schwenter, 2005; Otheguy, Zentella, & Livert, 2007). Thus, speakers’ choice to use a null or overt form in examples (1) and (2) would be determined by several linguistic and extra-linguistic factors, all present in the discourse context at the same time.

1. Ø Habla español (*(She) speaks Spanish*)
2. Ella habla español (*She speaks Spanish*)

In fact, research on subject expression and the factors that condition form selection is quite extensive. The most commonly studied constraints on subject expression in Spanish are the person and number of the referent (Bayley & Pease-Álvarez, 1996; Bayley et al., 2012; Otheguy et al., 2007), tense, mood, and aspect (TMA) of the verb (Bayley & Pease-Álvarez, 1996; Bayley et al., 2012; Cameron, 1994; Erker, 2005; Hochberg, 1986; Silva-Corvalán, 1994), lexical content of the verb (Otheguy & Zentella, 2007), switch reference (Bayley & Pease-Álvarez, 1997; Bayley et al., 2012; Bentivoglio, 1987; Cameron, 1995; Otheguy et al., 2007; Shin & Otheguy, 2009), linguistic priming (Abreu, 2012; Cameron, 1994; Cameron & Flores-Ferrán, 2004; Flores-Ferrán, 2005; Travis, 2005, 2007; Torres Cacoullos & Travis, 2011), discourse cohesion (Bayley & Pease-Álvarez, 1997; Geeslin & Gudmestad, 2011; Paredes Silva, 1993), frequency (Erker & Guy, 2012), and continuity of TMA (Bayley & Pease-Álvarez, 1997; Geeslin & Gudmestad, 2011) among others. Of these factors, only a subset will be manipulated in the current study, while the remaining factors will be held constant. Those that play a central role in our analysis are 1) person and number of the referent, 2) switch reference, 3) TMA of the verb, and 4) continuity of TMA. Each of these will be described in greater detail here.

---

<sup>1</sup> Overt subject expression is not limited to subject pronouns. It also includes full lexical noun phrases, demonstrative pronouns, interrogative pronouns, and indefinite pronouns. This study, much like the majority of research on subject expression in Spanish, limits the analysis to null and overt personal subject pronouns in order to maintain consistency with previous sociolinguistic research. We do, however, agree with Geeslin and Gudmestad (2010) that these additional forms are an important facet of the developing learner grammar and merit examination in their own right.

The first factor, person and number of the subject referent, is often found to be the most influential constraint in predicting subject pronoun use (cf. Otheguy et al., 2007). In general, studies find that singular subjects are expressed overtly more often than plural subjects (Bayley & Pease-Álvarez, 1996). In addition, within singular subjects, the first-person singular is found to be associated with the highest frequency of overt subject forms in oral data (Bayley & Pease-Álvarez, 1996, 1997; Cameron, 1992; Enríquez, 1984; Geeslin & Gudmestad, 2008; Shin, 2012). The second constraint, switch reference, considers whether or not the referent of a verb is the same or different from the referent of the previous finite verb in the discourse. Research shows that overt subject pronouns are more frequent in contexts where the referent changes between verbs (“switch-reference” contexts) than in contexts where the referent is the same between verbs (“same-reference” contexts) (Bayley & Pease-Álvarez, 1997; Bentivoglio, 1987; Cameron, 1994, 1995; Otheguy et al., 2007; Shin & Otheguy, 2009). Thirdly, with regard to TMA, results of several studies indicate that “ambiguous” forms<sup>2</sup> co-occur with a higher rate of overt subject pronouns than unambiguous forms (Bayley & Pease-Álvarez, 1996; Cameron, 1994; Erker, 2005; Geeslin & Gudmestad, 2010; Hochberg, 1986; Silva-Corvalán, 1994). Thus, the imperfect indicative TMA, which has various potentially ambiguous forms, is typically accompanied by more overt subject forms than other TMAs, such as the present indicative, that do not contain inherently ambiguous forms. Finally, continuity of TMA takes into account whether or not the TMA of the verb is the same as or different from that of the immediately previous finite verb in the discourse. Studies considering this factor have found that speakers tend to use more overt subject pronouns when the TMA is different between verbs than when there is no change in TMA between verbs (Geeslin & Gudmestad, 2011).

## 2.2. Subject expression in second language Spanish

Most research on the acquisition of subject expression in Spanish has focused on the Null Subject Parameter and associated syntactic properties (Al-Kasey & Pérez-Leroux, 1998; Bini, 1993; Emberson, 1987; Galván, 1999; Isabelli, 2004; Licerias, 1989; Licerias, Maxwell, Laguardia, Fernandez, & Díaz, 1997; Phinney, 1987; White, 1985). While some research has explored the influence of discourse-pragmatic features on subject expression in L2 Spanish (Blackwell & Quesada, 2012; LaFond, 2002; LaFond, Hayes, & Bhatt, 2000; Montrul & Rodríguez Louro, 2006; Quesada & Blackwell, 2009; Rothman, 2007), the research focus is primarily on those contexts considered to be “categorical” (i.e., where null or overt subjects are obligatory), rather than variable.

Studies on the L2 acquisition of subject expression in Spanish conducted within a variationist framework remain fewer in number. Nevertheless, this body of research is growing. In a series of analyses of data collected through sociolinguistic interviews with 16 highly advanced English-speaking L2 learners of Spanish and 16 native speakers (NS) of Spanish, Geeslin and Gudmestad (2008, 2010, 2011) have shown that L2 speakers demonstrate native-like variation between subject forms in the sense that this variation is conditioned by the same factors as for the NS group. Among other factors, they found that like NS expression of subjects in Spanish, subject expression by these highly advanced learners of Spanish was constrained by the person and number of the verb form, specificity of the referent, TMA of the verb form, potential ambiguity of the verb form, the discourse context, discourse cohesiveness, and perseveration (i.e., the effect of the preceding subject form on that which follows). Despite the similarities in the factors that condition subject form expression, this research also showed that the non-native speakers produced null pronouns more frequently than their NS counterparts.

In a subsequent study, Geeslin and Linford (2012) sought to examine not only the patterns of subject form expression in Spanish by highly advanced learners, but also how these patterns developed across levels of proficiency. They analyzed subject form selection on a highly-controlled written contextualized task in which several linguistic factors were held constant and others, related to discourse-level features, were manipulated. Specifically, they examined the effects of perseveration and referent cohesiveness. This second variable combines the distance to the previous mention of the referent with measures of cohesiveness that include changes in the TMA of the verb across referents

---

<sup>2</sup> This factor has been defined differently in different studies with some referring to the individual forms and their potential to be ambiguous regardless of the context while others consider the forms to be ambiguous only in combination with a discourse context that may lead to ambiguity.

and changes in the form of the previous mention (i.e., subject vs. other grammatical functions). Each category of these two variables was evenly distributed across the written contextualized task. The response format required participants to select between sentences that were identical except for the subject form contained: null, overt, or a full noun phrase (NP). The participants were 150 English-speaking learners of Spanish from six levels of enrollment and 18 NSs of Spanish. The results showed that beginning with third-year learners, each learner level was significantly different from the preceding level in terms of the rates of selection of subject forms. As the level of proficiency between groups increased, there was an increase in rates of selection of null pronouns as well as a decrease in lexical NPs and overt pronouns. The highly advanced group did not differ significantly from the NSs. Furthermore, as L2 speakers of Spanish gained proficiency, they showed form selection that was conditioned by the same discourse-level linguistic constraints as NSs.

Most recently, Geeslin, Linford, and Fafulas (2012) examined the development of rates of form selection and the extent to which both syntactic and discourse-level constraints condition form selection across proficiency levels. This study built on the work of Geeslin and Linford (2012) in that it explored additional linguistic factors, such as person and number and TMA of the verb form, which were held constant in the previous study. Geeslin, Linford, and Fafulas (2012) analyzed data from 180 English-speaking learners of Spanish enrolled in six different levels of instruction as well as 27 NSs of Spanish. The elicitation instrument they employed was a contextualized questionnaire (i.e., a dialogue in which each response item was embedded), which required participants to indicate a preference for either a null subject or an overt subject pronoun. Unlike previous studies, the results showed that rates of selection of null pronouns first decreased across levels of enrollment then eventually increased toward the NS norms. While at the lower levels of enrollment, there was no effect for the linguistic factors, as the level of proficiency of the learner increased, both person of the verb form and switch reference entered the predictive model, and subsequently TMA was added as a relevant predictive factor. At the highest level of proficiency, the non-native speakers did not differ from the NSs with regard to the factors included as significant predictors of subject pronoun selection. These methods and findings will be explored in much greater detail in the description of the current study because they provide the basis for comparison of the group norms discussed in our previous work to be used in the original analysis of individual patterns in the present investigation.

### *2.3. The role of individual factors in second languages*

The role of individual factors in L2 acquisition is well documented and it is known that factors such as language aptitude (cf. Carroll, 1991; Robinson, 1997; Skehan, 1998, 2002; Sternberg 2002), motivation (cf. Dörnyei, 2001; Gardner, 1985; Gardner & Lambert, 1972; Noels, Pelletier, Clément, & Vallerand, 2001), and anxiety (cf. Bailey, 1983; Horwitz, Horwitz, & Cope, 1986; MacIntyre & Gardner, 1994), to name only a few, have an impact on the acquisition of second languages (cf. Ellis, 2004 for an overview). Historically, the study of individual factors in SLA has reflected the growing interest in empirically investigating differences in the variable levels of language achievement demonstrated by learners (cf. Ellis, 2004). Differences have also been posited to be related to social factors, and there has been a strong call in recent years, preceding the seminal article of Firth and Wagner (1997) and continuing to the present day, for theories of SLA that incorporate social factors into the explanations and descriptions of L2 learning and use (see Adamson, 1988 (prototype theory); Atkinson, 2011; Larsen-Freeman, 2011 for example). Such studies also demonstrate that patterns of use may vary at the individual level in addition to levels of overall attainment. What remains to be investigated, however, is the extent to which individual L2 learners differ from each other within their corresponding groups, particularly in relation to variable structures of the target language. Another key distinction within variationist research, which may also interact with the degree to which individuals in the same group differ from one another, is that between Type I variation, which is associated with L2 development and describes variation between a native-like form and one or more non-native forms, and Type II variation, which is the variation between two or more native-like forms in response to linguistic, social, and contextual features of the discourse. Only this second type of variation is found in adult native speech. It may be the case that the degree to which individuals vary or the way in which individuals vary is different, depending on the type of variation under examination. For either type of structure, if differences at the level of the individual are found, further research into the specific factors associated with individual variation is certainly warranted as is a re-examination of current methods of

analysis that generally examine aggregate data. The current study expands on the few studies that have investigated the role of the individual in acquiring variation in a second language.

One of the first studies to explore the role of the individual as compared to group norms in L2 variation was Bayley and Langman (2004). Their work explored the acquisition of verbal morphology in English and Hungarian by Chinese-speaking learners, which is variable in learner language as development progresses (Chinese does not have verbal inflections), but not in the target languages (i.e., vertical or Type I variation<sup>3</sup>). Data were obtained from sociolinguistic interviews conducted with 20 learners divided into two proficiency levels (low intermediate/intermediate and advanced). Their results showed that the individual patterns of variation between native and non-native forms closely matched group patterns for constraints such as grammatical aspect, regardless of the learners' proficiency level. Perceptual salience and frequency were also found to affect verb marking by individual learners of both languages similarly. The results of the VARBRUL analyses at the level of the individual showed very similar weights across individuals for the variable 'aspectual constraint' (perfective or imperfective). They concluded that using quantitative results for groups in L2 research is valid from a theoretical standpoint. This finding is encouraging because it demonstrates that current practices in the field, such as using aggregate data from multiple participants, provides an accurate picture of individual patterns of use. Nevertheless, there are a number of reasons that make additional investigations important. Firstly, Bayley and Langman analyzed data from two levels of proficiency, but studies that explore greater spans of development, including both novice and highly advanced learners, are necessary before these results can be generalized. Additionally, Bayley and Langman examined variation between native and non-native forms in the acquisition of categorical rules (Type I variation), but it is possible that variation between various native forms and the acquisition of their respective constraints (Type II variation) would not show the same consistency between individual and group patterns. Our study aims to address these issues in greater detail.

To our knowledge, the only investigation examining both individual and group patterns in a single study of L2 sociolinguistic variation suggests that there may be greater divergence at the level of the individual from group norms on structures that are variable in the target language than on those that are not (Geeslin & Gudmestad, 2012). In a study of future-time reference, Geeslin and Gudmestad analyzed form selection by 151 English-speaking learners of Spanish representing five instructional levels (first year to graduate level), as well as 22 NSs (all residing in the US at the time of the study). They compared rates of selection for the present indicative, the morphological future (e.g., *hablaré* 'I will speak'), and the periphrastic future (e.g., *voy a hablar* 'I am going to speak') on a written contextualized questionnaire designed to present items representing every combination of the categories of the following three linguistic variables: presence of a lexical temporal indicator (+/-), temporal distance (immediate, today, <week, <month, >year), and certainty markers (certainty marker, uncertainty marker, no marker). Their analysis of aggregate rates of selection and subsequent comparison of individual and group patterns of selection demonstrated that the participants selected the forms at similar rates but there was considerable inter-speaker variability with regard to the constraints on the selection of these future-time forms. Geeslin and Gudmestad first examined the rates of selection of each of the three response choices for several proficiency levels and then examined the individual rates, assessing the standard deviation and range for each of the three response forms. On this measure, they found differing degrees of variability for each learner level and for each of the three future-time response forms. For example, the present indicative showed the least variability and the periphrastic future showed the greatest, but within rates of selection of the periphrastic future, the highest proficiency group showed less variability than the next lower learner level and the NS group. In a second measure of individual differences, Geeslin and Gudmestad examined the hierarchy of form use (most frequent to least frequent) in each of the categories of the temporal distance variable (see Gudmestad & Geeslin, 2013, who found this variable to be the most complex and varied across developmental levels). This second measure demonstrated additional variability at the level of the individual, with the morphological future showing greater variability than the present indicative. Geeslin and Gudmestad summarized these findings to compare both rates of selection and patterns of selection across the temporal distance variable. They found that individuals could differ from the

---

<sup>3</sup> Type I variation more generally refers to alternation between non-native or non-target forms (e.g., learner errors), whereas Type II variation refers to the alternation between native or target forms also found in native speech (i.e., sociolinguistic variation).

group norm on one or both measures, but very few individuals differed from group norms on all three forms on both measures, or were the same as the group on all three response forms on both measures. Likewise, they found less variability in the frequency of selection of the three response forms than in the distribution across categories of the temporal distance variable. Taken together, these findings suggest that not only may the acquisition of sociolinguistic variation create a context of greater individual divergence, but also that this divergence may not be the same for all measures such that frequency of use of a form or the constraints of that use may vary at the level of the individual while the other measures do not. Clearly, future research has much to explore in answering these questions.

#### *2.4. Assessment of existing research and goals for the current study*

The preceding review demonstrates that much is known about the factors that condition subject form use and selection in Spanish across native and learner populations. We know that such use is conditioned by both discourse-level and syntactic factors, and that L2 learners of Spanish are able to acquire these constraints. We also know that learners tend to use and select null pronouns at greater rates than NSs even at very high levels of proficiency. Furthermore, individual and social factors do affect SLA and these must be included in models of SLA. Nevertheless, research into the role of the individual from a variationist viewpoint is quite new and there is much that remains unknown. For example, it has been suggested that while Type I variation does not show great differences between individual and group patterns, Type II variation may be a case where greater variation at the level of the individual exists. To date, however, the research on both types of structures is too limited to make generalizable claims of any sort. To that end, the current study seeks to expand our knowledge base in this area.

### **3. The current study**

#### *3.1. Research questions*

Given the need for additional research on the role of the individual described in the preceding review, the current study was designed to answer the following questions:

1. When considering the linguistic factors (both syntactic and discourse-level) generally thought to be most important for determining subject expression,
  - a. Do individual rates of selection match group patterns?
  - b. Does the predictive model improve (i.e., predict additional variation) when individuals are included as a factor in the model along with the linguistic factors?

#### *3.2. Participants*

The participants in the current study were 180 native English-speaking learners of Spanish enrolled in Spanish courses ranging from first year to graduate-level. Our study also included 27 NSs of Spanish from a range of countries of origin (Argentina, Colombia, Mexico, Peru, Puerto Rico, Spain, and the United States), all of whom were pursuing graduate study in the United States at the time of the investigation. The NS group includes only those speakers of Spanish born into Spanish-speaking households in which both parents were NSs, and had received formal education in Spanish. The characteristics of all participants are summarized in Table 1. The table includes the mean score for each group and the standard deviation of those scores on a formal grammar proficiency test, which will be described in greater detail below. The last column of the table denotes the percentage of participants in a given group with more than three weeks experience in a Spanish-speaking country.

Table 1. Description of participants

Group	N	Mean proficiency score	Standard deviation on proficiency test	Percentage of participants with experience abroad
1st year	30	8.70	2.13	0
3rd semester	30	11.03	2.43	0
5th semester <sup>4</sup>	30	11.24	2.75	6.7
7th semester	30	13.43	3.53	20
4th year	30	17.60	3.73	73.3
Graduate	30	22.20	2.00	100
NS	27	23.42	0.86	n/a

Table 1 shows that there was an even number of participants in each group, except for the NSs where participants were more difficult to identify. Likewise, the scores on the proficiency test, which was used to corroborate the assignment of participants to groups based on levels of enrollment, show a general increase over time. A statistical analysis of those scores showed that all groups were significantly different from each other with two exceptions: the third and fifth semester groups were not different from each other and the graduate-level learners and the NSs were not significantly different from each other. Finally, one notes that the number of participants with experience in a Spanish-speaking country increases with level of enrollment.

### 3.3. Elicitation tasks and procedure

All participants in the current study completed three written instruments. The first was a detailed questionnaire that elicited background information concerning social factors and language learning experience. This questionnaire was used to ensure that all participants were comparable in terms of socio-economic status and to provide additional information that might be useful in assessing the role of individual differences. For example, the information regarding study abroad might be useful in explaining certain differences within groups. The second instrument completed by all participants was the proficiency test. This test might be more aptly named a test of formal grammatical knowledge in that it examines participant knowledge of the formal properties of the Spanish language that tend to be taught in formal classroom settings. The instrument had a multiple choice format and contained a total of 24 items, which were contextualized in a narrative about two students and an invitation to a party.

The third and primary data elicitation instrument was a written contextualized task that required participants to select the subject form they preferred in a given context. Each item was contextualized in a dialogue between two NSs of Spanish talking about college life. The two response choices were clauses that differed only in that one contained a null subject pronoun and the other contained an overt subject pronoun. Four independent linguistic variables were manipulated in this written instrument. Firstly, half of the contexts contained situations of switch references (i.e., the subject of the preceding tensed verb was different from the one in the response) and half contained contexts of same reference. Secondly, half of the items contained verb forms in the imperfect and the other half contained verb forms in the preterit in order to examine the effect of TMA on form selection. These two forms were chosen because the imperfect is potentially ambiguous in that the first- and third-person singular forms are the same, whereas the preterit forms hold no such ambiguity. Thirdly, the continuity of TMA was manipulated such that half of the items contained a switch between the preceding verb form and the form in the response, and half did not. Finally, the person of the verb form was evenly divided between first-person singular and third person singular forms. The categories of these variables were distributed in such a way that all combinations of each category of the independent linguistic variables are included on the instrument. Given the extensive research on the other linguistic factors that also influence subject form selection, several linguistic variables were held constant across items. Thus, all verb forms were singular, no TMA forms other than preterit or imperfect were included, all responses were found in main clauses, no fixed or idiomatic expressions were included in the response items, and

<sup>4</sup> This course is the first course beyond the basic language requirement and generally represents a shift from those seeking to meet the language requirement to those with an interest in minoring or majoring in Spanish. By the 7<sup>th</sup> semester course, students are all majors and minors and some may have completed a significant portion of their preparation prior to this level at the secondary level.

the previous mention of the referent in the response was always null. In order to ensure that all participants understood the contexts on the instrument, 22 less common words or phrases also contained an English translation in parentheses. Figure 1 presents a sample item from the written contextualized instrument. It contains a response context that has same-reference, the same TMA, and a first-person singular form in the imperfect.

Figure 1.

**Pedro:** Pues, anoche estábamos mirando el partido de fútbol en la tele pero yo estaba muy aburrido y como no tenía tarea,...

- a) ...quería hacer algo diferente de lo normal.
- b) ...yo quería hacer algo diferente de lo normal.

### 3.4. *Data coding and analysis*

The dependent variable in the current study was the form selected. The two categories of this variable were null and overt. As is implied by the design of the task above, each response item was coded for the following linguistic variables: switch reference, TMA, continuity of TMA, and person of the verb form. Additionally, the extra-linguistic variable ‘participant’ was used to examine the degree to which individual differences influenced the predictive model. As is generally found in research on variable structures, our analysis used a cross-tabulation to determine the distribution of these forms across participant groups as well as a regression analysis to determine the degree to which each of the independent variables examined contributes to the selection of subject forms for each group. Thus, because our study contains seven participant groups, we report on the distribution of forms selected by each group and then compare the results of seven different regression analyses, one for each group. Additionally, in order to examine the role of individual behavior for each group, we present the range of rates of selection of these forms, the minimum and maximum rates of selection for each group, and the standard deviation for each group all in comparison with the group mean determined by the initial cross-tabulations. In order to assess the role of the individual in the predictive models, we enter the factor ‘participant’ in a second set of regression analyses and report any changes in the model, including whether the percent of variance predicted by the model improves with this additional factor or not.

In sum, to answer research question 1a, we first present the distribution of forms selected by level, followed by a summary of ranges and group standard deviations of rates of form selection in order to determine whether or not individual selection rates match group patterns. To answer question 1b, we present the results of regression analyses conducted for each participant group to establish the predictive model, followed by an additional analysis including ‘participant’ as an independent variable to determine the extent to which individuals impact the predictive power of that model.

## 4. Results

### 4.1. *Distribution of forms selected*

We begin the presentation of the results with a summary of the rates of selection of null and overt pronominal forms for each of the six learner groups as well as for the NSs. These group norms, presented in Table 2, are the same as those reported in Geeslin, Linford, and Fafulas (2012) described in the previous review of literature.



Table 2. Distribution of forms selected by level

Level	Null Pronoun		Overt Pronoun		Total
	#	%	#	%	
1st year	316	65.8	164	34.2	480 (100%)
3rd semester	292	61.1	186	38.9	478 (100%)
5th semester	266	55.5	213	44.5	479 (100%)
7th semester	257	53.7	222	46.3	479 (100%)
4th year	293	61.6	183	38.4	476 (100%)
Graduate	317	66.0	163	34.0	480 (100%)
NS	291	67.5	140	32.5	431 (100%)

In addition to the number of null and overt forms selected by each group and the percentages that these numbers represent, we submitted the data of overall selection of null pronouns to an ANOVA test to determine if the differences in frequencies of selection of null pronouns by group differed significantly from one group to another. The results showed significant differences between the groups as a whole [ $F(6,200)=2.17$ ,  $p=.015$ ], but post hoc Tukey tests showed that only the 7th semester group and the native speaker group were significantly different from one another ( $p=.05$ ).

Nevertheless, the results summarized in Table 2 demonstrate that learners initially select null subjects, even at very low levels of proficiency. Recalling research on the Null Subject Parameter, which focuses on whether or not learner grammars allow null subjects, it is clear that allowing null subjects is not an acquisitional challenge for learners. Instead, it appears that acquiring constraints on these null pronouns is what truly demonstrates development across levels of proficiency. The rates of selection of null subjects gradually decline until the 7th semester and, in fact, the rates of selection are not statistically different between the 5th semester and the 7th semester. From the 7th semester, moving toward the graduate level, we see a gradual increase in rates of selection of null pronouns toward the NS norms. Again, there are no significant differences between learners at the 4th year and above and the NS group. Prior to examining the degree to which individuals follow these trends, it is worth noting that on this particular elicitation instrument all response contexts were at least the second mention of the referent. In other words, no context represents new information and all contexts should allow a null pronoun. This must be taken into account in the interpretation of the rates of selection of null pronouns since these numbers are higher than would be anticipated across a wider variety of discourse contexts.

#### 4.2. Individual variance in rates of forms selected

Having established the norms for group rates of selection of null and overt pronominal forms, we continued with an analysis of the degree to which individuals within those groups reflected the same patterns. In order to measure this, we calculated the range and standard deviation of the group rates of selection of the null pronouns. The results of this analysis are summarized in Table 3, with the column “group average” matching the rates of selection of null pronouns reported in Table 2. Table 3 also shows the highest rate of selection (maximum) and the lowest rate of selection (minimum) demonstrated by at least one individual in each participant group.

Table 3. Summary of individual ranges and group standard deviations of selection of null pronouns

Group	Minimum	Maximum	Range	Standard deviation	Group average
1st year	37.5	100	62.5	17.4	65.8
3rd semester	0	100	100	22.5	61.1
5th semester	0	100	100	19.4	55.6
7th semester	6.2	93.8	87.6	15.8	53.7
4th year	31.2	100	68.8	18.3	61.6
Graduate	31.3	93.8	62.5	13.8	66.0
NS	43.8	100	56.2	14.1	67.5

One interesting result shown in Table 3 is that there are individuals who select only one type of form in all contexts on this task. Specifically, in four of the six learner groups there is at least one

participant who selected null pronouns in every context. Recalling that all contexts were possible for null pronouns, this result may not be all that surprising: it simply shows that some participants do not vary their selection of null and overt forms according to the factors manipulated on the current written task. In fact, there is even one NS participant from Spain that selected only null forms on the task. In contrast, however, there are also learners that never select the null form in two of the three lowest-level groups. In these cases, the learners have yet to allow null pronouns into their grammars, at least for the contexts included on the current elicitation task, and this is not attested in the NS group. Thus, in describing development we can say that even at the level of the individual, all learners allow null subjects after the 5th semester in at least one context on the task. One further notes that the standard deviations begin to decrease after the 5th semester, with the graduate level and NS groups showing the smallest deviations. Thus, despite some spikes as development progresses, it appears that acquisition of subject form selection in this case is characterized by a decrease in individual differences over time. Although there is still individual variation in rates of selection for the L2 learners at the highest level as well as for the NSs (observe results for range in Table 3), it can be said that with increased L2 proficiency the rates of selection of subject forms show a similar pattern to the results presented in Bayley and Langman (2004). In both cases, the degree of individual difference within learner groups decreases as proficiency increases.

#### 4.3. Predictive models of subject form selection

In addition to our analysis of rates of form selection, we also assessed the factors that predicted form selection for each group by conducting a regression analysis for each participant group separately. The results from these seven regression analyses are summarized in Table 4 and, again, these results are consistent with those reported in Geeslin, Linford, and Fafulas (2012), which established the group norm to which the role of the individual will be compared. Table 4 contains an 'X' for each factor that was included in the predictive model for that participant group and uses asterisks to represent the degree of significance of each factor. Again, the factors that were entered into the statistical tests for each group were the same: switch reference, person of the verb form, TMA of the verb form, and continuity of TMA of the verb form.

Table 4. Summary of regression analyses for each group with independent linguistic variables

Group	Switch reference	Person	TMA	Continuity of TMA
1st year				
3rd semester				
5th semester	X*			
7th semester	X**	X**		
4th year	X***	X***	X**	
Graduate	X***		X***	
NS	X***		X***	

Note. \*= $p < .05$ , \*\*= $p < .01$ , \*\*\*= $p < .001$

The results summarized in Table 4 show that the factor continuity of TMA was never included in the predictive model for any participant group. All other factors were significant predictors in the model for at least one group. The factor switch reference enters into the predictive model at the 5th semester and becomes increasingly significant as proficiency increases. The factor person of the verb form (it will be recalled that all verb forms were either first- or third-person singular) enters into the predictive model at 7th semester, but disappears at the graduate level. The TMA of the verb form enters into the model during the 4th year and increases in significance at the graduate and NS levels. Finally, as seen in Table 4, only the graduate student group shows selection of subject pronouns based on the same constraints as the NS group.

#### 4.4. The role of the participant in predicting subject form selection

Using the regression analyses reported in Geeslin, Linford, and Fafulas (2012) and summarized in Table 4 as a basis for comparison, the current analysis sought to explore the degree to which individual patterns of behavior were consistent with group patterns. To this end, we entered the factor

‘participant’ into the model, thereby allowing each individual participant to be considered as a predictive factor. A model that identifies this factor as a significant predictor of form selection indicates that when considering individual behavior, a greater amount of variation can be predicted. On the other hand, if the participant factor is not included in the predictive model, this means that patterns of selection within a group are not greatly determined by variation caused by individuals. Table 5 summarizes the results of these additional regression analyses. The first four columns are the same as those reported in Table 4 for the tests conducted without the participant factor. The next column reports the percentage of all tokens predicted by the model that did not include the participant factor. The last two columns summarize the results for the models where the participant factor was included in the statistical run and they show whether or not the factor was included in the model (with an X), the degree of significance of this factor when it was included in the model (with asterisk(s)) and, in the final column, the amount of variance predicted by the model.

Table 5. Summary of regression analyses for each group with independent linguistic variables

Group	Switch reference	Person	TMA	Continuity of TMA	% predicted	Model with participant	% with participant
1st year					65.8	X*	67.5
3rd semester			X*		61.3	X***	71.5
5th semester	X*				55.5	X	66.4
7th semester	X**	X**			57.8	X	64.9
4th year	X***	X***	X**		70.6	X	74.2
Graduate	X***		X***		77.3	X*	79.4
NS	X***		X***		79.1	X	81.9

Note. \*= $p < .05$ , \*\*= $p < .01$ , \*\*\*= $p < .001$

Table 5 provides quite a bit of information about the degree to which individual variation contributes to subject form selection. One notes that at the two lowest levels of proficiency, the participant factor is more important than any other linguistic predictor. It will be recalled that at these levels there were some individuals who still demonstrated categorical selection of null and/or overt forms, thus our predictive model, which accounts for these individuals, is able to capture and better explain overall patterns of selection at these proficiency levels. In fact, when ‘participant’ is included in the analysis, TMA becomes a significant predictor of subject pronoun selection for the 3rd semester group which means that the effects of individual participants are quite possibly obscuring the importance of TMA when ‘participant’ is excluded from the analysis.<sup>5</sup> However, the effect of TMA is in the opposite direction of what would be predicted. Whereas for other speaker groups, including NSs, the imperfect forms are more likely to occur with overt subject forms, for the 3rd semester group, the preterit forms actually occur with more overt subject forms. After these first two levels, the participant factor is included in the model but does not reach significance, even at the .05 level, with the exception of the graduate level. Possible reasons for this exception at the highest level will be addressed in the discussion that follows. A good indicator of the importance of the individual participant factor is the percentage of the tokens that the model can predict with this additional factor as compared to the percentage predicted by a model that does not contain such a factor. In other words, the percent predicted tells us how much the inclusion of this factor “helps” the statistical model predict a greater percentage of the overall tokens. At the beginning and intermediate levels (i.e., up through 7th semester) the inclusion of the participant factor improves the predictive ability of the model between seven and 11 percentage points. By 4th year, we see a sharp decline in the importance of inclusion of this factor. For the 4th year learners, the change in the predictive model is just over three percentage

<sup>5</sup> Although regression analyses do not account for multiple tokens produced by the same participant (which is a violation of the assumption of independent observations), it is still common practice among sociolinguists to run this type of test. That said, the second model (with participant included; see Table 5) approaches a more appropriate repeated measures test and shows that there is an additional linguistic variable that should be taken into account for the 3rd semester participants (i.e., TMA). In fact, another binary logistic regression was run using a Generalized Estimating Equation in which the subject (i.e., participant) was incorporated in order to avoid the violation of independent observations. The results of this test confirmed that TMA is significant for this level ( $p < .05$ ).

points and for the graduate level and the NSs, the change in the model is even less than three percentage points. As a further exploratory measure, we also ran these same predictive models with extra-linguistic factors such as gender and experience abroad in order to determine whether certain extra-linguistic characteristics could actually explain the effect of the participant factor. Our additional analyses showed that for the 7th semester and beyond, only linguistic factors were included in the predictive model. In other words, the inclusion of the participant factor in the models summarized in Table 5 is not the result of a single, typically-examined extra-linguistic characteristic.

## 5. Discussion

It will be recalled that the current investigation sought to expand the body of research that explores the role of the individual in the study of L2 variation. Specifically, the study was designed to investigate whether rates of selection of subject forms and the linguistic predictors of that selection demonstrated by groups at several proficiency levels are also indicative of individual patterns within those learner groups. In considering the rates of selection of null and overt subject pronoun forms, we found a general decrease in individual variation as development progressed. Thus, at lower levels of proficiency, we see a greater range of rates of selection of null forms, including individuals who use only null or only overt forms across all contexts on the task, and greater standard deviations within groups. As proficiency increases, however, the number of participants that select only one form decreases, the range of rates of selection decreases and the standard deviations generally decrease (although there are a few spikes along the path of development). Finally, the results for the NSs and graduate-level learners are similar in that these groups show the smallest standard deviations and ranges. We note that although the lowest group shows a range that is equally small to these last two groups, the standard deviation for the beginners is greater. One interesting finding is that one of our NSs, a speaker from Spain, selects only null forms throughout the task. Nevertheless, this is not completely unexpected given that speakers from Spain have been shown to produce some of the lowest rates of overt subject pronouns (Cameron, 1994), and in light of the fact that every context on the written task was a potential context for a null form. For this reason we do not take exclusive use of null forms by any speaker as an indicator that their grammar does not include overt forms. Instead, we fully expect overt forms to appear in contexts where, for example, the referent has not previously been mentioned.

In addition to examining the degree to which rates of selection represent individual patterns within the group, we also examined the predictors of that selection. The predictive models for each group showed that at very early levels of development NNS selection is not constrained by the linguistic factors found to predict NS selection. At the 3rd semester level we see that TMA appears as a significant predictor but in the opposite direction than expected such that rates of selection of overt subject pronouns are higher with the preterit forms than the imperfect forms. While unexpected, this result may be due to the fact that while none of the imperfect forms in the WCT are accompanied by object pronouns, five of the eight items that have preterit forms included an object pronoun. Thus, the 3rd semester participants may be selecting more overt subject pronouns due to increased processing load when object pronouns are present. In fact, this result may be a reflection of the 'First Noun Principle' (cf. Malovrh, 2006 for a review of research on this topic). This principle describes the universal strategy by which learners tend to interpret the first noun as the subject of a sentence, regardless of its actual grammatical function. For example, when a learner confronts 'me dijo...' this may initially be interpreted as 'I said,...'. In the case of the current study, an item with an additional overt subject pronoun provides an additional cue that this is not an accurate interpretation<sup>6</sup>. However, this trend is short-lived and disappears by 5th semester. This is followed by the addition of the factor switch reference, then by person of the verb form, and then by the re-addition of the TMA factor to the model (this time in the expected direction). For the graduate level group and the NSs, the only two linguistic factors included in the predictive model are switch reference and TMA of the verb form and both of these are highly significant. Thus, similar to the ranges and standard deviations of the rates of

---

<sup>6</sup> We note that this is but one possible interpretation and our findings may also be related to verbal irregularity, lexical frequency, cognitive load, and patterns of processing and perception that fall outside the scope of our investigation. Each of these additional interpretations motivates future research, perhaps with on-line perceptual tasks to test the effects of these additional factors.

selection, which were native-like after the 4th year, only the graduate level group showed subject form selection on the written task that was constrained by the same predictors as the NSs. This finding corroborates the findings of Geeslin and Gudmestad (2008, 2010, 2011) who showed that the use of subject forms for the highly advanced learners and the NSs in their study was generally constrained by the same factors, although subtle differences across categories within those factors did exist. Thus, the current analysis shows that highly advanced speakers not only show an ability to vary their speech in native-like ways in oral tasks, but on written tasks as well.

Turning then to the issue of individual variation in comparison with these group patterns, we also analyzed our data using predictive models that included ‘participant’ as a factor in order to explore the degree to which it was significant and the impact this factor had on the overall predictive ability of the model. We found that the overall predictive ability of the model did, in fact, improve when the participant factor was included, but the degree of this impact was related to the level of the participant group. At lower levels of enrollment we saw greater levels of significance for the participant factor and a greater change in the overall percentage of the tokens the model was able to predict, whereas at higher levels, this effect was weaker. Indeed, it appears that failure to include ‘participant’ as a variable in the regression model skewed the results of TMA for the 3rd semester group, thereby underscoring the importance of the individual at this level. In sum, the participant factor went from having a strong enough effect to improve the percentage of tokens predicted by more than ten percent at lower levels of proficiency, down to changing the percent predicted by less than three percentage points for the graduate level and the NS group. One interesting contradiction to these findings is that despite the decrease in the change in the percentage of tokens predicted, the participant factor did reach significance in the predictive model for the graduate level. This is somewhat surprising given this factor’s lack of significance at lower levels (beginning at 7th semester) and for the NSs and also precisely because there is a relatively small change in the overall change in the percent predicted. One key difference between the graduate level model and those for lower levels is that the overall percentage of tokens predicted is relatively high and the model predicts more than three-fourths of the forms selected by this group. In this sense, a change in three percentage points is in some sense a greater change as one approaches the ceiling for each model. As was stated previously, experience abroad and gender were not found to be significant predictors and, thus, it is unlikely that the individual differences found for this group are the result of some extra-linguistic or social characteristic. Additionally, we note that the standard deviation on the proficiency measure employed in the current study is smaller for this highly-advanced group than for any other participant group. This leads us to suspect that it is not the case that a few individuals simply are not as proficient as the others in this group. Instead, our hypothesis is that although this group has moved to a model in which only the linguistic factors that also constrain NS form selection are included, as opposed to the model for the NNS level below them, which differs from the NS model and includes person of the verb form, there is still some degree of idiosyncratic patterning that separates the NSs from this highly advanced learner group. Further research is required to confirm this hypothesis and to explore in greater depth the learner characteristics or experiences that might explain such variability.

To summarize the trends demonstrated by our two analyses, the role of the individual in the rates of selection of null subject forms generally decreases as proficiency increases and this is shown through decreasing ranges and standard deviations. Additionally, the role of the individual in predicting the form selected also decreases across levels, as demonstrated by decreasing significance in the predictive model and a decrease in the change in the percentage of overall tokens predicted by the model. This second result meets with one important exception at the graduate level, where the individual factor is significant at the .05 level in the predictive model. In comparing these trends to findings in previous studies, our results pattern more consistently with the findings of Bayley and Langman (2004), who showed that individual patterns of use were similar to those for the group for Type I variation than with those from Geeslin and Gudmestad (2012), who showed that for Type II variation with future-time marking there was quite a bit of variation at the level of the individual, even at higher levels of proficiency.

The findings in the current study have several important implications. For example, although it was suggested that the differences between Bayley and Langman (2004) on the one hand and Geeslin and Gudmestad (2012) on the other hand may rest on the fact that the former examines Type I variation while the latter explores Type II variation, the current study calls this into question. Were the explanation simply that developmental variation shows fewer differences at the level of the individual

than sociolinguistic variation, we would anticipate that the current study would pattern better with Geeslin and Gudmestad, at least for the advanced speakers, but that is not the case. This means that additional research on a growing range of linguistic structures is necessary in order to refine our theories about the differences encountered in findings thus far. At present, we can suggest some alternative hypotheses. One such possible source of difference is that the methods of analysis differ across studies, with each analyzing the data in slightly different ways. Bayley and Langman (2004) conduct a regression analysis for the aggregate data and then repeat this procedure for each individual, comparing the factor weights across regression tests. Thus, they show that the predictive factors for each individual are similar to those for the group as a whole. In contrast, Geeslin and Gudmestad (2012) examined the rates of selection of three forms and the distribution of these forms across the categories of a single linguistic constraint, albeit the most important one for predicting forms of future-time reference. They found less variability on the rates of selection than on the distribution across categories but on both measures they showed considerably more variability at the level of the individual than that found by Bayley and Langman (2004). Finally, in the current study, we examine the rates of selection of two response forms and the impact of a single factor called 'participant' in our predictive models. Thus, the first part of our analysis is comparable to that of Geeslin and Gudmestad (2012) in that we show the range and standard deviation for each group, but unlike their findings, which show differing standard deviations for each form, we show a general decrease in variability as proficiency increases. The second part of our analysis is more comparable to that of Bayley and Langman (2004) in that we compare predictive models with and without taking into account the individual, but our analysis provides a single snapshot of the importance of the individual while theirs allows for a qualitative comparison across models for each individual. From this brief discussion it should be clear that each analysis in the research on individual differences in L2 variation conducted thus far differs in that the method of analysis alone could be seen to contribute to the differing results across studies.

There are other factors that may also be important sources of these differences and should be considered. Firstly, Bayley and Langman (2004) examine oral interview data whereas the current study and Geeslin and Gudmestad (2012) analyze data elicited using a written task. However, if this difference were the key explanation to differing findings, we would again expect that the latter two studies would pattern together and this is contrary to what was found. A second important factor is that the work by Bayley and Langman (2004) includes two proficiency groups, the highest of which is at the intermediate level. In contrast, the other two studies contain a greater range of proficiency groups and include learners with higher levels of proficiency. While the acquisition of sociolinguistic variation is often associated with advanced learners, this difference between previous investigations cannot explain the contradictory findings under examination in the current discussion. There is one final factor that we suggest may be a better hypothesis for future research to explore. That is, each of these studies investigates different linguistic structures. Even if we limit ourselves to a discussion of Type II variation, there may be key differences between future-time expression on the one hand and subject expression on the other. Firstly, Geeslin and Gudmestad (2012) have argued that a complete analysis must include the full range of forms selected and, in so doing, they have three response options rather than only two as in the current study. They may, in turn, be examining a larger window of variation. What is more, subject expression is often noted as a case of variation that is constrained by the same factors across dialects, even though frequencies of use of overt and null forms differs from one variety to another (e.g., Cameron, 1994). The results of the current study show that despite differing rates of form selection, our learners converge on the same linguistic predictors as are used by NSs. In contrast, however, future-time expression has been found to differ across varieties of Spanish, with some groups showing differing patterns of use (see Almeida & Díaz Peralta, 1998; Blas Arroyo, 2008; Díaz Peralta & Almeida, 2000; Orozco, 2005, 2007; Sedano, 1994 for research on future-time marking, noting that the variables across studies have been operationalized differently and may not yet be comparable). This may, in fact, show an important parallel to the findings of Geeslin and Gudmestad (2012) for language learners. One final consideration of note is the extent to which social factors influence variation of the linguistic structures under discussion. Social evaluation plays a decisive role in the variation encountered for future-time expression, as suggested by the work of Orozco (2007). The role of social evaluation and social factors in general for subject expression variation may be less pronounced, which may serve as another key difference between the findings for these two linguistic structures. Although additional research on these and other grammatical structures

is necessary, of the differences across studies cited here, we believe the examination of different linguistic structures is the most promising.

There is one additional factor that merits discussion. In the review of relevant literature we mention a host of individual factors that may influence learner language, both in terms of ultimate attainment and patterns of use. In the current study we were able to rule out the impact of factors such as differing L1s, gender, age of learning, and experience abroad, but in fact, our population is quite homogeneous. This makes sense for a preliminary investigation of individual-level variation precisely because additional sources of variation such as those mentioned here would make it impossible to identify the source of linguistic variation. Nevertheless, future research that examines the effect of known influences such as age and gender on patterns of variation in learner language remains in great need. What is more, there may also be factors that do differ across our participants that have not been examined. For example, Linford (2012) provides an account of the effect of individual levels of motivation on acquisition of subject pronoun use. He found that more motivated learners (as measured by a 12-item written questionnaire) appeared to acquire sensitivity to the factors person and number of the verb as well as switch reference before less motivated learners. Ellis (2004) points out that research investigating the relationship between language aptitude and language achievement has consistently revealed fairly high positive correlations (e.g., Sparks, Ganschow, & Patton, 1995). Likewise, we note that a host of research documenting the effects of other individual difference factors such as personality (e.g., Verhoeven & Vermeer, 2002), learning styles, and learning strategies on learner language exists (cf. Dörnyei, 2006; Ellis, 2004). Thus, it may well be the case that some of the individual variation found in the analysis reported here, as well as in the other two studies to which we compare our results, can be explained through a closer look at these additional factors. To that end, future research would do well to include carefully developed measures of additional individual factors.

## 6. Conclusions and future directions

The current study was designed to explore the degree to which variation at the level of the individual is present in patterns of subject form selection across several levels of L2 learners of Spanish and a NS comparison group. In general our findings show that rates of selection converge on group norms (and NS norms) as proficiency increases. Likewise, with the exception of the graduate participant group, the degree to which individual participants impact patterns of selection decreases with increased proficiency such that in a predictive model the factor ‘participant’ shows less significance and smaller gains in predictability as proficiency increases. This finding may well parallel the results for subject use and selection across varieties of Spanish, which seem to show that speakers’ use of subject forms is constrained by the same factors, despite varying rates of use of each form from one variety to another.

These findings contrast to some degree with each of the existing studies of individual variation as compared to group norms of use/selection conducted to date. These differences may be a result of the type of variation explored (Type I vs. Type II), the elicitation methods employed, the methods of analysis and ways in which group and individual patterns are compared or differing levels of proficiency across groups, and each of these differences merits attention in future work. We propose, however, that the best explanation of these differences may actually rest in the nature of the linguistic structure itself and that only through research on additional variable structures might this be confirmed. Additionally, because the graduate level group saw a resurgence in the significance of the individual factor in the predictive model, future research should explore the degree to which this is linked to a particular linguistic or extra-linguistic factor. Finally, we suggest that future research explore individual difference factors beyond those included in the current study (e.g., motivation, language aptitude) as these are known to impact learner attainment and may also contribute to differences in patterns of language use.

## References

- Abreu, Laurel. (2012). Subject pronoun expression and priming effects among bilingual speakers of Puerto Rican Spanish. In Kimberly Geeslin & Manuel Diaz-Campos (Eds.), *Selected Proceedings of the 14<sup>th</sup> Hispanic Linguistics Symposium* (pp. 1-8). Somerville, MA: Cascadilla Proceedings Project.

- Adamson, H. Douglas. (1988). *Variation theory and second language acquisition*. Washington, D.C.: Georgetown University Press.
- Adamson, H. Douglas & Regan, Vera. (1991). The acquisition of community speech norms by Asian immigrants learning English as a second language: A preliminary study. *Studies in Second Language Acquisition*, 13, 1-22.
- 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, NJ: Lawrence Erlbaum.
- Almeida, Manuel & Díaz-Peralta, Marina. (1998). Aspectos sociolingüísticos de un cambio gramatical: La expresión del futuro. *Estudios Filológicos*, 33(7), 22.
- Amaral, Patricia & Schwenter, Scott A. (2005). Contrast and the (Non-) Occurrence of Subject Pronouns. In David Eddington (Ed.), *Selected Proceedings of the 7<sup>th</sup> Hispanic Linguistics Symposium* (pp. 116-127). Somerville, MA: Cascadilla Proceedings Project.
- Atkinson, Dwight. (2011). *Alternative approaches to second language acquisition*. London: Routledge.
- Bailey, Kathleen M. (1983). Competitiveness and anxiety in adult second language learning: Looking at and through the diary studies. In Herbert W. Seliger & Michael H. Long (Eds.), *Classroom-oriented research in second language acquisition* (pp. 67-103). Rowley, MA: Newbury House.
- Bayley, Robert, Norma L. Cárdenas, Belinda Treviño Schouten, and Carlos Martín Vélez Salas. (2012). Spanish Dialect Contact in San Antonio, Texas: An Exploratory Study. In *Selected Proceedings of the 14<sup>th</sup> Hispanic Linguistics Symposium*, ed. Kimberly Geeslin and Manuel Díaz-Campos, 48-60. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2655.
- Bayley, Robert & Langman, Juliet. (2004). Variation in the group and the individual: Evidence from second language acquisition. *International Review of Applied Linguistics in Language Teaching*, 42(4), 303-318.
- Bayley, Robert & Pease-Álvarez, Lucinda. (1996). Null and expressed pronoun variation in Mexican-descent children's Spanish. In Jennifer Arnold, Renee Blake, & Brad Davidson (Eds.). *Sociolinguistic variation: Data, theory, and analysis* (pp. 85-99). Stanford: Center for the Study of Language and Information.
- 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.
- 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.
- Blackwell, Sarah E. & Quesada, Margaret Lubbers. (2012). Third-Person Subjects in Native Speakers' and L2 learners' narratives: Testing (and revising) the Givenness Hierarchy for Spanish. In Kimberly Geeslin & Manuel Díaz-Campos (Eds.), *Selected Proceedings of the 14<sup>th</sup> Hispanic Linguistics Symposium* (pp. 142-164). Somerville, MA: Cascadilla Proceedings Project.
- Blas Arroyo, José Luis. (2008). The variable expression of future tense in Peninsular Spanish: The present (and future) of inflectional forms in the Spanish spoken in a bilingual region. *Language Variation and Change*, 20(1), 85-126.
- 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), 1-27.
- 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.
- Carroll, John B. (1991). Cognitive abilities in foreign language aptitude: Then and now. In Thomas S. Parry & Charles W. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 11-29). Englewood Cliffs, NJ: Prentice-Hall.
- Díaz-Peralta, Marina. & Almeida, Manuel. (2000). Sociolinguistic factors in grammatical change: The expression of the future in Canarian Spanish. *Studia Neophilologica*, 72(2), 217-228.
- Dörnyei, Zoltán. (2001). *Motivational strategies in the language classroom*. Cambridge: Cambridge University Press.
- Dörnyei, Zoltán. (2006). Individual differences in second language acquisition. *AILA Review*, 19, 42-68.
- Dörnyei, Zoltán. (2009). Individual differences: Interplay of learner characteristics and learning environment. *Language Learning*, 59, 230-248.
- Ellis, Rod. (2004). Individual differences in second language learning. In Alan Davies & Catherine Elders (Eds.), *The Handbook of Applied Linguistics* (pp. 525-551). Malden, MA: Blackwell Publishing Ltd.
- Emberson, Jan E. (1987). The pro-drop parameter in the acquisition of Spanish as a second language (Doctoral



- dissertation, University of Texas at Austin). *Dissertation Abstracts International*, 47, 12.
- Enríquez, Emilia V. (1984). El pronombre personal sujeto en la lengua española hablada en Madrid. Madrid: Consejo Superior de Investigaciones Científicas.
- Erker, Daniel. (2005). *Functional Compensation for Morphological Ambiguity in New York City Spanish*. (Master's Thesis) The Graduate Center of the City University of New York, New York, NY.
- Erker, Daniel and Gregory R. Guy. (2012). The role of lexical frequency in syntactic variability: Variable subject personal pronoun expression in Spanish. *Language*, 88(3), 526-557.
- Firth, Alan & Wagner, Johannes. (1997). On discourse, communication, and (some) fundamental concepts in SLA research. *The Modern Language Journal*, 81(3), 285-300.
- 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.
- Galván, 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). *Dissertation Abstracts International*, 60, 5.
- Gardner, Robert C. (1985). *Social psychology and second language learning: The role of attitude and motivation*. London: Edward Arnold.
- Gardner, Robert C. & Lambert, Wallace E. (1972). *Attitudes and motivation in second-language learning*.
- Geeslin, Kimberly. (2011). Variation in L2 Spanish: The state of the discipline. *Studies in Hispanic and Lusophone Linguistics*, 4(2), 461-517.
- Geeslin, Kimberly & Gudmestad, Aarnes. (2008). Variable subject expression in second-language Spanish: A comparison of native and non-native speakers. In Melissa Bowles et al. (Eds.), *Selected Proceedings of the 2007 Second Language Research Forum* (pp. 69-85). Somerville, MA: Cascadilla Proceedings Project.
- Geeslin, Kimberly & 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 & Gudmestad, Aarnes. (2011). Using sociolinguistic analyses of discourse-level features to expand research on L2 variation in forms of Spanish subject expression. In Luke Plonsky & Maren Schierloh (Eds.), *Selected Proceedings of the 2009 Second Language Research Forum* (pp. 16-30). Somerville, MA: Cascadilla Proceedings Project.
- Geeslin, Kimberly & Gudmestad, Aarnes. (2012, October). Individual Patterns of Verb Selection: The Case of Future-Time Expression in L2 Spanish. Paper presented at the *31st Annual Second Language Research Forum*. Pittsburgh, PA.
- 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.
- Geeslin, Kimberly & Kanwit, Matthew. (2012). The interpretation of Spanish subjunctive and indicative forms in adverbial clauses: A cross-sectional study. Unpublished manuscript. Indiana University.
- Geeslin, Kimberly & Linford, Bret. (2012). A cross-sectional study of the effects of discourse cohesiveness and perseveration on subject expression. Paper presented at the *6<sup>th</sup> International Workshop on Spanish Sociolinguistics*, Tucson, Arizona.
- Geeslin, Kimberly, Bret Linford & Stephen Fafulas. (2012). The L2 development of subject form variation in Spanish: The individual vs. the group. Paper presented at the *New Ways of Analyzing Variation*. Bloomington, Indiana.
- Gudmestad, Aarnes. (2012). Acquiring a variable structure: An interlanguage analysis of second-language mood use in Spanish. *Language Learning*, 62(2), 373-402.
- Gudmestad, Aarnes & Geeslin, Kimberly. (2013). Second-language development of variable future-time expression in Spanish. *Selected Proceedings of the 6<sup>th</sup> International Workshop on Spanish Sociolinguistics*. Somerville, MA: Cascadilla Proceedings Project.
- Guy, Gregory R. (1980). Variation in the group and the individual. In William Labov (Ed.), *Locating language in time and space* (pp. 1-36). New York: Academic.
- Hochberg, Judith G. (1986). /s/ deletion and pronoun usage in Puerto Rican Spanish. In David Sankoff (Ed.), *Diversity and Diachrony* (pp. 199-210). Amsterdam: John Benjamins.
- Horwitz, Elaine K., Horwitz, Michael B., & Cope, Joann. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, 70, 125-132.
- Howard, Martin. (2012). The Advanced Learner's Sociolinguistic Profile: On Issues of Individual Differences, Second Language Exposure Conditions, and Type of Sociolinguistic Variable. *The Modern Language Journal*, 96(1), 20-33.
- 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.
- Killam, Jason. (2012). *An Interlanguage Analysis of Differential Object Marking in L2 Spanish*. Doctoral dissertation, Indiana University.
- Labov, William. (1963). The social motivation of a sound change. *Word*, 19, 273-309.

- LaFond, Larry. (2002). The pro-drop parameter in second language acquisition revisited: A developmental account (Doctoral dissertation, University of South Carolina), *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.
- Larsen-Freeman, Diane. (2011). A complexity theory approach to second language development/acquisition. In Dwight Atkinson (Ed.), *Alternative approaches to second language acquisition* (pp. 48-72). London: Routledge.
- Liceras, 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: Cambridge University Press.
- 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 Ana Teresa Pérez-Leroux & William Glass (Eds.), *Contemporary perspectives on the acquisition of Spanish, Volume 1: Developing grammars* (pp. 99-132). Somerville, MA: Cascadilla Press.
- Linford, Bret. (2012). Individual differences and subject pronoun variation in L2 Spanish. Paper presented at the 31st Annual Second Language Research Forum. Pittsburgh, PA.
- MacIntyre, Peter D. & Gardner, Robert C. (1994). The subtle effects of induced anxiety on cognitive processing in the second language. *Language Learning*, 44, 283-305.
- Malovrh, P. (2006). L2 sentence processing of Spanish OVS Word order and direct object pronouns: An analysis of contextual constraints. In N. Sagarra & A. Toribio (Eds.), *Selected Proceedings of the 9th Hispanic Linguistics Symposium* (pp. 169-179). Somerville, MA: Cascadilla Press.
- 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 Vincent Torrens & Linda Escobar (Eds.), *The Acquisition of syntax in Romance languages* (pp. 401-418). Amsterdam: John Benjamins.
- Noels, Kimberly A., Pelletier, Luc G., Clément, Richard, & Vallerand, Robert J. (2001). Why are you learning a second language? Motivational orientations and self-determination theory. *Language Learning*, 50, 57-85.
- Orozco, Rafael. (2005). Distribution of future tense forms in Northern Colombian Spanish. In David Eddington (Ed.), *Selected Proceedings of the 7th Hispanic Linguistics Symposium* (pp. 56-65). Somerville, MA: Cascadilla Proceedings Project.
- Orozco, Rafael. (2007). Social constraints on the expression of futurity in Spanish-speaking urban communities. In Jonathan Holmquist et al. (Eds.), *Selected Proceedings of the Third Workshop on Spanish Sociolinguistics* (pp. 103-112). Somerville, MA: Cascadilla Proceedings Project.
- Otheguy, Ricardo & Zentella, Ana Cecilia. (2007). Apuntes preliminares sobre el contacto lingüístico y dialectal en el uso pronominal del español en Nueva York. In Richard Cameron & Kim Potowski (Eds.), *Spanish in contact: Policy, social and linguistic inquiries* (pp. 275-95). Amsterdam: John Benjamins.
- Otheguy, Ricardo & Zentella, Ana Celia. (2012). *Spanish in New York: Language contact, dialectal leveling, and structural continuity*. Oxford: Oxford University Press.
- Otheguy, Ricardo, Zentella, Ana Celia, & Livert, David. (2007). Language and dialect contact in Spanish in New York: Toward the formation of a speech community. *Language*, 83(4), 770-802.
- 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.
- Quesada, Margaret Lubbers & Blackwell, Sarah E. (2009). The L2 acquisition of null and overt Spanish subject pronouns: A pragmatic approach. In Joseph Collentine, Maryellen Garcia, Barbara Lafford, & Francisco Marcos Martín (Eds.), *Selected Proceedings of the 11th Hispanic Linguistics Symposium* (pp. 117-130). Somerville, MA: Cascadilla Proceedings Project.
- Robinson, Peter. (1997). Individual differences and the fundamental similarity of implicit and explicit adult language learning. *Language Learning*, 47, 45-99.
- Rothman, Jason. (2007). Pragmatic solutions for syntactic problems: Understanding some L2 syntactic errors in terms of discourse-pragmatic deficits. In Sergio Baauw, Frank Drijkoningen, & Manuela Pinto (Eds.), *Romance Languages and Linguistic Theory 2005: Selected papers from 'Going Romance'* (pp. 299-320). Amsterdam: John Benjamins.
- Sedano, Mercedes. (1994). El futuro morfológico y la expresión ir a + infinitivo en el español hablado de Venezuela. *Verba*, 21, 225-240.
- Shin, Naomi Lapidus & Otheguy, Ricardo. (2009). Shifting sensitivity to Continuity of reference: Subject pronoun use in Spanish in New York City. In Manel Lacorte & Jennifer Leeman (Eds.), *Español en Estados Unidos y en otros contextos: Cuestiones sociolingüísticas, políticas y pedagógicas* (pp. 111-136). Madrid: Iberoamericana.

- Shin, Naomi Lapidus. (2012). Variable use of Spanish subject pronouns by monolingual children in Mexico. In Kimberly Geeslin & Manuel Díaz-Campos (Eds.), *Selected Proceedings of the 14<sup>th</sup> Hispanic Linguistics Symposium* (pp. 130-141). Somerville, MA: Cascadilla Proceedings Project.
- Silva-Corvalán, Carmen. (1994). *Language Contact and Change: Spanish in Los Angeles*. Oxford: Oxford University Press.
- Skehan, Peter. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.
- Skehan, Peter. (2002). Theorising and updating aptitude. In Peter Robinson (Ed.), *Individual differences and instructed language learning* (pp. 69-93). Amsterdam: John Benjamins.
- Sparks, Richard L., Ganschow, Leonore, & Patton, Jon. (1995). Prediction of performance in first-year foreign language courses: Connections between native and foreign language learning. *Journal of Educational Psychology*, 87, 638-655.
- Sternberg, Robert J. (2002). The theory of successful intelligence and its implication for language aptitude-testing. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 13-43). Amsterdam: John Benjamins.
- Tarone, Elaine & Liu, Guo-Qiang. (1995). Situational context, variation, and second language acquisition theory. In Guy Cook & Barbara Seidlholfer (Eds.), *Principle and Practice in Applied Linguistics: Studies in Honour of H. G. Widdowson* (pp. 107-124). Oxford: Oxford University Press.
- Torres Cacoullous, Rena, & Travis, Catherine E. (2011). Testing convergence via code-switching: Priming and the structure of variable subject expression. *International Journal of Bilingualism*, 15(3), 241-267.
- Travis, Catherine E. (2005). The yo-yo effect: Priming in subject expression in Colombian Spanish. In Randall Gess & Edward J. Rubin (eds), *Theoretical and experimental approaches to Romance linguistics: Selected papers from the 34th Linguistic Symposium on Romance Languages 2004* (pp. 329-349). Amsterdam/Philadelphia: John Benjamins.
- Travis, Catherine E. (2007). Genre effects on subject expression in Spanish: Priming in narrative and conversation. *Language Variation and Change*, 19(2), 101-135.
- Verhoeven, Ludo & Vermeer, Anne. (2002). Communicative competence and personality dimensions in first and second language learners. *Applied Psycholinguistics*, 23, 361-374.
- White, Lydia. (1985). The "pro-drop" parameter in adult second language acquisition. *Language Learning*, 35(1), 47-62.
- Young, Richard. (1991). *Variation in interlanguage morphology*. New York: P. Lang.

# Selected Proceedings of the 16th Hispanic Linguistics Symposium

edited by Jennifer Cabrelli Amaro,  
Gillian Lord, Ana de Prada Pérez,  
and Jessi Elana Aaron

Cascadilla Proceedings Project Somerville, MA 2013

## Copyright information

Selected Proceedings of the 16th Hispanic Linguistics Symposium  
© 2013 Cascadilla Proceedings Project, Somerville, MA. All rights reserved

ISBN 978-1-57473-459-1 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, Bret Linford, Stephen Fafulas, Avizia Long, and Manuel Díaz-Campos. 2013. The L2 Development of Subject Form Variation in Spanish: The Individual vs. the Group. In *Selected Proceedings of the 16th Hispanic Linguistics Symposium*, ed. Jennifer Cabrelli Amaro et al., 156-174. Somerville, MA: Cascadilla Proceedings Project. [www.lingref.com](http://www.lingref.com), document #2932.