Exploring the Roles of Redundancy and Ambiguity in Variable Subject Expression: A Comparison of Native and Non-native Speakers

Aarnes Gudmestad and Kimberly L. Geeslin
Virginia Polytechnic Institute and State University and Indiana University

Historically, our knowledge of the process of learning to vary language use like a native speaker (NS) has been based primarily on research on second-language (L2) French (e.g., Mougeon & Dewaele, 2004; Mougeon, Rehner & Nadasdi, 2004) and English (e.g., Adamson & Regan, 1991; Bayley & Preston, 1996; Young, 1991). Nevertheless, studies on variation in L2 Spanish have recently appeared and have shed light on variation in copula selection, mood use and variation across tasks (Geeslin, 2003; Geeslin & Gudmestad, 2008a; Geeslin & Guijarro-Fuentes, 2006; Gudmestad, 2008). What these studies have in common is that they indicate that the second language acquisition (SLA) of variation, as well as contrasts between NSs and non-native speakers (NNSs), can be described by the frequency of selection of a variant and the linguistic and social factors that condition the use of that variant. Thus, the process of SLA involves gaining the ability to identify those contexts where variation is permissible along with the ability to use such variants with the appropriate frequency and in response to the appropriate conditioning factors (linguistic and/or social).[1] The overarching purpose of the present investigation is to further this body of research on L2 variation.

In order to advance our knowledge of the SLA of variation, the current study focuses on subject expression in Spanish. Variable subject expression provides a rich avenue for better understanding and explaining L2 variation because, unlike the aforementioned studies of Spanish, it has already been studied extensively using contemporary sociolinguistic analyses. This existing research on subject expression among NSs has allowed us to delve more deeply into the effects of specific variables for NSs and NNSs, rather than beginning with a search for relevant predictive factors. In addition, our research differs from most existing research in two important ways, each of which allows us to make additional contributions to the field. Firstly, it has been noted that the 'envelope of variation' (e.g., Cameron, 1994) for subject expression, or the distinction between categorical and variable contexts, is not quantitatively documented for NSs and that this issue may even be more complex for learners because interlanguage can exhibit variability where a first language system does not (Geeslin & Gudmestad, 2008b; Otheguy & Zentella 2007).[2] Therefore, examining subject expression in the full range of tensed verbal contexts and its relationship to individual linguistic factors allows us to examine subject expression choices in all contexts where variability may exist and furthers our knowledge of the 'envelope of variation' because it enables us to empirically identify obligatory contexts for NSs and NNSs in any category of the independent variables we examine. Secondly, whereas most previous sociolinguistic investigations have focused on animate subjects whose subject-expression forms were null or personal pronouns (e.g., Bayley & Pease-Álvarez, 1997; Cameron, 1994; Comajoan, 2006; Flores-Ferrán, 2004), we broaden the scope of our analysis to include all forms produced in the subject position of tensed verbs. This approach is especially appropriate when dealing with language learners

1 The issue of whether or not this meets the traditional definition of 'sociolinguistic variation' (i.e., variation between two variants with equal semantic meaning) is rarely addressed in the L2 literature. We acknowledge, however, that the extension of such terminology beyond the level of phonology remains controversial.

2 Otheguy, Zentella and Livert (2007) present a detailed list of contexts that are included (and excluded) from the envelope of variation and add that their methods avoid unprofitable discussions about the frequency and (un)grammaticality of occurrences of variation outside this envelope. We agree that their methods are principled and that they are not limited to a list of lexical items, but we argue that there are likely to be additional contexts worthy of study that have been excluded from their analysis because obligatory contexts are defined a priori.

who may also demonstrate variability in their use of forms such as the full lexical noun phrase (NP), even though most sociolinguistic research has not explored whether NSs show variation between lexical NPs and pronouns. Expanding the dependent variable of subject expression from a binary one (i.e., null vs. personal pronoun) to a multinomial one (including lexical NPs and 'other' pronouns) means a possible divergence from the binary regression analyses that have become a hallmark of variationist investigations. In order to ascertain how multiple factors conditioning subject expression can best be analyzed when the dependent variable is expanded from the null/personal pronoun distinction, detailed examinations of subsets of data are a necessary precursor to developing new ways of describing and predicting language variation. Only once this has been completed is it possible to explore the use of more advanced multi-level predictive models. We begin this process by performing a closer look at subject expression in the context of each of the categories of the three independent variables examined in our study: tense/mood/aspect (TMA) of the verb, verb-form ambiguity, and discourse redundancy. In so doing, we take major strides toward understanding the pervasiveness of the effects of TMA across sociolinguistic studies while at the same time providing a methodological model that does not require an a priori definition of the envelope of variation nor the assumption that variation in subject expression only occurs between two pronominal forms. Thus, we are able to move the larger research program forward while simultaneously providing the first exploration of the effects of these factors on variable subject expression that addresses the subtle similarities and differences between NSs and advanced NNSs of Spanish.

1. Background

1.1. Subject expression in Spanish

The dependent variable in the current study is the form used to express the subject of a tensed verb. In Spanish, such forms include lexical NPs, overt pronouns of various types (e.g., subject, demonstrative and interrogative pronouns), or null subjects. In the latter category the subject of the verb is understood through information provided by the tensed verb itself as well as other features in the discourse, rather than an overt subject. Because there has been little research on the use of lexical NPs among NSs (see Silva-Corvalán, 1994 for an exception), the bulk of existing research focuses on the factors that condition the contrast between personal pronouns and null subjects in contexts that permit such variation. Specifically, most sociolinguistic studies have examined the factors that condition the variation between personal pronouns and null subjects in contexts where the subject is animate and the alternation between these two forms is permissible. Within such contexts, a great deal of information is available about the linguistic factors related to pronoun use (Bayley & Pease-Álvarez, 1997; Bentivoglio, 1987, 1988; Cameron, 1993, 1994; Comajoan, 2006; Davidson, 1996; Flores-Ferrán, 2004, 2005; Otheguy, et al., 2007; Serrano, 1996; Silva-Corvalán, 1982). The review that follows summarizes what is known about subject pronoun use in Spanish, from a sociolinguistic point of view, with a particular focus on the variables in the current study: ambiguity and discourse redundancy.

In early work on subject pronoun expression, Silva-Corvalán (1982) found that personal pronouns were most likely to occur when new information was presented, in contexts where a contrast was expressed, when new topics of discussion were established and when the verb form was one that might correspond to more than one person and number. This last variable was what Silva-Corvalán called 'ambiguity' and it contained three categories: forms that were never ambiguous, forms that could be ambiguous but were not so in a given context and forms that were ambiguous in the context in which they occurred. She found that personal pronouns were expressed 69 percent of the time when the form was ambiguous, 39 percent of the time when the form was potentially ambiguous but was not ambiguous in the context in which it occurred and 38 percent of the time when the form had no potential to be ambiguous. A chi-square test showed that these results were significantly related to variability in subject-form expression and the VARBRUL results for the absence of a personal pronoun showed that ambiguous forms disfavored null subjects, contextually-unambiguous forms slightly favored null subjects and consistently-unambiguous forms favored null subjects. Thus, this early work indicates that ambiguity is an important predictor and that the potential of a form to be ambiguous and the context in which such forms occur both play a role in predicting subject expression. In a related study Bentivoglio (1987) distinguished between morphological ambiguity and contextual ambiguity. The former is a feature of the verb that identifies its potential for ambiguity whereas the

latter refers to whether or not the form is actually ambiguous in the discourse context in which it appears. She found that contexts where the form had the potential to be ambiguous demonstrated a 50 percent use of the null subject whereas this use dropped to 25 percent in contexts which were ambiguous. In keeping with the close examination of the ways in which a form might be ambiguous, Bayley and Pease-Álvarez (1997) demonstrated that the degree of cohesiveness of a narrative, contexts of switch reference, the person and number of the verb, the surface ambiguity of the verb, and the verb type were all significant predictors of personal pronoun use. It was found that the strongest factor in the model was the person and number of the verb. The surface ambiguity of the verb which distinguished the first and third person singular forms of the imperfect, the conditional and the subjunctive (i.e., those which create the potential for ambiguity) from all others was also significant. It was found that the ambiguous forms favored the use of personal pronouns. A third variable, verb type, distinguished the preterit and *ser* in the present indicative, which are not ambiguous in any variety of Spanish, from the present indicative, which can be ambiguous in some varieties of Spanish (i.e., those with /s/-weakening), from the imperfect, conditional and subjunctive (which have the potential to be ambiguous in any dialect). The results for this variable showed that the imperfect, conditional and subjunctive favored the use of personal pronouns and that the present indicative favored null subjects more than the preterit. Finally, one particularly interesting finding is that in comparing different runs of statistical analyses (to avoid including two overlapping variables related to the ambiguity of the verb form in a single run), the TMA of a verb seemed to explain more than the surface ambiguity. This was true even despite the fact that the relationship between the present and the preterit was counter to expectations, because the person and number ordering held across all verb types. Cameron (1993, 1994) expanded this work by focusing on the role of functional compensation in personal pronoun expression. Like Bayley and Pease-Álvarez (1997), he distinguished surface ambiguity of the verb form using three categories that corresponded to forms that could be ambiguous in any variety of Spanish, those that were only ambiguous in /s/-weakening varieties of Spanish and those that were never ambiguous. He found that the effect of this variable was the same across varieties of Spanish (those with and without /s/-weakening) and that only the rates of null subject use were different. Most interestingly from the point of view of the current study is that the effect of verb type, as he called the aforementioned variable, disappeared in contexts of same reference. In connecting these studies, one notes that the effect for the TMA variable is quite strong across studies. The authors of these studies all connect this effect to various forms of ambiguity and redundancy. Nevertheless, these two concepts are defined in a variety of ways, ranging from inherent features of verbal morphology (e.g., the potential of a form to permit ambiguity) to features of the discourse (e.g., switch reference), and each appears to play some role in explaining the variability seen across categories of the TMA variable. We will further discuss this issue following a brief review of the use of subject pronouns in L2 Spanish.

1.2. Subject expression in second-language Spanish

The largest body of research on the acquisition of subject expression in Spanish focuses primarily on the null subject parameter and the properties associated with it (e.g., Al-Kasey & Pérez-Leroux 1998; Bini, 1993; Emberson, 1987; Galván, 1999; Isabelli, 2004; Liceras, 1989; Liceras, Maxwell, Laguardia, Fernandez & Díaz, 1997; Phinney, 1987; White, 1985). Although these studies generally focus on whether or not a learner's grammar allows null subjects, some work informed by generative and optimality theory has shown that discourse-pragmatic features do play a role in the L2 acquisition of subject expression (e.g., LaFond, 2002; Lafond, Hayes & Bhatt, 2000; Montrul & Rodríguez Louro, 2006; Rothman, 2007). Nevertheless, to our knowledge, ours is the first series of SLA studies aimed at addressing the use of subject expression forms and its relationship to the linguistic factors known to condition variable language use for NSs. In Geeslin and Gudmestad (in press) we conducted a large-scale analysis of five different potential contexts of variation based on interview data from 32 participants, half of whom were English-speaking learners of Spanish and the other half of whom were NSs. For each of the five grammatical structures under examination, we assessed the distribution of the various forms used in that context and compared this distribution across our two participant groups. For subject expression, each token was defined as the subject position of any tensed verb and we found that both groups produced lexical NPs, null subjects, and personal, indefinite, interrogative and demonstrative pronouns. The frequency of production of each form was significantly different between groups and these differences were not the result of a single category of that variable. One of

the greatest innovations in that study was the way in which we defined a token, leading to the examination of subject expression well beyond the animate subject of a tensed verb and also to a broader range of categories in the dependent variable than was included in most previous sociolinguistic research. In a second study using the same participant pool (Geeslin & Gudmestad, 2008b), we examined the person and number of the verb form and the specificity of the referent. Both factors were significantly related to subject expression for both groups, and the intersection of these two variables provided qualitative descriptions of the differences between NSs and NNSs. These earlier studies move the study of the use of subject expression forms forward but they also identify work left to be done.

2. The current study

Taking into account the previous research in both sociolinguistics and SLA there are several insights that guided the design of the current study. Firstly, the forms of subject expression appear to have differing distributions across categories of the TMA variable for NSs of Spanish and it is clear from the published reports of these studies as well as the other variables employed that researchers believe this effect is related in some way to ambiguity. Nevertheless, ambiguity has been examined in numerous ways, leading to the study of surface ambiguity, form ambiguity, context ambiguity and the like, making it difficult to draw conclusions across studies. We hypothesize that the effect of the TMA variable may be related to the resolution of ambiguity but it may also have other explanations. The goal of the current study is to examine the relationship of the TMA variable to two types of ambiguity: the potential ambiguity of the verb form and ambiguity in the discourse context (henceforth discourse redundancy) as indicated by the contrast between same and switch reference. In addition to asking whether or not these two variables are useful in describing subject expression for our data, we ask whether these factors may also lead to a greater understanding of the similarities and differences between NSs and NNSs. Finally, if it is the case that these factors are not related to the effects of the TMA variable, we seek to propose additional hypotheses to be tested in future studies. In keeping with these goals, the current study was guided by the following research questions:

1. What is the relationship of the TMA of the verb to the forms for subject expression produced by NSs and NNSs?
2. Do the differences found across the categories of the TMA variable correspond to the potential ambiguity of the verb form for one or both participant groups?
3. Do the differences found across the categories of the TMA variable correspond to switch reference for one or both participant groups?

Our study is different from most previous research in that we do not limit our study to animate referents and we include all subject forms, rather than focusing exclusively on variation between the null subjects and personal pronouns. In addition, we use the person/number variable to address form ambiguity, rather than a judgment about what is or is not ambiguous in the discourse. Finally, we do not look at the interaction between the two types of ambiguity examined here (form ambiguity and discourse redundancy as indicated by same reference), although we believe this is likely to be a reasonable follow-up to the research reported here.

2.1. Participants

The NNS and NS participants were graduate students living in the United States at the time of data collection. They were also instructors of undergraduate Spanish classes or had been in the past. The NNSs were English-speaking learners of Spanish (N=16).[3] They ranged in age from 22 to 35 years (mean=26.4); ten were male and 6 were female. Their scores on a 25-item language proficiency test (to be discussed in the following section) ranged from 16 to 25 (mean=20.8). All NNSs had spent time abroad in at least one of the following Spanish-speaking countries: Argentina, Chile, Colombia, Costa Rica, Ecuador, Mexico, Nicaragua, and Spain. This time ranged from three months to 11 years

[3] We recognize that analyzing one group of NNSs does not permit us to make observations about development.

(mean=18.1 months). One participant had spent 11 years abroad, with the next longest stay being 27 months. When the mean was calculated without this speaker it equaled 11.2 months. The NNSs had studied Spanish formally for 3 to 17 years (mean=8.07 years). Our NNSs were selected because they are deemed to have "near-native proficiency" according to institutional hiring practices and we anticipated that, although this group is often treated as homogeneous, it would contain a variety of language learning experiences and abilities. Thus, we are careful to document these differences here and through the use of the proficiency test in recognition of this diversity. Clearly, individual variation is also an important fact of this dataset and worthy of further exploration in future research. The NSs (N=16) were from Argentina, Chile, Colombia, Ecuador, Mexico, Spain, the United States (Puerto Rico and the Southwest), and Uruguay, were evenly divided between genders, and ranged in age from 24 to 37 years (mean=29.8 years). Their scores on the language proficiency test ranged from 22 to 25 (mean=23.6).

2.2. Elicitation tasks

Each participant completed three tasks: a background questionnaire, a language proficiency test, and a sociolinguistic interview. The background questionnaire contained short-response questions through which the participants provided the information summarized previously. The language proficiency test was a 25-item, multiple-choice task in which items contextualized in a story covered a range of grammatical concepts in Spanish (see Appendix A for sample items). This task served to objectively compare the two participant groups apart from the comparisons made about their use of subject expression variable in the present study. The third task was a digitally-recorded, 30-minute sociolinguistic interview conducted by two NSs of Spanish (one female and one male, from Latin America and Spain). All speakers were asked the same interview questions, which covered a range of topics including opinion questions, plans for the future, and recent and past experiences (see Appendix A for sample items).

2.3. Data coding and analysis

The dependent variable in the current study was the form used to express the subject of every finite verb produced in the interview. The categories included null (e.g., *ø habla* 'ø speaks'), lexical NP (e.g., *Clara habla* 'Clara speaks'), personal pronoun (e.g., *ella habla* 'she speaks'), demonstrative pronoun (e.g., *ésta habla* 'this one speaks'), indefinite pronoun (e.g., *alguien habla* 'someone speaks'), and interrogative pronoun (e.g., *¿quién habla?* 'who speaks?'). Contexts, such as the impersonal/passive *se* and subject-headed relative clauses, which were found to be categorically null for all speakers in previous analyses of this same dataset, were excluded from the current analysis (Geeslin & Gudmestad, in press).[4] Each token of subject expression was coded for three independent linguistic variables, TMA of the verb, potential form ambiguity, and switch reference, and one independent non-linguistic variable, speaker group. TMA of the verb initially consisted of all TMA forms that the speakers produced in the data sample. Because some TMA forms were produced infrequently, the current analysis was limited to TMA forms that each participant group produced at least 50 times in the interview.[5] The seven categories that met this condition were the conditional, imperfect, periphrastic future, present indicative, present perfect, preterit, and present subjunctive. The potential form ambiguity variable distinguished between finite verb forms that only correspond to one person and those that correspond to more than one person. Potentially ambiguous forms for all speakers were first and third-person singular verbs and contrastive forms were second-person singular,

[4] By identifying these contexts as categorically null in the current dataset, we have begun the process of defining the envelope of variation for these speakers through an objective analysis of the data themselves. Moreover, as the investigation of the relationship between subject expression and independent variables continues, it is possible that other categorical contexts will emerge from the data. For example, if a subject were always null in a context of switch reference, then this independent linguistic factor would serve to further delineate empirically the scope of categoricity for these participants.

[5] This decision was made to address the fact that each participant group did not produce sufficient examples of certain TMA forms for proper statistical analysis.

first-person plural, and third-person plural verbs.[6] The switch reference variable, which is one way of measuring redundancy in the discourse, identified whether the subject of the preceding finite verb was the same as the subject of the current token. Switch tokens were those for which the subject of the preceding finite verb was not the same as the current token, and same tokens were those for which the subject was the same for both. Tokens were coded by one researcher, using objective definitions for all variables and their categories. While these factors are motivated by previous research, the review of the literature demonstrated that many have not been investigated uniformly across studies. As a result, we sought to define them in such a way that they can be objectively coded and replicated by other researchers. The procedure employed by researchers in some studies of discussing any questionable items until consensus was reached was unnecessary because no such cases arose (i.e., the variables are objectively defined). Finally, the independent variable of speaker group allowed us to analyze NS and NNS speakers of Spanish separately.

The analysis consisted of a series of cross-tabulations and chi-square (X^2) tests. We began by comparing the distribution and frequency of subject expression forms used by NSs and NNSs (reported previously in Geeslin & Gudmestad, in press). Secondly, we conducted a cross-tabulation and a X^2-test comparing the distribution of TMA forms across the two participant groups. Thirdly, a cross-tabulation and a X^2-test for the TMA of the verb and the subject forms used were carried out for each participant group separately. Fourthly, we conducted a cross-tabulation and a X^2-test for the potential for form ambiguity and the subject form used within each TMA category (NSs and NNSs analyzed separately). This analysis allowed us to compare TMA categories in which there was a contrast between potentially-ambiguous forms and forms that had no such inherent ambiguity to TMA categories where no forms were potentially ambiguous. If the former showed a significant difference between the two categories of the potential form ambiguity variable but the latter did not, this could be taken as evidence of the importance of form ambiguity in explaining the TMA effects. The final cross-tabulation and X^2-test set analyzed the TMA of the verb and the subject form used within the contexts of same reference and within the contexts of switch reference (NS and NNS groups analyzed separately). If this analysis demonstrated no TMA effects within contexts of same reference but significant differences in contexts of switch reference, there was evidence of the importance of discourse redundancy.

3. Results

The presentation of the results begins with the distribution of subject expression forms used by NNSs and NSs. It will be recalled that subject forms used by the participants included null, lexical NP, and personal, demonstrative, indefinite, and interrogative pronouns and the frequency of use of each form was previously reported in Geeslin and Gudmestad (in press). In the current study, demonstrative, indefinite, and interrogative pronouns have been collapsed into one category called other pronouns because both speaker groups used these subject forms relatively infrequently. Whether all subject expression forms were analyzed separately (see Geeslin & Gudmestad, in press) or four categories are analyzed as they are in the current study, the distribution of subject expression forms is significantly different between NNSs and NSs (X^2=73.897, df=3, $p<0.001$, Cramer's V=0.076). As seen in Table 1, the largest differences between groups are seen in the categories of null and other pronouns. NNSs used more null subjects and fewer other pronouns, whereas the rates of production of lexical NPs and personal pronouns were similar.

[6] *Vosotros*, *usted* and *ustedes* (second-person subjects) forms were excluded from the analysis because the participants used them infrequently in the interview. The infrequent use of these forms was not unexpected due to the interview format.

Table 1. Distribution of Subject Expression Forms

Form	NNSs		NSs	
	#	%	#	%
Null	3792	70.2	5150	67.6
Lexical NP	770	14.3	1042	13.7
Personal pronoun	589	10.9	793	10.4
Other pronoun	249	4.6	637	8.4
Total	5400	100	7622	100

Note. X^2=73.897, df=3, p<0.001, Cramer's V=0.076

Just as the distribution of subject expression forms is significantly different between groups, the distribution of tokens across TMA categories is also significantly different between NNSs and NSs (X^2=1.751E^2, df=6, p<0.001, Cramer's V=0.118). Table 2 illustrates that while the rates of use of the conditional, imperfect, periphrastic future, preterit and present subjunctive were similar, larger differences were observed with the present indicative and the present subjunctive. NNSs used a larger proportion of present indicative and a smaller proportion of the present subjunctive than the NSs.

Table 2. Distribution of TMA Forms

Form	NNSs		NSs	
	#	%	#	%
Conditional	183	3.4	249	3.4
Imperfect	228	4.3	432	5.9
Periphrastic future	154	2.9	173	2.4
Present indicative	4270	80.2	5405	73.5
Present perfect	111	2.1	251	3.4
Preterit	310	5.8	484	6.6
Present subjunctive	71	1.3	362	4.9
Total	5372	100	7354	100

Note. X^2=1.751E^2, df=6, p<0.001, Cramer's V=0.118

Table 3. Distribution of Subject Expression According to TMA for NNSs

Form	Null		Personal pronoun		Lexical NP		Other pronoun		Total
	#	%	#	%	#	%	#	%	
Conditional	94	51.4	26	14.2	61	33.3	2	1.1	183
Imperfect	157	68.9	47	20.6	18	7.9	6	2.6	228
Periphrastic future	119	77.3	15	9.7	13	8.4	7	4.5	154
Present indicative	3012	70.5	430	10.1	611	14.3	217	5.1	4270
Present perfect	86	77.5	11	9.9	12	10.8	2	1.8	111
Preterit	219	70.6	45	14.5	37	11.9	9	2.9	310
Present subjunctive	44	52.0	8	11.3	13	18.3	6	8.5	71

Note. X^2=1.143E^2, df=18, p<0.001, Cramer's V=0.085, 1 small cell

Table 4. Distribution of Subject Expression According to TMA for NSs

Form	Null		Personal pronoun		Lexical NP		Other pronoun		Total
	#	%	#	%	#	%	#	%	
Conditional	141	56.6	32	12.9	52	20.9	24	9.6	249
Imperfect	288	66.7	76	17.6	41	9.5	27	6.2	432
Periphrastic future	136	78.6	6	3.5	20	11.6	11	6.4	173
Present indicative	3612	66.8	519	9.6	781	14.4	493	9.1	5405
Present perfect	180	71.7	33	13.1	25	10.0	13	5.2	251
Preterit	335	69.2	64	13.2	50	10.3	35	7.2	484
Present subjunctive	245	67.7	31	8.6	56	15.5	30	8.3	362

Note. X^2=84.580, df=18, p<0.001, Cramer's V=0.062

Continuing with the analysis of TMA of the verb, Tables 3 and 4 above show the distribution of

subject expression forms across categories of the TMA variable for NNSs and NSs, respectively. The relationship between TMA of the verb and the dependent variable is significant for each group (NNSs: X^2=1.143E^2, df=18, p<0.001, Cramer's V=0.085, 1 small cell and NSs: X^2=84.580, df=18, p<0.001, Cramer's V=0.062). When the distributions for each group are compared, many similarities, as well as some differences, between NNSs and NSs are observed. Within the category of the conditional and present subjunctive, both groups used null subjects most often, followed by lexical NPs personal pronouns and, finally, other pronouns, although the rate of use of personal and other pronouns within the category of present subjunctive for the NSs was similar (0.3% difference). Within the imperfect and preterit, NNSs and NSs used null subjects most frequently, followed by personal pronouns, lexical NPs, and, lastly, other pronouns. Within the periphrastic future, null subjects were the most common subject-expression form. However, the remaining forms were produced at a different proportion between groups. NNSs used personal pronouns and lexical NPs at a similar rate, followed by other pronouns, and the NSs produced lexical NPs more often than other pronouns and, lastly, personal pronouns. Within the present perfect the two participant groups used null subjects most frequently and other pronouns least frequently. A difference was observed between the personal pronouns and lexical NPs, in which the NNSs used the two forms at a similar rate and the NSs produced more personal pronouns than lexical NPs.

Thus far, our analysis has confirmed two important findings. Firstly, as with our previous research, there are striking similarities and differences in the distribution of forms and of TMA categories between NSs and NNSs. Secondly, the variable TMA is significantly related to the distribution of subject expression forms for both groups, and this is consistent with previous sociolinguistic research. The remainder of our analysis explores two potential explanations for those TMA effects.

In order to investigate the role that potential form ambiguity may play in explaining the effects of TMA of the verb on subject expression, it will be recalled that within each TMA category (e.g., conditional), the relationship between the potential form ambiguity and subject expression variables was examined through individual X^2-tests. The results of these X^2-test are summarized in Table 5 (see Appendix B for cross-tabulation results). The conditional, imperfect, and present subjunctive have ambiguous verb forms in the first and third-person singular forms, whereas the first and third-person singular forms of the periphrastic future, present indicative, present perfect, and preterit are contrastive forms. Thus, in order to conclude that the potential form ambiguity variable helps to explain the effects of TMA of the verb, there should only be a significant relationship between potential form ambiguity and subject expression for the conditional, imperfect, and present subjunctive, since the potential form ambiguity variable can only explain differences between ambiguous and contrastive forms with TMA categories that have ambiguous forms in the first and third-person singular. These results were not borne out in the data. For each speaker group, both TMA categories that have potentially ambiguous forms and those that do not were shown to be significantly related to subject expression. In other words, potential form ambiguity does not appear to explain the effects of TMA.

Table 5. Relationship between Potential Form Ambiguity and Subject Expression within TMA Categories

TMA category	NNSs	NSs
Conditional	X	
Imperfect	X	X
Periphrastic future		
Present indicative	X	X
Present perfect	X	
Preterit		
Present subjunctive		X

Note. An 'X' denotes forms for which there is a significant difference between potentially ambiguous and contrastive forms.

The final component of the analysis examined the relationship (i.e., X^2-tests) between TMA of the verb and subject form used, with contexts of same and switch reference analyzed separately in order to determine whether or not switch reference, as one measure of discourse redundancy, explains some of the effects of TMA. For both NNSs and NSs, the distribution of forms of subject expression was significantly different across the categories of the TMA variable in contexts of switch reference (NNSs: X^2=90.389, df=18, p<0.001, Cramer's V=0.090, 3 small cells and NSs: X^2=69.836, df=18,

p<0.001, Cramer's V=0.120) but not same reference (NNSs: X^2=22.431, df=18, p=0.213, Cramer's V=0.118, 15 small cells and NSs: X^2=28.799, df=18, p=0.051, Cramer's V=0.108, 10 small cells). With regard to contexts of same reference for both NNSs and NSs, the null subject was clearly the dominant form within each category of TMA. NNSs used null subjects between 79.0 percent of the time (83/105 contexts) with the imperfect and 91.3 percent of the time (42/46 contexts) with the conditional, and NSs' use of the null subject ranged from 80.6 percent of the time (79/98 contexts) with the present perfect to 93.3 percent of the time (70/75 contexts) with the periphrastic future. In contrast to contexts of same reference, in contexts of switch reference the null subject was produced less than 70.0 percent of the time with each TMA form. NNSs produced the null subject between 38.0 percent of the time (52/137 contexts) with the conditional and 69.6 percent of the time (48/69 contexts) with the present perfect. NSs' use of the null subject ranged from 46.0 percent of the time (80/174 contexts) with the conditional to 67.3 percent of the time (66/98 contexts) with the periphrastic future. Furthermore, while the difference in use of the null subject across TMA categories in contexts of same reference was 12.3 percent for the NNSs and 12.7 percent for the NSs, the range of use was larger across TMA categories in contexts of same reference: 31.6 percent for the NNSs and 21.3 percent for the NSs. Thus, not only did the range of use of the null subject vary to a greater extent across TMA categories in contexts of switch reference than in contexts of same reference, but the null subject was used less often (and, in turn, the lexical NPs and personal and 'other' pronouns were used more frequently). This finding illustrates that discourse redundancy helps to explain the relationship between TMA and subject expression use because in redundant contexts (i.e., those of same reference) the effect of TMA categories disappears.

4. Discussion

The current study examined the relationship between TMA of the verb and subject expression for NNSs and NSs and whether or not the effect for TMA corresponded to potential form ambiguity and switch reference variables. Firstly, we showed that the TMA of the verb is significantly related to subject expression for these NNSs and NSs. Each TMA category showed a different distribution of subject-form use. Although we observed differences in subject-expression use across the seven TMA categories within each speaker group, NNSs and NSs exhibited more similarities to each other than differences. The distributions of subject forms across the conditional, present subjunctive, imperfect and preterit were alike between groups, while the periphrastic future and the present perfect differed. In other words, the NNSs in the present investigation show relatively native-like distributions of the subject expression forms across various TMA categories. Secondly, the data showed uneven effects for potential form ambiguity across the categories of TMA for both participant groups. Because significant differences in form distribution did not disappear in contexts where form ambiguity was absent, this variable was not useful in explaining subject expression for our NNS and NS populations. Finally, the results revealed evidence for the importance of discourse redundancy for both speaker groups. We observed TMA effects for contexts of switch reference but not for same reference, which means that in redundant discourse contexts there is no significant relationship between TMA and subject expression (cf. Cameron, 1994). Thus, the NNSs and NSs are similar in that discourse redundancy helped to explain the effects of TMA on subject expression but potential form ambiguity did not.

In light of the results for the potential form ambiguity and switch reference variables, it becomes clear that there must be other factors that explain the TMA effects on subject expression. Although no effect was found for the potential form ambiguity variable, it may be that form ambiguity is related to subject expression in another way. For example, if first and third-person singular forms can be potentially ambiguous in some TMA categories, it may be possible that they are more likely to be marked with an overt subject (pronoun or lexical noun phrase) in any TMA category. Since switch reference, as one measure of discourse redundancy, explains some of the TMA effects, it seems reasonable to hypothesize that other features of the discourse are also influential. For example, certain TMA categories (e.g., present subjunctive) may appear more frequently in contexts of switch reference, or they may occur more often in contexts where multiple subjects (e.g., characters in narratives) are at play (e.g., preterit and imperfect in narratives) and these issues may lead to further discoveries about the role played by discourse redundancy in subject expression.

5. Conclusions and future directions

The current study was designed to further explore subject expression among NNSs and NSs of Spanish by specifically examining whether the potential form ambiguity and switch reference variables corresponded to the effects of the TMA of the verb on the forms of subject expression used. Although NNSs and NSs in the present investigation were significantly different in their use of subject expression forms and some qualitative differences were observed between groups with respect to TMA and potential form ambiguity, the effects found for TMA were largely similar. For these NNSs and NSs, while potential form ambiguity did not explain the effects of TMA for either group, the switch reference variable did. The findings for switch reference, a measure of discourse redundancy, suggest that further exploration of the relationship between other discourse features and subject expression is needed. Perhaps the greatest contribution of the current study is that we have operationalized the concepts of ambiguity and redundancy in ways that can be easily replicated. This is essential because these concepts have been addressed in so many different ways in previous sociolinguistic research that it is currently difficult to determine whether variability in the results (and in most cases the explanation of the strength of the TMA effects) are a result of methodological differences from one study to the next. Future research will serve to replicate our results, and, more importantly, to address other ways in which ambiguity and redundancy might be explored to further determine their relationship to subject expression.

In addition to the specific goals of the investigation concerning the relationships between TMA, potential form ambiguity and switch reference on subject expression for NNSs and NSs, the current study also continues the line of research that employs important methodological changes. The current study has demonstrated that the examination of the full range of forms that occur in syntactically-defined contexts (i.e., the subject position of all tensed verbs), rather than only those included on the list of 'potentially-variable contexts' (i.e., the envelope of variation), is profitable. The present investigation sheds light on how this group of NNSs and NSs vary in their use of subject expression across TMA categories and that switch reference, but not potential form ambiguity, helps to explain these variable effects. Determining whether these findings generalize beyond our participant group is an additional benefit to pursuing this issue in future studies. Our detailed analysis of the relationship between subject expression and three independent linguistic factors provides a clearer understanding of how such factors are related to each other and to subject expression and provides information beyond the identification of significant predictors of the use of a given form. Because we did not identify obligatory and variable contexts a priori, the type of analysis employed in the current study not only allows us to more thoroughly characterize variable contexts but also enables us to identify obligatory contexts in new ways because any category of our independent variables or any cross-section of those categories that demonstrated categorical use by one or both speaker populations could then be considered a non-variable context. In addition to the information gleaned from this study, we have identified several avenues for important future research and provided a model for the continuation of this investigation.

Appendix A: Sample items and their translations from the grammar test and the sociolinguistic interview

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

_____ vegetariana.
a. soy
b. estoy
c. tengo

'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.
a. I am (verb form that implies that this is the norm)
b. I am (verb form that implies a change)
c. I have'

¿Podrías explicarme cómo se prepara tu especialidad? 'Could you explain to me how to prepare your specialty?'
¿Dónde te ves en cinco años? 'Where do you see yourself in five years?'
¿Quién es el mejor instructor de idiomas, un hablante nativo o no-nativo? 'Who is the best language teacher, a native or non-native speaker?'
¿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)? 'What are the ideas or principles that you believe should motivate or guide the United States government (or your country of origin)?'

Appendix B: Cross-tabulations for potential form ambiguity and subject expression

Cross-tabulation for Potential Form Ambiguity (PFA) and Subject Expression within Each TMA Category in percentages for NNSs

Subject expression	TMA													
	Conditional		Imperfect		Periph. future		Pres. indic.		Pres. perf.		Preterit		Pres. subjc.	
	+PFA	-PFA	+PFA	-PFA	+PFA	-PFA	+PFA	-PFA	+PFA	-PFA	+PFA	-PFA	+PFA	-PFA
Null	**84.6**	**48.8**	**47.8**	**71.2**	86.3	72.8	**72.0**	**70.3**	**69.6**	**79.5**	75.9	69.5	80.0	54.9
Personal	**0**	**15.3**	**21.7**	**20.5**	2.9	13.6	**5.7**	**10.8**	**4.3**	**11.4**	9.3	15.6	5.0	13.7
Lexical NP	**7.7**	**35.3**	**30.4**	**5.4**	9.8	7.8	**18.9**	**13.5**	**26.1**	**6.8**	13.0	11.7	15.0	19.6
Other	**7.7**	**0.6**	**0**	**2.9**	2.0	5.8	**3.3**	**5.4**	**0**	**2.3**	1.9	3.1	0	11.8
Total	**100**	**100**	**100**	**100**	100	100	**100**	**100**	**100**	**100**	100	100	100	100

Note. Distributions where there is a significant difference are shown in bold.

Cross-tabulation for Potential Form Ambiguity (PFA) and Subject Expression within Each TMA Category in percentages for NSs

Subject expression	TMA													
	Conditional		Imperfect		Periph. future		Pres. indic.		Pres. perf.		Preterit		Pres. subjc.	
	+PFA	-PFA	+PFA	-PFA	+PFA	-PFA	+PFA	-PFA	+PFA	-PFA	+PFA	-PFA	+PFA	-PFA
Null	66.7	56.2	**70.9**	**66.0**	77.8	78.9	**73.2**	**65.5**	70	71.9	79.5	67.2	**68.3**	**67.4**
Personal	11.1	12.9	**1.8**	**19.9**	6.7	2.3	**10.1**	**9.5**	16.7	12.7	9.0	14.0	**13.5**	**6.6**
Lexical NP	11.1	21.2	**20.0**	**8.0**	13.3	10.9	**13.6**	**14.6**	13.3	9.5	9.0	10.6	**17.3**	**14.7**
Other	11.1	9.6	**7.3**	**6.1**	2.2	7.8	**3.1**	**10.3**	0	5.9	2.6	8.1	**1.0**	**11.2**
Total	100	100	**100**	**100**	100	100	**100**	**100**	100	100	100	100	**100**	**100**

Note. Distributions where there is a significant difference are shown in bold.

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. Martohardijono, & 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.

Bayley, Robert J., & Preston, Dennis. R. (Eds.) (1996). *Second language acquisition and linguistic variation*. Amsterdam: John Benjamins.

Bentivoglio, Paola. (1987). *Los sujetos pronominales de primera persona en el habla de Caracas*. Caracas: Universidad Central de Venezuela.

Bentivoglio, Paola. (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.

Bini, Milena. (1993). La adquisición del italiano: Más allá de las propiedades sintácticas del parámetro *pro-drop*. In J.M. Liceras (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.

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. (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. (2008a). 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* (pp. 64-77). Somerville, MA: Cascadilla Proceedings Project.

Geeslin, Kimberly, L., & Gudmestad, Aarnes. (2008b). Variable subject expression in second-language Spanish: A comparison of native and non-native speakers. In M. Bowles, R. Foote, S. Perpiñán, & R. Bhatt (Eds.), *Selected proceedings of the 2007 Second Language Research Forum* (pp. 69-85). Somerville, MA: Cascadilla Proceedings Project.

Geeslin, Kimberly, L., & Gudmestad, Aarnes. (in press). 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).

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. (2008). Acquiring a variable structure: An interlanguage analysis of second-language mood use in Spanish. (Doctoral dissertation, Indiana University, 2008). *Dissertation Abstracts International, 69*, 8.

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.

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; & Bhatt, 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.), *Lingusitic 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* (pp. 99-132). Somerville: 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 Torrens, Vincent and Linda Escobar (Eds.), *The Acquisition of syntax in Romance languages* (pp. 401–418). Amsterdam: John Benjamins.

Mougeon, Raymond, & Dewaele, Jean-Marc. (2004). Preface to variation in the interlanguage of advanced second language learners. *IRAL, 42*(4), 295-301.

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.

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.

Otheguy, Ricardo; Zentella, Ana Celia; & Livert, David. (2007). Language dialect contact in Spanish in New York: Towards the formation of a speech community, *Language, 83*, 1-33.

Phinney, Marianne. (1987). The pro-drop parameter in second language acquisition. In T. Roeper & E. Williams (Eds.) *Parameter Setting* (pp. 221-238). Dordrecht: Reidel.

Rothman, Jason (2007). Pragmatic solutions for syntactic problems: Understanding some L2 syntactic errors in terms of discourse-pragmatic deficits. In Baauw, Drijkoningen, & Pinto (Eds), *Romance Languages and Linguistic Theory 2005: Selected papers from 'Going Romance'* (pp. 299-320). Amsterdam: John Benjamins.

Serrano, María José. (1996). El análisis del discurso en variación sintáctica. *Hispanic Linguistics, 8*, 154-177.

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.

Silva-Corvalán, Carmen. (1994). *Language contact and change: Spanish in Los Angeles*. Oxford: Clarendon Press.

Young, Richard. (1991). *Variation in interlanguage morphology*. New York: P. Lang.

White, Lydia. (1985). The "pro-drop" parameter in adult second language acquisition. *Language Learning, 35*(1), 47-62.

Selected Proceedings of the 12th Hispanic Linguistics Symposium

edited by Claudia Borgonovo, Manuel Español-Echevarría, and Philippe Prévost

Cascadilla Proceedings Project Somerville, MA 2010

Copyright information

ISBN 978-1-57473-440-9 library binding

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:

Gudmestad, Aarnes and Kimberly L. Geeslin. 2010. Exploring the Roles of Redundancy and Ambiguity in Variable Subject Expression: A Comparison of Native and Non-native Speakers. In *Selected Proceedings of the 12th Hispanic Linguistics Symposium*, ed. Claudia Borgonovo, Manuel Español-Echevarría, and Philippe Prévost, 270-283. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2423.