Future Directions in the Acquisition of Variable Structures: The Role of Individual Lexical Items in Second Language Spanish

Kimberly L. Geeslin
Indiana University

1. Introduction

Accounts of second language (L2) acquisition that recognize a social dimension of language use have long been in existence and have received additional attention since the publication of the seminal work by Canale and Swain (1980), who argued that communicative competence in a L2 entails not only knowledge of the grammar of a language but also of how to organize discourse, make requests for clarification and use language appropriately in a given context (e.g., Li 2010; Mougeon, Nadasdi & Rehner 2010). In other words, it is agreed that a competent language user interprets language according to context and varies his or her own production based on features of the discourse setting (albeit subconsciously). Moreover, it is generally recognized that social contexts may affect the provision of input and/or feedback, language processing and sequences of development and it has been argued that cognitive models of development simply cannot ignore these facts (Firth & Wagner 1997, Tarone 2007). One approach to identifying and understanding the role of social factors in L2 learning and use comes from variationist linguistics. The potential contributions of this field include the ability to study cross-linguistic influence, provide a realistic view of the target, examine restructuring in L2 grammars, and explore the process of moving beyond the classroom variety (Bayley & Preston 2008). The purpose of the current paper is to highlight some of the recent advances in variationist research on L2 Spanish and to provide a more detailed look at one particular case of such developments: the examination of the role that individual lexical items play in the use of variable structures, such as the copula contrast, by L2 speakers of Spanish.

Research conducted within the variationist framework has examined a wide range of grammatical structures in L2 Spanish, including the copula contrast, mood contrast, forms of subject expression, object pronouns, the present progressive, the present perfect, forms of future time reference, intonation, the interdental fricative and /s/-weakening (see Geeslin 2011 for a detailed review). Clearly, one of the greatest developments in variationist research on L2 Spanish has been the rapid growth in this field of inquiry, which until recently was more limited to work on L2 French and English (see Geeslin & Gudmestad 2010 for a review). In addition, however, there are several substantial findings and methodological advances that merit attention. For example, research on L2 acquisition of variable structures such as the copula contrast (Geeslin 2000, 2003), the mood contrast (Gudmestad 2012b), object pronouns (Geeslin, García-Amaya, Hasler Barker, Henriksen & Killam 2010), differential object marking (Killam 2011) and the perception of aspiration (Schmidt 2011) has connected variationist research to mainstream research on the stages of L2 acquisition, demonstrating that variable structures are indeed acquired in a similar manner to those that are not variable. Additionally, this field has been bolstered by the inclusion of native speaker baselines which provide a more accurate view of the target toward which L2 learners are moving (e.g., Geeslin, García-Amaya, Hasler Barker, Henriksen & Killam 2010). Variationist studies have also served to move the field forward in the definition and operationalization of existing linguistic factors that serve to describe the use of variable structures. For example, Woolsey (2008) developed a methodology through which the effects of contexts of comparison on the copula contrast could be examined, Gudmestad and Geeslin (2010) explore competing definitions of the term ambiguity and show how these different types of

ambiguity affect forms of subject expression in distinct ways, and Gudmestad (2012a) provides a quantitative exploration of the effects of form regularity, which was not defined consistently across previous studies. Each of those investigations provides new information about the acquisition of variable structures as well as a model for applying the linguistic factor under examination in future studies of native or L2 use of that structure. Variationist research has also benefitted recently from advances in acoustic measurement technology (e.g., Henriksen, Geeslin & Willis 2010) and advances in statistical analysis (e.g., Gudmestad, House & Geeslin 2011). Finally, recent developments in the study of collocations in sociolinguistic research and lexical acquisition in L2 acquisition research have shed new light on the acquisition of the copula contrast in Spanish. It is this final case that is the focus of the current paper. We begin with a brief description of the copula contrast in Spanish and continue with a critical summary of the developments in both sociolinguistic research and second language acquisition research that inform the current investigation. A description of the methodology, analysis and results of the current study are then provided prior to discussing the implications of these findings for the body of research on L2 variation in Spanish as a whole.

2. The research context

The copula contrast in Spanish is one of the most widely examined variable structures, both in first and second language use. Although certain functions of the copula are not variable, such as the use of *estar* 'to be' in the progressive forms (e.g., *está hablando* "he/she is talking") and the use of *ser* 'to be' to tell time (e.g., *son las tres* "it is three o'clock"), when the copula is used to fulfill the attributive function in the [copula + adjective] syntactic structure, there are a number of contexts in which both *ser* and *estar* are acceptable and the degree of acceptability of each may be determined by a range of social and linguistic factors. The attributive function refers to contexts where the copula connects a referent with an adjective and that adjective performs the function of qualifying, describing or characterizing the referent in some way. The contrast between copulas in these contexts is illustrated in example (1), where both *ser* and *estar* are acceptable although they may yield slightly different meanings for some speakers but not for others.

(1) Marta es/está muy guapa
     Marta is (ser/estar) very attractive

In fact, contemporary sociolinguistic studies have documented variability in the copula contrast in pre-adjectival contexts for US Spanish (Gutiérrez 2003; Salazar 2007; Silva-Corvalán 1994), Mexican Spanish (Gutiérrez 1992, 2003; Cortes-Torres 2004), several varieties of Peninsular Spanish (Geeslin & Guijarro-Fuentes 2008), Puerto Rican Spanish (Brown & Cortes-Torres 2012; Ortiz-López 2000) and Venezuelan Spanish (De Jonge 1993; Díaz-Campos & Geeslin 2011a,b; Malaver 2000). Together, these studies have contributed to our understanding of the factors that predict use of *estar*, the path along which use of *estar* has been extended, and the relative importance of additional factors such as language contact on the copula contrast. For example, we know that an individual frame of reference (i.e., when the referent is compared to itself at another point in time) is a highly effective predictor of the use of *estar* (Geeslin 2003; Gutiérrez 2003; Silva-Corvalán 1994). We further know that certain adjective classes, such as size, physical appearance, age, and evaluation favor the extension of the use of *estar*, even though many of these contexts were formerly limited to *ser* (Gutiérrez 1992; Silva-Corvalán 1986). Additionally, it appears that the initial hypothesis that bilingualism accelerates language change (Silva-Corvalán 1986) does not hold across contact situations and may actually be an indication of the importance of other factors, such as access to the written standard (Geeslin & Guijarro-Fuentes 2008; Ortiz López 2000). This continually growing body of research on the copula contrast has explored a range of speech communities, many different linguistic variables and has employed differing methods of analysis. In general, however, the most recent research demonstrates several common features which ought to set the standard for future work in this area. For example, such studies analyze tokens according to function (i.e., attributive function) rather than seeking to demonstrate equivalent meaning within a given envelope of variation (see Geeslin 2011 or Geeslin & Gudmestad 2010 for discussion). Likewise, these studies analyze "use" of the copulas, rather than

"innovation" as determined by a single linguistic factor such as frame of reference (Brown & Cortés-Torres 2012, Geeslin & Guijarro-Fuentes 2008). Finally, these studies recognize that the contexts where native speaker use is categorical are not determined by a single factor, or even a single combination of categories of several factors and, thus, the envelope of variation for this structure cannot be determined a priori (Geeslin & Guijarro-Fuentes 2007). The present review of recent sociolinguistic research on the copula contrast takes these assumptions as a starting point and focuses exclusively on those studies that have built on this foundation and explored the role of individual lexical items in native speaker variation in the copula contrast.

2.1. Recent developments in native speaker variation in the copula contrast

As stated in the previous review, recent research focuses on copula use in attributive contexts and examines variation, rather than innovation. One example of such a study is Díaz-Campos and Geeslin (2011a), who conducted a variationist analysis of interview data taken from the *Estudio sociolingüístico de Caracas* "The Sociolinguistic Study of Caracas" (Bentivoglio and Sedano 1987), which includes 30-minute interviews with speakers across several different socioeconomic classes, four different age groups and both genders. Their Goldvarb analysis showed that *estar* was used 25.8 percent of the time and that the linguistic factors resultant state, adjective class, predicate type, experience with the referent, susceptibility to change and frame of reference, as well as the social factors age and socioeconomic level all contribute to the prediction of *estar* use, whereas the social factor gender did not. That study built on earlier work, such as Geeslin & Guijarro-Fuentes (2008) and Geeslin (2003), by extending variationist analyses based on contexts defined by function to a monolingual Spanish-speaking population and further confirmed the efficacy of this type of analysis.

Brown & Cortés-Torres (2012) also followed this model while at the same time moving the field forward by addressing the degree to which individual adjectives actually participated in the variability found in previous analyses. Their analysis of interview data from 31 Puerto Rican Spanish speakers yielded 2566 tokens of the [copula + adjective] structure. In addition to confirming the predictive value of the variables explored in previous research (i.e., susceptibility to change, frame of reference, experience with the referent, adjective class, gender and degree of bilingualism), they analyzed the degree to which frequent individual adjectives exhibited co-occurrence with *ser* and *estar*. Unlike the oft-cited research by Mesa Alonso, Domínguez Herrera, Padrón Sánchez, and Morales Aguilera (1993), who stated that a full 80 percent of all adjectives show variability with *ser* and *estar*, Brown and Cortés-Torres (2012) found that only 16 percent of the adjectives in their corpus occurred with both *ser* and *estar*. Their analysis provided a list of adjectives that co-occurred with *ser* and *estar* and which were used twenty or more times in the corpus. These adjectives along with the rates at which they occurred with *estar* are summarized in Table 1.

Table One. Adjectives that co-occurred with both copulas in Brown and Cortés-Torres (2012)

Rate of Occurrence with *Estar*	Adjectives in this Category
Less than 25%	*fuerte* "strong", *grande* "big", *chiquito* "small", *difícil* "difficult", *nuevo* "new", *diferente* "different"
Between 26 and 50%	*lindo* "pretty", *igual* "equal", *caro* "expensive", *raro* "rare", *tranquilo* "calm", *bonito* "pretty", *bueno* "good"
Between 51 and 75%	*flaco* "skinny", *gordo* "fat", *rico* "rich", *loco* "crazy", *malo* "bad"
More than 75%	*brutal* "brutal", *seguro* "certain"

What is clear from the results of Brown & Cortés-Torres (2012) is that variation in their corpus was found in a relatively small group of highly frequent adjectives. One also notes that these adjectives were nearly entirely linked to evaluative descriptions. These findings signal the importance of looking at collocations, or the co-occurrence of individual lexical adjectives with the copular verbs in order to better understand the degree to which variation between *ser* and *estar* occurs.

In a related study, Geeslin (forthcoming) analyzed interview data collected from 16 US-based Spanish-speaking bilinguals from a variety of countries of origin. The analysis included 764 [copula +

adjective] tokens and, in addition to exploring the degree to which variation was found at the level of individual lexical items, the study also sought to corroborate previous *a priori* coding schemes designed to classify adjectives according to the degree to which they allow variation between *ser* and *estar*. Examples of linguistic factors which classify the degree to which a given adjective allows co-occurrence with both copular verbs include Silva-Corvalán's (1986) Semantic Transparency, which distinguished adjectives that could only be paired with one copula from those that allowed a contrast (see also Silva-Corvalán 1994; Geeslin 2002). The variable further distinguishes adjectives that created very subtle meaning differences when combined with *ser* and *estar* from those that led to complete meaning changes or, in the opposite case, were nearly synonymous. A second example is Geeslin's (2003) Copulas Allowed, which maintained the categories *ser* required and *estar* required but collapsed the three contexts where a contrast was possible into a single category for all variable adjectives (see also Geeslin 2005; Geeslin & Guijarro-Fuentes 2007, 2008). Although logically flawed because these variables determined the degree of variability an adjective permitted prior to the analysis of the data, these variables were also highly significant in predictive models, leading Geeslin (forthcoming) to argue that the degree of accuracy of such classifications warranted further attention. Her analysis showed that of the 764 tokens analyzed, 204 included adjectives that were used only once in the corpus. In the remaining tokens, it was shown that only 26 adjectives were used 5 or more times and only 9 of these 26 occurred with both *ser* and *estar* in the corpus. These 9 adjectives included the lexical items *abierto* 'open', *casado* 'married', *difícil* 'difficult', *feliz* 'happy', *grande* 'big', *ingenuo* 'naïve', *mejor* 'better', *seguro* 'certain/safe' and *tranquilo* 'calm', and constituted only 5 percent of the total adjectives in the corpus and 40 percent of the frequent adjectives. Thus, Like Brown and Cortés-Torres (2012), Geeslin (forthcoming) found that the previously projected number of adjectives that allow variability was likely far too high. Moreover, her analysis of the classification of each adjective according to earlier factors, such as Semantic Transparency and Copulas Allowed, as compared to actual patterns of co-occurrence in her corpus, demonstrated that such classifications made incorrect predictions, including not only adjectives that were predicted to co-occur and did not, but also adjectives that were less conservative than predicted. Although the former could be explained as simply a consequence of the task in that certain tasks are unlikely to elicit all uses of every adjective, the latter cannot. Finally, a comparison of the patterns of co-occurrence between Brown and Cortés-Torres (2012) and Geeslin (forthcoming) showed several differences. While four of the frequent adjectives that occurred with both *ser* and *estar* are the same across studies, four more of the adjectives occurred with both copulas in the former but only with *ser* in the latter and a final adjective occurred with both in the former and with *parecer* 'to seem' and *ser* in the latter. Although we believe that many of these differences are likely the result of a much smaller token count in Geeslin (forthcoming), leading to fewer opportunities for the full range of use of individual adjectives to occur, these results also point to the possibility that different speech communities allow variation with different lexical items. This, of course, provides further demand for studies that examine patterns of use at the level of the adjective.

One final study that examined collocations of adjectives with *ser* and *estar* is Díaz-Campos & Geeslin (2011b). Analyzing the same Caracas Corpus data as Díaz-Campos & Geeslin (2011a), these authors sought to determine the degree to which individual adjectives varied in attributive contexts and the role that frequency played in determining such variation. Additionally, they compared the factors that predicted copula choice in contexts where both copulas are allowed (as determined by actual analysis of the data) to those identified when all adjectives were included in the analysis. For the 1669 tokens they analyzed, they found that 72 percent of those tokens were comprised of adjectives that only occurred with *ser*, 20 percent of the total tokens contained adjectives that occurred only with *estar* and 8 percent occurred with both. In other words, the number of adjectives that occurred with both copulas was lower than for those that did not demonstrate such variability. Of the total number of adjectives, 32 percent demonstrated co-occurrence with both copulas. Nevertheless, the rates of use of adjectives that occurred only with *ser* or only with *estar* were approximately 2 percent (2.1 and 2.3 percent, respectively) whereas the adjectives that did occur with both copulas were used an average of 11.9 times each, showing that variable adjectives were also used with much greater frequency than others. The list of adjectives that were produced with both *ser* and *estar* included *bueno* 'good', *malo* 'bad', *grande* 'big', *chévere* 'great', *chiquito* 'small', *pequeño* 'small', *tranquilo* 'calm', *bonito*

'pretty', *horrible* 'horrible', *barato* 'cheap', *igual* 'equal', *viejo* 'old', *tremendo* 'terrible/tremendous', *casado* 'married', *solo* 'alone', *joven* 'young' and *feo* 'ugly', and an analysis of lexical frequency showed that the percentage of adjectives that occurred with both copulas was significantly higher for high frequency adjectives than for low frequency adjectives. Finally, the regression analysis performed in Goldvarb that was limited to only those adjectives that were produced with both *ser* and *estar* demonstrated superior log likelihood and model significance to models that contained all [copula + adjective] tokens. Thus, these results were consistent with previous studies that showed that the number of adjectives that demonstrated co-occurrence with both copular verbs was smaller than previously predicted and that the lexical items that co-occur were highly frequent and may have differed from one speech community to another (as compared to the previously reviewed studies). As a whole, this body of research shows that variables or methods of analysis based on expected norms or patterns of co-occurrence are inferior in accuracy Geeslin (forthcoming) and predictive power (Díaz-Campos & Geeslin 2011b) to those that are based on empirical data from the actual corpus under analysis and that the number of adjectives that participate in variation may be more limited than previously thought. The implication of these findings is that much greater attention ought to be paid to individual lexical items in future studies of NS copula use.

2.2. Variation in the copula contrast in L2 Spanish

The variationist framework has also been used extensively to analyze the use of copular verbs with adjectives (i.e., in attributive functions) among L2 learners of Spanish (see Briscoe 1995, Gunterman 1992, Ryan & Lafford 1992 and VanPatten 1987 for research on acquisition of all functions of the copula). In an early study, Geeslin (2000) elicited data using a guided interview, a picture-description task and a written contextualized questionnaire from 77 English-speaking learners of Spanish at 4 levels of proficiency, finding a gradual increase in the rates of use of *estar* over time as learners moved away from the exclusive use of *ser* to perform all functions and integrated *estar* into attributive contexts. The analysis also used a regression analysis to identify the predictive factors that contribute to *estar* use at each level. The factors examined in these models included adjective class, modality (i.e. degree of meaning contrast), animacy of the referent, susceptibility to change of the referent + adjective combination, dependence on experience, directionality, dynamicity, perfectivity, frame of reference and telicity, as each of these was found to be important in describing this contrast for NSs (Clements 1988, De Jonge 1993, Delbecque 1997, De Mello 1979, Falk 1979, Gutiérrez 1992, 1994, Lema 1992, 1995, 1996, Luján 1981, Navas Ruíz 1963, Roldán 1974a, 1974b, Sanz & González 1995, Schmitt 1992, Silva-Corvalan 1986, 1994 and Vañó-Cerdá 1982). The results from the predictive models identified some factors that were significant across all levels, others that were less effective in describing change (and have since been dropped from more recent studies), and some factors that showed clear change over time. For example, the factor *susceptibility to change* was significant in predicting use of *estar* at lower levels of proficiency whereas *frame of reference* was significant at the higher levels.

Since that early study, this research has been extended in a variety of directions, including the study of higher proficiency levels (Geeslin 2003), learners with first languages other than English (Cheng, Lu & Giannakouros 2008; Geeslin & Guijarro-Fuentes 2005, 2006a), additional linguistic variables (e.g., Geeslin 2003), the in-depth examination of particular linguistic constraints (Woolsey 2008), and task-based variation (Geeslin 2006; Geeslin & Gudmestad 2008). In addition, analyses of the frequency of use of *estar* and the predictors of that use have been complemented by detailed investigations of the contexts in which variation occurs, for both NSs and NNSs. For example, Geeslin (2001) addressed the contrast between obligatory and variable contexts in an analysis of data from 72 English-speaking learners of Spanish at four levels of proficiency and 10 NSs of Spanish from five different countries of origin. NS responses on a 15-item written contextualized task were used to identify obligatory contexts (those where response was unanimous) and variable contexts (those where NSs did not agree unanimously) and the learner data were then evaluated to determine average accuracy rates for each proficiency level in the obligatory contexts. Learners at the lowest level actually showed a higher accuracy rate (68 percent) in contexts that unanimously elicited *ser* from the NSs than learners at higher levels (e.g., Level 4 showed 52 percent accuracy.). In contrast, contexts in

which NSs unanimously selected *estar* had a 45 percent accuracy rate at the lowest proficiency level and a 58 percent accuracy rate at the highest level. Thus, Geeslin (2001) showed that *estar*-required contexts were better indicators of development over time than the *ser*-required contexts because of the initial overgeneralization of *ser*. In a related study, Geeslin & Guijarro-Fuentes (2006b) examined longitudinal data from seven absolute beginners and found that the greatest changes in rates of use of *estar* actually occurred in the variable contexts, rather than the *estar*-required or *ser*-required contexts on their written instrument. Finally, Geeslin & Guijarro-Fuentes (2007), which is based on data from 11 Portuguese-speaking learners of Spanish, 19 NSs of Spanish and 11 NSs of Portuguese tested in Portuguese compared the unanimous and variable items across participant groups. They found that the learners were more variable in their responses than the NSs of either background. Each context was also coded for the independent linguistic variables predicate type, susceptibility to change, adjective class, frame of reference, animacy, experience with the referent and copulas allowed in order to assess whether or not obligatory contexts shared a certain group of features. The analysis of the linguistic variables demonstrated that obligatory contexts could not be defined as consistently possessing a certain list of characteristics because no combination of the categories of these variables was always present in a given type of context. In other words, it was shown to be impossible to determine whether a context is variable based only on the linguistic constraints present in that context and prior to analysis of NS data.

Of the developments in recent sociolinguistic research, some have been incorporated into research on L2 acquisition and use of the copulas. For example, Geeslin & Gudmestad (2010) adopt the approach of defining their tokens for analysis by function (i.e., attributive contexts) and all of the studies cited here analyze use rather than innovation (or precision, as it might be termed among NNSs). What this work has not done, however, is to explore the role of individual lexical items in L2 variation, even though Brown and Cortés-Torres (2012), Geeslin (forthcoming) and Díaz-Campos and Geeslin (2011b) have all shown that for NSs the analysis at the level of the lexical item is profitable and, we know more generally, that models of frequency and lexical diffusion posit that language change does not affect all items in a grammar at the same time. This void in the L2 literature is even more serious because of the fact that research on obligatory and variable contexts has demonstrated that no generalizable linguistic constraint examined to date serves to identify contexts of potential variability (Geeslin & Guijarro-Fuentes 2007). Moreover, research on L2s other than Spanish has shown that constructs such as lexical density and lexical deployment have an important relationship to development in L2s. In fact, Zyzik and Gass (2008) suggested several years ago that vocabulary development is likely an important part of understanding variation in the [copula + adjective] structure. Given the exciting research taking place on individual lexical items and NSs copula use, the current study was designed to apply this same line of analysis to L2 data.

3. The current study

The recent developments in sociolinguistic research on the copula contrast showing the importance of analyses of co-occurrence at the level of individual lexical items as well as the more general developments in lexical acquisition in second languages, motivate the current study whose goal is to move research on the second language acquisition of variation in the copula contrast in these same directions. The current research will provide an analysis of the occurrence of adjectives with *ser*, *estar* and any other copular verb that appears in the [copula + adjective] attributive context, in interview data produced by highly advanced non-native speakers in order to determine whether the extent of variation present at the level of individual lexical items is yet another source of difference between highly advanced NNSs and NSs of Spanish. Specifically, the current study will answer the following research questions:

1. How many adjectives do NNSs of Spanish produce in a 30-minute interview?
    a. How many of these adjectives occur only with *ser*?
    b. How many of these adjectives occur only with *estar*?
    c. How many of these adjectives occur with both *ser* and *estar*?

2. How do the patterns of use for NNSs compare to those for NSs?

The first research question will be answered by analyzing data from a group of 16 NNSs of Spanish, who will be described in greater depth in the following sections. The second research question will be answered by comparing the findings of the current analysis with that reported for NSs in Geeslin (forthcoming), and to a lesser degree, to the findings reported for Brown and Cortés-Torres (2012) and Díaz-Campos and Geeslin (2011b). The reader will be reminded of the relevant details of each of those studies where appropriate.

3.1. Participants

The participants in the current study are 16 graduate-level English-speaking learners of Spanish. The group includes 10 males and 6 females who range in age from 22 to 35 years of age (mean = 26.4). These highly advanced second language speakers of Spanish have all spent time in a Spanish-speaking country and their average length of stay is 18.1 months (range = 3 months – 11 years), although this average is reduced to 9.7 months when the one speaker who spent 11 years abroad is removed from the calculation. Additionally, these learners have an average of 8.6 years of formal study of the Spanish language (range = 3-17 years) and all were currently employed as instructors of Spanish at the time of the study. Participants scored between 16 and 25 points on a 25-point proficiency test (to be described below) and represent a range of relatively advanced proficiencies. This group of learners represents a highly-functional, highly-motivated learner population with a wide range of experiences, both domestic and abroad, that make-up their language acquisition history. All participants in the study had regular, daily contact with the Spanish language and were part of a bilingual speech community that also includes NSs of Spanish from a variety of countries of origin, including Argentina, Chile, Colombia, Ecuador, Mexico, Spain, United States (Puerto Rico & the Southwest) and Uruguay.

In order to answer the second research question, data elicited from the participants in the current study will be compared to data elicited from a NS group from this same speech community, the analysis of which is reported in Geeslin (forthcoming). This NS group is divided evenly between males and females (total N=16) and ranges in age from 24 to 37 years. These native speakers scored between 22 and 25 points on the same 25-item proficiency task administered to the NNSs (average = 23.6). This group represents educated speech from a variety of countries of origin which may have been influenced by contact with English and/or contact with other dialects of Spanish. Despite the likely influence of these factors, this is a more appropriate target group for the highly advanced learners in this study because (1) these are the speakers with whom the NNSs are in contact, (2) monolingualism is not a reasonable target for second language learners because they are acquiring a second language rather than replacing the first, and (3) the NNSs in the current study have contact with a range of varieties of Spanish and the arbitrary selection of a single monolingual speech community would not be profitable for the purposes of the current analysis.

3.2. Elicitation tasks

Each of the participants completed the same three activities. The first task was a detailed background questionnaire from which the previous description was taken. The second task was a 25-item proficiency exam with a multiple-choice response format and items covering a range of discrete-point formal properties of Spanish grammar. Each item was contextualized as part of a story (see the Appendix for a sample item). As reported earlier, not all NSs scored 100 percent on this test so it is likely that despite the attempt to survey formal properties, some of these structures are indeed variable. The final task was a sociolinguistic interview conducted by two research assistants, one male and one female, from Spain and Latin America, respectively. The list of interview questions was the same for each participant although interviewers followed the interest of the participants with follow-up questions and conversation. Interviewers tried to limit their contributions to the content to the degree possible. Sample topics covered in the interview include plans for the future, opinion questions and recounting of recent and past experiences (see the Appendix for sample questions). Each interview lasted approximately 30 minutes and was digitally recorded in a language laboratory.

3.3. Analysis

The interview data were transcribed from the digital recording and then checked by at least one other researcher. In most cases a third research assistant, a NS of Spanish, was hired to triple check the accuracy of the transcription. Next, each [copula + adjective] structure that fulfilled an attributive function was identified for analysis. For each token, the researcher noted the copula that was used and the adjective used. Next a total token count and total number of adjective counts were calculated for all adjectives, those used only one time, those used two or more times, those used 3 or more times, and those used 5 or more times. This last category will be referred to as 'frequent' adjectives and will be the basis of the additional analyses. The last step of the analysis was to look at the collocations for all frequent adjectives, noting which of the copular verbs appeared with each. The results of this analysis are summarized below and then compared to the results of an identical analysis of NS data (reported in Geeslin (forthcoming)).

4. Results

As stated in the preceding description of the analysis, each [copula + adjective] token was identified and the adjective and its collocations with the copular verbs were noted. This section begins with a presentation of the results of the analysis of all adjectives and then continues with the results of the analysis of collocations with frequent adjectives (i.e., those that were used 5 or more times by the NNSs). Following the presentation of the original analyses of the NNS data, a second sub-section will repeat the results for the NSs in Geeslin (forthcoming) on the same tasks and compare those results to the patterns of use for NNSs.

4.1. Results for NNSs

Once each [copula + adjective] token was identified, calculations were made to find the total token count, the number of adjectives used only once in the corpus, the number used two or more times, the number of adjectives used 3 or more times, and the number of adjectives used 5 or more times. The total number of tokens in each of these categories was divided by the total number of adjectives in each category in order to calculate the average rates of use of each adjective in that category. An adjective was considered to be the same regardless of gender or number markings found on the lexical stem. The percentage of the total corpus represented by tokens in each category was also calculated. The results of this analysis are summarized in Table Two.

Table Two. Summary of rates of use of adjectives for NNSs

	Token count	Adjective count	% of corpus	Rate of use
Total tokens	611	236	100	2.59
Adjectives used only once	145	145	23.7	1
Adjectives used 2 or more times	466	91	76.3	5.12
Adjectives used 3 or more times	396	56	64.8	7.07
Adjectives used 5 or more times	282	24	46.2	11.75

Table Two shows that a total of 611 [copula + adjective] tokens were used in attributive contexts in the interview data elicited from the 16 advanced learners of Spanish. Additionally, a total of 236 different lexical adjectives occurred, meaning that each adjective was used approximately two and a half times. One further notes that only 23.7 percent of the corpus contains tokens with adjectives that occurred only once and the group that contains the most frequent adjectives, those used at least five times, constitutes nearly half of the total tokens in the interview data. Another striking result is that in this last category of frequent adjectives the average rate of use is actually well above 5, at nearly 12

times per adjective. This shows that the adjectives used most often in the corpus are used with a notably higher frequency than those in the other categories.

This group of highly frequent adjectives was further analyzed to determine the collocations with the different copular verbs. In addition to *ser* and *estar*, adjectives occurred in attributive contexts with *parecer* 'to seem' and *sentir* 'to feel' and these are included in the analysis of co-occurrence between adjectives and copulas. A distinction is made, however, between adjectives that occur with both *ser* and *estar* and those that occur with *ser* or *estar* and another copular verb, as this is not typically the variation that is the focus of sociolinguistic research. Thus, those adjectives that occurred only with *ser* or *estar* are listed as such and then listed a second time if they occurred with additional copulas. The results of this portion of the analysis are summarized in Table Three, which shows the 24 frequently-occurring adjectives, their English translations and the number of times and copulas with which they occurred. When an adjective occurred with more than one copula the additional counts are noted in parentheses.

Table Three. Summary of frequent (5 or more uses) adjectives and occurrence with copulas

Copular verb	Adjectives used with copula	Translation	Count
Ser only	Bonito	Pretty	5
	Bueno	Good	18 (+1 parecer)
	Diferente	Different	12
	Difícil	Difficult	54
	Fácil	Easy	21
	Grande	Big	7 (+1 parecer)
	Importante	Important	35
	Ingenuo	Naive	8
	Inteligente	Intelligent	6
	Interesante	Interesting	11
	Malo	Bad	6
	Nativo	Native	4 (+1 parecer)
	Pequeño	Small	5
	Posible	Possible	9
	Tonto	Foolish	6
Estar only	Acostumbrado	Accustomed	5
	Contento	Content	5
	Dispuesto	Available/willing	7
	Seguro	Certain	6
	Nervioso	Nervous	5 (+1 sentir)
	Solo	Alone	6 (+1 sentir)
Ser + estar	Amable	Friendly	6 (1 estar, 5 ser)
	Mejor	Better	9 (1 estar, 8 ser)
	Feliz	Happy	21 (9 estar, 12 ser)
Parecer + ser	Bueno	Good	1 (+18 ser)
	Grande	Big	1 (+7 ser)
	Nativo	Native	1 (+4 ser)
Sentir + estar	Nervioso	Nervous	1 (+5 estar)
	Solo	Alone	1 (+6 estar)

One of the most striking results shown in Table Three is that only three of the frequent adjectives occur with both *ser* and *estar* in the interview data. In other words, there are only 3 adjectives used five or more times that demonstrate the potential for variation in this corpus. In comparison with the 80 percent figure cited for NSs (Mesa Alonso et al. 1993) prior to this line of inquiry, this is a notable contrast. Furthermore, although five other adjectives occur with *ser* or *estar* and another copular verb, these adjectives also show very limited variability because those adjectives that occur with *parecer*

only occur with *ser* and those that occur with *sentir* only occur with *estar*. In other words, there are no additional cases of co-occurrence with *ser* and *estar* and these adjectives. In looking at the count, it can be further seen that in most cases the adjectives that occur with more than one copula do so infrequently such that there is one preferred copula and alternatives are used quite infrequently. For example, although *mejor* is used with both *ser* and *estar* it occurs with the former 8 times and the latter only once (see below for a discussion of the link between this distribution and elicitation tasks). Only the adjective *feliz* shows a more even split between occurrences with *ser* and *estar*. Thus, for the frequently used adjectives in this interview data, variation at the level of the individual lexical item appears to be quite limited.

One might be curious as to whether this lack of variation is a characteristic of frequent adjectives and, thus, an analysis was also done to identify adjectives used between two and four times to see how many additional cases of co-occurrence were present in the data. Table Four summarizes the analysis of adjectives that occurred with more than one copula which were used less than five times each.

Table Four. Summary of adjectives used two to four times that occur with more than one copula

Copular verb	Adjective	Translation	Count
Ser + estar	Accesible	Accesible	1 ser (+1 estar)
	Animado	Animate/enthusiastic	1 ser (+1 estar)
	concentrado	Concentrated	1 ser (+1 estar)
	Viejo	Old	1 ser (+1 estar)
	Abierto	Open	3 ser (+1 estar)
	Aislado	Isolated	1 ser (+ 1 estar, +1 sentir)
	Casado	Married	1 ser (+ 3 estar)
	Listo	Ready/ intelligent	1 ser (+3 estar)
	Ocupado	Occupied / busy	1 ser (+ 2 estar)
	Pragmático	Pragmatic	3 ser (+1 estar)
	Tranquilo	Calm	1 ser (+2 estar)
Parecer + other	Natural	Natural	1 (+ 1 ser)
	Normal	Normal	1 (+ 3 ser)
Sentir + other	Inútil	Useless	1 (+ 1 ser)
	Saludable	Healthy	1 (+ 1 ser)
	Aislado	Isolated	1 (+1 ser, + 1 estar)
	Estresado	Stressed (out)	2 (+ 2 estar)
	Relajado	Relaxed	2 (+ 1 ser)

The data summarized in Table Four do suggest that in a larger corpus there would be a greater number of adjectives that demonstrate co-occurrence with both *ser* and *estar*. This is because although it would be possible that in a larger corpus the adjectives used between two and four times would not be used with greater frequency and only the highly frequent adjectives would have greater rates of occurrence, this seems unlikely. Thus, with additional participants or longer conversations about the same topics, it appears that several other adjectives would be added to the group of adjectives that permit alternation between copulas. One other observation can be made about this group of adjectives. In cases where there are only two uses of an adjective, little can be said about overall tendencies, but in those cases where an adjective was used three or four times, it appears that several of these adjectives are actually more often used with *estar* than with *ser*. This contrasts with the adjectives used five or more times reported in Table Three and may suggest that there are important differences between these groups. For example, adjectives such as *listo* actually change meaning when contrasted with *ser* and *estar* and, thus, are not likely to be in true variation in the sense that both are used to fulfill the same function. Nevertheless, there are adjectives such as *tranquilo* and *pragmático* in this group that are very much like the more frequently used adjectives *amable* and *feliz* in that they describe a personal characteristic. Thus, although future studies would need to confirm this hypothesis with larger data sets, it does appear that there are additional adjectives that would join the ranks of variable adjectives in a larger corpus. One final observation is possible looking at these slightly less frequent adjectives.

That is that there appears to be a difference between *sentir* and *parecer* that was not evident only with the most frequently used adjectives in that *parecer* occurs only with adjectives that occur with *ser*, whereas *sentir* is used with adjectives that are also paired with *ser* or *estar* or both. The potential of *sentir* to be paired with adjectives that also occur with *ser* was not present in the analysis of most frequently used adjectives alone.

4.2. Comparison to NS use in Geeslin (forthcoming)

It will be recalled that in addition to seeking to examine the patterns of use of adjectives with the copular verbs among learners, the current study was also designed to evaluate NNS use in light of the tendencies already found for NSs in a variety of speech communities. Because the NS data from Geeslin (forthcoming) were available for the current analysis, we begin here with a presentation of the findings of that study in comparison to those from the current study in order to assess the similarities and differences between NSs and NNSs on the same interview task. Table Five summarizes the results for overall token counts, adjective counts and rates of adjective use for all tokens in the interview data as well as for varying levels of frequency of use. The results presented in Table Two are repeated here for the ease of comparison.

Table Five. Summary of frequency of use of adjectives for NNSs and NSs

	Token count		Adjective count		% of corpus		Rate of use	
	NSs	NNSs	NSs	NNSs	NSs	NNSs	NSs	NNSs
Total tokens	764	611	329	236	100	100	2.3	2.59
Adjectives used once	204	145	204	145	26.7	23.7	1	1
Adjectives used 2+ times	560	466	125	91	73.3	76.3	4.48	5.12
Adjectives used 3+ times	436	396	61	56	57.1	64.8	7.15	7.07
Adjectives used 5+ more	314	282	26	24	41.1	46.2	12.08	11.75

Table Five demonstrates several similarities between the highly advanced NNSs in the current study and NSs of Spanish in the same speech community on the same task. For example, while the NSs have 26 adjectives that are used five or more times, the NNSs have 24 and, taking into account the slightly smaller token count for the latter group, this leads to nearly identical rates of use among these frequently-used adjectives (i.e., 12.08 vs. 11.75). Likewise, these highly frequent adjectives constitute just over 40 percent of the total token count for both groups of speakers. These same tendencies appear to hold true even when looking at adjectives used three or more times. As one moves up the table toward adjectives used two or more times a subtle shift is noted in that the rates of use for the NNSs are very slightly higher than for the NSs, which is the opposite of the direction of the difference for more frequent adjectives. It is important to note that these differences are not likely to be significant for any of these categories of frequency. Looking at the adjectives used only once, we note the greatest differences between the NSs and NNSs in the number of adjectives and the rates of use, where NNSs have a less diverse lexical inventory than the NSs. Nevertheless, when the token counts are considered we find only a three percent difference in the amount of the total dataset between groups. In other words, while it is the case that NSs use more adjectives overall than the NNSs, when the slightly lower token count of the NNSs is accounted for, there is very little difference between groups. This may be best seen in the overall rates of use where NSs use adjectives an average of 2.3 times each and NNSs use each adjective an average of 2.59 times. In other words, NSs tend to have slightly more adjectives and use them slightly less than the NNSs. Nevertheless, these differences are quite small and not likely to be statistically significant.

A final means of comparison between NSs and NNSs is the set of adjectives used most frequently in the interview and the patterns of co-occurrence of those adjectives. The 24 adjectives used by NNSs

5 or more times and the 26 adjectives used by NSs five or more times are listed in Table Six and categorized according to the copular verbs with which they occurred. The adjectives in regular type are unique to each group, those in italics are the same for each group, both in terms of their frequency of use and the collocations found in the analysis, and those adjectives in bold face represent adjectives that were used five or more times by both groups of speakers but with different patterns of use. The adjectives that occurred with *parecer* and *sentir* as well as with any other copular verb are reported twice in the table.

Table Six. Summary of frequent (5+ uses) adjectives and occurrence with copulas for NSs and NNSs

	Adjectives used with each copula	
Copular verb	Non-native speakers	Native speakers
Ser only	Bonito, *bueno, diferente,* **difícil**, *fácil*, **grande**, *importante*, **ingenuo**, inteligente, *interesante*, *malo*, *nativo*, *pequeño*, *posible*, tonto	*Bueno*, capaz, chico, *diferente*, duro, fuerte, *fácil*, *importante*, *interesante*, *malo*, moreno, *nativo*, *pequeño*, *posible*, sociable
Estar only	Acostumbrado, *contento*, dispuesto, **seguro**, nervioso, solo	*Contento*, dividido
Ser + estar	Amable, *mejor*, *feliz*	Abierto, casado, **difícil**, *feliz*, **grande**, **ingenuo**, *mejor*, **seguro**, tranquilo
Parecer	Ser: bueno, grande, nativo	Ser: diferente, pequeño, interesante; Both: ingenuo
Sentir	Estar: nervioso, solo	Both: feliz

Table Six shows that the largest number of adjectives used five or more times by either group occurred exclusively with *ser,* or with *ser* and *parecer* only. One difference between the NSs and the NNSs is that there are a greater number of adjectives that can be seen to allow use with *ser* and *estar* among the NS group, whereas the NNSs appear to be more conservative in this regard. This is further confirmed by the fact that three of the four adjectives that are shown in bold in the both category for the NSs (i.e., those that were frequently used by both groups but classified differently according to collocations), were used exclusively with *ser* by the NNSs. In other words, where variation with *estar* is allowed under certain circumstances for the NSs, the NNSs do not exhibit such variability. One other difference shown for the NSs is that in the case of both *parecer* and *sentir*, there is one adjective that occurs with both copulas as well. It will be recalled that even when slightly less frequently-used adjectives were included in the analysis for NNSs, that *parecer* was only paired with adjectives that also occurred with *ser*, rather than with *estar* or both. Given this comparison, in conjunction with the slight differences demonstrated for lexical diversity in Table Five, it appears that another source of difference between highly advanced NNSs and NSs is in the degree to which variability is demonstrated among frequently-used adjectives. These findings will be summarized and their implications assessed in the section that follows.

5. Discussion

It will be recalled that the current study was designed to answer two research questions. The first sought to assess the range of adjectives produced in a 30-minute interview by highly advanced NNSs of Spanish and to identify the copulas with which each lexical adjective occurred. The second research question guided the comparison between this group of NNSs and the NSs analyzed in previous research in order to determine whether lexical diversity of adjectives or patterns of co-occurrence could be said to contribute to differences between NSs and NNSs even at this high level of proficiency. The first set of analyses showed that just over 20 percent of the total tokens produced in the NNS interview data contained adjectives that were used only once and their overall average rate of use was about 2 and one-half times per adjective. An analysis of the most frequently-used adjectives, those used five or more times, showed that these frequent adjectives were used nearly twelve times each and

that this group of adjectives constituted just over 40 percent of the total tokens. Further analysis showed that of the 24 adjectives used five or more times, only three demonstrated co-occurrence with both *ser* and *estar* (*amable, mejor,* and *feliz*). All other adjectives occurred only with one or the other, or with *parecer* or *sentir* and one other copular verb. In sum, highly advanced NNS interviews are characterized by a relatively wide range of lexical adjectives (a total of 236 different adjectives), very frequent use of a small subset of those adjectives (24 adjectives used an average of 11.75 times each), and very little evidence of variation between *ser* and *estar* at the level of the individual lexical item (only 3 adjectives even occurred with both copulas across the entire NNS corpus).

The previous analysis reported in Geeslin (forthcoming) of interview data from NSs in the same speech community as the NNSs analyzed in the current project allows for a comparison between these highly advanced NNSs and NSs of Spanish. It was shown that the NSs used slightly more adjectives and therefore had slightly lower overall rates of use of individual lexical adjectives but that these differences were quite small. Likewise, the number of frequently-used adjectives and rates of use of those adjectives for both groups were very similar (NSs produced 26 adjectives five or more times and used these adjectives an average of 12.08 times each). The only true difference between NSs and NNSs in this speech community appears to be the degree to which these highly frequent adjectives co-occur with both *ser* and *estar*. While the NNSs had only three frequently-used adjectives that exhibit this pattern, the NSs produced 9 different frequently-used adjectives with both *ser* and *estar*. Moreover, they also showed that adjectives that can be paired with *parecer* or *sentir* can also be paired with both copulas, a pattern that was not found among NNSs. In sum, the comparison of NSs and NNSs revealed that while lexical diversity is not likely to be an important distinction between groups, the degree to which variation is permitted at the level of the individual lexical item may be.

This possible difference between NSs and NNSs has important implications for the way we study the development of second languages. It is generally accepted that early stages of language acquisition are characterized by one-to-one associations between a given form and a given meaning (Andersen 1984). For example, L2 learners of Spanish associate the preterit forms with accomplishment verbs and this is the context in which they first emerge (Cadierno 2000; Liskin-Gasparro 2000). As development continues, learners begin to make more subtle distinctions between contexts and also to recognize that a single form can fulfill multiple functions and a single function may be fulfilled by more than one form. Such is the case of the attributive function in Spanish where the [copula + adjective] structure may be filled with either *ser* or *estar*, and where in several cases both copulas are permissible with the same adjective and the choice between the two forms is complex and influenced by linguistic and social factors. Thus, in order to acquire a variable structure in a L2, a learner must not only recognize that multiple forms may fulfill a given function but also that the selection of these forms is a complex phenomenon which may be variable for some speakers and is determined by multiple social and linguistic factors. The comparison of NSs and NNSs in the current study suggests that learners may acquire a lexical item (i.e., a vocabulary word) and learn to use that word invariably, even at very high levels of proficiency and, that the knowledge that a particular lexical item permits variation may be acquired quite late. This result is consistent with research on variable structures in other second languages, such as French (e.g., Mougeon et al. 2010) and Chinese (e.g., Li 2010), which has shown that NNSs are often more conservative (i.e., show less variability or use of non-standard forms) than their NS counterparts. Taken along with previous research, the findings in the current study further suggest an exciting new direction for research on variable structures where lexical development is charted across proficiency levels in relation to variability and where analyses of individual lexical items provide new insights into the development of variation in L2s.

When casting a wider net and comparing the NSs and NNSs in the speech community examined here to those in other speech communities, two other issues arise that are worthy of consideration. The first is the degree to which the frequently-used adjectives that demonstrate co-occurrence with *ser* and *estar* overlap from one speech community to another. When comparing the work from Brown & Cortés-Torres (2012) and Díaz-Campos & Geeslin (2011b) to the NSs analyzed in Geeslin (forthcoming), there are only two adjectives, *grande* and *tranquilo*, included in the group of frequently used adjectives that occur with both copulas common to all three corpora. When looking at only one of the other corpora and comparing them to the NSs analyzed here, there are only three adjectives common to each pair (adding *casado* for Díaz-Campos & Geeslin 2011b and *difícil* for Brown &

Cortés-Torres 2012). Furthermore, there are some adjectives which are clearly particular to certain geographic regions or speech communities, such as *chévere* in the Caracas corpus and *brutal* in the Puerto Rican dataset. Additionally, the rates of co-occurrence may also be an artifact of the speech community or the elicitation task itself. For example, one might expect a higher incidence of '*está mejor*' than '*es mejor*' in general but in the current corpus the types of comparisons made, such as between academic programs or teaching styles, leant themselves to greater use of *ser*. Despite these facts, it does appear that many of the adjectives belong to the category of 'descriptions', 'personal traits' or 'evaluative comments' and this semantic category may be useful to identify and refine as future studies are conducted. For now, however, the important insight to be drawn is that co-occurrence with *ser* and *estar* may be a characteristic of a given lexical item but that the most frequently-used lexical items will vary from one speech community to another. The second issue that arises when looking at the frequently-used lexical items across speech communities is that this is not equivalent to looking at the frequency of a given lexical item in general. The items that occur most often in a corpus may well be an indication of the speech community as already noted but they may also be a reflection of the task, topic of conversation or formality of the relationship between the interviewer and the participant, among other factors. For example, the adjectives most frequently used to describe Caracas and how it has changed over the years are not necessarily the same set that were used to describe academic life in the US, nor are they likely to be the same as the list of most frequent adjectives in Spanish overall. The implication of this fact is that looking at lexical frequency of a given adjective may provide additional insights into variability at the level of the adjective. This is something that was considered by Díaz-Campos and Geeslin (2011b), who found that adjectives with high lexical frequency showed greater rates of co-occurrence than less frequent adjectives, and is consistent with theories of language change that hypothesize that frequent lexical items are the starting-point for such changes (Bybee 2002, Phillips 2001). Thus, examining lexical frequency in general will likely provide additional insights for both NSs and NNSs of Spanish.

There is one final issue that must be addressed in discussing the results of the current study and that is the difference between co-occurrence and variation. In the current analysis, the reported results include those adjectives that occurred at least once with each copula. In other words, these are adjectives that permit co-occurrence. Using the definition of a token based on function, it can be said that in these contexts both copulas fulfilled the same function with the same adjective. Nevertheless, most would hesitate to claim that the mere fact of co-occurrence is evidence of variation. This is because at times the pairing of one copula or the other with an adjective leads to clear differences and would not be said to be a variable context. For example, the meaning change that occurs with *listo* ('ready' with *estar* and 'intelligent' with *ser*) gives this context a different status than the use of the adjective *grande*, which means 'big' when paired with either copula. Some would further argue that even when *grande* is used with *estar* in a context of comparison (i.e., an individual frame of reference) that this is not a case of a variable context because the use of *estar* is linked to the function/meaning of indicating comparison. The purpose of the current analysis is not to make a claim regarding the definition of variation but rather to identify those adjectives that may permit variation as evidenced by their use with both *ser* and *estar* in the interview data. In so doing, it is important to maintain the distinction between co-occurrence and variation because even adjectives that are used with both *ser* and *estar* may exist outside a true envelope of variation.

6. Limitations and future directions

The current project represents one of several expansions and developments in the field of L2 variation in that it connects recent work on variation in the copula contrast among NSs of Spanish to progress in the field of L2 acquisition. Specifically, the degree to which individual adjectives permit variation in the [copula + adjective] structure among NSs can be seen to connect firmly to the recent push to explore the role of vocabulary acquisition and the individual lexical item in L2 acquisition. The current project is the first to examine the degree to which individual adjectives co-occur with *ser* and *estar* in oral interview data produced by highly advanced NNSs of Spanish. The results of the analysis showed that these NNSs exhibit lexical diversity similar to that of NSs on the same task and that the number of adjectives used frequently on this task is also similar, as are the rates of use of each

frequency-based classification of adjectives. The difference that was identified between groups is that the number of adjectives that occur with more than one copula was much smaller than for NSs, indicating less potential for variation on the part of the NNSs. This indicates that learners may acquire the semantic meaning of an adjective well before they acquire the knowledge that a given adjective permits variation, further supporting the idea that variation in a L2 is late-acquired. This finding points to several exciting directions for future research. For example, the analysis of patterns of co-occurrence across NS speech communities and levels of NNS proficiency will allow researchers to narrow the envelope of variation and to refine methods of classifying adjectives according to the degree to which they permit variation. Additionally, future research should continue to examine the role of individual lexical items such that enough detail exists to confirm, refute and develop theories of language use and acquisition that do (not) include a role for individual lexical items. Finally, it can be hypothesized that the degree to which learners show co-occurrence with adjectives and different copulas may correlate with other measures of vocabulary knowledge such as lexical density or deployment, further strengthening the connection to research in mainstream SLA.

It is clear from the comparison across studies that the lexical items that are used frequently may vary from one speech community to another and this author believes this is also the case across tasks and research settings. Thus, prior to generalizing these findings it will be important to expand this research to include a wider range of tasks, learners, and native speech communities. Given the fact that the analysis of adjectives used between two and four times greatly expanded the findings made when only those adjectives used five or more times were analyzed, it is likely that a larger corpus should be used in future research. This can be attained by including more participants or by collecting more tokens from each participant. A reasonable goal in this regard would be to match Brown and Cortés-Torres (2012), who analyzed just over 2500 tokens, although the nearly 1700 in Díaz-Campos & Geeslin (2011b) proved effective in identifying a large enough group of frequent adjectives to run a new regression analysis with only those adjectives that allow co-occurrence.

Appendix

Sample translated items from the grammar test:
I think it is very interesting _____ (I talk / to talk / talking) about people's eating habits. As for me, _____ (I am (*ser*) / I am (*estar*) / I have) a vegetarian.

Sample interview questions:
What plans do you have for the weekend?
Where do you see yourself in five years?
Who is the best language teacher, a native or non-native speaker?

References

Andersen, Roger. 1984. What's gender good for, anyway? In Roger Andersen (Ed.) *Second Languages: A Cross-Linguistic Perspective*, 77-99. Rowley, MA: Newbury House.

Bayley, Robert & Preston, Dennis. 2008. Variation and second language grammars. *Studies in Hispanic and Lusophone Linguistics*, 1(2), 385-397.

Bentivoglio, Paola and Sedano, Mercedes. 1987. Estudio Sociolingüístico de Caracas. Unpublished material.

Briscoe, Gregory. 1995. *The acquisition of ser and estar by non-native speakers of Spanish*. PhD Dissertation: University of Pennsylvania.

Brown, Esther & Cortés-Torres, Mayra. 2012. Syntactic and pragmatic [*ser/estar* + adjective] use in Puerto Rican Spanish: ¡Está brutal! In Kimberly Geeslin. & Manuel Diaz-Campos (Eds.), *Selected Proceedings of the 14th Hispanic Linguistics Symposium*. Somerville, MA: Cascadilla Press.

Bybee, Joan. 2002. Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change*, *14*, 261-290.

Cadierno, Teresa. 2000. The acquisition of Spanish grammatical aspect by Danish advanced language learners. *Spanish Applied Linguistics, 4*, 1-53.

Canale, Michael & Swain, Merrell. 1980. Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*(1), 1-47.

Cheng, An Chung, Lu, Hui-Chuan, & Giannakouros, Panayotis. 2008. The uses of Spanish copulas by Chinese-speaker learners in a free writing task. *Bilingualism: Language and Cognition, 11*, 301-317.

Clements, J. Clancy. 1988. The semantics and pragmatics of the Spanish <copula + adjective> construction. *Linguistics, 26*, 779-822.

Cortés-Torres, Mayra. 2004. ¿*Ser* o *estar?* La variación lingüística y social de *estar* más adjetivo en el español de Cuernavaca, México. *Hispania, 87*, 788-795.

De Jonge, Bob. 1993. (Dis)continuity in language change: Ser and estar + age in Latin American Spanish. In F. Drijkoningen & K. Hengeveld (Eds.), *Linguistics in the Netherlands*, 69-80. Amsterdam: John Benjamins.

Delbecque, Nicole. 1997. The Spanish copulas SER and ESTAR. In M. Verspoor, K. Dong Lee & E. Sweetser (Eds.), *Lexical and syntactical constructions and the construction of meaning*, 247-270. Amsterdam: Benjamins.

De Mello, George. 1979. The semantic values of ser and estar. *Hispania, 62*, 338-341.

Díaz-Campos, Manuel & Geeslin, Kimberly. 2011a. Copula Use in the Spanish of Venezuela: Is the pattern indicative of stable variation or an ongoing change? *Spanish in Context, 8*(1), 73-94.

Díaz-Campos, Manuel & Geeslin, Kimberly. 2011b. Copula choice in Spanish: An analysis of adjective co-occurrence and frequency of collocations. *New Ways of Analyzing Variation 40,* Georgetown University, Washington, D.C.

Falk, Johan. 1979. *Ser y estar con atributos adjetivales*. Uppsala: Almqvist & Wiksell.

Firth, Alan & Wagner, Johannes. 1997. On discourse, communication, and (some) fundamental concepts in SLA research. *Modern Language Journal, 81*, 285-300.

Geeslin, Kimberly. 2000. A new approach to the second language acquisition of copula choice in Spanish. In Ronald Leow & Cristina Sanz (Eds.), *Spanish Applied Linguistics at the Turn of the Millennium: Papers from the 1999 Conference on the L1 & L2 Acquisition of Spanish and Portuguese*, 50-66. Somerville, MA: Cascadilla Press.

Geeslin, Kimberly. 2001. Changing norms, moving targets and the SLA of copula choice. *Spanish Applied Linguistics, 5*(1-2), 29-55.

Geeslin, Kimberly. 2002. Semantic transparency as a predictor of copula choice in SLA. *Linguistics: An Interdisciplinary Journal of the Language Sciences, 40*(2), 439-468.

Geeslin, Kimberly. 2003. A comparison of copula choice in advanced and native Spanish. *Language Learning* 53(4), 703-764.

Geeslin, Kimberly. 2005. *Crossing Disciplinary Boundaries to Improve the Analysis of Second Language Data: A Study of Copula Choice with Adjectives in Spanish*. Munich: LINCOM Publishers.

Geeslin, Kimberly. 2011. Variation in L2 Spanish: The state of the discipline. *Studies in Hispanic and Lusophone Linguistics, 4*(2), 461-517.

Geeslin, Kimberly. Forthcoming. An analysis of co-occurrence of the Spanish copulas with adjectives: How variable are they really? In Lui, Feng-Hsi (Ed.), *Festricft for Rudolph Troike*. Taiwan: Academia Sinica.

Geeslin, Kimberly, García-Amaya, Lorenzo, Hasler Barker, Maria, Henriksen, Nicholas & Killam, Jason. 2010. The SLA of direct object pronouns in an immersion environment where use is variable. In C. Borgonovo, M. Español-Echevarría & P. Prevost, (eds.), *Refereed Proceedings from the 12th Hispanic Linguistics Symposium*, pp. 246-259. Somerville, MA: Cascadilla Press.

Geeslin, Kimberly & Gudmestad, Aarnes. 2008. Comparing interview and written elicitation tasks in native and non-native use: Do speakers do what we think they do? In Joyce Bruhn de Garavito and Elena Valenzuela (Eds.), *Selected Proceedings of the 10th Hispanic Linguistics Symposium (HLS),* 64-77. Somerville, MA: Cascadilla Press.

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: Native and Non-Native Contrasts in Forms of Spanish Subject Expression. In Luke Plonsky and Maren Schierloh (Eds.), *Selected Proceedings of the Second Language Research Forum* 2009.

Geeslin, Kimberly & Guijarro-Fuentes, Pedro. 2004. Ultimate attainment by Portuguese learners of Spanish copula choice: A study of frequency and contexts of use. Unpublished manuscript. Indiana University.

Geeslin, Kimberly & Guijarro-Fuentes, Pedro. 2005. The acquisition of copula choice in instructed Spanish: The Role of Individual Characteristics. In David Eddington (Ed.), *Selected Proceedings of the 6th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages*, 66-77. Somerville, MA: Cascadilla Proceedings Project.

Geeslin, Kimberly & Guijarro-Fuentes, Pedro. 2006a. The Second Language Acquisition of Variable Structures in Spanish by Portuguese Speakers. *Language Learning, 56*(1), 53-107.

Geeslin, Kimberly & Guijarro-Fuentes, Pedro. 2006b. A longitudinal study of copula choice: Following development in variable structures. In Nuria Sagarra & Almeida Jacqueline Toribio (Eds.), *Selected Proceedings of the 9th Hispanic Linguistics Symposium*, 144-156. Somerville, MA: Cascadilla Proceedings Project.

Geeslin, Kimberly & Guijarro-Fuentes, Pedro. 2007. The SLA of variable structures: analyzing obligatory and variable contexts. *International Journal of Applied Linguistics (ITL), 153*, 25-51.

Geeslin, Kimberly & Guijarro-Fuentes, Pedro. 2008. Variation in contemporary Spanish: Linguistic predictors of *estar* in four cases of language contact. *Bilingualism: Language and Cognition, 11*, 365-80.

Gudmestad, Aarnes. 2012a. Toward an understanding of the relationship between mood use and form regularity: Evidence of variation across tasks, lexical items, and participant groups. In Kimberly Geeslin & Manuel Diaz-Campos (Eds.), *Selected Proceedings of the 14th Hispanic Linguistics Symposium.* Somerville, MA: Cascadilla Press.

Gudmestad, Aarnes. 2012b. Acquiring a variable structure: An interlanguage analysis of second-language mood use in Spanish. *Language Learning*, 62(2), 373-402.

Gudmestad, Aarnes & Geeslin, Kimberly. 2010. Exploring the roles of redundancy and ambiguity in variable subject expression: A comparison of native and non-native speakers. In Claudia Borgonovo, Manuel Español-Echevarría & Philip Prevost (eds.), *Refereed Proceedings from the 12th Hispanic Linguistics Symposium*, 270-283. Somerville, MA: Cascadilla Press.

Gudmestad, Aarnes, House, Leanna & Geeslin, Kimberly. 2011. A statistician to the rescue: What a Bayesian analysis can do for the sociolinguistic study of subject expression in Spanish. *American Association of Applied Linguistics*, Chicago, Illinois.

Gunterman, Gail. 1992. An analysis of interlanguage development over time: Part II, *ser* and *estar*. *Hispania, 75*, 1294-1303.

Gutiérrez, Manuel. 1992. The extension of *estar*: A linguistic change in progress in the Spanish of Morelia, Mexico. *Hispanic Linguistics, 5*, 109-41.

Gutiérrez. Manuel. 1994. Simplification, transfer and convergence. *Bilingual Review, 19*, 111-121.

Gutiérrez, Manuel. 2003. Simplification and innovation in US Spanish. *Multilingua, 22*, 169-184.

Henriksen, Nicholas, Geeslin, Kimberly & Willis, Erik. 2010. The Development of L2 Spanish Intonation During a Study Abroad Immersion Program in León, Spain: Global Contours and Final Boundary Movements. *Studies in Hispanic and Lusophone Linguistics*, *3*(1), 113-162.

Killam. Jason. 2011. *An Interlanguage Analysis of Differential Object Marking in L2 Spanish.* PhD Dissertation: Indiana University.

Lema, José. 1992. Tiempo y aspect, correlatos sintácticos y semánticos: los auxiliares *ser* y *estar*. In José Pascual (Ed.), *Estudios Lingüísticos de México y España*, 279-292. Salamanca, Spain: Universidad de Salamanca.

Lema, José. 1995. Distinguishing copular and aspectual auxiliaries: Spanish ser and estar. In Jon Amastae, Grant Goodall, Mario Montalbetti, Marianne Phinney (Eds.), *Contemporary research in Romance linguistics*, 257-274. Amsterdam: John Benjamin.

Lema, José. 1996. Aspectual parametric effects in Spanish. In Claudia Parodi, Carlos Quicoli, Mario Saltarelli & María Luisa Zubizarreta (Eds.), *Aspects of Romance linguistics*, 279-292. Washington, D.C.: Georgetown University Press.

Li, Xiaoshi. 2010. Sociolinguistic variation in the speech of learners of Chinese as a second language. *Language Learning*, 60(2), 366-408.

Liskin-Gasparro, Judith. 2000. The use of tense-aspect morphology in Spanish oral narratives: Exploring the Perceptions of Advanced Learners. *Hispania, 83*, 830-844.

Luján, Marta. 1981. The Spanish copulas as aspectual indicators. *Lingua, 54*, 165-210.

Malaver, Irania. 2000. "Cuando Estábamos Chiquitos": *Ser* y *estar* en expresiones de edad. Un fenómeno americano. *Boletín de Lingüística, 16*, 44-65.

Mesa Alonso, M., Domínguez Herrera, M., Padrón Sánchez, E., and Morales Aguilera, N. 1993. Ser y estar: Consideraciones sobre su uso en español. *Islas, 104*,150-56.

Mougeon, Raymond, Nadasdi, Terry & Rehner, Katherine. 2010. *The Sociolinguistic Competence of Immersion Students*. Bristol, UK: Multilingual Matters

Navas Ruíz, Ricardo. 1963. *Ser y estar: Estudio sobre el sistema atributivo del español*. Salamanca, Spain: Acta Salmanticensia.

Ortiz López, Luis A. 2000. Extensión de *estar* en contextos de *ser* en el español de Puerto Rico: ¿evaluación interna y/o contacto de lenguas? *BAPLE,* 98-118.

Phillips, Betty (2001). Lexical diffusion, lexical frequency, and lexical analysis. In Joan Bybee & Paul Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure*. Philadelphia: John Benjamins.

Roldán, Mercedes. 1974a. Toward a semantic characterization of ser and estar. *Hispania, 57*, 68-75.

Roldán, Mercedes. 1974b. On the so-called auxiliaries 'ser' and 'estar'. *Hispania, 57*, 292-295.

Ryan, John & Lafford, Barbara. 1992. The acquisition of lexical meaning in a study abroad environment: Ser + estar and the Granada experience. *Hispania, 75*, 714-722.

Salazar, Michelle. 2007. Está Muy Diferente a Como Era Antes: *Ser* and *estar* + adjective in New Mexico Spanish. In Kim Potowski and Richard Cameron (Eds.), *Spanish in Contact: Policy, social and linguistic inquiries*, 345-355. Amsterdam: John Benjamins.

Sanz, Cristina. & González, María José. 1995. *Ser* and *estar* in Tortosí Catalan: Language contact, language variation, and language change. *Sintagma, 7*, 5-25.

Schmidt, Lauren. 2011. *Acquisition of dialectal variation in a second language: L2 perception of aspiration of Spanish /s/.* PhD Dissertation: Indiana University.

Schmitt, Cristina. 1992. *Ser* and *estar*: A matter of aspect. *Proceedings of the North East Linguistic Society* 22, 411-425.

Silva-Corvalán, Carmen. 1986. Bilingualism and language change: The extension of *estar* in Los Angeles Spanish. *Language, 62*, 587-608.

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

Tarone, Elaine. 2007. Sociolinguistic approaches to second language acquisition-1997-2007. *Modern Language Journal*, *95*(5), 837-848.

Vañó-Cerdá, Antonio. 1982. *Ser y estar + adjetivos*. Tübingen, Germany: Narr.

VanPatten, Bill. 1987. The acquisition of *ser* and *estar*: Accounting for developmental patterns. In Bill VanPatten, Trisha Dvorak & James Lee (Eds.), *Foreign language learning: A research perspective*, 61-75. New York: Newbury House.

Woolsey. Daniel. 2008. From theory to research: contextual predictors of estar + adjective and the study of the SLA of Spanish copula choice. *Bilingualism, Language and Cognition, 11*(3), 277-296.

Zyzik, Eve & Gass, Susan. 2008. Epilogue: A tale of two copulas. *Bilingualism: Language and Cognition, 11*(3), 383-385.

Selected Proceedings of the
15th Hispanic Linguistics Symposium

edited by Chad Howe, Sarah E. Blackwell, and Margaret Lubbers Quesada

Web access and citation information

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

This paper can be cited as:

Geeslin, Kimberly L. 2013. Future Directions in the Acquisition of Variable Structures: The Role of Individual Lexical Items in Second Language Spanish. In *Selected Proceedings of the 15th Hispanic Linguistics Symposium*, ed. Chad Howe et al., 187-204. Somerville, MA: Cascadilla Proceedings Project. www.lingref.com, document #2885.