

SECOND LANGUAGE VARIATION OF *SER* AND *ESTAR*: A COMPARATIVE
ANALYSIS OF ADVANCED SECOND LANGUAGE LEARNERS

By

DORIAN DORADO

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2011

© 2011 Dorian Dorado

To my wonderful family

ACKNOWLEDGMENTS

I would like to take this opportunity to first thank my dissertation committee. Dr. Gillian Lord has worked with me from the very beginning of my academic career. She is an amazing professional, and she has been one of my main inspirations to study linguistics. Dr. Jessi Aaron, I am forever indebted to, she taught me the quantitative analysis used in this dissertation, without her, the process would have been a lot more challenging. Dr. Ana De Prada Pérez gave me valuable input on my methodological choices and was always there for moral support. Dr. Hélène Blondeau always encouraged me and provided assistance in regards to my quantitative analysis, and methods. Dr. Maria Coady always had something positive to say and always had smile on her face. I sincerely thank each committee member for their time, comments, and encouragement. I truly believe that I had one of the best dissertation committees in the world!

I would also like to thank the UF office of Graduate Minority Programs, for their financial assistance. Mainly, I would like to thank Dr. Alexander, Mr. Wade, and Ms. Broiles. They are truly remarkable people who genuinely want to see students succeed. They are warm-hearted and just simply beautiful people all around. I am forever grateful to each one of them. I absolutely love them.

I like to also thank my family for loving me unconditionally. Also part of my family is Dr. Chalmers and his beautiful wife, Mrs. Jean, an amazing couple who opened their home for me and treated me like a daughter, with constant encouragement and love. I love them! Norma Augustine, she was always available, and was never too busy to listen. She is an angel. I would also like to thank my sweet friends, particularly Angelica Montoya who was always there for me and would always find a way to make me laugh

during the writing process of my dissertation. Finally, I also like to thank my wonderful uncle Dean. He was always available to read and edit my chapters, absolutely love him.

TABLE OF CONTENTS

	<u>page</u>
ACKNOWLEDGMENTS.....	4
LIST OF TABLES.....	9
LIST OF FIGURES.....	11
ABSTRACT	12
CHAPTER	
1 INTRODUCTION	14
General Tendencies of <i>Ser</i> and <i>Estar</i> Found in Native Use	16
Copula Use among Monolingual Native Spanish Speakers.....	19
Uses of <i>Ser</i> and <i>Estar</i> Taught in the Classroom.....	23
Copula Use in Spanish/English Bilingual Speakers	26
Second Language Acquisition of <i>Ser</i> and <i>Estar</i>	28
Copula Use in Second Language Learners	32
Previous SLA/Variation Studies	34
2 METHODOLOGY	40
Research Questions and Hypotheses.....	40
Speech Communities.....	44
Participants	48
Materials	55
Proficiency Assessment	56
Social Network	58
Narratives	61
Analysis	64
Linguistic Variables	65
Extralinguistic Variables	68
Statistical Procedure	70
3 RESULTS FOR ADVANCED L2 LEARNERS FROM A MONOLINGUAL SPEECH COMMUNITY	84
Results for Extralinguistic Factors.....	84
Individual Differences	85
Gender	86
Age.....	86
Years of Formal Instruction	87
Time Abroad.....	88
Spanish Self-Evaluation	89

Marginal Results for Linguistic Factors	90
Tense, Aspect and Mood.....	90
Collocations.....	91
Adjective Class.....	94
Subject Class	96
Person/Number	97
Complements	97
Multivariate Analysis Results for Linguistic Factors	98
Grammatical Person and Animacy	99
Collocations.....	100
Tense, Aspect and Mood.....	102
4 RESULTS FOR ADVANCED L2 LEARNERS FROM A BILINGUAL SPEECH	110
Results for Extralinguistic Factors.....	110
Participants.....	111
Gender	113
Acquisition Age.....	114
Years of Formal Instruction	115
Study Abroad.....	116
Spanish Proficiency Self-Evaluation	116
Marginal Results for Linguistic Factors	117
Tense, Aspect and Mood.....	118
Collocations.....	119
Adjective Class.....	119
Person/Number	123
Complement	124
Multivariate Analysis Results for Linguistic Factors	126
Collocations.....	126
Grammatical Person and Animacy	127
Tense, Aspect and Mood.....	128
Concluding Remarks	128
5 DISCUSSION AND CONCLUSION	137
Advanced Learners from Two Speech Communities: A Comparison of Linguistic Factors.....	137
Collocations.....	138
Grammatical Person and Animacy	140
Tense, Aspect, and Mood.....	143
Research Questions Revisited.....	144
Research Question 1: Frequency of <i>Estar</i>	144
Research Question 2: Extralinguistic Factors.....	145
Gender.....	146
Number of years of formal instruction in Spanish.....	147
Age of formal instruction in Spanish.....	149
Duration of study abroad in a Spanish speaking country	149

Spanish proficiency	151
Research Question 3: Linguistic Factors	152
Limitations and Directions for Future Study	154
Concluding Remarks.....	156

APPENDIX

A INFORMED CONSENT	164
B DEMOGRAPHICS	165
C THE LANGUAGE CONTACT PROFILE	167
D INSTRUCTIONS FOR NARRATIVES.....	169
LIST OF REFERENCES	170
BIOGRAPHICAL SKETCH.....	178

LIST OF TABLES

<u>Table</u>	<u>page</u>
1-1 Prescriptive uses of <i>ser</i> and <i>estar</i>	38
2-1 L2 learners from Spanish/English bilingual speech communities (BISC)	74
2-2 L2 learners from monolingual English speech communities (MONOSC)	75
2-3 Coding of independent linguistic variables for copula choice in advanced L2 learners	80
3-1 Overall distributions of <i>estar</i> and <i>ser</i> by L2~monolingual community	104
3-2 Overall distributions of <i>estar</i> and <i>ser</i> according gender.....	104
3-3 Overall distributions of <i>estar</i> and <i>ser</i> according to age of formal instruction.....	105
3-4 Overall distributions of <i>estar</i> and <i>ser</i> according to years of formal instruction in Spanish.....	105
3-5 Overall distributions of <i>estar</i> and <i>ser</i> according to time abroad	105
3-6 Overall distributions of <i>estar</i> and <i>ser</i> according to Spanish self-evaluation	106
3-7 Marginal results for advanced L2 Learners from a monolingual speech community. <i>Total N 869</i>	107
3-8 Multivariate analysis of the contribution of linguistic factors selected as significant to the probability of the copula.....	109
4-1 Overall distributions of <i>ser</i> and <i>estar</i> by L2 ~ bilingual community	129
4-2 Overall distributions of <i>ser</i> and <i>estar</i> according to gender.....	130
4-3 Overall distributions of <i>ser</i> and <i>estar</i> according to age of formal instruction in Spanish	131
4-4 Overall distributions of <i>estar</i> and <i>ser</i> according to years of formal instruction in Spanish.....	131
4-5 Overall distributions of <i>estar</i> and <i>ser</i> according to time study abroad	131
4-6 Overall distributions of <i>estar</i> and <i>ser</i> according to Spanish proficiency self-evaluation	132
4-7 Preliminary results for advanced L2 Learners from Spanish/English bilingual speech communities <i>Total N 1173</i>	133

4-8	Preliminary results for advanced L2 Learners from the BISC and MONOSC Groups.....	135
4-9	Multivariate analysis of the contribution of linguistic factors selected as significant to the probability of the copula.....	136
5-1	Multivariate analysis of the contribution of linguistic factors selected as significant to the probability of the frequency of estar.....	159

LIST OF FIGURES

<u>Figure</u>	<u>page</u>
2-2 Hispanic census data for English monolingual communities in Florida.....	76
2-3 Self-reported proficiency level in Spanish (scale 0-3) for listening.....	76
2-4 Self-reported proficiency level in Spanish (scale 0-3) for speaking	77
2-5 Self-reported proficiency level in Spanish (scale 0-3) for reading.....	77
2-6 Self-reported proficiency level in Spanish (scale 0-3) for writing	78
5-1 Frequency of <i>estar</i> in monolingual speech community groups and bilingual speech community groups.....	160
5-2 Frequency of <i>estar</i> according to gender	160
5-3. Frequency of <i>estar</i> according to number of years of formal instruction in Spanish	161
5-4 Frequency of <i>estar</i> according to age of formal instruction in Spanish	161
5-5 Frequency of <i>estar</i> according to time abroad	162
5-6 Frequency of <i>estar</i> according to proficiency level in Spanish	163

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

SECOND-LANGUAGE VARIATION OF *SER* AND *ESTAR*: A COMPARATIVE
ANALYSIS OF ADVANCED SECOND LANGUAGE LEARNERS

By

Dorian Dorado

August 2011

Chair: Gillian Lord

Major: Romance Languages

The current study is part of a relatively new strand of research that describes and accounts for linguistic variation between second language learners. The focus is not on the errors produced in the learners' speech but instead on the variability among the learners. This paper investigates the variation of *ser* and *estar* (both of which mean 'to be') use in two groups of advanced L2 learners who are comparable except for one factor: one group has received ample contact with Spanish outside of the formal classroom setting, and the other has not. In other words, in one group the participants lived continuously in a bilingual Spanish/English speech community (e.g., Miami, Florida) and the other in a primarily monolingual English community (e.g., Gainesville, Florida).

The present study also looks to see if linguistic factors such as tense, aspect and mood (TAM), collocations, adjective type, subject class, complements and person and number have any effect on the variation of copula use. Moreover, various extralinguistic factors (e.g., age at which participants began formal instruction of Spanish, age at which participants were first exposed to Spanish, length of time that participants studied

abroad, proficiency level in Spanish) are considered as relevant to variation in copula frequency.

The present investigation has shown that not only certain linguistic factors but also type of speech community can affect the frequency use of *estar*. We interpret these findings to mean that contact with the target language beyond the exposure received in a traditional classroom setting is in fact relevant to L2 acquisition.

As it relates to the extralinguistic factors, we offer marginal results indicating the number and percentages of instances of *ser* and *estar* in the data of the dependent variable. These numbers and percentages serve as a preliminary glimpse at copula use. The descriptive data also revealed the distribution of the variants in the dependent variable, allowing us to see the contrast between the categories with a factor group and between the two groups of second language learners.

CHAPTER 1 INTRODUCTION

There is a growing body of research investigating native speakers' alternation between two or more variants that convey similar notions (e.g., *ser/estar* 'to be' in Spanish, was/were variation in English, (t/d) deletion in English, negatives in African-American English, auxiliary and main verb forms in Canadian French¹) (Mougeon & Rehner 2001: p. 398). The current study, however, is part of a relatively new strand of research that describes and accounts for linguistic variation between second language learners. It is recognized that "L2 speech indeed demonstrates variation that is highly systematic and subject to a range of social factors" (Bayley & Regan 2004: p. 323), but few have systematically studied this variation. This dissertation project investigates the variation of Spanish *ser* and *estar* (both of which mean 'to be') use in two groups of advanced L2 learners who are comparable except for one factor: one group has received ample contact with Spanish outside of the formal classroom setting, and the other has not; the former group lived continuously in a bilingual Spanish/English speech community and the latter in a primarily monolingual English community.

The focus of this research is not on accuracy or any "errors" the learners may make but rather the learners' interlanguage variability itself. As such, we offer no correct/incorrect judgment regarding learners' copula choice, but rather investigate the frequency of the copula, as an indicator of whether these learners favor or disfavor *estar* in innovative contexts. These are contexts that were previously limited to *ser* (Silva-Corvalán 1994: p. 92). This phenomenon is referred to as the extension of *estar* or

¹ Each example of variable type mentioned here serves only for the purpose of illustration rather than comprehension.

innovative *estar*, which is furthered, discussed in later chapters. Because the use of the copula has been changing gradually for some time, and the prescriptive sources are inadequate given native-speaker variation, in the communities concerned, the norm is yet unknown. It is appropriate at this point to limit ourselves to observation, and to stay away from any form of subjective assessment. Díaz-Campos & Geeslin (forthcoming) agree “it is difficult to document a given prescriptive norm at any single point in time” (p. 5). Nonetheless, in this introduction a presentation of the traditional prescriptive contexts of copula use is offered in order to get a better understanding to the sort copula information that L2 learners receive in their classrooms, as well as to understand how *estar* use is evolving.

Unlike most other copula studies, which have been limited to [copula + adjective] contexts, the present study also includes other independent factors such as tense, aspect and mood, collocations, adjective type, subject class and person and number and complement. Moreover, various extralinguistic factors (e.g., age at which participants began formal instruction of Spanish, age at which participants were first exposed to Spanish, length of time that participants studied abroad, proficiency level in Spanish) are considered to examine the variation found in more detail.

This study investigates Type II variation by L2 learners. There are two types of variation: “Type I variation refers to the existence of forms that are not found in native varieties but instead vary in nonnative varieties with the target like form”, (Geeslin & Guijarro-Fuentes, 2006: p. 57). “This type of variation shows that the learners are on their way to acquiring an invariant target form” (Mougeon, Rehner & Nadasdi, 2004: p. 409). On the other hand, “Type II variation is that which occurs between two forms, both

of which exist and vary in the speech of native speakers” (Geeslin & Guijarro-Fuentes, 2006: p. 57). Type II variation shows that L2 learners are knowledgeable of the full range of native variants, indicating successful acquisition (Mougeon, Rehner & Nadasdi, 2004: p. 58). Moreover, to acquire this type of variation, it is generally assumed that the L2 learners need to be in an environment in which they have ample access to native speech (Mougeon, Rehner & Nadasdi, 2004: p. 58).

This introductory chapter is divided as follows: The next section presents the general tendencies of *estar* and *ser* use which also help characterize some of the difficulties in learning the copula as an L2 learner, and is followed by a review of previous findings on copula use in native Spanish monolinguals. We then continue with a review of the prescriptive uses of *ser* and *estar*. Later, copula uses in heritage speakers is presented, followed by a review of second language acquisition studies, investigating L2 copula choice and second language learners.

General Tendencies of *Ser* and *Estar* Found in Native Use

According to relevant literature, there are features that are particularly relevant in determining native speakers’ choice of the copula: frame of reference, dependence on experience, and susceptibility to change. Cortés-Torres (2004), Geeslin (2003) and Silva-Corvalán (1986) affirmed that frame of reference can either be classified as class frame or individual frame: a class frame compares a referent to a group of like objects and an individual frame compares a referent to itself at another point of time (Geeslin 2003: p. 723). Traditionally, class frame is associated with *ser* and individual frame is associated with *estar* (Cortés-Torres 2004: p. 790). For example, Silva-Corvalán (1994) provides the following examples of class and individual frames:

(1) (a) *Juan es alto.*

‘Juan is tall.’

(b) *Juan está alto.*

‘Juan is tall.’

(1a) is an example of class frame of reference: Juan is tall because he has a set of characteristics (assuming that all masculine humans have a certain height), of which being ‘tall’ which has set him as a member of the human male species. Conversely, (1b) is an illustration of an individual frame of reference, which credits Juan’s tallness within a specific human being on the basis of the speakers’ previous knowledge of Juan (p. 591). 1 (b) would imply a context such as the following: Juan was 10 years old when the speaker first saw him, now 10 years later he is older and much taller. In this context, there is a clear indication that the referent is compared to a past knowledge or attribute of when he was younger. Note that frame of reference can also be created by lexical means. For example, phrases such as *ya* (‘right now’) or *ahora* (‘now’) allow for the speaker to contrast the referent’s present description to a previous state (Silva-Corvalán 1986: p. 598), although the preferred copula use would not change.

Dependence on experience is another variable that accounts for the choice between *ser* and *estar* among Spanish speakers. Generally “*ser* is associated with defining something abstract, and is independent of immediate experience, while *estar* is associated with concrete and/or immediate experiences” (Silva-Corvalán 1986: p. 590). In review of previous research, dependence on experience is often mentioned (e.g., Gutierrez 1992), but is rarely incorporated into empirical studies of copula choice. To date the only study that has investigated this feature is Geeslin (2003), although her findings revealed it to be a statistically insignificant factor, (p. 742).

The third feature considered by many linguists as an influencing factor in choice of copula is susceptibility to change, *ser* being most often used with referents that are not susceptible to change and *estar* with those are. Susceptibility to change is the variable that differentiates between contexts that are changeable from those that are not (Geeslin 2003: p. 723). A clear example of this would be the size of a car opposed to the size of a boy. Following Geeslin's example, the car will remain the same size and therefore it is not susceptible to change, it would thus use *ser*. The boy will grow into a bigger size as time passes, thus making him susceptible to change, *estar* would more likely be used in this context. According to Silva-Corvalán (1986) susceptibility to change is confirmed to be a factor that influences the choice between *ser* and *estar* (p. 591).

In summary, the semantic distinction between *ser* and *estar* is seen in terms of whether a connection to a locus or another situation is assumed (Maienborn 2005: p.167). In contrast, the pragmatic distinction between *ser* and *estar* may be restricted to a specific topic situation (p. 171). Thus copula choice is strongly dependent on both pragmatics and semantics, more often than on its purely standard forms. This is one of the major causes for the complexity in second language acquisition of the copula, as learning the semantics and pragmatics of a second language is often regarded as more difficult than standard lexical choice. Nonetheless, the following section provides a presentation of the uses of *ser* and *estar* that are generally taught in the classroom. While insufficient for achieving native like use, this information is relevant to the current research because it establishes the information to which the participants of the current

study have been exposed during their formal language instruction. The following section discusses several studies that have looked at copula choice in native Spanish speakers.

Copula Use among Monolingual Native Spanish Speakers

The following paragraphs address various variation studies on copula use found in native speakers of Spanish. In general, the research presented here illuminates the uses of the copula in [adjective + copula] contexts and the corresponding extension of *estar*. Clear evidence of the uses of *ser* and *estar* among native Spanish speakers in monolingual areas are found in Díaz-Campos & Geeslin (forthcoming), Cortés-Torres (2004), and Gutierrez (1992). While Cortés-Torres (2004) and Gutierrez (1992) investigate the speech of Mexicans in monolingual regions (e.g., Cuernavaca, Mexico and Morelia, Mexico respectively), Díaz-Campos & Geeslin (forthcoming), look at the Spanish of speakers from a non-contact area in Caracas, Venezuela.

In some varieties of Spanish (e.g., Mexican, Costa Rican, Puerto Rican, Madrileño and Venezuelan), *estar* has come to be used in non-traditional contexts among native Spanish speakers, leading to the suggestion that the choice between *ser* and *estar* may be more semantic than syntactic in Spanish monolingual speakers (Geeslin 2002: p. 424). Also of note is that the innovative use of *estar* is an ongoing process, meaning that it has not reached a stopping point; *estar* continues to take over functions that have belonged to *ser* (Sterwart 1999: p. 99).

The occurrence of the innovative copula found in native speakers of Spanish is particularly important in the current research because it raises questions about the impact that such variation may have on L2 acquisition, if any. For example, if L2 learners are only exposed to the traditional uses of the copula in a formal setting, can

exposure to such innovative use in other contexts impact their frequency use of the copula verbs?

Díaz-Campos & Geeslin (forthcoming) looked at the effects of several social variables on copula selection in [copula + adjective] contexts, such as age, socioeconomic class and gender on copula selection. They applied a regression analysis, using the Goldvarb statistical program to the data gathered from the Estudio sociolingüístico de Caracas “The Sociolinguistic Study of Caracas”, which is comprised of interviews conducted with 160 speakers that were born in Caracas (p. 14). The participants in this study were divided into four age groups (e.g., 14-29, 30-45, 46-60, and 61 and over) (p. 14).

Their results with regards to these extralinguistics factors were the following: the lower socioeconomic group favored *estar* with a weight of .574 while the upper socioeconomic group neither favored nor disfavored *estar* with a weight of .504 (p. 18). The Goldvarb statistical program reveals that a probabilistic weight of above .5 favors the appearance of the application value (in this case *estar*) and a weight below .5 disfavors it (Tagliamonte 2006). Conversely, the data revealed that the middle class disfavored *estar*. As it relates to age, the results demonstrated that the older speakers (e.g. 61 years old or older) favored *estar*, while the younger group disfavored it (e.g., 14-29 years old) (p. 19). However, when age and socioeconomic level were combined it was found that younger speakers from both upper and lower socioeconomic backgrounds favored *estar* while the older speakers who favored *estar* were only from the lower socioeconomic background (p. 20). In contrast to age and socioeconomic level, the data revealed that neither gender favors or disfavors *estar*.

The authors also analyzed the effects of various linguistic and social variables on copula choice. These included resultant state, adjective class, predicate type, experience with the referent, susceptibility to change, and frame of reference. Interestingly, the results revealed that resultant state, adjective type, predicate type, experience with the referent, susceptibility to change, frame of reference, age and socioeconomic level were strong predictors of *estar* use, however gender (of speaker) is not (p. 24). With attention to adjective type, the results revealed that mental/physical states (e.g., *enojado* 'angry,' *contento* 'happy,' *triste* 'sad,' *parado* 'standing,' *sentado* 'seating,' etc.) favored the use of *estar*. Similarly, status adjectives (e.g., *divorciado* 'divorced,' *casado* 'married,' *soltero* 'single,' and so forth) also favored *estar* (p. 17). Conversely, adjectives that describe observable traits (e.g., *alto* 'tall,' *chaparro* 'short,' *bonito* 'pretty,' *feo* 'ugly,' *grande* 'big') disfavored *estar* (p. 16). Furthermore, the data revealed that those adjectives that were susceptible to change favored *estar*, while those that were not susceptible to change disfavored *estar* (p. 17). The next two studies to be discussed, Cortés-Torres (2004) and Gutierrez (1992), also concentrated on the innovative copula looking only at [adjective +copula] contexts, although with a slightly more detailed classification of adjective types.

Cortés-Torres (2004) considered the following adjective types in her examination of copula use: physical appearance, age, evaluation, and characteristics. Gutierrez' (1992) classification of adjectives was considerably more detailed, and included: age, size, physical appearance, description (*repleto* 'full,' *vacío* 'empty'), moral characteristics (*generoso* 'generous,' *cruel* 'cruel'), class (*español* 'Spaniard,' *católico* 'catholic'), perception (*dulce* 'sweet,' *ruidoso* 'noisy'), color, social status, evaluation

(*barato* 'cheap,' *caro* 'expensive') and others (mental estates and physical states such as *contento* 'pleased,' *falso* 'phony') (Gutierrez 1992: p.117).

Gutierrez (1992) and Cortés-Torres (2004) shared similar results, although with slight differences, on the production of the innovative copula with adjective type. Cortés-Torres (2004) showed that the type of adjective with the most occurrences with the innovative copula was physical appearance with 41% of uses. On the other hand, Gutierrez (1992) revealed a 33% use of *estar* with physical appearance. The second adjective type to appear with a high frequency with the innovative copula was age, which showed 40% use of *estar* in Cortés-Torres (2004) and 43% in Gutierrez (1992).

In regards to *ser*, Gutierrez (1992) affirmed that it has remained consistent among native speakers. *Ser* is almost always used in its traditional use while *estar* is used in its innovative use (p. 121). Nonetheless, *estar* is still more commonly used in its traditional uses than its innovative uses in spite of an evident change in progress. Gutierrez (1992) stated that only 12.5% of the *estar* uses were related to the prescriptive uses of *ser* (p. 120). Similarly, Cortés-Torres (2004) concluded that the use of the innovative copula (*estar*) accounts for a total of only 23% of all uses. Given the results in the above studies, it can be assumed that the innovative copula is in fact present in monolingual Spanish communities in Mexico and Venezuela, even though its diffusion is resistant, and is progressing very slowly among these communities.

To summarize, previous research on monolingual native Spanish speakers has found that the innovative copula is not as prominent among Spanish native speakers. Instead, it has been shown that Spanish native speakers are more likely to use the prescriptive forms of *ser* and *estar* than the innovative copula. Even with this being the

case, it has been found that the innovative copula is used more frequently with certain types of adjectives, (e.g., age, size, physical appearance). However, it is worth noting that the research dealing with choice of copula in Spanish monolingual communities is still relatively limited, and the field would benefit from continued work in monolingual areas.

Uses of *Ser* and *Estar* Taught in the Classroom

The purpose of this section is to show a division between the prescriptive uses of *ser* and *estar* ('to be'). The prescriptive uses presented in this section are extracted from King and Suñer's *Gramática Española: Análisis y práctica* (1999), which is considered a representative example of the kind of textbook L2 learners often use. Table 1-1 demonstrates the prescriptive uses of *ser* and *estar*, which are taught to L2 learners of Spanish.

The contrast between *ser* and *estar* is generally introduced to novice L2 learners in the first year of formal instruction, and it is taught often. Thus we assume that the learner acquires this contrast within the first few years of language learning (Geeslin 2003: p. 703). However, this assumption is false; in reality, research (Briscoe, 1995; Guntermann 1992; Geeslin 2001) has shown that learners have difficulty in acquiring the various contexts of *ser* and *estar*, in particular locatives and pre-adjectival contexts that require *estar* (Geeslin 2003: p. 704). Even after many years of instruction, the L2 learner often still finds himself choosing the erroneous copula (Geeslin 2003: p. 703) for the given context.

Some of these learner problems stem from the way textbooks present the copula. Prescriptive uses of the copula tend to be either too oversimplified or too detailed (Mason 1990: p. 506), both of which can prevent the learner from completely grasping

the contexts of *ser* or *estar*: too much information is problematic because it provides the student with too many facts to process all at once, while oversimplification is not helpful either because it does not explain all the different *ser* and *estar* contexts a learner may encounter (p. 506). An example of oversimplification is the taxonomy of the copula that is often made in traditional grammar as was presented above: it is taught that the copula represents either 'permanent' (*ser*) vs. 'temporary' (*estar*) characteristics, or 'essential' (*ser*) vs. 'accidental' (*estar*) events (Luján 1980: p. 166). At first glance, these divisions may seem to work well, but when analyzing all the possible *ser/estar* contexts it is clear that this dichotomy fails in more than one context (Mason 1990: p. 560), for example, the Spanish adjective *muerto/muerta* ('dead'). Death is something that once it occurs, it cannot be reversed, (i.e., death is permanent); however, in Spanish the adjective *muerto/muerta* ('dead') is used with *estar* (the temporary copula). This example highlights the problem in simplifying the contextual explanations as such 'temporary' versus 'permanent' or 'essential' versus 'accidental'.

To further explain, *estar* and *ser* tend to be presented in opposition, so again, permanent vs. temporary characteristics or essential vs. accidental properties (Luján 1981: p.166). Thus, it has been taught that *estar* is used with non-permanent characteristics, or that have the ability to change (Schmitt, Holtheuer & Miller 2004: p.1). Roldán (1974) concurred, stating that "a sentence having *estar* is normally understood to imply that a change has taken place" (p.72). The opposite is true for *ser*, which is generally applied with adjectives that are considered permanent, or an inherent characteristic (Luján 1981: p. 166). Take for example, Juan *está gordo* ('John looks fat') and Juan *es gordo* ('John is fat'). Lujan (1981) explains that in the example with *estar*

one already knows John, therefore knowing that there was a change in his weight, whereas the example with *ser* could only be said by a person who sees and describes John for the first time (Roldán 1974: p. 72). It should be noted here that traditional explanations of when and how to use *ser* and *estar* with adjectives are not always accurate (Luján 1981: p. 167). Luján (1981) and Querido (1976) claimed that “it is always possible to find contexts where the terms that are normally used with *ser*, are acceptable with *estar*”, although “by contrast, the terms that are restricted to *estar* are without exception, unacceptable with *ser*” (Lujan 1981: p. 172). Some additional illustrative examples are provided in Table 1-2.

To further complicate matters, it has been claimed that 80% of Spanish adjectives may either be used with *ser* or *estar* (Woolsey 2006: p. 180). Further, Silva-Corvalán (1986, 1994) suggests that while either copula could be used appropriately with most adjectives, the copula choice may change the meaning of the utterance, thus indicating that *ser* and *estar* are not interchangeable. For example observe the following examples:

(2) *Mi abuelo es muy viejo.*

‘My grandfather is very old.’

(3) *Mi abuelo está muy viejo.*

‘My grandfather is very old.’

While the use of *ser* identifies the grandfather as an older person, a normal identification for a person that is aging throughout the years (King & Suñer 1999: p. 228), the use of *estar* is more complex because there have to be distinctive contexts for its use. These possible contexts, assuming the use of *estar*, are as follows: (1) the

grandfather recently went through surgery and the speaker went to see him at the hospital or (2) the speaker just visited his/her grandfather after 15 years of not seeing him. In the first context the use of *estar* indicates that the grandfather looks older than before he went into surgery, that a sudden change has taken place. For the second context the speaker is comparing her grandfather to a previous state (of 15 years ago), thus using *estar* to express this comparison. The speaker has to consider the context in order to express the true essence of the utterance by either using *ser* or *estar*. Essentially, the speaker has to create the meaning, which cannot be found in the plain sense of words or structures (King & Suñer 1999: p. 228-229).

The following section presents literature on copula choice among Spanish/English bilinguals. This section is important because it provides information on general tendencies found among bilinguals, thus allowing for a better understanding of copula choice in bilingual speech. (Recall that in the current study, one of the two groups lived continuously in a Spanish/English bilingual speech community).

Copula Use in Spanish/English Bilingual Speakers

This section presents studies of Spanish/English bilinguals and/or heritage speakers (non L2 learners) in the United States and Canada. Valdés (2000) defines a heritage speakers as an individual who is raised in a home where a non-English language is spoken, who speaks or merely understands the heritage language, and who is to some degree bilingual in English and the heritage language (p. 1).

The studies presented here highlight some *ser/estar* patterns found in Spanish/English bilinguals, revealing that they may manifest different patterns than non-contact native Spanish speakers. These studies include those by Silva-Corvalán (1986, 1994), Gutierrez (1994, 2003) and Bruhn de Garavito &Valenzuela (2006). Silva-

Corvalán (1986, 1994) focused on the three generational groups of Mexican-Americans in Los Angeles. Gutierrez (2003) also studied the Mexican-American communities in Houston and Los Angeles. In contrast, Bruhn de Garavito &Valenzuela (2006) studied Spanish/English bilinguals in Canada. All four studies focused on adjective types.

Silva-Corvalán's (1986) data revealed that the type of adjective that produced the highest percentage of innovative copula use among Mexican-Americans was size with 84%, followed by physical appearance with 74%, then age (i.e., *viejo* 'old' vs. *joven* 'young') with 73%. Silva-Corvalán's later (1994) study also corroborated the innovative use of *estar* among the Spanish/English community in Los Angeles: a total of 1,686 tokens of *ser* and *estar* were coded with 49% of all the *estar* constructions considered innovative (p.105).

Similarly, Gutierrez (1994, 2003) looked at two communities one in Houston and the other in Los Angeles, and discovered that the innovative copula was used in [copula + adjective] contexts. In Los Angeles the results revealed that certain types of adjectives appear more frequently with the innovative copula. For example, age (e.g., *viejo* 'old' vs. *joven* 'young') revealed the highest frequency of innovative copula use, with 79% of all occurrences of age appearing with *estar*. Size (e.g., *grande* 'big,' *pequeño* 'small') appeared with the innovative copula 53% of the time, and perception (e.g., *dulce* 'sweet,' *ruidoso* 'noisy') occurred 47% of the time with the innovative copula. Likewise in Houston, age also appeared to be linked closely to the innovative use of *estar*, 57% of the references to age appearing with *estar*. Physical appearance (e.g., *guapo* 'handsome,' *suave* 'soft,' *hermosa* 'gorgeous') followed with 53%, and then perception, 50%. These results indicate that while there are some slight differences between the

production of innovative uses between Houston and Los Angeles, the presence of the innovative copula is indeed evident in these Spanish/English bilingual communities.

Previous research reveals that new generations of Spanish/English bilingual speakers show a high frequency of the innovative use of *estar*, and more importantly that this innovation process is accelerated by the linguistic contact situation (Gutierrez 2003: p. 170). Silva-Corvalán (1994) states “language contact and lack of formal instruction contribute to the propagation of a change in the secondary language” (p.115). In other words, the progression of the innovative copula has no correlation to the comparison of verb forms (e.g., *estar*, *ser*, ‘is’) between English and Spanish (p.115). “Language contact manifests itself not exclusively as transfer from the contact language, but perhaps in simplification or selective acquisition (p.115). Others concur that contact with the English language “is not a trigger for change in itself but is merely accelerating a change in application of the copulas, which is already underway” (Stewart 1999: p. 99).

The following section discusses various studies on the L2 acquisition of *ser* and *estar*. The section starts with Van Patten (1985, 1987), who was among the first to introduce certain stages of copula acquisition. This is followed by other studies that have also indicated that for the most part these stages are accurate, while providing additional information.

Second Language Acquisition of *Ser* and *Estar*

Van Patten (1985, 1987) carried out a longitudinal study where he looked at the stages of acquisition of *ser* and *estar* among six college level adults of various proficiency levels all of whom were enrolled in the same communicative based class. The data were gathered through a series of oral interviews and picture description/story

telling. Van Patten's findings revealed a series of acquisition stages of *ser* and *estar*, which are summarized in example (4) below:

(4) Van Patten's order of acquisition of *ser* and *estar*

1. The omission of both copulas

Ella ___ bonita. ('She pretty.')

Rosie _____ China. ('Rosie Chinese.')

2. *Ser* is the copula of choice by L2 learners

Rosie es relajando sola. ('Rosie is relaxing alone.')

Los padres son gordos. ('The parents are fat.')

3. *Estar* appears with the progressive

Rosie está haciendo su tarea. ('Rosie is doing her homework.')

Rosie está llevando su suéter. ('Rosie is wearing her sweater.')

4. *Estar* appears with location

Rosie está en la biblioteca. ('Rosie is at the library.')

La comida está en el plato. ('The food is on the plate.')

5. *Estar* appears with adjectives of condition

Rosie está pobre. ('Rosie is poor.')

Rosie está penosa. ('Rosie is shy.')

Ryan & Lafford (1992) set out to test Van Patten's stages of copula acquisition, which had been based on classroom data. They carried out a similar longitudinal study, but using data from 16 L2 learners on a study abroad program in Granada, Spain. These students lived with Spanish families and met for three hours a day of formal instruction in Spanish throughout their stay in Granada. The authors only investigated

lexical choice of *ser* and *estar*, and not the correct/standard conjugation of the verbs, (i.e., for person, tense, or number). However, they did determine if the use of *ser* and *estar* was required in contexts, that is to say, as long as some recognizable form of either of *ser* or *estar* was chosen, then it was considered accurate and if neither copula appeared in a required context then it was marked as wrong (Ryan & Lafford 1992: p.715). Their findings supported Van Patten's (1985, 1987) results, with the exception of two stages: locatives and adjectives of condition with *estar*. While Van Patten (1985, 1987) had found that the locative with *estar* is acquired first, and then the conditional (p. 403), Ryan & Lafford (1992) found that the conditional with *estar* was acquired first, then the locative (p. 720).

Gunterman (1992) also analyzed the stages of acquisition of the copula, and found similar results to Ryan & Lafford (1992). Her data were gathered from oral examinations evaluating the Spanish proficiency of twenty Peace Corps volunteers. Her results also revealed that the adjectives of condition are acquired before the locatives. The authors of both studies (Ryan & Lafford 1992; Guntermann 1992) attribute this difference to the fact that the Spanish learners were exposed to more natural input than those in Van Patten's (1985, 1987) study: the L2 learners in Granada, as well as the Peace Corp volunteers, were exposed to numerous situations where they had to make sense of unfamiliar settings and learn from their mistakes to adapt to the new environment (Ryan and Lafford 1992: p.721). For example, many students in Granada were unaware of the different schedules of various businesses, and may have had to return several times to it on the same day, always to see the sign *cerrado* 'closed' or *abierto* 'open,' thus this

sort of regular contact is credited for the early acquisition of the conditional use of *estar* in L2 learners who are immersed in the target language (p. 721).

Other studies that have looked at the acquisition stages of *ser* and *estar* include Briscoe (1995) and Ramírez-Gelpi (1995). Briscoe (1995) data is from 77 L2 learners who were enrolled in different semesters of Spanish, ranging from beginning level to advanced level learners. The students completed a picture-based oral story-telling activity. Briscoe's findings seem to support Van Patten (1985, 1987), although with one slight difference: not all functions of *ser* were over-generalized at first but instead were acquired after some functions of *estar* (p. 711).

Another author that also looked at the stages of acquisition was Ramírez-Gelpi (1995). His data were elicited through student compositions. Unlike the other studies mentioned thus far, which only coded for [*estar* + locative], Ramírez-Gelpi coded for prepositional phrases and classes of adjectives. This methodological difference makes direct comparisons to Van Patten (1985, 1987), Ryan and Lafford (1992), Gunterman (1992) and Briscoe (1995) almost impossible (Geeslin 2003: p. 713). Even so, Ramírez-Gelpi's (1995) findings indicate also that L2 learners are not accurate with *estar* in adverbial or adjectival contexts.

In review, Van Patten (1985, 1987), Ryan & Lafford (1992), Gunterman (1992), Briscoe (1995) and Ramírez-Gelpi (1995) have confirmed that most learners acquire *ser* and *estar* through a series of stages. These stages begin with the omission of both copula, and then result in *ser* in most functions (even when it is erroneous). In regards to *estar*, only those students in the more advanced stages of acquisition begin to incorporate that verb in obligatory contexts. We can then assume that the L2 learners in

this study have passed all the acquisition stages of *ser* and *estar*, since they are in their advanced stages of acquisition.

Copula Use in Second Language Learners

The present section discusses various studies that have investigated L2 copula choice. This section discusses L2 learners' copula choice and compares their choices with those of native Spanish speakers. The purpose of this section is to bring forth the similarities and differences of copula use among L2 learners and Spanish native speakers, as documented by previous research in order to set the scene. The present study, takes the investigation of L2 copula choice further than previous work has in the hopes of addressing unanswered questions.

Geeslin (2000, 2003), Geeslin & Guijarro-Fuentes (2006) and Woolsey (2006) all compare copula choice between native speakers and advanced learners of Spanish. Geeslin (2003) analyzed adjective type. Interestingly, her results revealed no significant difference in the overall use of *ser* and *estar* between the Spanish native speakers and the advanced L2 learners (p. 738). She affirmed that advanced learners of Spanish, use the copula like native speakers of Spanish, 45.4% and 43.6%, respectively. It should also be noted that these overall percentages do not tell us if speakers are actually using them in the same way, but only at about the same rate. On the other hand, the results for variation in the L2 learner group were inconclusive.

Geeslin (2000) and Geeslin & Guijarro-Fuentes (2006) established that copula choice is related to L2 proficiency level. These studies looked at moderate to low proficiency learners rather than advanced. Their results revealed that the learners' selection of *estar* tends to be lower than that of native speakers of Spanish (Guijarro-Fuentes 2006: p. 151). However, as time passes learners do seem to progress in their

accuracy of the copula, indicating that over time L2 learners are able to approximate native copula choice. Woolsey (2006), following up on previous studies, looked at L2 learners at four levels of proficiency in Spanish, from beginning to moderate excluding the advanced level. Overall, the data confirmed again that *ser* is the most frequently used, representing a rate of 76.5% among the data collected, while *estar* tokens were only at a rate of 8.4% among all four levels of proficiency in L2 learners (p. 185). Moreover, the results show that in the 3rd and 4th levels of proficiency, *estar* and *ser* are commonly applied when there are explicit prompts to compare the referent with itself (p. 186).

Woolsey compared 'now' frames versus 'before' frames. The results revealed that there is a progression between the first three proficiency levels in favoring *estar* in 'now' frames. For example, level one barely produced *estar* in 'now' frames but as the level increases, the production of *estar* is more frequent (p. 189). Recall that Geeslin & Guijarro-Fuentes (2006) also noted that as proficiency in Spanish improves so does the native-like selection of *ser* and *estar*. In sum, the results revealed that *ser* is the copula of choice by second language learners of Spanish who are native speakers of English and as proficiency increases so does the use of *estar*.

Also relevant is Bruhn de Garavito & Valenzuela's (2006) study, in which they compared copula choice between native Spanish speakers from various countries, Spanish/English bilingual speakers, and second language learners of Spanish. Unlike the other studies described above, participants in this study did not produce the copula verbs but instead were asked to rate their use by means of a grammaticality judgment task. Participants were provided with two sentences, which only differed in their use of

ser or *estar*. The speakers judged each sentence on a scale from 1 (ungrammatical) to 5 (grammatical). Results revealed that the L2 learners accepted all sentences to a greater degree than the other two groups, regardless of copula choice, although there was no significant variation across participants. The researchers did note an obvious difference between the Spanish monolingual speakers and the second language learners. On the other hand, the difference between the Spanish/English bilingual speakers and Spanish monolingual speakers was relatively small, indicating native-like tendencies among the former. Taken as a whole, the results indicate that generally all L2 learners still have not acquired the copula to the degree of a Spanish monolingual speaker or the Spanish/English bilingual.

Previous SLA/Variation Studies

The current section reviews SLA research with a variationist approach that adopt “quantitative techniques to deal with the variability among learners and to uncover the regular differences between speakers, often associated with extralinguistic variables” (Mougeon & Dewaele, 2004: p. 295). The research included here includes studies of *ser/estar* use among L2 learners of Spanish (Geeslin & Guijarro-Fuentes, 2006; Geeslin, 2003; and Geeslin, 2002) as well as, research outside Spanish and copula choice (Bayley, 1996; Regan, 1996; and Dewaele 2004).

Geeslin & Guijarro-Fuentes (2006) carried out an extensive study of Spanish copula choice among L1 Portuguese speakers. Their findings revealed that L2 learners of Spanish and native speakers of Portuguese (Spanish and non-Spanish learners) selected *estar* more often than native Spanish speakers. Portuguese also has a similar copular system, which implies that the same factors might predict the choice of *estar* in both languages; however the frequency with which *estar* is chosen is not the same (p.

90). The results revealed that Portuguese learners of Spanish incorporate *estar* in their speech more often than native Spanish speakers.

In contrast, Geeslin (2003) focused on the differences between Spanish native speakers from 10 different countries (e.g., Spain, Puerto Rico, Argentina, Chile, Colombia, Mexico, Costa Rica, and Uruguay and U.S) who were studying in the United States at the time of the study. Out of the 25 participants, 14 were in a Master's program and 10 in a doctorate program (p. 730). She considered the following social factors: age, gender, level of study and field of study (p.731). In addition, she looked at several linguistic factors: predicate type, susceptibility to change, adjective class, copulas allowed, resultant state, frame of reference and experience with reference (p. 732). Her results revealed no significant difference in the overall use of *ser* among participants. She states "that it is unlikely that there is a statistically significant difference between 45.4% use of *ser* among native speakers and the 43.6% use of *ser* among advanced learners" (p. 738). Moreover, she also confirmed that there is no significant difference found in *estar* frequency between non-native speakers and native speakers, with 48.5% and 43.7% respectively (p. 738).

In other words, the L2 learners from Geeslin (2003) used the copula in the same contexts as native speakers did, and there is no variation found between L2 learners and native Spanish speakers. Yet in a later study, Geeslin & Guijarro-Fuentes (2006) discover contradictory findings: they affirm that the advanced learners from Geeslin (2003) in fact do not select *ser* and *estar* in the same manner that native speakers do (2006: p. 71). Relevant here, variation among advanced L2 learners may perhaps

depend on social and contextual factors such as age, gender, education level, number of years abroad, syntactic, semantic and pragmatic features (Geeslin, 2000: p. 52).

Since the current research on advanced second language learners of Spanish and copula choice is limited, we turn also to other L2 research to expand understanding of L2 variation, Bayley (1996), Regan (1996) and Dewaele (2004) are three such studies. While Bayley (1996) considered Chinese low and advanced learners of English, Regan (1996) and Dewaele (2004) investigated advanced learners of French. Regan (1996) and Dewaele (2004) concentrated on ne deletion or retention between advanced second language learners of French, while Bayley (1996) focused on [t/d] deletion and the effects of competing rules on past tense marking among L2 English learners.

Bayley (1996) considered the social networks of the participants as a variable in the deletion of [t/d] on consonant clusters (p. 98). He claimed that speakers with little opportunity for acquisition, that is speakers who interact mainly with other Chinese, appeared to be more advanced in the acquisition of inflectional [t/d] than speakers with plenty of opportunities for acquisition (p. 114).

Dewaele (2004), on other hand, found that L2 learners use informal variants less frequently than native speakers of French. “However, prolonged contact with native speakers of French and active use of the language stimulate the use of vernacular speech and lead to an increase in frequency of informal variants” (p. 444). In comparison, Regan (1996) claimed that advanced L2 learners acquire the details of variability, which exist in the speech of the native community fairly often (p. 19).

Worth reiterating, current study explores the likelihood of interlanguage variation in Spanish as being systematic and not random (Geeslin 2006: p. 58). Particularly, it

investigates the importance of contact with Spanish with respect to copula choice of advanced L2 learners, and additionally answers questions regarding language variation in advanced learners of Spanish.

The following chapter describes the methodology used to obtain the data for the current study. Again, this SLA study takes a variationist approach to examining the linguistic variable of Spanish copula choice. A variable rule analysis (Goldvarb) is used to separate, quantify and test the significance of the effects that social and linguistic factors have on the linguistic variables (i.e., Poplack 1993: p. 273). The goal of this study is to find out which linguistic and extralinguistic factors favor the frequency of *estar* and, more specifically, to see if the type of community in which the L2 learners lived continuously has any impact on the frequency of copula choice. Subsequent chapters present and discuss the results obtained.

Table 1-1. Prescriptive uses of *ser* and *estar*

<i>Ser</i>	<i>Estar</i>
1. With a noun/ pronoun that identifies the subject	1. To indicate the location of objects and persons
2. With adjectives or nouns that identify the nationality, religious, and political affiliations, or occupation of the subject	2. With progressive expressions
3. With adjectives to express characteristics of the subject such as size, color and shape	3. With adjectives to express a physical or mental/emotional state or condition of the subject
4. With the preposition <i>de</i> to indicate origin or possession, and to tell what material something made of	4. With a participle to describe the resultant condition of a previous actions
5. To indicate where and when events take place	5. To express change from the norm, whether perceived or real
6. To express dates, days of the week, months, and seasons of the year	
7. To express time	
8. With the preposition <i>para</i> to tell for whom or for what something is intended In impersonal expressions	
9. With the past participle to express the passive voice	

Table 1-2. Examples of *ser* and *estar* uses with adjectives

Adjective	Meaning	<i>Ser</i> example	<i>Estar</i> example
<i>Bonita</i>	'pretty'	<i>Ella es bonita</i> She is pretty	<i>Ella está bonita.</i> She looks pretty. (Context needed)
<i>Gorda</i>	'fat'	<i>Ella es gorda.</i> She is fat.	<i>Ella está gorda.</i> She looks fat. (Context needed)
<i>Redonda</i>	'round'	<i>La luna es redonda</i> The moon is round	<i>La luna está redonda.</i> The moon is round. (Context needed)
<i>Grande</i>	'big'	<i>La casa es grande.</i> The house is big.	<i>La casa está grande.</i> The house is big. (Context needed)
<i>Católico</i>	'catholic'	<i>Él es Católico.</i> He is catholic.	***
<i>Cauto</i>	'cautious'	<i>Él es cauto.</i> He is cautious.	***
<i>Capaz</i>	'capable'	<i>Él es capaz.</i> He is capable	***
<i>Cortes</i>	'polite'	<i>Él es cortes.</i> He is polite.	***
<i>Vacio</i>	'empty'	***	<i>La taza está vacía.</i> The cup is empty.
<i>Descalzo</i>	'bare-footed'	***	<i>Él está descalzo.</i> He is bare-footed.
<i>Satisfecho</i>	'satisfied'	***	<i>Él está satisfecho.</i> He is satisfied.
<i>Desnudo</i>	'naked'	***	<i>Él está desnudo.</i> He is naked.

Note: *** The adjective cannot be used with copula.

CHAPTER 2 METHODOLOGY

This chapter presents an account of the methodological design of the current study, providing descriptions of the procedure and justifications for the methods chosen. The first section discusses the research questions that guided the experiment and presents the hypotheses. The second section provides definitions and distinctions for the terms ‘speech community’ and ‘social network,’ two concepts that are often mistaken as interchangeable and that are crucial to the present work. Likewise, a description of the relevant populations of Florida is given, particularly those regions where the L2 learners of the current study lived continuously. The third section provides a detailed description of the participants, followed in the fourth section by a discussion of the materials and procedures for data collection. The last part of the chapter explains and justifies the process of data analysis (e.g., coding and statistical program).

Research Questions and Hypotheses

The following questions guided the current research.

1. How does frequency of *estar* use among advanced second language learners of Spanish who have lived in a monolingual English area (MONOSC Group) compare to that of similar learners who have lived in a Spanish/English bilingual area (BISC group)?

Hypothesis: As discussed in Chapter 1, previous research confirms that advanced L2 learners have the ability to appropriately select between *ser* and *estar*; their choices can be compared to those found in native speakers. With this in mind it can be hypothesized that the participants from the BISC group will evidence different frequency levels of copula use from their counterparts. It also seems reasonable to note that L2 learners who do not have contact with native speakers of Spanish generally might

select *estar* less often than those that do (although it should be noted that native speakers of Spanish do not accept *estar* everywhere) since Gutierrez (2003) and Silva-Corvalán (1994) affirm that situations of language contact may accelerate changes that have been originated in monolingual environments (p.175).

2. How do extralinguistic factors (e.g., gender, number of years of formal instruction in Spanish, age of formal instruction, and duration in a study abroad program language) correlate to the frequency of *estar* among L2 learners from a Spanish/English bilingual community? Do these factors vary in L2 learners from an English monolingual speech community?

Hypothesis: Based on limited previous research and the fact that previous work has often found contradictory results it is difficult to hypothesize the impact of extralinguistic factors on the frequency of the copula choice, particularly the choice of *estar*.

Gender: Previous research has shown that women use *estar* more frequently in Spanish/English bilingual communities whereas men show a higher frequency of *estar* in monolingual areas, therefore, it is hypothesized that gender will influence the frequency of the copula choice in both groups (e.g., BISC and MONOSC group). It is hypothesized that in the MONOSC group, women will have a lower rate of *estar* uses while the BISC group will show women with higher rates of *estar* uses.

Number of years of formal instruction in Spanish and age of formal instruction: As addressed above, there is a limited number of previous studies investigating extralinguistic factors and copula choice; (Guijarro-Fuentes & Geeslin, 2003, Geeslin, 2003, Geeslin 2005 and Guijarro-Fuentes & Geeslin, 2006) nonetheless the existing

research supports the following hypothesis. It is hypothesized that because the participants in the current study are advanced L2 learners, the number of years of formal instruction in Spanish and age of formal instruction should not affect frequency of copula choice because after a certain number of years of study, evidence indicates that learners no longer continuously modify their grammars (Guijarro-Fuentes & Geeslin 2003: p. 102).

Duration in a study abroad program: It is hypothesized that duration in a study abroad program will have an effect on the frequency of copula choice, regardless community type. It is hypothesized that all the participants, who have studied abroad for considerable periods of time, will have higher rates of *estar* than those participants who experienced shorter durations abroad. It is worth noting that there are very few studies dealing directly with frequency of copula use and time abroad (namely Van Patten 1987; Dekeyser 1990; Ryan & Lafford 1992; Geeslin 2005). As addressed above, all the latter studies show some sort of effect on copula production, with the exception of Geeslin (2005).

3. Which linguistic factors (e.g., adjective class, collocations, TAM, subject class, person and number) trigger the frequency of *estar* use among L2 learners from a Spanish/English bilingual community? Do these factors vary in L2 learners from a monolingual English speech community?

Hypothesis: Most of the studies on copula use only look at [copula + adjective] contexts, whereas the current study includes various contexts (e.g., prepositions, adverbs, pronouns, conjunctions, negations etc.). Based on the very limited research for some of the linguistic factors (namely TAM, grammatical person and animacy,

collocations and subject type) it is rather difficult to hypothesize which of them should affect the frequency of the copula. Nonetheless, the following can be proposed:

Adjective class: It is hypothesized that adjective class will influence frequency of copula use. Previous research (i.e., Gutierrez 1992, Cortés-Torres 2004, Silva-Corvalán 1994 and Gutierrez 1994) shows that adjectives of age, size, and physical appearance generally result in the highest rates of *estar*. Other types of adjectives generally produce lower numbers of *estar*.

Grammatical person and animacy: It is hypothesized that animacy and person will not influence the frequency of *estar*, regardless of community type. Previous research (Geeslin & Guijarro-Fuentes 2000 and Silva- Corvalán 1986) revealed that animacy does not predict frequency of copula choice. **Collocations and TAM:** Given the lack of previous research in this area, it was difficult to hypothesize whether collocations and TAM will have an effect on the frequency of *estar*. Nonetheless, it can only be assumed that collocations and TAM should have some sort of influence on the frequency of copula choice.

In order to address these research questions, the experiment described in the following sections was designed to collect data from L2 advanced learners from higher-level Spanish courses at a large research university in the United States.

In studying language change and variation (geographical or social), reference to the speech community is “inescapable, yet there is remarkably little agreement or theoretical discussion of the concept in sociolinguistics, though it has often been defined” (Patrick 2002: p. 574), and as a result the following section presents and discusses literature on the importance of speech communities within variationist studies.

The section below also highlights the difference between social network and speech communities.

Speech Communities

The L2 learners in this study come from two speech communities: Spanish/English bilingual and English monolingual speech communities. In order to obtain a clear representation of the social network structures of the individuals in the current study, a description of the communities is provided here. The large number of Hispanics in Florida is reflected, particularly in the Spanish/English bilingual community, the influence that Spanish has had on the participants of the current study.

Arriving at a general definition of a speech community is a complex task, as seen in previous research (e.g., Weinreich 1953; Ferguson 1959; Stewart 1962; Gumperz 1968; Fishman 1971; Hymes 1972; and; Romaine 1982). Romaine (1982) acknowledges it is “counter-productive, to question a concept like speech community, or to even suggest that there is a problem in defining it” (p. 13), although it is a term that is often considered too difficult to investigate and is frequently ignored in variation studies (Patrick 2002: p. 574). However, various linguists (e.g., Weinreich 1953; Ferguson 1959; Stewart 1962; Gumperz 1968; Fishman 1971; Hymes 1972; and; Romaine 1982) have provided their own explanation with an attempt to contribute a universal definition to the field.

One issue in classifying a speech community is deciding whether to identify it based purely on social criteria, linguistic criteria, or both. Even more difficult is determining how the extralinguistic and linguistic factors are linked, and to what degree they impact each speaker in a particular setting (Romaine 1982: p. 13). A more important issue is establishing if there is homogeneity in a speech community, and if we

can treat it as such. Hymes (1974) suggests we recognize that there are both “kinds of language” and “uses of language” within any speech community (p. 50). For example, while all the members of a speech community share the same kinds of language (e.g., choice of *ser* and *estar*, or the use of centralized variants of /aw/ and /ay/ from Martha’s Vineyard (Romaine 1982: p. 13)), not all members of the speech community use these variants in the same ways in their speech.

Despite the complexities found in developing a definition of a speech community, it would be inadequate to completely dismiss it from the research at hand; therefore the following definitions will be adopted for this study. Fishman (1971) describes a speech community “as a subtype of community of whose members share at least a single speech variety and the norms for its appropriate use (Fishman 1971: p. 28). Patrick (2002) affirms as well that, “speech communities are collectivities of social networks” (p. 581). These two definitions are those employed in this research, as they take into account that a speech a community is a combination of social networks in which the speakers share a common way of speaking.

Bayley (1996) defines a person’s social network as his place of work, friends and random everyday contact with speakers of the target language. Milroy (2002) adds that a social network is a “boundless web of ties which reaches out through a whole society, linking people to one another, however remotely” (p. 550). It seems reasonable to suppose that the members from a social network most likely share the same kind of language and as result also use it in their speech since social networks are composed of more restricted associations from speaker to speaker. There are two types of social networks described by Milardo (1988), exchange and interactive networks. Exchange

networks consist of “relatives and close friends with whom the ego interacts regularly, and exchanges advice, criticism, and support” (p.27). On the other hand, interactive networks consist of “persons with whom the ego interacts frequently and perhaps over prolonged periods of time but on whom he/she does not rely for material or symbolic resources, such as the interaction between a storeowner and customer” (Milardo 1988: p. 29). These definitions are applied to the study at hand, although exchange social network has been operationalized to include only friends and not family members who spoke Spanish, otherwise this would indicate that participants are heritage speakers and, as explained above, heritage speakers are excluded from the present study.

The L2 learners of this study lived continuously in the following bilingual Spanish/English speech communities, hence making it important for these to be described in detail. The Spanish/English bilingual communities in the current study include Miami-Dade county area, Hillsborough County (Tampa) and Fort Lauderdale-Broward County. It should also be noted here that the numbers for the Hispanic population presented here are gathered from ‘designated market area’ or DMA references, the Hispanic population in these areas (e.g., Miami-Dade county area, Hillsborough county (Tampa) and Fort Lauderdale- Broward county) is determined by television and radio station ratings and viewing areas thus providing a more detailed number of native Spanish speakers.

Some of the largest Hispanic markets in the United States are Los Angeles, Houston and Miami-Ft. Lauderdale area. Los Angeles is projected to reach a Hispanic population of 8,419,831 by the year 2014, while Houston’s Hispanic population is estimated to reach 2,397,462, in the same time period. Similarly, Miami-Ft. Lauderdale

is estimated to reach 2,121,584 by the same year. The majority of participants in the second group are from the Miami-Ft. Lauderdale area. Language use in this area is highly influenced by Spanish thus providing the participants with continuous contact with Spanish. This exposure to Spanish is dependent on interpersonal ties (e.g., acquaintances) but not necessarily. An estimated total of 494, 292 people in this area are Spanish dependent, while only 112, 715 people are English dependent. This means that of the entire population in this area 112, 715 people more likely only speak English while 494, 292 people are only able to speak Spanish with very little English knowledge. While the Hispanic population in the Hillsborough County (Tampa) is not as large as that in the Miami-Ft. Lauderdale, it is relatively high and therefore capable of producing social networks with Hispanic ties. The total population in the Tampa area is 303, 447 people; out of these 58, 522 are Hispanics (of any origin). Most of the Hispanic population in this area is Spanish/English bilingual, unlike the Hispanic population in Miami-Ft. Lauderdale, which remains predominantly Spanish monolingual.

This brings up the next point, which is that although there are a large number of Spanish/English bilinguals, there is an even greater number of Hispanics who generally only speak Spanish in the Miami-Ft. Lauderdale area. These numbers illuminate the undeniable, continuous contact that L2 learners in this area have with the Spanish language and realistically have no choice but to be part of heavily Spanish influenced social networks. The Miami-Ft. Lauderdale area reflects the strong Spanish contact that these L2 learners experienced growing up. As a side note, there is no accurate way of measuring exactly how much contact the participants received as they were growing up in these speech communities, it can only be assumed that they were in fairly consistent

contact with the target language due to the presence of Spanish in the everyday lives of people in the area.

In the following sections a detailed review of previous research on copula use among monolingual Spanish speakers, Spanish/English bilingual speakers and L2 Spanish learners is presented. Most of the existing literature deals with the [copula +adjective] context and the extension of *estar* in innovative contexts that once prescriptively belonged to *ser* (Silva-Corvalán 1994: p. 91). Generally, studies with a focus on the extension of *estar* are based on Spanish/English bilinguals in the United States (e.g., Silva-Corvalán, 1986 and 1994) and monolingual speakers of Spanish (e.g., Gutierrez 1992; Geeslin 2003; Cortés-Torres 2004; Diaz-Campos & Geeslin (forthcoming)). These studies provide evidence that the innovative copula is more advanced in Spanish/English bilingual communities as well as Spanish monolingual communities.

The following section discusses some of the findings on copula use among monolingual native Spanish speakers. Note that these are non-contact areas, meaning that only Spanish is spoken in these areas with no connection to English. The previous research presented here is important because it highlights that innovative copula use is not necessarily dependent on contact with English but instead it may be related to other social and linguistic variables. At the same time the existing literature can also be indicative of the copula use that is found in the speech of Spanish native speakers.

Participants

The L2 learners were divided into two experimental groups according to the type of speech community in which they lived continuously, namely monolingual or bilingual. Each experimental group contained thirteen participants. As the study also examines

gender, an equal number of males and females were included in order to have a balanced representation. All the participants are undergraduate advanced L2 Spanish learners. One group is composed of advanced L2 learners of Spanish who lived continuously in a monolingual English community while the other group had L2 Spanish learners who lived continuously in a Spanish/English bilingual community.

All participants were volunteers, chosen on the basis of three prerequisites for inclusion. First, learners with a Spanish heritage background (i.e., Spanish speaking family) were not eligible, due to the nature of the factors under investigation. Nevertheless, some participants' parents were born in other countries and therefore spoke a language other than English as their native language (see Tables 2-1 and 2-2); however, since participants themselves were all English dominant, they were included. Second, participants had to be L2 learners of Spanish, studying in a formal classroom setting. It is important to highlight here that all the participants have been exposed to considerable formal classroom input regardless of speech community. Finally, all participants were required to be enrolled in advanced university Spanish classes at the time of data collection, although the number of years of prior formal instruction could and did vary among the participants from both groups. It should also be emphasize here that the participants from both groups, MONOSC and BISC had to have learned Spanish in a formal setting. However in the case of the L2 learners from the MONOSC group, because of their minimum contact with native speakers of Spanish, it is hypothesized that these learners have been exposed to only the "standard" linguistic forms learned in the classroom, as has been discussed previously.

Age of the participants is not considered a relevant factor here since all the participants fall between 19-22 years of age. Even so the age at which formal instruction in Spanish began is considered as a potential factor to affect the frequency of *estar*. In the BISC group the ages of acquisition varied between 7 to 17 years old while in the MONOSC group they varied between 10 to 16 years.

All participants were pursuing either a major or minor in Spanish. While seven of the participants were pursuing a minor in Spanish, nineteen of them had declared majors in other areas (e.g., zoology, psychology, English, microbiology, cell science, economics, neuroscience) with Spanish being the second major. At the time of data collection students were enrolled in 4000-level (i.e., fourth year) Spanish courses and were thus considered advanced in their formal instruction. As corroboration of this categorization, there was no major indication that these learners lacked the ability to speak exclusively in Spanish to a native speaker of Spanish (following, for example, Geeslin 2003). Moreover, the majority of the participants (all but 5) had studied abroad in a Spanish-speaking country, although it was not required in order to be included in the present study.

Tables 2-1 and 2-2 summarize the demographic information provided by the participants for the BISC group and MONOSC group, respectively. The tables contain the following information: participant label; chronological age; gender; parents' mother tongue; whether the participants speak that language; the number of years of formal instruction in Spanish; the age at which they began formal instruction in Spanish; and the amount of time, if any, spent in a study abroad program in a Spanish speaking country.

One should recall that speech community and social network are not considered the same construct and these terms are not used interchangeably. Bayley (1996) defines social network as the place of work, friends, and random everyday contact with Spanish speakers that provides the L2 learner with additional input of the language of interest (p. 110). Similarly for this study, any place or activity within the community that gives the L2 learner access to Spanish is part of their social network. Even if the participants who were lived continuously in bilingual speech communities choose not to speak Spanish to native speakers, they were still placed in the bilingual speech community group because of the input to which they have been exposed. In order to determine each participant's social network within the speech community, they filled out the Language Contact Profile (LCP). This form and process will be further explained in the following section.

The first group is made up of advanced L2 learners from Spanish/English bilingual communities in Florida. For purposes of simplicity, this group will be referred from now on as the BISC group. In comparison, the second group will be referred to as the MONOSC group, is composed of advanced L2 learners from predominately monolingual English communities. The speech communities were determined to be either bilingual or monolingual through percentages acquired through the 2009 Census Bureau but also through each participant's self-evaluation. One should also note that in order to maintain accurate assessments, the participants were given two definitions to follow when defining their respective communities as either monolingual or bilingual.

A Spanish/English bilingual community was defined for participants as a community that is heavily influenced by speakers of Spanish. This means that there are

a noteworthy number of public places such as restaurants, grocery stores, shops, gyms, malls, and gas stations, where the language of preference is Spanish. English is the secondary language in these locales and only spoken when absolutely necessary, for example, when an English monolingual enters the premises. Likewise, a bilingual community was also defined as a community where a large number of places of employment are generally owned or managed by Spanish speakers. This means that the employers as well as the employees are Spanish speakers, and the language of preference again is Spanish. In a bilingual community the main goal of the L2 learner is to have the ability to function, as much as possible, as a member of the speech community (Regan, 1996: p. 178). Thus, although the participants in the present study learned Spanish in a formal setting, it is possible, even probable that their interlanguage is impacted by the frequent interaction with Spanish native speakers.

Participants in the BISC group are mainly from the Miami-Dade County area, ten of them claiming this area as their home community. According to the 2009 Census Bureau, this area has 66% Spanish speakers of Hispanic descent, implying that Spanish is the dominant language in this area. The current sample contains two additional participants from Hillsborough County in Tampa, which is also heavily influenced by Hispanics. According to the 2009 Census Bureau, this area is 23% Hispanic; although this is a much smaller percentage than in Miami-Dade County, it is nonetheless much higher than the other communities that are populated by Hispanics in Florida. Lastly, only one person was from West Palm Beach, which reports 18% Hispanic population according to the 2009 Census data.

Generally, the participants in the current study that belonged in the Spanish/English bilingual speech community had interactive social networks. Nonetheless there were some participants who also were part of an exchange social network; although in these cases the exchange social network only included friends and not family members who spoke Spanish

It is worth noting that Spanish/English bilinguals in the United States show diverse levels of proficiency in both English and Spanish. Silva-Corvalán (1994) describes the proficiency continuum in Spanish/English bilingualism in the United States with the following:

“These situations of societal bilingualism an oral proficiency continuum may develop between the two languages in contact. This continuum resembles in some respects a Creole continuum, in as much as one can identify a series of lects, which range from standard or unrestricted Spanish to an emblematic use of Spanish and, vice versa, from unrestricted to emblematic English. At the individual level, these lects represent a wide range of dynamic levels of proficiency in the subordinate language. Speakers can be located at various points along this continuum depending on their level of dominance in one or other of the languages or in both, but it is in principle possible for an individual to move or be moving towards one or the end of the continuum at any given stage of his life” (Silva-Corvalán 1994: p.11).

Thus there is wide range of ability levels among the bilinguals with whom the participants interacted, making it a difficult task to measure their proficiency. At any rate, because there was no accurate method or tool to measure levels of proficiency, it was not considered in the present research. Instead, only the quantity of contact that

participants received was taken into consideration. Figure 2-1 illustrates the percentages of Hispanics in the areas mentioned above.

The second definition provided to all participants was for a monolingual English community. A monolingual community was defined as a community where all businesses (namely restaurants, shopping malls, gas stations, grocery stores and shops) are in English, and the preferred language is English. For members of this community to hear spoken Spanish throughout their daily routine is not common but may occur from time to time. However, it is feasible to go through daily activities with only the use of the English language.

Therefore, the participants in the MONOSC group are from communities with a very low percentage of Hispanic population, where English is obviously the dominant language. Four of the participants are from East Orlando, a major city with 3% Hispanic population, an essentially monolingual area. One participant is from Tallahassee, which has a 4% Hispanic population. There are five participants who claimed Gainesville as their community, a city with 6% Hispanic population. Lastly, there are three participants who indicated Clearwater as their speech community, which has 4% Hispanic population. Figure 2-2 illustrates the percentages of Hispanics in the areas mentioned above.

In an effort to accurately categorize the participants in the communities involved, their social networks within the speech communities are considered. In other words, participants who were part of the Spanish/English bilingual speech community were those who indicated strong social networks with constant contact with Spanish native speakers outside the home. In contrast, those participants who were part of the English

monolingual speech community indicated that they belonged in social networks with minimal exposure to native speakers of Spanish.

In order to assure that all participants were placed in the correct group, and to maintain homogeneity in each group, each participant filled out a demographics questionnaire (see Appendix B) and the language Contact Profile (see Appendix C).

Similar to the BISC group, all of the participants in the MONOSC group learned Spanish in a formal setting. However in this case, because of their minimum contact with native speakers of Spanish, it is hypothesized that the learners have been exposed to only the 'standard' linguistic forms learned in the classroom, as has been discussed previously.

Materials

Data collection began with each participant completing an informed consent form (Appendix A), which explained the purpose of the study as well as the role of the participant. Detailed information was not provided about the variable of interest or about anything else that could have impacted participants' production. Only at the end of each individual session was the participant informed of the specifics of the study.

After the completion of the consent form, participants completed a demographics questionnaire (Appendix B), whose main purpose was to elicit social and linguistic background information, and, as in Geeslin (2001), to ensure that the participants in each group had similar backgrounds with other members of their groups. The questionnaire included a detailed set of demographic questions regarding age, gender, ethnicity, study abroad experience, type of community where they lived continuously, number of years of formal instruction in Spanish, major, minor, and languages spoken.

Proficiency Assessment

First, however, the participants' were asked the following questions: (a) what language do you speak at home? And (b) what is your proficiency level in Spanish? The first question was asked to ensure that no person in the present sample spoke Spanish at home and that they have parents whose mother tongue is not Spanish. The second question was asked to ensure that their level in Spanish was advanced enough to be considered for inclusion in the study. Obviously, this reliance on self-reported data and course enrollment (all the participants were enrolled in advanced courses) is a potential limitation that will be discussed further below. There was no external standard proficiency measure to evaluate participants' language skills, so as part of the questionnaire, each participant self-evaluated his or her own proficiency level in the four Spanish skills (i.e., speaking, writing, reading, listening), ranging on a scale of 0 to 3 (0=poor, 1=good, 2= very good, 3=native/native like).

All participants rated themselves as "native/ native-like" in English, which was expected since this sample includes only participants that are considered native L1 speakers of English. Conversely, when the students were asked to self-evaluate their proficiency level in Spanish, only one participant (B1) in the BISC group said that she was a "native/native-like" in all four skills. Generally participants evaluated their proficiency in Spanish in all four skills as "good" or "very good".

Figure 2-3 displays the self-reported proficiency levels in Spanish for listening, on a scale of 0-3, for the BISC group and MONOSC group, respectively. There were eight participants in the BISC group that self-evaluated their proficiency as "very good" and in the MONOSC group there were seven. In other words, the majority of the participants from both groups self-reported their proficiency level as "very good"; nonetheless there

are four participants who also self-reported their proficiency as “native-like”. Globally, the group with the highest self-rated average in listening is the BISC group at 1.92, although in the MONOSC group its average of 1.85 is not drastically different.

The information provided in Figure 2-4 shows that the majority of the participants from both groups self-reported to have a proficiency level of “very good” in speaking, with the exception of six participants who self-reported their proficiency level as “good”. Moreover, the majority of the participants who self-evaluated their proficiency level as “good” belonged in the MONOSC group, with only one in the BISC group. On the whole, the BISC group reported a higher level of proficiency in speaking, with an average of 1.85 while the MONOSC group reported an average of 1.54 (from a 0-3 scale).

Figure 2-5 displays the self-reported proficiency levels in Spanish for reading, on a scale of 0-3, for both groups. Typically, most of the participants from both groups self-reported their proficiency level as “very good”, however, there were some that also self-reported their proficiency level in reading as “good”. There were only two participants (e.g., B1 and M2) who reported their level as “native-like”. Generally, the MONOSC group shows an average of 1.92 while the BISC exhibits a slightly lower average of 1.77.

Figure 2-6 reveals the self-reported proficiency level data for writing, on a 0-3 scale, for both the MONOSC and BISC groups. It should be noted that the majority of the participants in both groups self-reported their proficiency level as “very good” or better. Nonetheless, there are five participants in the BISC group and three in the MONOSC group that self-reported their proficiency level as “good”. Worth highlighting are the two participants from the BISC group who self-reported their proficiency level as

“native-like”. This sort of evaluation is not common in any of the four skills, however participant B1 is the only one to self-evaluated herself as “native-like” in all skills. Overall, according to the identical averages, 1.77 for both groups, there was no difference found in the groups

The main motivation for having the participants self-evaluate all four skills was to maintain a balanced representation of the data. In other words, it is important to only include individuals who self-evaluated all four skills as good or better. If any individual self-evaluated any of the four skills as “poor” they were to be taken out of the sample. However, given their enrollment in upper level courses, no individual self-reported their proficiency as “poor”, thus no elimination was needed. Overall, the averages presented above show that the participants in both groups perceive their proficiency level in all four skills somewhat the same; even so, it is important to observe the slight dissimilarities which may indicate that type of community perhaps may influence listening, speaking, writing and reading skills differently.

To summarize, the participants who lived continuously in Spanish/English bilingual communities generally rated themselves slightly higher in listening and speaking than in the MONOSC group. These slight differences may be correlated to the amount of Spanish contact that participants have received throughout their lives and perhaps the participants who experienced more input (active or passive) in Spanish may feel more at ease or secure with the language than those participants who have been exposed to little Spanish input in their lifetime, (outside the classroom).

Social Network

As mentioned previously, the social networks of the participants were also considered to better understand the extent to which the participants were exposed to

Spanish in their daily lives. This type of information could not fully be obtained from the demographics questionnaire alone; consequently a second form was included. The Language Contact Profile (LCP) was modified from the form used by Freed, Dewey, Segalowitz, and Halter (2004) and is presented in Appendix C. According to the authors, their LCP “was specifically designed to gather demographics, language learning history, contact with native speakers, and use of the language in the field information” (p. 349). It should also be stated that that because this form has generally been used in study abroad contexts, it was extensively modified to serve the nature of the current study.

The 21-statement questionnaire contained specific statements regarding participants’ amount and type of Spanish contact. Statements were presented to the participants in written form to which they responded with a number (i.e., 1-7) indicating their agreement where 7 equaled “strongly agree”, and 1 was “strongly disagree” Some example statements from the modified LCP questionnaire are given below.

1. I speak Spanish outside class almost always
2. I almost always speak Spanish at home
3. I almost always read Spanish books
4. My significant other is a native Spanish speaker
5. Most people in my community do not speak English

The information provided on this questionnaire allowed participants to evaluate the amount of contact that s/he had experienced with Spanish on a day-to-day basis, and in turn allowed the researcher to explore the influence that Spanish contact may have had in the interlanguage of the L2 learner. It is hoped that future research can shed light on additional input from native Spanish speakers and may shape the way L2 learners apply linguistic variables, such as the copulative verbs, as in the present study.

The percentages were calculated by multiplying the number of total statements, 18, by 7 assuming the highest possible answers in the LCP. A score of 126 would indicate maximum contact with Spanish in an informal setting. A minimum score of 18 (18 statements X 1 = strongly disagree) would reflect minimal contact with native speakers. Participants' responses were added, and then divided by the maximum of possible (126) to result in a percentage that can be loosely interpreted as indicative of the amount of native speaker contact participants experienced. The participants in the BISC group indicate consistently higher percentages of contact with Spanish than the MONOSC group, although there is a certain degree of individual variation.

Participants B4 (37%), B6 (36%) and B9 (38%) have comparatively lower percentages; even so, they are relatively higher than those percentages in the MONOSC group. This may be explained by the fact that the three participants indicated to shy away from speakers of Spanish. Also all three said that they do not feel comfortable speaking Spanish with native speakers (although all three participants self-evaluated their Spanish proficiency level in speaking as (very good), and avoid speaking it unless completely necessary. Nonetheless, one should recall that since these participants lived in a Spanish/English bilingual community they were exposed to Spanish even if this exposure was not actively sought out.

Recall that participants were placed either in the BISC group or in the MONOSC group on the basis of where they lived continuously. Moreover, their groups were determined by census information; nonetheless, having additional information such as the social network of the participants helps paint a broader picture of their overall language environment.

Generally the participants in the MONOSC group reported relatively less contact with Spanish, ranging somewhere below 32% percent. However, as with the BISC group, individual variation exists: participants M3 and M6 have relatively high contact rates because they have significant others who are native speakers of Spanish. Others, like participant M5, M11 and M12 indicated that they always watch television, listen to the radio and read novels in Spanish, which resulted in a comparatively higher percentage of contact.

The LCP form assisted in getting a description of the type of social network (i.e., strong or weak) for each participant. Several of the statements in the LCP form were related to access to Spanish through other types of contact, so an L2 learner who lived in a monolingual speech community could still have considerable contact through other means such as the internet, books, music, television and so forth. However, these sorts of situations still did not surpass the 35% cut-off bracket for inclusion in the monolingual speech community. As previously mentioned, the information provided by the 2009 Census Bureau, and the DMA references had the most weight in deciding which type of community each participant was placed.

However, the description of each student about their speech community as well as the LCP form served as additional information to confirmed that in fact those who lived continuously in a Spanish/English bilingual community were exposed to more contact to Spanish than those who lived continuously in a monolingual English community.

Narratives

The linguistic data were elicited orally through narratives from a children's storybook, *My Family is Forever* by Nancy Carlson (2004). The book contains 28 pictures with different backgrounds, characters and actions. So as to not overwhelm the

participants, only eight pictures were chosen for the description narration task. Each picture was chosen because of the numerous possibilities that it presented in creating a story. The English text on these pictures was hidden so that participants would not be influenced by it, allowing participants the freedom to create their own narrative in Spanish. Ultimately the goal was to create as many contexts of *ser* and *estar* as possible. An IC Recorder ICD-SX57 was used to record each participant's narratives.

The method used to collect data in this project was loosely based on Woolsey (2006), whose procedure prevents the researcher from having to guess the intended meaning of the participants. Silva-Corvalán (1986) agreed that the researcher on the basis of the relevant sentence alone might not determine the acceptability of either copula; the extended discourse and shared knowledge among the interlocutors must also be considered (p. 591). In essence, this type of procedure allows for accurate results because there is no need to hypothesize as to participants' meaning.

Unlike Woolsey (2006), the purpose in the current study is to look at all *estar* and *ser* contexts and not only [copula + adjective] contexts. The participants were expected to create a full story that included a plot, descriptions of the characters, weather, home, food, pets and so forth. Again, this method not only decreases the chance for data to be interpreted inaccurately by the researcher but it also provides means for natural language production. The main purpose of storytelling is to create a lot of descriptive details and natural language. As an added bonus, it has been suggested that when the participant gets into a story telling mode, s/he is more likely to produce vernacular speech (Tagliamonte 2006: p. 38).

One should also note that semantic/pragmatic considerations of the choice of the copula by native speakers of Spanish could be classified as one of the most complicated issues that a researcher must investigate. This is because it is complicated to carry out an investigation in which the researcher is able to know the intended meaning of the speaker. Silva-Corvalán (1986) affirmed that the researcher on the basis of the relevant sentence alone might not determine the acceptability of either copula; the extended discourse and shared knowledge among the interlocutors must be considered (p. 591). As seen in the following example:

(5) *Elena está bonita.*

'She looks pretty (today).'

(6) *Elena es bonita.*

'She is pretty.'

Often the researcher needs to deduce the intended meaning of such examples as the above if he/she does not know the referent; in this case it is Elena. Situations like these will often alter the data. Researchers such as Cortés-Torres (2004), Geeslin (2003), Silva-Corvalán (1986, 1996), and Gutierrez (1992) admit at times to deducing the intended meaning of the speaker. Because of their presumptions in the intended meaning, it can be assumed that it had an effect on the results: however; it is not clear to what degree.

Woolsey (2006) considered the fact that as a researcher one might deduce an erroneous intended meaning of the speakers' responses. As a result he designed a data-elicitation instrument that freed him from having to deduce the participants' meaning because the speaker himself revealed it. To further explain, Woolsey (2006)

made a case for confirming the speaker's intent in [copula + adjective] contexts through a picture description task that explicitly prompts comparisons within an individual frame (p.182). Familiarity with the referent is controlled with the use of famous celebrities, such as Britney Spears, Harrison Ford, and Michael Jackson.

To avoid any potential discrepancies in the data, the participants were provided with the plot as well as with the name of the main character, Rosie. Also they were given specific guidelines to follow as they created their narratives (Appendix D). However, the guidelines were kept to a minimum so as not to intervene in the natural process of language production. The guidelines included: (1) must speak approximately 20 minutes; (2) must only speak Spanish; (3) must include specific details of each picture provided; and (4) must create a complete narrative. The fourth was included in order to avoid having a list of descriptions rather than a story with characters, actions, and descriptions.

The researcher was not present at the time of the recording, in an effort to ensure that the participant felt at ease so that language production would not be altered by feelings of discomfort to the extent possible. The participants were in a private conference room at the University of Florida campus and the primary role of the researcher was to greet them, explain instructions and turn the recorder on and off.

Analysis

The current section describes the factor groups that were included in the analyses and also explains the process used in deriving the final statistical model. A total of eight hours and 40 minutes (520 minutes) of recordings were collected from then participants in the two experimental groups. Each recording was transcribed in its entirety and each occurrence of *ser* and *estar* was extracted. There were a total of 2,174 situations where

ser or *estar* were applied: this number includes the production of *ser* and *estar* as well as all possible contexts where either could occur. It is worth recalling that the focus of the present study was to look at the frequency of *estar* and not any supposed errors produced by the participants. As a result, every context where *estar* and *ser* appeared was part of the analysis and therefore coded; this means that even the erroneous contexts of copula use were part of the analysis. Like Geeslin & Guijarro-Fuentes (2006) there was no coding for accuracy or any other type of variable that required the researcher to determine if something was target-like or not. The decision if a copula is being used in its correct form is too arbitrary and generally reflects the personal judgments of the researcher (Geeslin & Guijarro-Fuentes 2006: p. 56). Coding for this study was done following Geeslin & Guijarro-Fuentes' (2006) coding scheme, with some slight modifications. Namely, here every context of *estar* and *ser* was coded, while Geeslin & Guijarro-Fuentes (2006) coded only for [copula + adjective] contexts.

Linguistic Variables

The dependent variable is the use of *estar* and *ser*. The independent linguistic variables are grammatical person and animacy, adjective class, tense, aspect and mood (TAM), collocations, and subject class, as was discussed in the research questions/hypotheses section. Table 2-3 displays the independent linguistic variables that were included in the analysis, with the categories and criteria for each.

Each variable was coded with as much detail as possible in order to obtain the best linguistic model. As will be described later, some of the variables initially included were either erased completely or collapsed with other variables to maintain the validity of the statistical model. Each variable is described in detail below.

The first linguistic variable coded for was person/number; thereafter, the contextual feature, animacy, was also coded. Animacy distinguishes living referents from non-living referents. In previous research there are some contradictions on the influence that animacy has on copula choice. Recall that Silva-Corvalán (1986, 1994) found no effect for animacy; however, Geeslin (2000) found that there was an effect but only for adjective class (p. 54). It is also worth noting that animacy was coded differently from Geeslin (2000) and Silva-Corvalán (1986, 1994) in order to maintain detail in the statistical model. In the present study, a new variable was created that is referred to as grammatical person and animacy. To explain, due to the type of statistical analysis involved, the variable animacy was highly interactive with person/number; which meant that either animacy or person/number would have been completely excluded from the model. To avoid the exclusion of either, the variable was operationalized by coding animacy as a part of the independent variable person/number. The examples in (7) illustrate:

(7) (a) *Rosa está durmiendo.*

'Rosa is sleeping.'

Third person singular animate

(b) *El cielo es azul.*

'The sky is blue.'

Third person singular inanimate

(c) *Ellos están enamorados.*

'They are in love.'

Third person plural animate

(d) *Son libros graciosos.*

'They are funny books.'

Third person plural inanimate)

In the cases of 1st person singular, 2nd person singular, 1st person plural and 2nd person plural, animacy was not coded for, given that they are always used to refer to animate referents. In other words, animacy was assumed in these cases.

Tense, aspect and mood (TAM) was also coded. Tense, aspect and mood had a total of 19 categories. As many categories were included in the initial coding in order to obtain as much detail for the final linguistic model. The fourth linguistic variable included was adjective class. Adjectives are an interesting phenomenon in their own right because of how traditional grammar has tried to establish a definite dichotomy between *estar* and *ser* (Luján 1981: p. 164).

Collocations and complements were also coded. Collocations are the words that immediately appear after *ser* or *estar*, for example, *La casa es muy grande* 'The house is very big.' In this case it would be coded under adverb since the word that immediately follows the copula is an adverb. Another example, *Ella está leyendo un libro* 'She is reading a book,' this collocation would be coded under present participle since that is what is immediately after the copula. Overall there were 18 categories under collocations. On the other hand, a complement is the part of a sentence that comes after the verb and is needed to make the sentence complete. There were 9 categories or types of complements total; namely, clauses, pronouns, nouns, adverbs, locatives, adjectives, prepositions, gerunds and other.

In this project, adjective class was divided into semantic classes, based on Silva-Corvalán's division scheme. The adjectives were divided as age (*Ella está 7 o 8 años* 'She is 7 or 8 years old'), size (*La foto es grande* 'The picture is big'), physical appearance animate (*Ella está pelirroja* 'She is a redhead'), description inanimate (*La casa es cuadrada* 'The house is square'), Evaluation (*El libro está interesante* 'The book is interesting,' *Todo es perfecto* 'Everything is perfect'), moral value (*La mamá de Rosie es muy honesta* 'Rosie's mom is very honest,' *Ellos están buenos padres* 'They are good parents), color (*El pájaro en su cabeza es rojo* 'The bird in his head is red'), sensory character (*La torta de Rosie es deliciosa* 'Rosie's cake is delicious'), social status (*La familia de Rosie son muy ricos* 'Rosie's family is very wealthy') and emotional state (*La familia está contenta* 'The family is happy').

The last linguistic variable coded was subject class. This variable had 10 categories total; namely, tangible things (e.g., *libros, fotos, vasos* 'books, photos, glasses'), nature (e.g., *día, árbol, flor* 'day, tree, flower'), tools (e.g., *martillo, clavo* 'hammer, nail'), food (e.g., *hamburguesa, salchicha* 'hamburger, sausage'), galaxy (e.g., *cielo, estrellas, nubes, sol, luna* 'sky, stars, clouds, sun moon'), clothing (e.g., *pantalones, blusa, bufanda*, 'pants, blouse, scarf'), people (e.g., *Rosie, niña, bebe, mujer, Juan* 'Rosie, girl, baby, woman, Juan'), animals (e.g., *perro, gato* 'dog, cat'), holidays (e.g., *navidad, Pascua* 'Christmas, Easter') , and sports (e.g., *futbol, correr* 'soccer, run').

Extralinguistic Variables

In addition to the five linguistic variables included in the initial model, extralinguistic variables were also considered. These included gender, community type (social network), years of formal instruction in Spanish, age at which formal instruction in

Spanish began, duration on study abroad program in a Spanish speaking country, English language proficiency, and overall Spanish language proficiency. Table 2-8 illustrates these variables, their categories and criteria

Here gender refers to biological sexual characteristics: female or male, following Aguilar-Sanchez (2009) and others. Gender has been acknowledged to be an important factor to influence speech and, previous research has shown that women speak differently than men. According to Silva-Corvalán (2001) women tend to apply linguistic variants in their speech more often than men (p. 97). What's more, it is shown that women tend to autocorrect more often than men in formal contexts, but in informal situations they tend to use the innovative forms more often (p. 97).

The second extralinguistic variable was speech community type. Within the speech community, the social network of the participants is also considered. It has been found that knowledge of the social network of each participant provides a clearer picture of their contact to the target language, in this case Spanish (Milroy 1982: p 554). In the present research the social network, the population percentages (2009 Census Bureau) and the descriptions given by the participants of their communities were considered before placing them in the bilingual speech community or monolingual speech community.

The third and fourth independent variables are interrelated: years of formal instruction in Spanish and age when formal instruction in Spanish began. Age of formal instruction was considered since, according to Lenneberg (1967) and Silva-Corvalán (2002) among others, the 'critical age' by which the structures of one's native language are firmly acquired is around 11-12 (Silva-Corvalán 2002: p. 15). Furthermore, the

length of time that learners have been in a formal setting also provides more data for the model that will hopefully lead to an analysis that might shed light on how formal instruction might relate to the frequency of *estar* and if this acquisition can be altered through contact with speakers of Spanish.

Duration on a study abroad program in a Spanish speaking country was the fifth social variable. Most participants had studied abroad in Spain or Mexico with program's duration ranging from six weeks to six months. There were also five participants who had never studied abroad, four from the BISC group and one from the MONOSC group. This variable was coded because any type of contact with Spanish can also influence the interlanguage of the L2 learner, as was discussed in Chapter 1.

Participants' knowledge of a third language was also coded, as such ability could potentially impact their frequency of *estar* and *ser* depending on how closely related the third language is to Spanish. Only three participants in the BISC group spoke a third language (i.e., Malayalam, Korean and Marathi), whereas the MONOSC group had only one participant who spoke a third language (i.e., German). Even so their knowledge of the third the language was somewhat below moderate and even more importantly the four languages involved were not Romance languages and did not have a copula system similar to Spanish so as a result this social variable was subsequently discarded from the analysis.

Statistical Procedure

Goldvarb X was used to facilitate the raw numbers and percentages and perform the variable rule analysis for the extralinguistic and linguistic factors. Goldvarb X is a program that allows the researcher to provide a quantitative model of interlanguage variation (Young & Bayley 1996: p. 253). Interactions were found in the linguistic and

extralinguistic factor groups and were further tested with cross-tabulations that showed numerous empty cells. The empty cells indicate that the data are distributed unevenly or contexts are not equally represented, or that interactions are present. Interactions are often fixed by collapsing factor groups together or simply by their removal (Tagliamonte 2006: p. 209).

All the extralinguistic factor groups were revealed to be highly interactive, meaning that all the factors interacted with each other. These interactions were so severe that neither collapse nor removal fixed the problem. As a result, a regression analysis was not run. Recall that Goldvarb X does not provide testing for interactions (Paolillo 2002: p. 65) and the model needs to be free of interactions for a regression analysis, nonetheless, the raw numbers and percentages (referred to here as marginal results) for the extralinguistic factors in both groups are presented and discussed in Chapters 3 and 4. Although such a discussion cannot show the influence of each factor independent of the others, it can indicate how often each factor occurred with *estar*, or with *ser*, and the total number of tokens and percentages of the frequency of the variants (Young & Bayley 1996: p. 269). In other words, “the marginals provide relative frequencies and percentages of the dependent variable, allowing for the factor-by-factor correlations to be revealed in the data” (Tagliamonte 2006: p. 135). The marginal results alone can be misleading at times, especially when the distribution of contexts is not equally represented (Tagliamonte 2006: p.137); nonetheless, they are used to gain knowledge of the frequency of *estar* and *ser* with each extralinguistic factor.

Likewise, the linguistic model also revealed interactions. However, they were not as severe as those found in the extralinguistic factors. Cross-tabulations were again

carried out for the linguistic factors to identify the regularity in the application of *ser* and *estar* as well as any interaction violations (Tagliamonte 2006: p. 208). Several statistical runs took place to arrive at a final analysis free of knockouts, singletons and interactions. A knockout shows no variation in data. For example, in the factor group tense, aspect and mood, there were four occurrences of the present subjunctive. All four occurred with *ser* and none with *estar*. This produces a knockout because 100% or 0% of the data occurred with one of the dependent variables. A variable rule analysis cannot be run in this context; therefore the knockout factors must be removed or collapsed with others that are linguistically or mathematically similar (Tagliamonte 2006: p. 152). A singleton group, however, shows that there is only one factor group (Tagliamonte 2006: p. 153). In other words, “there is no variation in the factor, and so it is meaningless and as a result they have to be eliminated through recoding” (Paolillo 2002: p. 74). Recoding will sometimes produce other knockout factors and singleton groups, and as result it is crucial to repeat the recoding process until a cell file appears free of knockout factors or singleton groups. Deleting some of these factors may be necessary to eliminate these sorts of violations. Fortunately, there were no singleton groups in the current data set.

Finally, interactions within the factors are not permissible because the rule of thumb in a variable rule analysis is for all factors included in the analysis to be independent in order to ensure an accurate analysis with the specific weights of each factor (Young & Bayley 1996: p. 273). Worth noting is that interacting factor groups are often found through cross-tabulations, possibly in multiple small or empty cells (Paolillo 2002: p. 89). Interactions of factors can be eliminated the same way as singletons and

knockouts: by removal or by collapsing factors with others that are linguistically or statistically similar. Unlike the extralinguistic factors, to achieve results for the linguistic factors a regression analysis was run. This step-up/step-down analysis assigns weights to each factor and calculates whether each factor contributes significantly to the variation in the dependent variable (Young & Bayley 1996: p. 268), providing three forms of evidence for interpreting the results: (1) statistical significance; (2) relative strength; and (3) constraint ranking of factors (Tagliamonte 2006: p. 140). Statistical significance refers to which factors favor the application value. Weights over .50 favor the application of the rule while those under .50 disfavor it (Paolillo 2002: p. 34). Relative strength refers to which factor is the most significant; as a result the range of each factor group is presented to find the most significant factor (Tagliamonte 2006: p. 145). With this in mind, “the range or magnitude of effect enables us to situate factor effects with respect to each other” (Poplack & Tagliamonte 2001: p. 93)

Finally, constraint ranking refers to the order (from greater to lower probability weights) of the categories within a factor group (Tagliamonte 2006: p. 237). In other words, “the constraint ranking determines the relative contribution of factors” (Poplack & Tagliamonte 2001: p. 7). In this case the constraint ranking shows the relative contribution of factors to the frequency of *estar*.

All of these statistical procedures and findings are presented in the next chapters. Chapter 3 presents and discusses the marginal results as well as the statistical results for the extralinguistic and linguistic factors for the advanced L2 learners from a Spanish/English bilingual speech community (BISC). Again due to the problems with the model, there are no statistical results for the extralinguistic factors; however, the

linguistic factors do have marginal as well as statistical results. Chapter 4 then presents the same data, but for the learners from a monolingual speech community (MONOSC). In Chapter 5, the conclusion, these findings are compared and contrasted together in order to propose answers to the research questions that guided the study.

Table 2-1. L2 learners from Spanish/English bilingual speech communities (BISC)

Participant	Age	Gender	Parents' mother tongue	Speaks parents' mother tongue	Years of formal instruction in Spanish	Age began formal instruction	Study abroad duration
B1	21	F	Russian	No	7	14	2 months
B2	20	F	Malayalam	Yes	8	12	1 months
B3	21	M	Korean	Yes	7	14	0
B4	22	M	Marathi	Yes	10	12	0
B5	21	F	English	Yes	10	11	6 weeks
B6	22	M	English	Yes	7	15	0
B7	22	M	English	Yes	5	17	2 months
B8	21	F	English	Yes	10	11	3 months
B9	19	M	Italian	No	5	14	1 months
B10	21	F	English	Yes	14	7	5 months
B11	22	F	English	Yes	8	14	3 months
B12	21	F	English	Yes	9	12	6 weeks
B13	22	M	Bosnian	No	9	13	0

Table 2-2. L2 learners from monolingual English speech communities (MONOSC)

Participant	Age	Gender	Parents' mother tongue	Speaks parents' mother tongue	Years of formal instruction in Spanish	Age began formal instruction	Study abroad duration
M1	22	F	English	Yes	6	16	4 months
M2	21	F	English	Yes	7	14	1 months
M3	22	F	English	Yes	10	12	6 months
M4	21	M	English	Yes	8	13	9 months
M5	21	F	English	Yes	11	10	5 months
M6	21	F	English	Yes	11	10	4 months
M7	20	F	English	Yes	9	11	2 months
M8	22	M	English	Yes	10	12	6 weeks
M9	22	M	English	Yes	10	12	6 weeks
M10	21	F	English	Yes	5	16	6 weeks
M11	19	M	German	Yes	8	11	1 months
M12	20	M	English	Yes	4	16	2 months
M13	22	M	English	Yes	7	15	0

Figure 2.1 Hispanic Census Data for English/Spanish Bilingual Communities in Florida

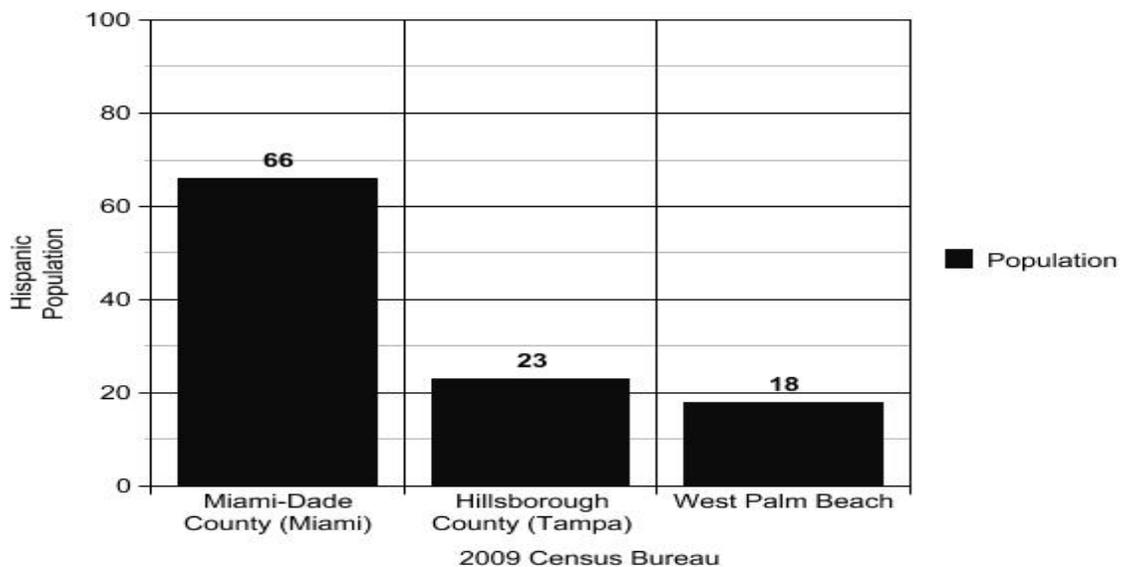


Figure 2-1. Hispanic census data for English/Spanish bilingual communities in Florida

Figure 2.2 Hispanic Census Data for English Monolingual Communities in Florida

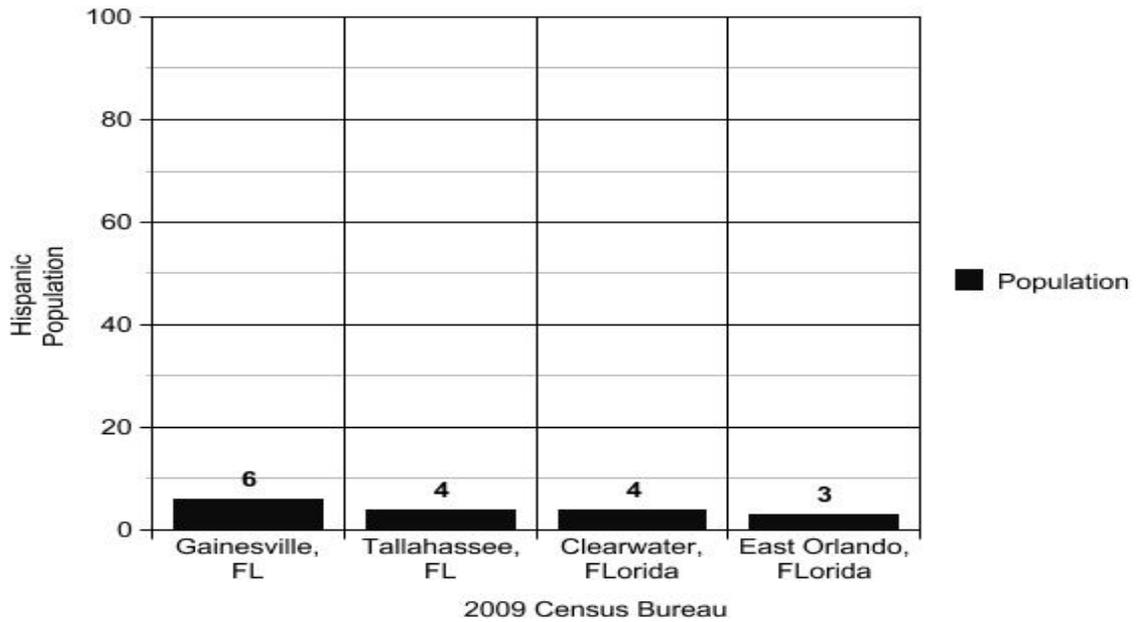


Figure 2-2. Hispanic census data for English monolingual communities in Florida

Figure 2.3 Self-Reported Proficiency Level in Spanish (scale 0-3) for Listening

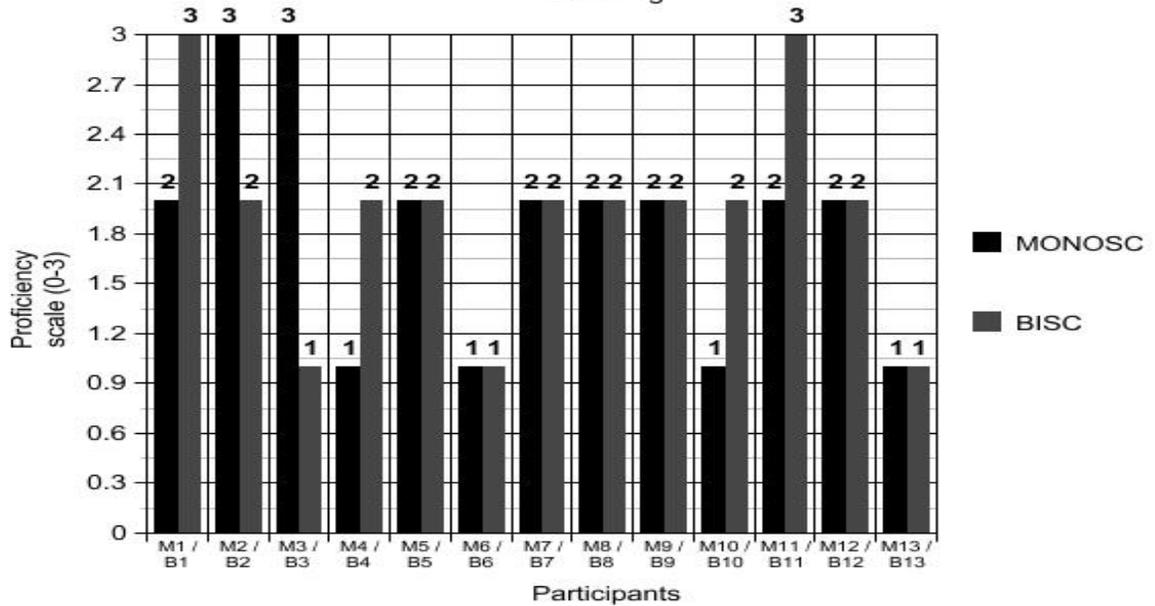


Figure 2-3. Self-reported proficiency level in Spanish (scale 0-3) for listening

Figure 2.4 Self-Reported Proficiency Level in Spanish (scale 0-3) for Speaking

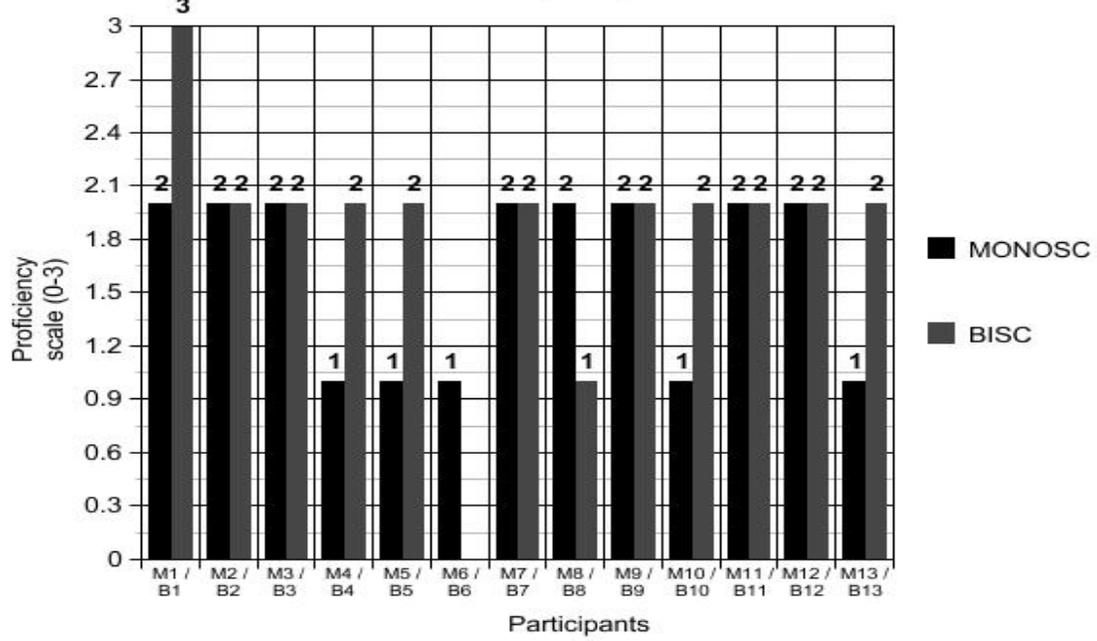


Figure 2-4. Self-reported proficiency level in Spanish (scale 0-3) for speaking

Figure 2.5 Self-Reported Proficiency Level in Spanish (scale 0-3) for Reading

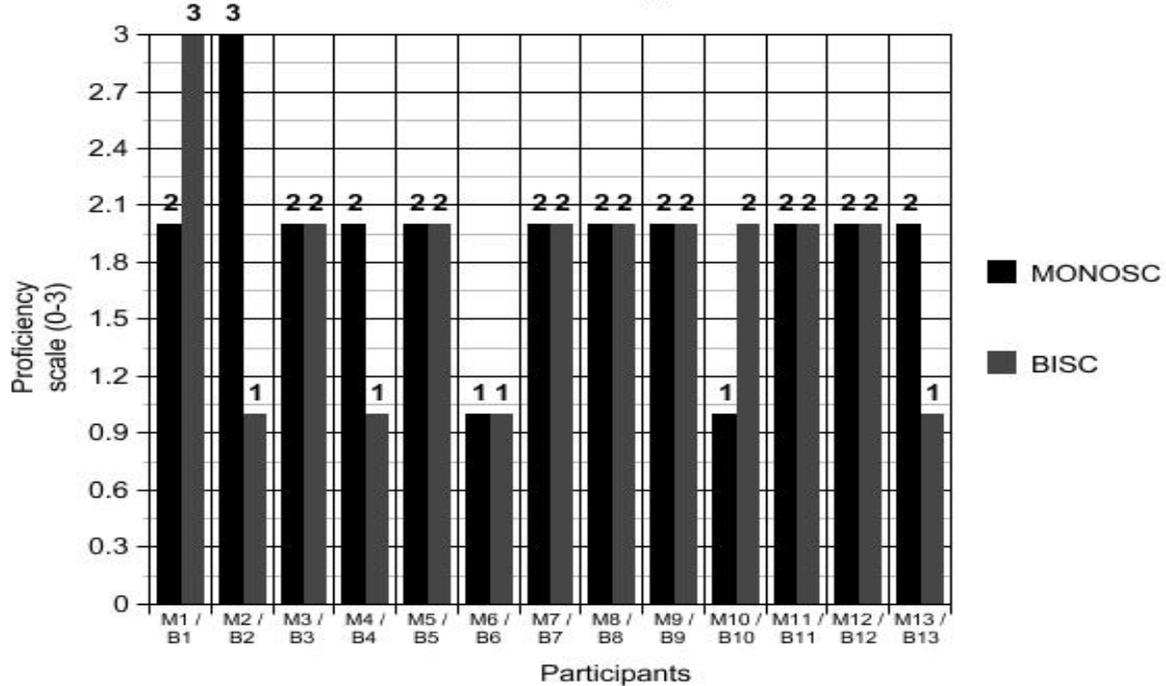


Figure 2-5. Self-reported proficiency level in Spanish (scale 0-3) for reading

Figure 2.6 Self-Reported Proficiency Level in Spanish (scale 0-3) for Writing

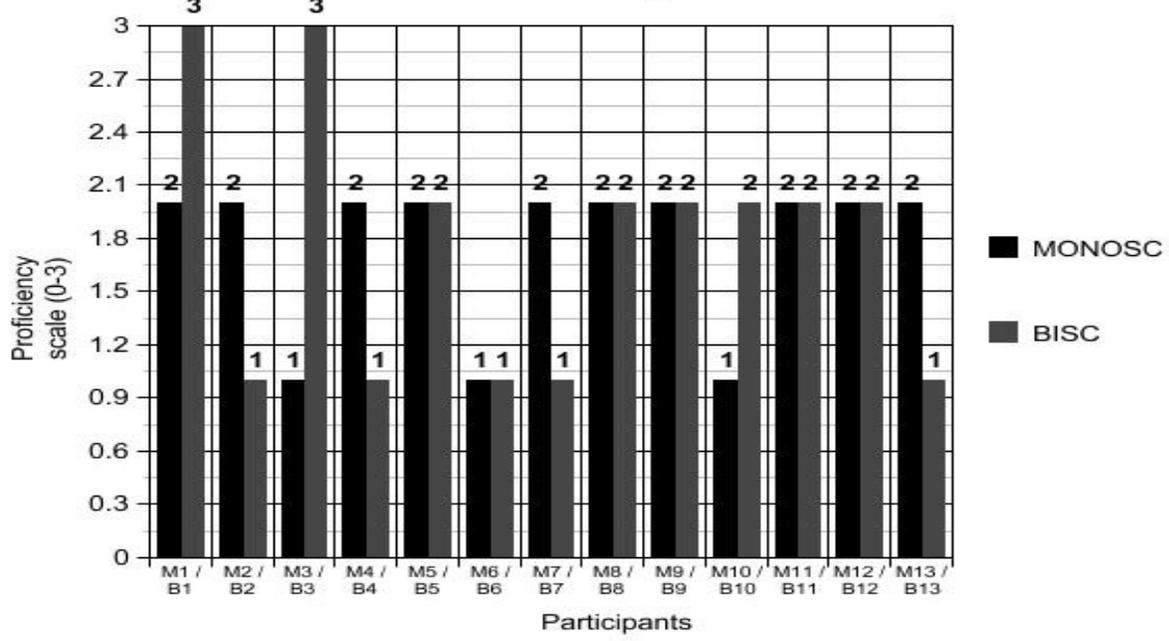


Figure 2-6. Self-reported proficiency level in Spanish (scale 0-3) for writing

Figure 2.7 Contact with Spanish for MONOSC and BISC groups

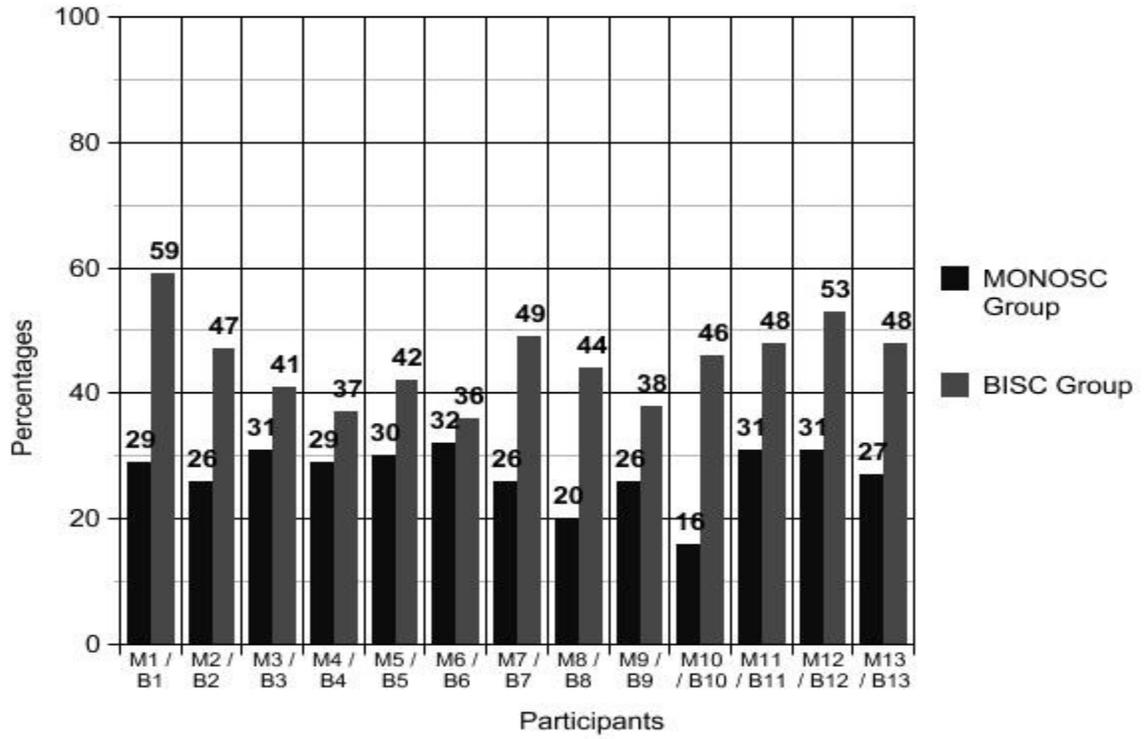


Figure 2-7. Contact with Spanish for MONOSC and BISC groups

Table 2-3. Coding of independent linguistic variables for copula choice in advanced L2 learners

Variable	Categories	Criteria
<u>Grammatical person and animacy</u>	[1]	1 st person singular
	[2]	2 nd person singular
	[3]	3 rd person singular animate
	[4]	3 rd person singular inanimate
	[5]	1 st person plural
	[6]	2 nd person plural
	[7]	3 rd person plural animate
	[8]	3 rd person plural inanimate
<u>TAM</u>	[a]	present indicative
	[b]	present progressive
	[c]	present subjunctive
	[d]	present perfect indicative
	[e]	present perfect subjunctive
	[f]	preterit
	[g]	preterit progressive
	[h]	imperfect
	[i]	imperfect progressive
	[j]	past perfect indicative
	[k]	past perfect subjunctive
	[l]	conditional
	[w]	future
	[X]	imperfect subjunctive
	[m]	conditional perfect
	[n]	command
	[o]	infinitive
	[p]	auxiliaries
	[q]	gerunds
<u>Collocations</u>	[V]	verb
	[A]	adjective
	[B]	location
	[C]	adverb
	[N]	noun
	[P]	preposition
	[Y]	conjunction
	[N]	pronoun
	[R]	relative pronoun
	[Z]	demonstrative adjective
	[O]	negation
	[E]	present participle

Table 2-3. Continued

Variable	Categories	Criteria
<u>Collocations</u>	[F]	past participle
	[J]	direct object pronouns
	[K]	blank
	[L]	English
	[I]	indirect object pronoun
	[T]	possessive adjective
<u>Adjective Class</u>	[a]	age
	[s]	size
	[p]	physical appearance
	[d]	description
	[d]	emotional state
	[m]	moral value
	[r]	color
	[y]	social status
	[e]	evaluation
	[w]	physical states
	[c]	sensory character
	[o]	other
	[x]	no adjective
<u>Subject Class</u>	[T]	tangible things
	[N]	nature
	[D]	tools
	[F]	food
	[G]	galaxy
	[C]	clothing
	[P]	people
	[A]	animals
	[H]	holidays
	[S]	sports
	[K]	parts of a home or building
	[J]	body parts
	[O]	other

Table 2-3. Continued

Variable	Categories	Criteria
<u>Subject Class</u>	[W] [Y] [Q] [U]	year seasons school subjects furniture dishes
<u>Complement</u>	[C] [P] [O] [A] [L] [J] [N] [G] [X]	clause pronoun noun adverb locative adjectival prepositional gerund other

Table 2-4. Coding of independent extralinguistic variables for copula choice in advanced L2 learners

<u>Variable</u>	<u>Categories</u>	<u>Criterion</u>
<u>Gender</u>	[m]	male
	[f]	female
<u>Speech community type</u>	[b]	bilingual community
	[m]	monolingual community
<u>Years of formal instruction in Spanish</u>	[1]	1-3 years
	[2]	4-7 years
	[3]	8-11 years
	[4]	12 or more
<u>Age at which formal instruction in Spanish began</u>	[0]	0-5 years of age
	[1]	6-9 years of age
	[2]	9-11 years of age
	[3]	12 or older
<u>Duration on study abroad program in Spanish speaking country</u>	[1]	0-6 weeks
	[2]	6-12 weeks
	[3]	12-18 weeks
	[4]	18 or more weeks
<u>Overall Spanish proficiency- self rated</u>	[0]	poor
	[1]	good
	[2]	very good
	[3]	native/ native like

CHAPTER 3 RESULTS FOR ADVANCED L2 LEARNERS FROM A MONOLINGUAL SPEECH COMMUNITY

This chapter discusses the analysis of the frequency of *ser* and *estar* in second language (L2) learners from a monolingual community speech community, whose access to Spanish has been limited to formal classroom input. These second language learners are those who were born and lived continuously in primarily monolingual areas of Florida such as in Gainesville and parts of Orlando. Recall that the participants indicated their place of birth in the demographics questionnaire as well as providing information regarding their speech community type, all of which helped to identify them as part of the Monolingual English speech community group (i.e., MONOSC group).

This chapter presents and discusses the results for the extralinguistic and linguistic factors of advanced Spanish participants who lived continuously in these monolingual speech communities. As with the results presented previously, the numbers for the extralinguistic and linguistic factors were derived from variable rule analysis using Goldvarb X.

Results for Extralinguistic Factors

The tables in the following subsections illustrate the raw numbers and percentages of *ser* and *estar* with each table representing a factor group. All tables contain the following information: Column one contains the categories of each factor group, column two contains labels for descriptive analysis, column three shows the number and percentage of occurrences of *estar*, and column four contains the number and percentages of *ser*. The fifth column contains the total of occurrences of both *ser* and *estar* in the data. It bears repeating that although this information is useful, it does not tell us the influence of each factor independently of the others (Young & Bayley 1996: p.

270), and that the current model is problematic because most factors involved overlap with each other. To further test the high number of interactions, data were also cross-tabulated; this further confirmed the interactions by showing a high number of empty cells and small cells. This shows that all the factor groups are inextricably linked

The first column in Tables 4-3 and 3-7 highlights the categories in each factor group, column 2 displays the number and percentage of the application value, and the third column displays the percentage and the number of *ser* uses.

Individual Differences

Table 4-1 shows the raw numbers and percentages for each participant in the group. These numbers illustrate the rate of occurrence of *estar* and *ser* as well as the overall production of the copula. The data show that the overall use of *estar* and *ser* does not vary among the participants.

The marginal results confirm that eight out of the thirteen participants use *estar* with a higher frequency than *ser*, while the other five participants favor *ser*. Participant 5, for example, use *estar* five times (8%), and *ser* 62 times (93%). To put it differently, the small percentage of *estar* production could perhaps imply that Participant 5 might be at a lower proficiency level than initially specified. As suggested in Ryan & Lafford (1992) and Van Patten (1985, 1987) L2 learners who use *ser* more often than *estar* may have not reached an advanced level of Spanish proficiency. Overall, the L2 learners from this group revealed an average frequency of 33 instances of *estar* per learner.

These raw data indicate that perhaps some participants have yet to reach the advanced stages of language acquisition. Moreover, Geeslin (2003) also shows that L2 advanced participants applied *estar* more frequently than *ser*. This is important because it implies that there is a correlation between the frequency of copula and the level of

proficiency. Recall that all of the participants in this group learned Spanish in a formal setting and had no or very little contact to native Spanish speakers with the exception of their study abroad experience, which is discussed furthered below.

Gender

Table 4-2 displays the raw numbers for gender. While these numbers do not specify which of the extralinguistic factor groups is significant to the frequency of *estar* and which are not, the numbers are nonetheless presented to show the percentage of times that each factor occurred with *estar*, the number of times it occurred with *ser*, and the total number of tokens as well as its percentages. Only generalizations are made here, with no intention to specify that these numbers are in fact true representations of any sort of variation that may exist within categories in each factor group.

The marginal results reveal a very slight difference between the use of *estar* among males and females, 50% and 46% respectively with only a slight difference in the frequency of *ser* between males and females, 50% and 54% correspondingly. Percentage values remain similar between male and female across the board, suggesting that there is no real difference in the uses of *ser* and *estar* based on gender.

Age

Table 3-3 illustrates the raw numbers for acquisition age, which is defined here as the age at which the participants first began to study Spanish. The participants, who learned Spanish after the critical period, which is generally puberty (e.g., Gass & Selinker 2001; Moyer 1999; Johnson & Newport 1991; Long 2005), have the highest frequency of copula use, 53%. In this case, the age of 11 was chosen as the cut off age because it is considered by many to be the age by which the structures of one's native language are firmly acquired (Silva-Corvalán 1994: p. 15).

Even though the group age 12 and older had the most instances of general copula use, there is only a slight difference in the frequency of *estar* (46%) versus *ser* (54%). On the other hand, note that in the 0 to 5 year old group there is a strong difference in which copula they use. *Estar* (67%) is more frequent than *ser* (33%) suggesting perhaps that an early age of learning allows a higher frequency of *estar*. Conversely, the second youngest group, 6 to 9 year olds, shows a higher frequency of *ser* (71%) than *estar* (29%). The high percentage of *ser* may indicate that they are in the lower stages of acquisition. These results, however, are inconclusive given the contradicting results and the variability across groups.

Years of Formal Instruction

Previous research shows that formal instruction may be needed to learn some target forms, (e.g., Rehner & Mougeon 1999; Sax, 2003) and that learners may also need opportunities to interact with native speakers. Nonetheless, if participants receive most or all of their language instruction in the classroom without greater access to TL speech communities, their target form will maintain its linguistic prescriptivism (Dewaele 2004: p. 437). In other words, they will most likely not acquire vernacular speech. By any means it is not suggested here that prescriptivism equals incomplete acquisition or that vernacular speech equals complete acquisition; instead it simply means that L2 learners are not able to learn the non-standard forms found in speech communities, if they lack access to native speakers of the target language.

Notably, there is a slight difference in the use of *ser* and *estar* between groups. For instance, the participants that studied Spanish in a formal setting between eight to eleven years lean slightly more towards the use of *ser* (54%), those that had four to seven years of formal instruction use *estar* (52%) slightly more often than *ser* (48%).

Here, the group that received the least formal instruction has a tendency to use *estar* more regularly than those with more time in the classroom. Even so, the choice between *estar* and *ser* in both groups seems to remain somewhat constant with only a slight difference, which would indicate that the number of years that participants study in a formal setting may have no direct influence on the variation of the target form, confirming that the more formal instruction an L2 learner receives the more likely he/she will be able to apply the prescriptive forms.

Time Abroad

Previous variationist research on the influence that study abroad may have on variation (e.g., Regan 1996; 2002; Dewaele 2004; Sax 1999, 2003; Mougeon & Rehner 2001) reveals that those participants who have studied abroad show substantial differences in the target language from those who do not.

The marginal results reveal that when comparing the frequency use of *estar* and *ser*, there is only a very slight distinction, 46% and 54%, correspondingly. On the other hand, the participants who studied abroad 12 to 18 months have a higher frequency of *estar*, 75% than *ser*, 25%, indicating that in order for the L2 learner to acquire the variation present in the TL community, he/she must be part of the community at least 12 or more months. This again is suggestive that the more contact L2 learners have with the target language in a natural setting; the more likely it is to stimulate interlanguage variation, confirming Sax's (2003) suggestions that the time spent in a French-speaking environment was the most significant factor for variation to occur (p. 97).

Complete immersion in the target language does appear to assist the L2 learner to acquire the variants found within the speech community. Nevertheless it needs to be reemphasized that it certainly does not indicate that they will acquire the target

language identically to a native speaker. Nonetheless, even more important at this point is to acknowledge here that because we do not have a control group, similarities to native language input cannot be made, and as result we can only compare both groups in this study and the influence that speech community could have in the frequency of *estar*. In the future, it may be of interest to further develop this study by incorporating a control group composed of native speakers of Spanish by doing so we can then look at possible parallels to native language.

Spanish Self-Evaluation

This section discusses the marginal results for self-evaluation in Spanish. According to Woolsey (2006), Geeslin (2002) and Silva-Corvalán (1996), as L2 proficiency level increases, production of *estar* should be more evident. In the current study there were originally four different levels of proficiency, poor, good, very good, and native-like, however because all participants ended up evaluating their Spanish as good or very good and none used the two extreme categories, the other two were excluded from the initial run.

Table 3-7 illustrates that there is almost no difference in the frequency between *estar* and *ser* in the very good group, 51% and 50% respectively. In the good group there is a greater difference in the use of *estar* and *ser*, 44% and 56% correspondingly. As Geeslin (2002) and Van Patten (1985) point out, *ser* is the copula of choice for those participants who are in the earlier stages of acquisition, while *estar* begins to emerge in the more advanced stages of acquisition.

The following section presents the marginal results for the linguistic factors, followed by the results from the variable rule analysis.

Marginal Results for Linguistic Factors

The following section includes the marginal results of the linguistic factors. Recall that a variable rule analysis was not run due to interactions among factor groups, leaving us with only the raw numbers which only show the distribution of the variant forms without indicating which factor groups (independently) contribute significantly to their variation. The following section highlights these results along with the procedure taken to arrive at the three statistically significant factors. Table 4-8 shows these results, it includes column 1 which highlights the categories in each factor group, column 2 displays the number of *estar*, column 3 displays the percentage of *estar* and the fourth column displays the number of applications as a percentage of the total number of occurrences of each factor (refer to Chapter 2 for original list of categories that were coded). Due to the knockouts in the very first run, recoding was necessary to collapse these knockout factors with other non-knockout factors. However, if they were not characteristically or statistically similar to any of the others they were simply removed.

Tense, Aspect and Mood

Table 3-8 exhibits the distribution of *ser* and *estar* according to tense, aspect and mood. Originally there were a total of 17 categories, although only the types produced are part of the marginal results: present perfect indicative, present perfect subjunctive, past perfect indicative, past perfect subjunctive, future, conditional perfect, and commands did not show up in the speech of the L2 learners during data elicitation. The present progressive reveals to have the highest frequency of *estar* (99%) but cross-tabulations show that the present progressive is irrelevant because it is categorical. In other words, the present progressive must always appear with *estar*, and because of

this it was removed from the final run. The present indicative is revealed to be the second highest verb form to favor *estar* (34%) for the marginal results.

The past tenses consist of the imperfect and preterit. Only a very small percentage of the overall data consisted of the preterit with a 4% and the imperfect with a 10%, these small numbers may be directly related to the data elicitation task. Given that the participants were given a set of open-ended pictures, there was no real indication of when an event occurred in each picture and as a result it was fairly simple to narrate a story completely in the present tense and even easier to avoid any other verb form. In any case, it appears that the participants in this group particularly prefer the present indicative. Due to the small percentages that stemmed from the preterit and imperfect, they were recoded under past tenses. However, because of crossovers, they were recoded a second time under the 'other' category to fix crossovers. The other tenses that were recoded under the 'other' category include conditional, imperfect subjunctive, present subjunctive and the infinitive. Recoding allowed for small percentages and crossovers to be eliminated.

Collocations

Collocations are the parts of speech that immediately follow the copula or the context in which *ser* and *estar* appear. The most favorable predictor of *estar* is the present participle with a 98%. Below, example [1] illustrates examples of participles from our data.

(8) (a) *Se están relajando juntos.*

'They are relaxing together.'

(b) *El papá está tomando un cafecito.*

'The dad is drinking a coffee.'

However, because present participles always appear with the progressive form, they were also labeled as categorical. Due to its categorical characteristic, they were removed from the final run. Prepositions, 65%, other category 56% and adverbs 37% were revealed to appear more frequently with *estar*. The participants in this group produced the examples seen below, preposition examples are those in (a) and (b), and in (c) and (d) are examples of adverbs

(9) (a) *Estar en la nieve.*

‘To be in the snow.’

(b) *...está para su niña.*

‘...for her daughter.’

(c) *Ella es muy inteligente.*

‘She is very intelligent.’

(d) *Las flores son muy hermosas.*

‘The flowers are very beautiful.’

Conversely, nouns are the primary context to appear the least with *estar*, only 14 occurrences (see example (a) and (b)), followed by pronouns (see example (c)) with one occurrence.

(10) (a) *Rosie es una estudiante.*

‘Rosie is a student.’

(b) *Colorado es un estado.*

‘Colorado is a state.’

(c) *¿Es ella muy inteligente?*

‘Is she very intelligent?’

The other category includes all pronouns (e.g., indirect, direct, relative pronouns), negation, conjunctions and verbs.

(11) (a) *Se estar se dan*

'To give themselves'

(b) *...y como no seria lo mismo.*

'...and since it would not be the same.'

(c) *El hombre es que quiere Rosie.*

'The man is what Rosie wants.'

(12) (a) *La familia está no...*

'The family is not...'

(b) *La familia tan gorda están no sé.*

'The family very fat, I don't know.'

(13) (a) *Sus vidas están y que diferentes...*

'Their lives are and how different...'

(b) *No sé si es porque...*

'I don't know if it is because...'

(14) (a) *La madre es viene de la cocina.*

'The mother comes from the kitchen.'

(b) *Rosie es ponen sus pantalones...*

'Rosie puts on her pants...'

All of these were originally coded separately. However, the preliminary run revealed that these parts of speech were either knockouts or the overall production was

too small to be considered in the variable rule analysis. As a result, they were collapsed into the other category to remove knockouts and to increase percentages.

The preliminary results reveal that adverbs and adjectives have similar percentages, 37% and 34%, respectively, equally disfavoring *estar*. Moreover, because of crossovers, adverbs and adjectives were collapsed together. This new category is referred to as the modifier group. Demonstrative and possessive adjectives were also coded under modifiers (see examples below). In the preliminary run possessive and demonstrative adjectives produced a knockout with occurrences appearing only with *ser*.

(15) (a) *¿Está ese hombre allí?*

‘Is that man there?’

b) *Están sus comidas muy calientes.*

‘Your food is very hot.’

Adjective Class

Adjective class has been studied extensively with copula choice (i.e., Geeslin 2000, 2002, 2003a, 2003b; Silva-Corvalán, 1986, 1996; & Ramirez-Gelpí, 1995 among others). All of these studies have coded for adjective type under the idea that *estar* with different types of adjectives may be acquired at different rates since they tend to allow different amounts of variability (Geeslin 2000: p. 53). In the current research twelve categories or types are included in the factor group of adjective class.

Although adjective class was revealed to be statistically non-significant, the marginal results showed that several of these categories disfavored the use of *estar*, while, only four adjective classes favored it: emotional state with a 73%, physical state

with a 100% followed by physical appearance with the lowest percentage, 59%. See examples below of emotional state, physical state and physical appearance.

(16) (a) *Sí están celosos de otra pareja.*

‘Yes they are jealous of the other couple.’

(b) *Juan está parado.*

‘Juan is standing.’

(c) *La madre de Rosie no está gorda.*

‘Rosie’s mother is not fat.’

These results confirm Geeslin’s (2003) analysis, which indicate that the two main categories in adjective class to favor *estar* are mental and physical states. While adjectives were produced among the participants in this group, their production was very low. 66% of all the data did not have any adjectives; instead they used other parts of speech such as adverbs, verbs, nouns, prepositions, pronouns, participles and so forth. Although the frequency of adjectives is slight, this was not the reason for their removal of the variable rule analysis. Instead they were removed because they were highly interactive with subject class and grammatical person and animacy. As a result it could not be included in the statistical run. Even then, it is important to show that while it was removed from the present research, other research such as Geeslin (2003) indicates that adjective class is in fact a significant predictor of *estar* in advanced L2 learners, however, she points out that individual categories of adjective class are not significant (p. 742).

Adjective class is furthered discussed more detailed in Chapter 4 thus providing insight on the existing variation found within and between L2 learners of the present

study. In addition, a comparison of the current results to those found in Silva-Corvalán (1994) takes place. Refer to Table 4-9 for summary.

Subject Class

The motivation for including subject class in the analysis was to see how animacy could influence copula choice. Animacy distinguishes living referents from nonliving referents. Previous research commonly has coded animacy as, [+ animacy] (i.e., 'boy') and [- animacy] (i.e., 'car'). In the current analysis there is more detail in its categorization. To add detail, [+animacy] and [-animacy], were further extended to categories of like objects. For example, the category animals was recoded as +animals (i.e., the + indicating existing animacy), on the other hand the category food was recoded as -food (i.e., the - indicating no animacy).

The numbers show that people is the highest category to appear with *estar* with 388 occurrences. This number may suggest that +animacy favors *estar*. However, this is only an assumption since there was no variable rule analysis to conclude if +animacy in fact influences the use of *estar*. Interestingly enough all the other categories that are considered -animacy with the exception of animals appeared with very low rates of *estar*. The category of animals (+animacy) was the category with the second highest choice of *estar*, with a 38%. It should be obvious that even then its percentage is relatively smaller than the percentage for the people category that appeared with *estar* 388 times. Even then the current marginal results establish that +animacy most often appears with *estar* while -animacy appears the least with *estar*.

This factor group also appeared to be highly interactive with the person/number group. Because detail would have been entirely lost if this group had been removed, both groups were collapsed into a new group called grammatical person and animacy to

maintain some sort of detail. It should be noted here that even then all detail was lost in regards to subject class. To explain, we no longer have access to knowing which subject class specifically (e.g., animals, clothing, food, galaxy, holidays, people, nature, and so forth) motivates copula choice.

The following section presents the marginals for person/number.

Person/Number

Note that person/number also tests for animacy. However, as previously mentioned, this group was collapsed with subject class to remove interactions between factor groups for the variable rule analysis. Nonetheless, here we only present the marginals without the collapse of groups, which reveal that 3rd person plural animate forms appear the most with *estar* with an 82%. The second category to appear the most with *estar* is 3rd person singular animate with a 58%. Again these results also agree with subject class, which reveals that animate categories (e.g., people and animals) appeared the most with *estar*. Most instances of *estar* and *ser* were conjugated; notice that there were only 28 instances of non-conjugated verbs. This case could possibly be an example of transfer taking place from their L1 (English). One should note here that although there are conjugations in the English language, they tend to also be regular in occurrence, just not as rich in morphology.

Complements

Table 3-8 also shows the marginal results for the 'complement' factor group. The following are the preliminary results: The 'gerund' appears to have the highest percentage of *estar* with a 97%; nonetheless, this is purely categorical. Generally, *estar* + the gerund are taught together (i.e., the present progressive) and may explain the reason for its high occurrence with *estar*. The next category with the highest *estar*

percentage is the 'locative', which appears 95% of the time with *estar*, indicating that the L2 learners from the Spanish/English bilingual community group continue to use the standard copula (i.e., *estar*) with location. Interestingly, the 'preposition' category has one of the lowest *estar* percentages with a 17%; it would be expected for it to appear more often with *estar* since prepositions often concur with location. However, on a side note, in the current data set most of the prepositional complements were used with the present tense of *ser* and *estar* and with the preposition *en* or with the preterit tense of *ser* and the preposition *de*, it seems that the tense of the copula and the type of preposition had more of an impact on whether *ser* or *estar* was used. Moreover, the categories 'noun' and 'adverbs' had the lowest percentages of *estar*, 7% and 18%, respectively. Most adverb complements occurred with *ser* and the adverb *como* (comparative), for example, *No es como sus padres* 'She is not like her parents' or *Esta escena es como un sueño* 'This scene is like a dream', perhaps the use of the adverb *como* initiates the use of *ser*, indicating longevity in the likeness of one object to another. There were 212 instances where 'noun' complements appeared, however, only 7% are used with *estar*. It should be noted that a large number of the nouns were animate (e.g., *chica*, 'girl,' *profesor*, 'professor,' *persona*, 'person,' *hija*, 'daughter,' *familia*, 'family,' *perro* 'dog' etc.), this could be the reason for the large numbers of *ser* again indicating permanency.

In chapter 4 the 'complement' linguistic factor is further discussed and in addition any variation found between the MONOSC and BISC groups is also conferred.

Multivariate Analysis Results for Linguistic Factors

This section presents the statistical results performed by the step-up/step-down analysis. The factor groups below were revealed to contribute significantly to the

variation of the copula. Moreover, cross-tabulations demonstrated that the present model was free of interactions. The current research looks at very specific factor groups that have not been looked at in the same manner in previous research. Yet other research has looked at factors that may be associated with the ones presented here. In other words, there may be a connection between some of the current variables and the ones that we know to predict copula choice found in previous work (see Chapter 1).

The first row in Table 4-9 displays the corrected mean and the log likelihood. The corrected mean is the average frequency of occurrence of the application value of the dependent variable (Paolillo 2002: p. 79). The second row displays the total number of *ser* and *estar* occurrences in each linguistic group. The first column displays the linguistic group; the second column displays the factor weight plus the range. The range is calculated by subtracting the lowest factor weight from the highest factor weight (Tagliamonte 2006: p. 242). In the present research, the largest range indicates the most significant in the choice of *estar* while the lowest range indicates the least significance. The third and fourth columns display the number and percentage of *estar* occurrences and the fifth column displays the total number of both copula occurrences.

Grammatical Person and Animacy

In the present research animacy was collapsed with person/number. Recall that the adjective class was removed from final analysis because it overlapped with subject class and person/number. It is worth noting that subject class was also removed from the final analysis because it interacted with person/number. In the absence of adjective class and subject class, animacy was included with the person/number factor group as to not lose all detail. Grammatical person and animacy is the factor group with the largest range, 69. The first two categories have the strongest factor weights, 3rd person

plural animate, .86 and the 'other' category .62, indicating that they are significant to the frequency of *estar*. Conversely, tangible 3rd person singular/plural inanimate and other 3rd person singular/plural inanimate do not favor *estar* since they show to have the weakest factor weights, .30 and .17, respectively.

Despite the results revealing that the category, 3rd person plural animate, has the strongest weight, it should be reiterated that there is a possibility that this result is directly related to the story telling used to gathered data. If the task had been different (i.e., open interview), there is a strong chance that 1st person singular and plural animate would have appeared to have stronger factor weights since the participants would be talking about themselves and no other characters. So while the multivariate analysis reveals that grammar person and animacy contributes the strongest to the variation of copula choice, the task effects must not be forgotten

Collocations

Again, the purpose of this factor group was to sort out all of the various parts of speech and to see which of these favor *estar*. The group collocations are the second group to have the highest range, 68. Notice that there is nearly no difference in range between the first group, grammatical person and animacy, 69. The preposition category had the highest factor weight, .84; this means that they strongly favor *estar*. It is worth noting that prepositions often introduce locations, which would favor their use with *estar*. All instances of *ser* + preposition were used to express possession, origin or composition, for example,

(17) (a) *El libro es de los esposos.*

'The book belongs to the married couple.'

(b) *El pastel es para el cumpleaños de Rosie.*

'The cake is for Rosie's birthday.'

On the other hand, most of the instances of *estar* + preposition expressed location, 64%; for example,

(18) (a) *Rosie está en la biblioteca.*

'Rosie is at the library.'

(b) *Ella está en la cocina.*

'She is in the kitchen.'

Franco (1984) points out that most textbooks teach *estar* + location but do not teach [*ser* + location], the current results possibly indicate that because most of these L2 learners received most of their Spanish contact in the classroom, they tend to use *estar* + location more often. Effectively, prepositions should favor *estar* since locatives are often used with *estar* (except for events) and those are often expressed with prepositional phrases (Geeslin 2000: p. 60).

The second category with the strongest factor weight is the 'other' category, .72. Recall that the 'other' category is composed of all pronouns (indirect, direct and relative pronouns), negation, conjunctions, and verbs. It possibly may be that this category has one of the strongest weights because it is composed of different parts of speech; therefore collectively they reveal a strong weight, however individually they produce insignificant percentages.

The modifier group revealed to have the third highest factor weight, .55. As previously mentioned, the modifier group is composed of adjectives and adverbs. Although not the highest of the group, modifiers reasonably favor *estar*. Aguilar-Sanchez (2010) is the only study to investigate the effect of the adverb on copula

frequency. Although he differentiates between cover and overt adverbs, he concludes that both types disfavor *estar* (35% and 27%, respectively) (p. 158). While this differentiation was not duplicated in the current research (i.e., overt vs. covert adverbs), adverbs were also revealed to disfavor *estar* here. Moreover, recall that the factor group of adjective class alone was not significant, although, once adverbs and adjectives are collapsed in the same group, their factor weight shows that they to slightly favor *estar*, .55. Conversely, the noun category had the lowest factor weight, .17. According to Geeslin (2000) constructions with nouns are often used with *ser*, therefore it makes sense for nouns to disfavor *estar*.

Tense, Aspect and Mood

One should note that the factor weights for this group are displayed in brackets in Table 4-9. The brackets indicate that the effect of the factor group did not reach statistical significance (Tagliamonte 2006: p. 237) so while tense, mood and aspect did not reach significance, it is included in Table 4-9 as it was part of the step-up/step-down analysis. In essence, cross-tabulations revealed that this group did not interact with other factor groups, consequently making it acceptable to include in statistical analysis. TAM does not appear to have an effect on the learners' choice of the copula, with a range of 5, or practically zero statistical significance to the frequency of *estar*.

Because the extralinguistic factors showed unavoidable interactions within social groups, a variable rule analysis was not performed and only marginal results were discussed for those factors. Conversely, both marginal results and statistical results are discussed for the linguistic factors. Initially there were five linguistic groups (i.e., TAM, collocations, adjective class, subject class, subject class and person/number) that were part of the model, however, after interactions and other violations only three are part of

the step-up/step-down analysis (e.g., TAM, collocations and grammatical person and animacy). The results reveal that grammatical person and animacy as well as collocations, are significant to the frequency of use of *estar*, while TAM was not.

The following chapter discusses and compares the findings from the two speech community groups in order to see the similarities and differences that speech community may have on the frequency use of *estar*. By the way of conclusion, answers to the research questions that guided the current study are provided. Lastly, limitations and implications of this work and directions for future research are presented.

Table 3-1. Overall distributions of *estar* and *ser* by L2-monolingual community

Factor Group	<i>Estar</i>	<i>Ser</i>
Individuals		
Participant M1	36 53%	32 47%
Participant M2	32 54%	27 46%
Participant M3	66 78%	19 22%
Participant M4	39 75%	13 25%
Participant M5	5 8%	62 93%
Participant M6	18 38%	29 62%
Participant M7	34 42%	48 59%
Participant M8	21 24%	66 76%
Participant M9	71 51%	69 49%
Participant M10	19 40%	28 60%
Participant M11	32 57%	24 43%
Participant M12	25 58%	18 42%
Participant M13	36 54%	31 46%
Total tokens 900	434	466

Table 3-2. Overall distributions of *estar* and *ser* according gender

Factor Group	<i>Estar</i>	<i>Ser</i>
Gender		
Male	224 50%	221 50%
Female	210 46%	245 54%
Total tokens 900	434	466

Table 3-3. Overall distributions of *estar* and *ser* according to age of formal instruction

Factor Group	<i>Estar</i>	<i>Ser</i>
Acquisition age		
0-5 years old	123 67%	61 33%
6-9 years old	37 29%	89 71%
9-11 years old	54 47%	60 53%
12- Older	220 46%	256 54%
Total tokens 900	434	466

Table 3-4. Overall distributions of *estar* and *ser* according to years of formal instruction in Spanish

Factor Group	<i>Estar</i>	<i>Ser</i>
Yrs. of formal instruction		
4-7 years	148 52%	136 48%
8-11 years	286 46%	330 54%
Total tokens 900	434	466

Table 3-5. Overall distributions of *estar* and *ser* according to time abroad

Factor Group	<i>Estar</i>	<i>Ser</i>
Time study abroad		
0-6 months	36 54%	31 46%
6-12 months	359 46%	422 54%
12-18 months	39 75%	13 25%
Total tokens 900	434	466

Table 3-6. Overall distributions of *estar* and *ser* according to Spanish self-evaluation

Factor Group	<i>Estar</i>	<i>Ser</i>
Spanish Self-evaluation		
Good	142 44%	181 56%
Very good	292 51%	285 50%
Total tokens 900	434	466

Table 3-7. Marginal results for advanced L2 Learners from a monolingual speech community. *Total N 869*

Factor Group	N <i>Estar</i>	% <i>Estar</i>	N
<u>TAM</u>			
Present progressive	211	99	214
Present indicative	173	34	505
Other	12	29	41
Past tenses	37	28	132
<u>Collocations</u>			
Participles	215	98	220
Prepositions	59	65	91
Other	9	56	16
Adverbs	67	37	181
Adjectives	64	34	187
Conjunctions	4	29	14
Pronouns	1	14	7
Nouns	14	8	176
<u>Adjective class</u>			
Physical states	4	100	4
Other	13	77	17
Emotional state	55	73	75
Physical appearance	23	59	39
No adjective	298	52	569
Feliz	9	43	21
Description	5	28	18
Sensory characteristics	1	25	4
Age	3	21	14
Color	4	11	36
Evaluation	6	9	67
Size	0	0	5
Social class	0	0	0
<u>Subject class</u>			
People	388	66	587
Animals	9	38	24
Clothing	2	18	11
Galaxy	4	18	22
Nature	11	18	60
Parts of home or building	7	17	41
Food	1	10	10
Holidays	1	8	13
Body Parts	3	3	9
Other	7	6	115

Table 3-7. Continued.

Factor Group	N <i>Estar</i>	% <i>Estar</i>	N
<u>Person/Number</u>			
1 st & 2 nd pers. Sing & plural	4	83	6
3 rd pers. Animate plural	147	82	179
3 rd pers. Animate	235	58	408
Infinitive	12	43	28
3 rd pers. Inanimate plural	6	26	23
3 rd pers. Inanimate	28	11	24
<u>Complement</u>			
Clause	0	0	5
Pronouns	16	31	51
Noun	15	7	212
Adverb	3	18	17
Locative	41	95	43
Adjective	123	41	300
Preposition	2	17	12
Gerund	219	97	225
Other	2	50	4

Table 3-8. Multivariate analysis of the contribution of linguistic factors selected as significant to the probability of the copula

L2 learners from a monolingual speech community				
Corrected mean				1.24
Log likelihood				317.787
Total N				684
	Factor weight	N Estar	%Estar	N
Grammatical Person and Animacy				
Person 3 rd person plural animate	0.86	69	69	
Other	0.62	21	41	51
Person 3 rd person singular animate	0.61	101	39	258
Animal 3 rd person singular animate	0.44	5	26	19
Tangible 3 rd person sing/ pl inanimate	0.3	23	15	149
Other 3 rd person sing/ pl inanimate	0.17	8	6	126
<i>Range</i>	69			
Collocations				
Prepositions	0.84	58	64	90
Other	0.72	25	50	50
Modifiers	0.55	127	34	370
Nouns	0.17	12	7	174
<i>Range</i>	68			
TAM				
Present indicative	[.51]	173	34	511
Other	[.46]	49	28	173
<i>Range</i>	5			

CHAPTER 4 RESULTS FOR ADVANCED L2 LEARNERS FROM A BILINGUAL SPEECH

This chapter discusses the findings of the analysis of the frequency of *estar* in second language learner (L2) from a bilingual speech community; that is, students who learned Spanish through formal language instruction but experienced frequent access to native Spanish speakers in their communities. As addressed in Chapter 2, second language learners from a bilingual community were defined as those who were born and lived continuously in primarily Spanish/English bilingual areas in Central and South Florida (e.g., Broward County, some areas in Tampa, and Orlando). Information gathered from the Language Contact Profile (LCP) form and demographics questionnaire was also used to obtain a detailed description of their respective community speech. The results of the LCP form revealed that generally all the participants in this group had approximately 45% to 70% contact with Spanish while residing in their respective communities. Even so, the participants in this group also received Spanish input in a formal setting as their predominant L2 influence to the prescriptive forms.

The findings for the extralinguistic and linguistic factors of these learners are presented and discussed in the following sections. The results for the sociolinguistic and linguistic factors were derived from a variable rule analysis using Goldvarb X, as were described previously.

Results for Extralinguistic Factors

Before proceeding, one should note that severe interaction violations were found in the data, which were further tested with cross-tabulations that showed numerous empty cells. The empty cells often indicate that the data are distributed unevenly or

contexts are not equally represented or that interactions are present (Tagliamonte 2006: p. 209). In this case, the interactions were so severe that neither collapsing nor removing fixed the interaction violation. One should remember that it could often be assumed that extralinguistic factors have a role in the variation of any application value (Paolillo 2002: p. 36). To further explain, because there were extreme dependencies among extralinguistic factor groups, a variable rule analysis did not take place. Nonetheless, this is no basis for dismissing the possibility that these extralinguistic factors are relevant to the frequency of *estar*. One should not forget that “much of variationist research is characterized as sociolinguistic factors often play an important role in its design” and cannot be ignored (Paolillo 2002: p. 35).

Given the origin of this study, instead of dismissing the extralinguistic factors in their entirety, only the marginal results facilitated by Goldvarb X are presented here. Although not statistically significant, the marginal results are important because they show the relative frequencies and percentages of the copula, thus providing information about learner proficiency levels as well as copula use. In essence, the marginal results reveal the factor-by-factor correlations of the data (Tagliamonte 2006: p. 135).

The following Tables (3-1 to 3-7) highlight the categories in the factor groups and display the number and percentage of the application value and *ser* uses. Moreover, the percentage of the total data is displayed.

Participants

The following section presents and discusses some of the differences found between the participants in this group. As can be seen the participants in this group were not socially identical, for example, not all had the same amount of study abroad time, or the same number of years of learning Spanish in a classroom setting and so

forth. Nonetheless, in the following paragraphs an effort is made to describe not only the similarities but also differences found in the participants and relates these potential differences in their copula choice.

Table 4-1 displays the raw numbers and percentages for copula choice for each participant in the group. These numbers illustrate the rate of occurrence of *estar* and *ser*. The marginal results reveal that more than half of the participants have higher rates of *estar* than *ser*. After examining their demographics questionnaires for any potential reasons in their higher frequency of *estar*, no apparent factor stood out from the questionnaires. In other words, the participants who had higher rates of *estar* showed to be relatively similar to the other half of the group.

Nearly all of the participants were speakers of only English and Spanish (i.e., their L1 is English and their L2 is Spanish, acquired in a formal setting). However, there were six participants who had knowledge of a third language. Of the four participants, with the highest percentages of *estar* use, three of these had knowledge of a third language. Notably, participant B2 who spoke Malayalam has one of the highest frequencies of *estar* use, 71%. Malayalam is one of twenty-two languages spoken in India. It is not a Romance language and does not have a copular system similar to Spanish. Followed by participant B3, who had knowledge of Korean, also had one of the highest frequencies of *estar* use, 68%, from the group. However, nor Korean does have a copular system similar to Spanish. Participants B4 (who had knowledge of Marathi) and B9 (who had knowledge of Italian) also exhibit some of the highest frequencies of *estar*, 66% and 55% respectively. In sum, it is evident that those participants who had knowledge of a third language show the highest degree of *estar* production. Although,

Geeslin & Guijarro-Fuentes (2005) suggested that knowledge of an additional language does not particularly seem to influence copula choice in Spanish (p. 74), the relevance of third language knowledge in these data cannot be dismissed entirely. Given the low number of participants who had knowledge of a third language in both groups in this study, though the present numbers do not provide enough data to make strong conclusive statements. However, future work might benefit from exploring this factor further.

In regards to the amount of contact to Spanish, although it is plausible that the amount of contact may be related to frequency of *estar* use, participants B5, B6, B7, B8, B10, and B11 exhibit otherwise. These participants show lower rates of *estar* in spite of having relatively similar amounts of contact to Spanish as the participants who favored *estar*. To clarify, the median percentage of contact with Spanish was taken from those participants who favored *estar* and those that favored *ser*, with 36% and 38% respectively. In essence, there is no major difference in the amount of contact found in those that favor *estar* and those that favor *ser*.

Generally the L2 learners from this group demonstrate a frequency average of 51 instances of *estar* per learner. However, since no definite conclusion can be made based on individual social differences and contact with Spanish or just the raw numbers alone, the following sections present the results for other possible factors, such as gender.

Gender

Table 4-2 displays the marginal results for gender. Recall here that previous literature (e.g., Boretti de Macchia & Ferrer de Gregoret, 1984; Valdivieso & Magaña 1991; Bonvillain 1993; Watt & Milroy 1999; Milroy, 1999; Mougeon and Rehner 2001;

Navas Sánchez-Élez 1997; Aguilar-Sanchez 2009 among others) has been successful in showing that women have a stronger tendency to adapt to innovative forms, as well as to be the linguistic trendsetters. In essence, differences are found in the way women and males use certain linguistic forms. In this case, however, it appears to be no difference in the copula use between males and females. With regards to *estar*, males show a 55% of *estar* occurrences, while the females have a 56%. There is relatively no difference in the percentages and as result it can be suggested that gender does not influence variation in copula choice. These numbers are somewhat surprising since it was initially posited that gender would be relevant to the frequency of *estar* use, which is furthered corroborated by other research (e.g., Labov 1972; Naro 1981; Gutierrez 1992; Silva-Corvalán 1994, 1986; Cortés-Torres 2004 among others). It is worth noting here, that previous research has often looked at gender in conjunction with other extralinguistic factors, that is to say, seldom is it analyzed independently, implying that this method may be the cause of the marginal results here.

Acquisition Age

Table 4-3 presents the marginal results for age of acquisition or the age in which the participants were first exposed to formal Spanish. The results reveal that the older category or the group who acquired Spanish at the oldest age (12 years old to older) has a higher frequency of *estar* to *ser*, 60% and 40% respectively. Simultaneously, we can also see that the youngest group, those who learned Spanish between the ages of to 5 have the lowest production of *estar*, 41% and the highest production of *ser*, 59%.

Previous research (i.e., Lenneberg 1967; Silva-Corvalán 1994; Guijarro-Fuentes & Geeslin 2003; Johnson and Newport 1991) reported that language learning declines after the age of twelve, so here we consider age of acquisition as a possible variable

related to variation in copula choice. In this case, the higher frequency of *estar* seems to be related to advanced proficiency level in the L2 learner. Since previous work (i.e., Van Patten 1985; Ryan & Lafford 1992) has shown that *estar* is acquired once the learner is at more advanced stages of acquisition, this finding is not necessarily surprising if we assume that the earlier the age of acquisition the greater number of years of formal instruction and thus the higher the proficiency level. Although somewhat expected it is nice to be able to confirm previous work to this extent.

The two groups in between, 6 to 9 years old and 9 to 11 years old have the same percentages; they minimally prefer *estar* to *ser*, 53% and 47% respectively. The marginal results suggest that there is no difference in frequency use of *estar* between the ages of acquisition of 6 to 9 years old and 9 to 11 years old. It appears that acquiring the copula between 6 to 11 years of age will have the same effect on copula choice. The results, however, suggest that the older 12 and older group has a somewhat higher frequency of *estar* use. It would be of interest for future research to continue to investigate this factor in relation to formal instruction and overall proficiency. In sum, these marginal results imply that age of acquisition of Spanish does not affect their frequency use of *estar*; it may be that in conjunction to other factors it may reveal a different outcome. The next section presents the results for the number of years of formal instruction.

Years of Formal Instruction

Table 4-4 displays the marginal results for the number of years of formal instruction that each participant has received throughout their schooling.

The group having 8 to 11 years of formal instruction shows the highest frequency of *estar*, 58%, while the participants who had the least amount of formal instruction (4 to

7 years) also have a relatively higher frequency of *estar*, 55%. Compared to those who had studied the most Spanish (12 or more years), which show 47%. The marginal results suggest that the number of years of formal instruction should be analyzed along with other extralinguistic factors and not alone. However, although the marginal results do not provide means to make conclusions, it should be noted that previous research (i.e., Guijarro-Fuentes & Geeslin 2003; Geeslin 2005) found that the number of years of study has no effect on variation of frequency of copula use (Guijarro-Fuentes & Geeslin 2003: p. 74).

Study Abroad

The marginal results for the factor of study abroad are illustrated in Table 4-5.

As can be seen the participants who studied abroad between 0 to 6 months show a 56% frequency use of *estar* while those individuals who studied abroad for longer periods of time, 18 or more months, exhibit a 55% frequency of *estar*. However, note that both groups in Table 3-5 show higher rates of *estar*. The slightly higher frequency of *estar* for both groups may be explained in part by their more advanced levels of proficiency. Perhaps if looked at in conjunction with other factors different results would have been attained. For now, we must tentatively conclude that the time spent abroad is not correlated to the frequency of *estar* choice.

Spanish Proficiency Self-Evaluation

Table 4-7 displays the marginal results for the Spanish proficiency self-evaluation that participants completed. Although one cannot trust such self-reported data entirely, these tentative findings may be of interest.

The results show that the participants who claimed to have a 'very good' proficiency level had the highest rate of *estar*, 61%. In the same way, the group who

claimed to have a 'native-like' proficiency also shows a high frequency of *estar*, 55%, while the group of participants who reported having poor proficiency in Spanish had the lowest rate of *estar*, 36%. According to Geeslin (2003), the less proficient the student is the lower the relative frequency of *estar*. Geeslin (2000), Ryan & Lafford (1992) and Van Patten (1985), also agree that frequency use of *estar* increases gradually with proficiency.

Marginal Results for Linguistic Factors

Having discussed the extralinguistic factors, the following section includes the preliminary results for the linguistic factors. Of course, again, the marginal results cannot claim that any of the linguistic factors are statistically significant in the variation of *estar* uses, although they do show the distribution of *estar* and *ser* in the overall data as well as the number and percentage of the copula. Therefore, they may allow for interesting observations. Moreover, this section also serves another purpose, and that is to explain the process of coding, recoding and elimination of factor groups and/or categories within factor groups, as an attempt to provide a clearer overall picture of how some factors groups and categories were re-categorized, and/or maintained and others were not for the variable rule analysis. Originally there were six linguistic variables (collocations; adjective class; grammatical person and animacy; tense, aspect, mood; complements and subject class); however, after collapsing and removing of categories as well as factor groups due to singletons, knockouts and interactions, only five linguistic variables remained tense, aspect and mood; collocations; adjective class; grammatical person and animacy and complements. These are presented in the following sections.

In Table 3-9, the first column displays the categories in each factor group, column 2 shows the raw number of uses of *estar*, and column 3 illustrates the percentage of *estar* in the data while the fourth column provides the number of applications as a percentage of the total number of occurrences of each factor (Young & Bayley 1996: p. 268). The following sections discuss each factor individually and present the preliminary results.

Tense, Aspect and Mood

The progressive not only includes the present progressive (*La chica está trabajando*, 'The girl is working') but also the preterit progressive (*Ellos estuvieron corriendo en el parque*, 'They were running in the park') and imperfect progressive (*La pareja que estaba corriendo*, 'The couple that was running'). The progressive has highest frequency of use with *estar*, 97%. The present indicative (*El perro está gris*, 'The dog is gray') had the second highest frequency of *estar*, 40%, while other verb forms were collapsed into the 'other TAM' category, which includes the subjunctive (*La mamá espera que ella esté feliz*, 'The mother hopes she is happy'), future (...*y estarán en la niebla*, '...and they will be in the fog'), conditional (...*serían muy felices*, 'they would be very happy') and infinitives (...*y querían estar padres*, '...and they wanted to be parents'). These forms were all placed in the 'other' category because in the preliminary run they all produced knockouts, except for the conditional, which had an extremely small percentage, 4%. The 'past tense' category includes both the preterit (...*estuvo perfecto* 'it was perfect') and the imperfect (...*era la mejor parte de su vida*, 'it was the best part of her life'). This category produced the lowest frequency of *estar*. In the preliminary run, crossovers were present, and as a result they were collapsed in the 'other' category to eliminate any sort of violations including crossovers in final run.

Collocations

Participles are also considered as a possible independent variable that may be related to the dependent variable. Participles have the highest frequency of *estar*, 97% in comparison to the other categories in this group (e.g., prepositions, other, adverbs, adjectives and nouns); however, because present participles always appear with the progressive, are considered as categorical and were also excluded from final multivariate analysis. On the other hand, prepositions show the second highest frequency of *estar*, 73%. Also it should be noted that prepositions include locations (Rosie está en la escuela 'Rosie is at school') and simple prepositions (Ellos están en un cuarto). On the other hand, adverbs and adjectives manifest the same frequency of *estar*, 38%. However, the cross tabulations in the preliminary run revealed that adjectives and adverbs were problematic, because they shared the same percentage, and as a result they were collapsed into one group to fix the problem. For the final run, this group containing adjectives and adverbs is referred to as the 'modifier' group. Likewise, in the preliminary run conjunctions, pronouns and verbs revealed extremely small percentages and as a result they were recoded under the 'other' category. Nouns were also placed in the 'other' category to solve the problematic cross tabulations that showed multiple empty cells.

Adjective Class

Adjective class was completely removed from final multivariate run because of strong interactions with the other factor groups. For instance, in the factor group grammatical person and animacy, tangible/ 3rd person singular/plural inanimate cannot have an emotional estate in the adjective class, since tangible things such as a book, cannot be sad or happy, etc. It should also be noted that adjectives were not as

frequent in the data, for instance in the MONOSC group there were 298 instances from a total of 870 instances where adjectives did not appear. This also holds true for the Spanish/English bilingual speech community group (i.e., BISC group) where more than half of all the possible instances (607/1174 total instances) did not have any type of adjective present.

Recall that even though these numbers are not statistically significant, they are nonetheless important because they illustrate the distribution of *estar* according to each category in each factor group. With this type of information, only assumptions can be presented in regards to the relationship between the application value and the factors. Nonetheless, some interesting trends emerged.

Table 4-8 demonstrates the types of adjectives with the highest frequency of *estar* for the L2 learners from the BISC group and the Monolingual English speech community group (i.e., MONOSC group). It also presents the numbers and percentages found in Silva-Corvalán's (1994) results, thus providing a line of comparison between the three sets of data and giving us an opportunity to assess any similarities or differences. One should recall that the participants in Corvalán's (1994) study are Mexican-American bilinguals living mainly in a Hispanic area of Los Angeles area. "Contact between Spanish and English in the bilingual communities of Los Angeles has been extensive" (Silva-Corvalán's 1994: p. 9). Her results revealed "the diffusion of *estar* is at a more advanced stage in the United States than in monolingual Spanish areas (e.g., Morelia), this acceleration may be contributed to several factors, one being the lack of formal education" (p. 114).

As Table 4-8 illustrates, there were no adjectives of size (*El libro es grande* 'The book is big; *Las montañas son muy grandes* 'The mountains are very big') in neither the MONOSC nor the BISC group that appeared with the innovative use of *estar*, however, Silva-Corvalán's (1994) data set showed a 53% of adjectives of size to appear with the innovative use of *estar*. In comparison, the number of tokens for the adjective of age (*Ella está joven*, 'She is young; *El es muy viejo*, 'He is very old') rather low in the MONOSC and BISC groups; however, we can see that in both groups there is only a 21 to 25% of the innovative use with adjectives of age as compared to Silva-Corvalán's (1994), whose data shows 78% of innovative uses of *estar* with adjectives of age. For adjectives of physical appearance (animate) (*Ella está rubia*, 'She is blond'; *Ella está pelirroja*, 'She is redheaded'), the MONOSC and BISC groups produced similar percentages for the innovative use of *estar*, 59% and 61% respectively. This means that there was no variation between groups in the use of the innovative *estar* with adjectives of physical appearance. In comparison, Silva-Corvalán's (1994) data showed only a 31% percent of innovative uses of *estar* with adjectives of physical appearance (animate). On the other hand, all adjectives of physical state in both groups appeared with the conservative use of *estar* this means that there was no variation within the L2 learners from both groups. It should be noted that all these types of adjectives were accompanied by past participles for example, *Ella está sentada en la silla*, 'She is seated on the chair,' or *Ella está rodeada de libros* 'She is surrounded by books'. As mentioned previously, adjectives of description refer to those for inanimate subject referents, for example, *El parque es muy hermosa* 'The park is very beautiful'; *Una flor que es muy bonita también* 'A flower that is also very pretty'. The marginals showed that

the MONOSC group used the innovative *estar* 28% with the adjectives of description while the BISC used it 46%. These percentages indicate that variation is present between the MONOSC and BISC groups on the innovative use of *estar*. Interestingly, Corvalán's (1994) data showed to be more consistent with the results found in the MONOSC, with a 25% of innovative uses of *estar*. Next, we have adjectives of evaluation (*El cielo fue muy especial* 'The sky was very special'; *El tiempo es perfecto* 'The weather is perfect'), the MONOSC and BISC groups used the innovative use of *estar* similarly, 9% and 11% respectively, while Silva-Corvalán's (1994) results had a slightly higher percentage of innovative uses of *estar*, 23%. Again with adjectives of emotional state (*Están muy alegres* 'They are very happy'; *Estaban completamente satisfechos* 'They were completely satisfied'), we see similarities in the MONOSC and BISC groups, 73% and 72% respectively. These percentages indicate that there is no variation in the innovative uses of *estar* with adjectives of emotional state between groups. Within the adjectives of emotional state, the L2 learners from both groups commonly used the adjective *feliz* 'happy'; prescriptively *feliz* should be used with *ser* although the other adjectives in the our data prescriptively should be used with *estar*. As a result all the *feliz* 'happy' adjectives were extracted and a separate run was done to see its variability between the MONOSC and BISC groups separate from the other adjectives of emotional state. It appears that the adjective of *feliz* 'happy' is variable, although more variation is found in BISC group, using *feliz* 'happy' more often with *estar* than *ser*, 62% and 38% respectively. There was also variation found in the MONOSC group, although not as prominent, *feliz* 'happy' was used more often with *ser*, 57% than *estar*, 43%. It should also be briefly mentioned, with regards to Silva-Corvalán's (1994)

data, she included emotional state adjectives under miscellany due to the rather low sample of adjectives of emotional state. Thus no comparison was made to Silva-Corvalán's results. The last type of adjective was color, the MONOSC and BISC groups as well as Silva-Corvalán's (1994) results showed similar percentages across the board, 11%, 15% and 11% respectively. These percentages indicate that there is no variation in the use of the innovative *estar* with adjectives of color. We can then assume that generally adjectives of color are used with the standard copula (*ser*).

Previous research (i.e., Gutierrez, 1992; Gutierrez, 1994; Silva-Corvalán, 1994; Gutierrez, 2003; Cortés-Torres, 2004) confirms somewhat similar results. In spite the fact that the participants in both groups are L2 learners, their frequency of adjectives with *ser* or *estar* was influenced by either the continuous or lack of contact to Spanish native speakers in within their speech communities. The marginals revealed that there is variation between the MONOSC and BISC groups although the level of variation differs according to the type of adjective type. However, this is a very cautious conclusion and it would be wise to carry out a study that solely looks at adjective type along with speech community. As previously mentioned, adjective class had to be removed from the final analysis due to strong interactions.

Person/Number

Subject class was initially considered for the current model; however, it showed to be highly interactive with the other factor groups. Unlike adjective type, it was not removed completely. Instead, to maintain detail subject class was recoded under grammatical person and animacy. There were only 21 instances of 1st and 2nd person singular in the data, which may be related to the type of task used (creating a story about characters in a book) which favors 3rd person. Therefore, the categories with the

highest frequencies of *estar* were generally those in the 3rd person category. For example, 3rd person plural animate (person), (*Ellos están muy ocupados*, 'They are very busy'), 75%, followed by 3rd person singular animate (person) (*Ella estaba joven*, 'She was young'), 74%, then followed by 3rd person plural inanimate (other), (*...y ahora sus sueños están completos*, '...and now their dreams are complete'), 57%.

The three categories with the lowest *estar* frequencies include only inanimate categories; 3rd person singular inanimate (other), (*La cosa está, no sé el nombre*, 'The thing is, I don't know the name'), 14%, 3rd person singular inanimate (tangible), (**Una lámpara está apagado en la esquina del cuarto*, '*In the corner of the room, a lamp is turned off'), 22%, and 3rd person plural inanimate (tangible)(**...y una camiseta que son rojos*, '*...and a shirt that are red'), 32%. It can be generalized that those categories with the lowest *estar* frequencies are those that are not living things, therefore allowing for the assumption, that *estar* is most likely to appear with living referents.

Complement

In this section we discuss the variability that may be present between the L2 learners from the MONOSC and BISC groups with regards to complement type and the use of *estar*. In the MONOSC and BISC groups, the percentage for the gerund complement is the same, 97%, indicating that there is no variation between the L2 learners from both groups. Locatives were the second category with the highest percentages of *estar* in both groups, (i.e., The MONOSC group with a 95% and the BISC group with a 90%). These percentages primarily indicate that there is no variation on how the L2 learners from both speech communities use locatives with *estar*, indicating that lack of contact with native Spanish speakers does not influence the traditional use of *estar* with location. And possibly the [*estar* + location] is learned in the

classroom. Moreover, the preliminary results revealed that there is also no variation with regards to the 'adjective' category. Note that the L2 learners from the MONOSC and the BISC produced equal percentages, 41% of *estar* with the adjective complements. It seems to be that the type of speech community is not relevant to any sort of copula variation, particularly in gerunds, locative, and adjectival. In other words, the L2 learners from both speech communities incorporate *ser/estar* similarly regardless of speech community. Meanwhile, variation is found in adverbs and prepositions. The L2 learners from the BISC group have a higher percentage of *estar* when accompanied by prepositional complements (52%) while the L2 learners from the MONOSC group, have a considerably lower percentage, 17%. In comparison, adverbial complements show a similar pattern, the L2 learners from the BISC group demonstrate a higher rate of *estar*, 55% while again in the MONOSC group, the percentage is low, 18%. Similarly, variation is also present in clauses, nouns and pronouns. In the BISC group, clauses appear with *estar* 14% of all occurrences while in the MONOSC they appear 0%, indicating some sort of variation between L2 learners from both types of speech communities, although this variation is not as strong as the one seen in the adverbial and prepositional complements. This sort of weaker variation can also be seen in pronoun complements in the BISC group with a 20% of *estar* occurrences while in the MONOSC group we can see slightly more variation with 32% of *estar* occurrences. The same occurs for the noun complements, in MONOSC group there is a 7% while in the BISC group there is a 14% of *estar* occurrences.

In essence we can assume that variation exists only in certain types of complements (i.e., adverbs, prepositional, noun and clauses) and not in others (i.e.,

gerunds, locatives and adjectival), a possible reason for the lack of variation in certain complement types could certainly be related to the classroom. Generally speaking gerunds and locatives are taught as combinations with the standard copula *estar*, for example in the present progressive or [location + *estar*]. The complements that do have variation may be because there are no specific combinations taught formally, as a result, L2 learners have more flexibility to experiment with *ser* and *estar* according to what they are exposed to within their speech communities.

Multivariate Analysis Results for Linguistic Factors

In the following subsections, the results for the step-up/step-down analysis are presented. Table 4-10 illustrates the best model for the appearance of *estar*, showing that collocations, grammatical person and animacy, and TAM are the only factor groups to contribute significantly to the frequency use of the application value. Overall, the factor weights point out that collocations (i.e., prepositions, modifiers and other) are most likely realized with the *estar*.

Collocations

Table 4-10 exhibits the statistical results for collocations. Collocations are the factor group with the largest range, 45. The range indicates that compared to other factors in the model it has the strongest influence on the frequency of the application value. A similar tendency is found in the constraint ranking within the categories: For example, prepositions have the strongest probability weight, .80, indicating that prepositions have the strongest contribution to the use of *estar*. The next category is the 'modifier' group, which has the second strongest probability weight; although at .49 it does not favor the application value (*estar*).

The last category, 'other,' is composed of all the categories that either produced small raw percentages or were part of any sort of violation in the preliminary run. The 'other' category has the weakest weight, .35 that is insignificant to the frequency of the application value. Overall, the constraint ranking is successful in demonstrating that prepositions have the strongest relevancy to the frequency of *estar*.

Grammatical Person and Animacy

Table 4-10 also displays the results for grammatical person and animacy. As a whole, grammatical person and animacy are the second factor group with the largest range, 40, comparable to the collocation factor; these factors have a somewhat equal impact on the occurrence of the application value.

In regards to the constraint ranking, 'person/3rd person plural animate' and 'person/3rd person singular animate' exhibit the same strong weight, .70. These weights suggest that both of these categories strongly favor the use of *estar*. Of note, both of these categories are animate and belong to the 3rd person, suggesting that 'living things' have an important influence on the frequency of the use of *estar*. This is further supported by the last two categories in this group: 'other' and 'tangible 3rd person singular/plural inanimate, which have relatively have weak weights, .40 and .30 respectively. From this, we can speculate that 'nonliving referents' are not significant to the frequency of *estar* and that person and number are also insignificant.

To summarize, the categories in this group show a notable division between them, note that person 3rd person plural and person 3rd person singular animate share the same probability weight while the other two categories clearly have lower weights, meaning that they do not favor the application value.

Tense, Aspect and Mood

Lastly, Table 3-10 also displays the statistical results for tense, aspect and mood (TAM), which are part of the current model, but with the lowest range in comparison to the other two factor groups, 16. This range implies that this group has the weakest relevancy to the frequency of *estar*. In regards to the constraint raking, it exhibits two important pieces of information; first it shows that the present indicative category slightly surpasses the weight over .50, indicating that it favors the use of *estar*. Secondly, note that the two groups pattern a clear division; the 'other' category has a slightly weaker probability weight of .36, indicating that it does not favor the occurrence of *estar*. Recall that the 'other' category is composed of all the categories (e.g., preterit, imperfect, present perfect, future, conditional), these appeared in very small percentages or had some violations (e.g., knockouts, singletons and crossovers). Overall, the results suggest that the present indicative was the only tense that had any sort of weight on the occurrence of the *estar*.

Concluding Remarks

Overall, the linguistic model is successful in showing the factor groups that influence the appearance of *estar*. Furthermore the probability weight of each category allows for a clear understanding of those categories that favor *estar* and those that do not. To summarize, collocations and grammatical person and animacy seem to have more influence on the frequency of *estar*, while TAM does not. Moreover, prepositions, person 3rd person plural, person 3rd person singular animate and the present indicative favor *estar* while the 'other' categories in the three factor groups and tangible 3rd person sing/pl inanimate do not.

Chapter 4 presents and discusses the marginal results as well as the statistical results for the extralinguistic and linguistic factors for the advanced L2 learners from an English monolingual speech community. Again, note that the absence of the statistical results for the extralinguistic factors; however, the linguistic factors do have marginal as well as statistical results.

Table 4-1. Overall distributions of *ser* and *estar* by L2 ~ bilingual community

Factor Group	<i>Estar</i>	<i>Ser</i>
Individuals		
Participant B1	31 55%	25 45%
Participant B2	70 71%	29 29%
Participant B3	56 68%	27 33%
Participant B4	52 55%	42 48%
Participant B5	29 42%	40 58%
Participant B6	30 36%	53 64%
Participant B7	28 46%	33 54%
Participant B8	23 40%	35 60%
Participant B9	63 66%	33 34%
Participant B10	78 47%	89 53%
Participant B11	41 41%	58 59
Participant B12	98 89%	12 11%
Participant B13	109 55%	90 45%
Total tokens 1274	708	566

Table 4-2. Overall distributions of *ser* and *estar* according to gender

Factor Group	<i>Estar</i>	<i>Ser</i>
Gender		
Male	338 55%	278 45%
Female	370 56%	288 44%
Total tokens 1274	708	566

Table 4-3. Overall distributions of *ser* and *estar* according to age of formal instruction in Spanish

Factor Group	<i>Estar</i>	<i>Ser</i>
Acquisition age		
No answer	31 55%	25 45%
0-5 years old	41 41%	58 59%
6-9 years old	171 53%	153 47%
9-11 years old	113 53%	100 47%
12- Older	352 60%	230 40%
Total tokens 1274	708	566

Table 4-4. Overall distributions of *estar* and *ser* according to years of formal instruction in Spanish

Factor Group	<i>Estar</i>	<i>Ser</i>
Yrs. of formal instruction		
4-7 years	208 55%	171 45%
8-11 years	422 58%	306 42%
12 or more years	78 47%	89 53%
Total tokens 1274	708	566

Table 4-5. Overall distributions of *estar* and *ser* according to time study abroad

Factor Group	<i>Estar</i>	<i>Ser</i>
Time spent studying abroad		
0-6 months	677 56%	541 44%
18 or more months	31 55%	25 45%
Total tokens 1274	708	566

Table 4-6. Overall distributions of *estar* and *ser* according to Spanish proficiency self-evaluation

Factor Group	<i>Estar</i>	<i>Ser</i>
Spanish Self-evaluation		
Poor	30 36%	53 64%
Good	189 49%	198 51%
Very good	458 61%	290 39%
Native like	31 55%	25 45%
Total tokens 1274	708	566

Table 4-7. Preliminary results for advanced L2 Learners from Spanish/English bilingual speech communities *Total N 1173*

Factor Group	N <i>Estar</i>	% <i>Estar</i>	N
<u>TAM</u>			
Progressives	350	97	361
Present indicative	321	40	795
Other	14	32	44
Past tenses	16	25	64
<u>Collocations</u>			
Participles	337	97	346
Prepositions	102	73	140
Other	45	38	118
Adjectives	108	38	284
Nouns	32	18	175
Adverbs	77	3	201
<u>Adjective class</u>			
Physical estate	18	100	18
Emotional state	41	72	57
Feliz	51	62	82
Physical appearance	11	61	18
Description	5	46	11
Other	9	45	20
Sensory character	4	33	12
Age	3	25	12
Color	14	15	92
Evaluation	7	11	64
Social class	0	0	3
Size	0	0	10
<u>Grammatical person & animacy</u>			
1 st & 2 nd person singular	19	91	21
3 rd person singular animate (person)	374	74	506
3 rd person plural animate (people)	206	75	274
3 rd person plural inanimate (other)	4	57	7
3 rd person singular animate (animal)	11	36	31
3 rd person plural inanimate (tangible)	12	32	37
Non-conjugated	11	34	32
3 rd person singular inanimate (tangible)	55	22	248
3 rd person singular inanimate (other)	16	14	118

Table 4-7 Continued.

Factor Group	<i>N Estar</i>	<i>% Estar</i>	<i>N</i>
<u>Complement</u>			
Clause	3	14	21
Pronouns	3	20	51
Noun	30	14	215
Adverb	12	55	22
Locative	68	90	76
Adjective	163	41	399
Preposition	29	52	56
Gerund	351	97	362
Other	6	86	7

Table 4-8. Preliminary results for advanced L2 Learners from the BISC and MONOSC Groups

<u>Adjective Class</u>	N Estar		SC 1994 Number	% Estar		SC 1994 Percentage
	MONOSC	BISC		MONOSC	BISC	
<u>Category</u>						
Age	3	3	136	21	25	78
Size	0	0	101	0	0	53
Sensory Character	1	4	8	25	33	47
Physical Appearance	23	11	43	59	61	31
Physical State	4	18	0	100	100	0
Description	5	5	12	28	46	25
Evaluation	6	7	90	9	11	23
Emotional State	55	41	0	73	72	0
Color	4	14	4	11	15	11
Other	13	9	10	77	45	18
No adjective	298	607	0	52	60	0

Table 4-9. Multivariate analysis of the contribution of linguistic factors selected as significant to the probability of the copula

L2 learners from a bilingual speech community				
Corrected mean				0.36
Log likelihood				-493.2
Total N				896
	Factor weight	N <i>Estar</i>	% <i>Estar</i>	N
Collocations				
Prepositions	0.8	99	72	137
Modifiers	0.49	175	37	475
Other	0.35	70	25	284
<i>Range</i>	45			
Grammatical Person and Animacy				
Person 3 rd person plural animate	0.7	97	60	162
Person 3 rd person singular animate	0.7	152	55	278
Other	0.4	47	24	194
Tangible 3 rd person sing/pl inanimate	0.3	48	18	262
<i>Range</i>	40			
TAM				
Present indicative	0.52	318	40	792
Other	0.36	26	25	104
<i>Range</i>	16			

CHAPTER 5 DISCUSSION AND CONCLUSION

The main objective of this study was to investigate whether Spanish contact is relevant to the frequency of *estar* use among advanced second language learners of Spanish. The first group of second language (L2) Spanish learners investigated lived continuously in a Spanish/English bilingual speech community, while the second was composed of second language learners (L2) from an English monolingual community. Although both groups were equally exposed to formal language instruction in the classroom, the investigation sought to determine what effect, if any; passive exposure to Spanish via speech community could have on the usage of the copular verbs among second language learners.

A subsequent objective was to explore linguistic (e.g., tense, aspect and mood (TAM), person and number, collocations, adjective class and subject class) and extralinguistic (e.g., gender, number of years of formal instruction in Spanish, amount of time in a study abroad in a Spanish speaking country, knowledge of third language, and proficiency level in Spanish) factors vis-à-vis their influence on copula choice. Finally, we explored if these factors, varied depending on speech community.

This chapter briefly summarizes the results presented in Chapters 3 and Chapter 4, in order to discuss and interpret them in light of the research questions that guided the study. Finally, we conclude with limitations, directions for future research, and final thoughts.

Advanced Learners from Two Speech Communities: A Comparison of Linguistic Factors

The table below illustrates the multivariate analysis results of the linguistic factors where the two participant groups exhibit certain similarities and differences in the factor

groups that appear to contribute to the frequency of *estar*. These include collocations, grammar person and animacy, and TAM. In addition, it includes the categories of each factor group, and their corresponding probability weights, as well as the range for all factor groups included in the statistical model.

The subsequent sections will discuss each of the factors selected as significant. Table 5-1 contains the following information: contains the factor group in column one, column two and three contain the weights and ranges for the BISC and MONOSC groups. The table also provides the input probability, the number of tokens and the log likelihood.

Collocations

In the step-up/step-down analysis in both participant groups, collocations appear as one of the factor groups with the strongest influence on the frequency of *estar*, although with different ranges. The range for collocations in each type of speech community is somewhat significantly different. The collocations in the Monolingual English speech community group (i.e., MONOSC group) show a stronger influence on the frequency of *estar* than in the BISC group. This difference indicates that the use of the copula is not the same among groups, meaning that perhaps this difference may be related to the amount of contact that the participants received while growing up.

With regard to the constraint ranking within the categories, there is an evident divide in both groups, although in the category of 'prepositions,' both groups have almost identical probability weights, indicating a strong favor to the frequency of *estar*. So perhaps the amount of contact that the L2 learner is exposed to outside the classroom may have little influence in this case. This of course suggests that on the

whole prepositions are very strong indicators for the appearance of *estar* regardless of speech community.

Other categories though show statistically significant differences between speech community populations. In the Spanish/English Bilingual community group (i.e., BISC group), modifiers have a somewhat weak probability weight, but in the MONOSC group, modifiers favor the appearance of *estar*. The same tendency can be seen in the other category, which is also significant to the occurrence of *estar* in the MONOSC group. Although in the BISC group, it appears with a relatively weaker probability weight. Overall, the categories in the BISC group do not favor the frequency of *estar* while in the MONOSC group generally all the categories strongly favor it. In this case, the higher rate of *estar* found in the categories belonging to the MONOSC may be related to the amount of contact to which the participants have been exposed throughout life, and also to other extralinguistic as well as social factors that often can influence how L2 learners may learn a mixture of forms. Of course, in this case further research would be needed to confirm these assumptions.

Also worth noting in the MONOSC group, is the noun category, which appears with the weakest probability weight, indicating that it does not favor the frequency of *estar*. The BISC group, however, the noun category does not appear in its statistical model; thus indicating that its absence may be related directly to the recoding process, which took place in order to arrive at the final statistical run. In other words, in the preliminary run there were cross-tabulations that showed empty cells, therefore, forcing this category to be recoded under the other category for the final run.

To summarize, both speech communities reveal that prepositions strongly favor the frequency of *estar*. In the MONOSC group, the modifier and the other categories appear to also strongly favor it, whereas in the BISC group neither the modifier nor other categories favor *estar*. Given the results, collocations can tentatively be suggested as part of the model that distinguishes the speech of monolingual communities and Spanish-English bilingual communities in Florida.

Grammatical Person and Animacy

In this study, although animacy is collapsed with grammatical person, it is worth noting that previous research has shown that animacy individually has varied results (Geeslin 2006: p. 13). For example, in Silva-Corvalán's (1986, 1994), she found that animacy does not affect copula choice; however, Geeslin (2000) found that animacy, when included in the statistical analysis individually, was a significant statistical predictor of copula choice (Geeslin 2000: p. 54).

In Table 5-1 we also see the results from the step-up/step-down analysis for grammatical person and animacy for both populations. This factor group appears to have the highest ranges, particularly in the MONOSC group, meaning that its magnitude of effect favors the frequency of *estar*. In the BISC group it appears as the factor group with the second highest range. These ranges suggest that in the MONOSC group, grammatical person and animacy is the most significant group to favor the frequency of *estar* while in the BISC group, it is the second group to favor it. The given ranges show a noticeable separation of both speech communities. Although in both speech groups the factor of grammatical person and animacy favors *estar*, it appears with a stronger magnitude of effect in the monolingual community.

Within this group, the category with the strongest probability weight in both populations is 'person/ 3rd person plural animate,' suggesting that 'person/ 3rd person plural animate' strongly favors the frequency of *estar* in both groups over other conjugations and the inanimate form. Nonetheless, there is a strong possibility that these results may be directly related to the task used to gather data in the present study; because participants were narrating a story about other people, they manifested a tendency to use the 3rd person. In other words, a different task (i.e., open interview) may have produced different results. For example, '1st person singular and plural animate' may have appeared to have stronger factor weights if the participants were talking about themselves and no other characters.

Similarly, the category of 'person/3rd person singular animate' has some of the strongest weights in the constraint ranking as well, suggesting that this category also has a significant influence on the occurrence of *estar*. The 'other' category, however, shows significant differences between the two speech community groups. In the MONOSC group the Other category favors the appearance of *estar*, with the second strongest weight within the group, while in the BISC group the Other category does not, having one of the weakest weights in the constraint ranking. Furthermore, the categories of 'animal/3rd person singular animate' and 'other/3rd person singular/plural inanimate' were not part of the model for the BISC group. Although they were part of the statistical model for the MONOSC group, both seem not to favor the appearance of *estar*, with relatively weak weights. In regard to the category 'tangible/3rd person singular/plural inanimate' also appears in both models, with equally low probability

weights, indicating that it does not favor the appearance of *estar* in either speech population.

To summarize, in the MONOSC group, the 'person/3rd person plural animate,' 'other,' and 'person/ 3rd person singular animate,' categories all favor the appearance of *estar* with very strong weights, .86, .62 and .61 correspondingly. In contrast, the categories animal/ 3rd person singular animate, tangible/3rd person singular/ plural inanimate, and 'other/3rd person singular/ plural inanimate' do not favor the occurrence of *estar*, appearing with considerably weaker weights, .44, .30, .17 respectively. This distinction is also present in the BISC group, in which the categories 'person/3rd person plural animate' and 'person/ 3rd person singular animate' appear with an equal probability weight, .70. On the contrary, the categories 'other' and 'tangible/3rd person singular/ plural inanimate' do not favor the appearance of *estar* with relatively weaker weights, .40 and .30 correspondingly. The results for the 'tangible 3rd person sing/pl inanimate' category suggest that the amount of exposure to the target language may essentially not have any sort of influence on the frequency of *estar* with regard to tangible things; that is to say, tangible things do not favor the frequency of *estar* regardless of the amount to contact to Spanish. According to the results, grammatical person and animacy form part of the model that distinguishes the speech of the English monolingual communities between the Spanish/English bilingual speech communities in Florida.

In conclusion, the difference between groups with regard to these categories may be directly associated to the amount of exposure that the participants have had with Spanish during their life. This notion can be assumed due to the clear difference in

probability weights between populations. The results suggest that the type of speech community may perhaps have a significant relevance to the grammatical person and animacy factor group given that there is a clear divide between speech community groups. We can see this divide within the categories, suggesting that the less exposure to Spanish produces weaker probability weights and the more contact producing stronger ones.

Tense, Aspect, and Mood

The last factor group that appears in the statistical model for both speech community groups is TAM (i.e., tense, aspect, and mood), whose results were also displayed previously in Table 5-1. The category TAM is part of the model for the BISC group, though it has the lowest range of all the factor groups in the model, indicating that it has very little influence on the frequency of *estar*. On the other hand, the MONOSC group has an even lower range, indicating nearly zero statistical significance with respect to the frequency of *estar* use. The categories within this group are discussed individually below.

With regard to the constraint ranking, the ‘indicative’ category has a relatively strong probability weight in the BISC group, which suggests that the indicative favors the frequency of *estar*. While the Other category appears to have a weaker weight, implying that it does not favor the frequency of *estar*. However, in the MONOSC group, TAM did not reach statistical significance, and are included in the statistical model for the sole purpose of representing data in its entirety, as part of the step-up/step-down analysis.

The results propose a clear difference between the MONOSC and the BISC groups, suggesting that the amount of exposure to Spanish that the participants have

had during their lives may be related to the frequency of *estar*, in regard to the TAM factor group. Overall, the only tense relevant to the frequency of *estar* is the present indicative, which is only for the BISC group. This denotes that perhaps the BISC population may be acquiring the copula patterns found in Spanish/English bilinguals.

Research Questions Revisited

In light of the results summarized above, we are now in a position to revisit the research questions that motivated the investigation. Recall that the present study was guided by three questions, each of which are restated and discussed in the following subsections.

Research Question 1: Frequency of *Estar*

The first research question asked how the frequency of *estar* among advanced learners who lived continuously in a monolingual area (MONOSC Group) compared to that of the advanced learners who lived continuously in a bilingual area (BISC group). While the numbers in Figure 5.1 were previously presented in Chapters 3 and 4, these numbers are repeated here in order to visualize a comparison between groups.

As can be seen, the frequency of *estar* among advanced L2 learners who were continuously in an English monolingual region are indeed different from the frequency seen in the L2 learners who lived continuously in a Spanish/English bilingual region. Overall, we can see that the L2 learners from the BISC group have a higher frequency of *estar* use than those from the MONOSC group. Even so, the difference in the frequency of *estar* exhibited by both groups cannot entirely be related to the type of speech community or the amount of contact. It is crucial to understand that the participants in each group were not completely identical in regards to their demographic background, they varied in gender, the amount of time they studied abroad, the number

of years of formal instruction, and the amount of exposure to native Spanish speakers. Therefore, a definite conclusion cannot be made regarding the relevance that speech community has on the frequency of *estar* in advanced L2 learners alone. At the same time, we must consider that other extralinguistic factors may also play a role in the frequency of *estar* use. Note for example, participants M3, M4, M6, M11, all of whom display a higher frequency than their counterparts, in the same group. This higher frequency cannot be related to the amount of Spanish contact have been exposed to because all of them have similar amounts of exposure to Spanish. The same can be noted for the participants from the Spanish/English bilingual community; for example, B6, B8 and B11, have comparable percentages of contact to Spanish yet they produced lower rates of *estar* than other members of their group, indicating again that speech community may not be the only factor involved in *estar* use.

Therefore, although the higher frequency of *estar* found in the BISC participants indicates that their frequent contact (solicited or unsolicited) with native Spanish speakers may have influenced their copula acquisition or use there are other factors to consider as well.

Previous research claims that L2 learners with ample access to native speakers tend to show variation (similar to Spanish native speakers) and generally these learners are relatively advanced (Geeslin, & Guijarro-Fuentes 2006: 6-7), as corroborated by the current results.

Research Question 2: Extralinguistic Factors

The second research question asked how extralinguistic factors (e.g., gender, number of years of formal instruction in Spanish, age of formal instruction in Spanish, duration in a study abroad program in a Spanish speaking country, and Spanish

proficiency) correlate to the frequency of *estar* use among advanced L2 learners from the experimental communities. A comparison of the raw data from both populations can provide relative frequencies and percentages as a preliminary glimpse at copula use. The descriptive data also revealed the distribution of the variants in the dependent variable, allowing us to see the contrast between the categories within a factor group (Tagliamonte 2006: p. 137).

The following sections review the similarities and differences found in the extralinguistic factors between the English monolingual speech community and the Spanish/English bilingual speech community and offer tentative explanations for these trends.

Gender

Figure 5.2 illustrates the results in percentages for the frequency of *estar* use according to gender of the learner from both speech communities.

The numbers revealed that males and females in the MONOSC group produce similar percentages of *estar* use. An even stronger similarity is found in the BISC group where males and females have almost exactly the same percentages. Because majority of variation studies (e.g., Boretti de Macchia & Ferrer de Gregoret, 1984; Valdivieso & Magaña, 1991; Bonvillain, 1993; Navas Sánchez-Élez, 1997; Watt & Milroy, 1999; Milroy, 1999; Mougeon and Rehner, 2001; Geeslin & Guijarro-Fuentes 2006; Aguilar-Sanchez, 2009;) have shown that gender is a predictor of variation, it was expected to also find it here. Instead the relatively equal percentages of *estar* found in males and females indicate that the gender does not seem to influence the use of *estar*. In essence, gender had no direct effect on the use of *estar* in this data. Moreover, Dewaele & Regan's, (2002), Salazar's, (2007) and Geeslin's (2006) studies also

corroborated that gender was not relevant variation. Therefore we turn to other factors that might play a role in the subsections that follow. Before proceeding, however, note the slight differences found in gender between speech communities, the males in the MONOSC group have a slightly lower frequency of *estar* use from those in the BISC population. This slight difference proves to be the same for females. For example, in the MONOSC group the females also show a slightly lower frequency of *estar* than the females from the BISC group. Overall, we see that the learners from the bilingual population show a slightly higher frequency of *estar*, which may suggest that the amount of exposure to Spanish may have some sort of correlation to it when considering gender, although this correlation may not be as strong as posited in Chapter 1.

Number of years of formal instruction in Spanish

Previous research (e.g., Dewaele and Regan 2002; Geeslin 2002) revealed that the length of formal instruction had no effect on the vernacular speech of the L2 learners, however, the amount of active use of the target language outside the classroom and amount of exposure through radio and television were positively correlated with the use vernacular forms (Dewaele and Regan 2002: p. 148). Figure 5-3 illustrates the percentages of *estar* use according to the number of years of formal instruction the learners had. These percentages include those from both speech communities In the MONOSC group there were no participants with 12 or more years of formal instruction in Spanish, therefore there is no column for this in Figure 5.3. Even so, in the MONOSC group, the L2 learners with 4-7 years of formal instruction exhibited slightly different percentages, (52%) than those with 8 to 11 years (46%), the same can be said of the BISC group, where the participants with 4 to 7 years of formal instruction

(55%) and 8 to 11 years (58%) also exhibited very slight differences, including those that received 12 or more years of formal instruction (47%). These slight differences can also be seen when comparing speech populations, for example, the L2 learners who received 4 to 7 years of formal instruction in the MONOSC group showed only a slight difference in the production of *estar* from those with the same amount of formal instruction in the BISC group, 52% and 55% correspondingly. However, the L2 learners from the BISC group who received 8 to 11 years of formal instruction exhibited a somewhat higher use of *estar* than those who also received the same amount of formal instruction from the MONOSC group, 58% and 46% respectively, suggesting that speech community may have some correlation to those L2 learners with more classroom time. Overall, though the amount of time in the classroom had little effect on the frequency of *estar* use, particularly in the BISC group this may be due to the fact that it had considerable exposure beyond classroom. However, when comparing speech communities, the L2 learners who received 4 to 7 years of formal instruction exhibited percentages, which indicate that the type of community may not have relevance to the production of *estar* among those learners who had less experience in classroom instruction. Conversely, the same may not be true for the participants who experienced 8 to 11 years of formal instruction in Spanish; they showed a somewhat higher difference between community types indicating that the type of community of the L2 learner may be relevant to the frequency of *estar*. We can then conclude that the type of community can affect the frequency of *estar* only in the L2 learners with the most classroom time.

Age of formal instruction in Spanish

Figure 5.4 illustrates frequency of *estar* use according to age at which formal instruction in Spanish first started for speech communities.

Looking at the populations separately, it is evident that in the MONOSC group the youngest learners are the ones that have the highest frequency of *estar*. Interestingly, though there is a good deal of variability among the other age groups. This may be due to the possibility that

once the age of nine is reached, it makes no real difference when formal instruction begins. In other words, similar levels of proficiency should be reached regardless of formal instruction beginning at nine years or older in this community. In contrast, in the BISC community the older group exhibits the highest frequency of *estar* while the younger group has the lowest. On the other hand the two in-between categories, 6 to 9 and 9 to 11 years old, exhibit identical frequency levels.

When comparing groups, there is a clear difference in the frequency of *estar* use between learner groups, which may be related to the type of speech community of the L2 learner. Overall the learners from the Spanish/English bilingual group showed higher frequencies of *estar* use across the board, possibly indicating correlation between *estar* use and exposure to Spanish outside of the classroom.

Duration of study abroad in a Spanish speaking country

The preliminary numbers for the frequency of *estar* according to time spent in a study abroad program in Spanish speaking country are illustrated for both groups in Figure 5-5. Before proceeding, attention should be drawn to Figure 5-5, notice that for the BISC group there are no columns between 6 to 18 months of study abroad experience, and for the MONOSC group there is no column for the category of 18 or

more experience abroad, meaning there were no participants belonging in these categories.

In the MONOSC group the L2 learners who studied between 0 to 12 months in an abroad program demonstrate the lowest frequencies of *estar* use. In contrast, there is a relatively higher frequency of *estar* for the L2 learners who studied abroad between 12 to 18 months. These preliminary numbers could potentially mean that the more time spent in an abroad setting the more likely that *estar* will be more frequent in L2 speech. Previous research e.g., Regan, 1999; Sax, 1999; Mougeon & Rehner 2001; Dewaele, and 2004; concurred, stating contact (i.e., study abroad) with native speakers of the target language stimulates the use of vernacular speech (Dewaele 2004: p. 444). This is important information because it suggests that once participants are immersed in the TL speech community, they will try to approximate the input that they receive (Regan 1996: p. 178), and as a result their speech might resemble the variation found in native speakers. It would be of interest, however, to further support this statement in the future by including L2 learners who have studied abroad longer than 18 months, to see if their production of *estar* is higher than the one seen in the L2 learners that studied abroad between 12 to 18 months.

Conversely, the BISC group only includes data for L2 learners who studied abroad between 0 to 6 months and 18 or more months. Interestingly, the frequency of *estar* in the L2 learners from the lesser time in an abroad program show almost identical percentages of *estar* to L2 learners who have studied abroad longer periods of time. This could mean that any immersion effects are nullified by their extensive community contact. It also seems like the L2 learners from the English monolingual communities

may need to go abroad to start using *estar* more, but those in the Spanish/English bilingual communities do not need that additional exposure since they already had. It is possible that living in a region with constant contact to Spanish may have more influence on the production of *estar* than a study abroad experience regardless of time spent abroad.

Spanish proficiency

Figure 5-6 illustrates the frequency of *estar* according to the self-reported proficiency level data.

In the current study, the majority of the participants self-reported their proficiency level as 'good' or 'very good'. The largest difference in the frequency of *estar* use is found in the BISC group, where those who self-rated themselves as 'very good' have the highest use of *estar* use while those who rated themselves as 'good' showed a somewhat lower percentage of *estar*. In this case, the marginal's imply that the L2 learners who have a higher frequency of *estar* have surpassed the beginning stages of Spanish acquisition because they use *estar* more often and those who favor *estar* less frequently are still in the beginning stages of learning Spanish (e.g., Van Patten (1985, 1987).

Also we can see a slight difference between percentages in the MONOSC group, where the L2 learners who evaluated themselves as 'good' have a slightly lower use of *estar*. However, when comparing community types, the L2 learners who self-reported a 'good' proficiency level from both communities demonstrate almost similar percentages. This similarity in percentages may indicate that regardless of amount of contact to Spanish, L2 learners will produce similar frequencies of *estar* use; that is to say, type of speech community may have little influence to the production of *estar* when students

have a 'good' proficiency level. On the other hand, the L2 learners who self-reported a 'very good' proficiency level from both groups, the BISC group exhibited a slightly higher frequency in *estar* than those from the MONOSC group, indicating that speech community may have some sort of relevance to the frequency of *estar* only when L2 learners have reached a 'very good' level of proficiency.

Research Question 3: Linguistic Factors

The final question that guided the current study asked which linguistic factors (e.g., adjective class, collocations, TAM, subject class, person and number) might affect the frequency of *estar* use among L2 learners from different communities. Recall that factors (e.g., collocations, grammatical person and animacy and TAM) formed part of the statistical models for both the BISC and MONOSC groups.

The TAM factor group, which was found to be insignificant to the frequency of *estar* use in both speech communities, appears to have no relevance to the use of *estar*. Even so one significant difference is found in its categories, which revealed that the present indicative favored the frequency of *estar* use in the BISC group while in the MONOSC group it did not. This might indicate that the L2 learners with more contact to Spanish tend to use *estar* more often; in essence the type of speech community does affect the production of *estar*.

Other similarities are found within the categories in factor groups. For example, both populations manifested preference for *estar* with 'prepositions,' 'person/3rd person plural animate' and 'person/3rd person singular animate'. Another parallel is seen in the 'tangible 3rd person sing/pl inanimate' category, in which both speech populations seem to disfavor the use of *estar*. Additionally, the identical weights with this category also indicate that exposure to Spanish outside the classroom does not make a difference

with regard to the 'tangible 3rd person sing/pl inanimate' category. Other categories that did not favor *estar* include: 'nouns,' 'other/3rd person sing/pl inanimate,' 'animal/3rd person singular animate' and 'other/TAM' regardless of populations

Nonetheless, other differences are found between population groups. The first is seen in the ranges: the MONOSC population shows larger ranges for the collocation and grammatical person and animacy factor groups than in the BISC population, indicating that these factors have a stronger influence on the frequency of *estar* in non-contact regions. Furthermore, some individual differences are found in the categories. For example, the modifiers, Other collocation and Other grammatical person and animacy categories have weights that reveal that the presence of these in a bilingual speech community population tend to strongly effect the frequency of *estar*, while in the monolingual speech community they do not. Again, this may imply that in non-contact regions, there is a stronger probability for higher rates of *estar* with regard to specific linguistic factors. In essence, it may suggest that the amount of contact that L2 receive after they have reached a certain level in Spanish may not have a strong effect in the frequency of *estar* use.

Given these results, we can propose some basic assumptions: The first is that some linguistic factors are more relevant than others to the frequency of *estar*, and, more importantly, that these factors are linked to the amount of exposure to Spanish. As shown in this study, there is an obvious division between groups with regards to factor groups and within most of the categories involved.

Advanced L2 learners from regions with less contact to Spanish appear to use *estar* particularly with specific linguistic factors, that is to say, "that the presence of a

given linguistic feature or category tends to affect rule frequency in a probabilistic uniform way in all the environments containing it” (Tagliamonte 2006: p.131).

Furthermore, it can also be assumed that the amount of natural exposure to Spanish that the advanced L2 learners receives may have some influence on which linguistic factors are more relevant to their use of *estar* in the speech of L2 learners. With this said, we can conclude here by suggesting that variation is present between both populations and there is a probability that this variation is caused by the type of speech community. However, also crucial to note is that other extralinguistic or linguistic factors cannot be disregarded completely as causes for this variation.

The following section discusses the some of the drawbacks and future development of this study to reconsider new approaches to investigate variation in L2 learners.

Limitations and Directions for Future Study

Although the current study was successful in revealing that linguistic and extralinguistic factors can affect frequency of *estar* use in two L2 speech communities, limitations are presented in the following paragraphs. The first limitation is related directly to the participant groups. Although there were two principal experimental groups, there was no control group because the main purpose of this study was to compare the use of *estar* in L2 learners rather than assess their native-like use of *estar*. This leaves us, however, with an important gap open for further research. A possible control group would have been composed of native monolingual speakers of Spanish within the same age group as the other participants in order to provide a point of reference with regard to native-like copula patterns. While an attempt was made to use previous literature as point of reference, it ended up being a difficult task given the

limited work on copula use among Spanish native speakers. Future research should allow for the incorporation of a control group thus providing further information on how comparable the frequency of *estar* use in the L2 learners of this study is to the use of *estar* in native speakers of Spanish.

Another limitation to the current study was the method used for data collection, which required participants to create a story from a set of pictures. For future research it may be of interest to also incorporate an open interview in which participants could use forms other than the predominantly 3rd person animate forms provided in the current data set.

A third limitation is related directly to the categorization of participants. It was a fairly complex process to group the participants due to numerous individual differences. The lack of homogeneity among the participants, in both groups, made drawing general conclusions extremely difficult. Additionally, the participants were placed in their corresponding groups based on their speech community descriptions and their responses to the Language Contact Profile (modified from Freed, Dewey, Segalowitz and Halter 2004), neither of which is a foolproof method. Future work should strive to find participants who share more similar social as well as linguistic backgrounds in order to be able to reach more definitive conclusions.

Lastly we recognize problems related to the method used to determine the proficiency level for the participants in the present study. (Recall that each participant self-reported his or her own proficiency level). We are aware that this sort of auto-assessment is problematic, as an all self-reported data measures, as participants may over- or under-estimate their abilities causing inconsistencies in the data. While all

learners were enrolled in courses at the same level (advanced courses), and the self-assessing were only used as an additional source of information, it would nonetheless be favorable for future research to measure proficiency through a more objective approach such as those proficiency tests given by DELE (*Diplomas de Español como Lengua Extranjera*), etc. These sorts of tests provide a more standardized measure of proficiency, and leave less room for error in the data reported.

In spite of these limitations, this study has provided a valuable step closer to understanding how our communities can impact the way we acquire a second language. With the current investigation we understand that there are differences between L2 learners of Spanish who have continuous contact with Spanish speakers and those who barely encounter Spanish speakers throughout their daily lives. Although these differences are subtle, we are able to explain them through a variable rule analysis.

Concluding Remarks

This is one of the few studies to look at variation in L2 in relation to type of community speech and also to incorporate linguistic as well as extralinguistic factors that have rarely been looked at as it relates to L2 learners of Spanish. With the current results, we were able to support previous research that variation is present in the interlanguage of the student, although we still do not know to what extent. We can only propose certain assumptions as a result to the preliminary and statistical results. Yet we can still be sure that the type of speech community does influence how the L2 learner incorporates the copula within certain linguistic and extralinguistic factors.

As the results have shown various linguistic factors as whole are relevant to the frequency of *estar* use among L2 learners, and that exposure to Spanish in their

community can also correlate to copula choice. The factor groups that appeared to be significant are collocations, grammatical person and animacy, and TAM, although depending on type of speech community; each factor group appeared to have different levels of significance. Likewise, we are able to see that each category within factor groups also presents different levels of variation within both groups of L2 learners. Accordingly, this further proves that variation is present and as a result we have to consider it as we continue to research second language acquisition.

Another key point is that the marginal results were also included here; these are successful in showing the number and percentages of instances of *ser* and *estar* in the data of the dependent variable (Tagliamonte 2006: p. 135). With this type of information we were able to derive to specific assumptions of the effects of the extralinguistic factors on the frequency of *estar* and provided information on the factor groups that were not able to be part of the statistical analysis. For example, with regards to adjective factor group we were able to conclude that some variation exists between L2 learners from the two speech communities from this study.

The current investigation has also shown that there are differences in the amount of variation that exists between the L2 learners from both types of speech communities. However, we still have many unanswered questions and future research would greatly benefit from carrying out similar studies. For example, we were able to look at several extralinguistic factors, however as repeated before there was no variable rule analysis, as a result, there was no statistical analysis indicating the strength of each factor, future studies would benefit from carrying out investigations where extralinguistic factors can be analyzed separately.

In sum, the present study has shown that not only certain linguistic factors but also type of speech community can affect the frequency of use of *estar*. We interpret these findings to mean that contact with the target language beyond the exposure received in a traditional classroom setting is in fact relevant to L2 acquisition. Furthermore, this study may serve as a platform for other L2 variation studies, to continue investigating the significance of speech communities, extralinguistic (e.g., social class, knowledge of a third language, age, etc.) and linguistic factors (e.g., adjective class, susceptibility to change, frame of reference, experience with referent etc.). Finally, the current research has substantially contributed to L2 variation studies, which is essential to develop new ways to look at second language and to understand linguistic variability in second language learners.

Table 5-1. Multivariate analysis of the contribution of linguistic factors selected as significant to the probability of the frequency of estar

Factor Group	Factor Weights	
	BISC	MONOSC
Collocations		
Prepositions	X	√
Modifiers	X	√
Other	X	√
Nouns	X	√
Grammatical Person and Animacy		
Person 3 rd person plural animate	X	√
Person 3 rd person singular animate	√	X
Other	X	√
Tangible 3 rd person sing/pl inanimate	√	√
Other 3 rd person sing/pl inanimate	X	√
Animal 3 rd person singular animate	X	√
TAM		
Present indicative	√	X
Other	√	X

√ Indicates favor

X Indicates disfavor

□ Indicates significance

L2 Learners from a BISC: Input probability .36, N= 896

L2 Learners from a MONOSC: Input probability .24, N= 684

Figure 5.1 Frequency of *Estar* in MONOSC and BISC groups

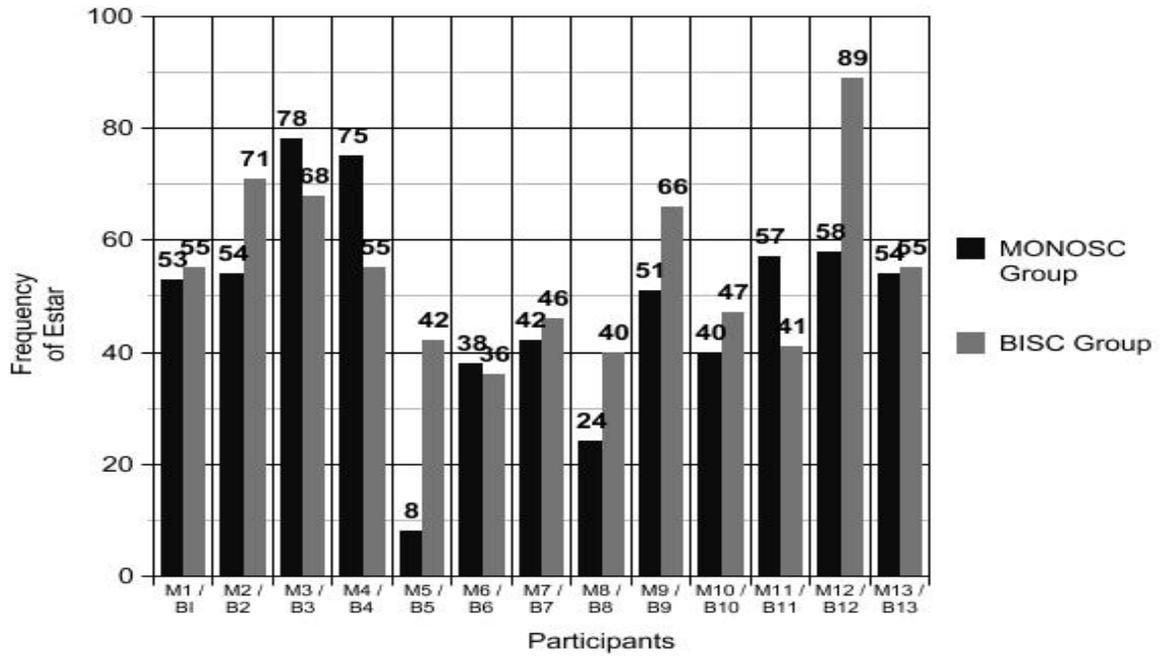


Figure 5-1. Frequency of *estar* in monolingual speech community groups and bilingual speech community groups

Figure 5.2 Frequency of *Estar* According to Gender

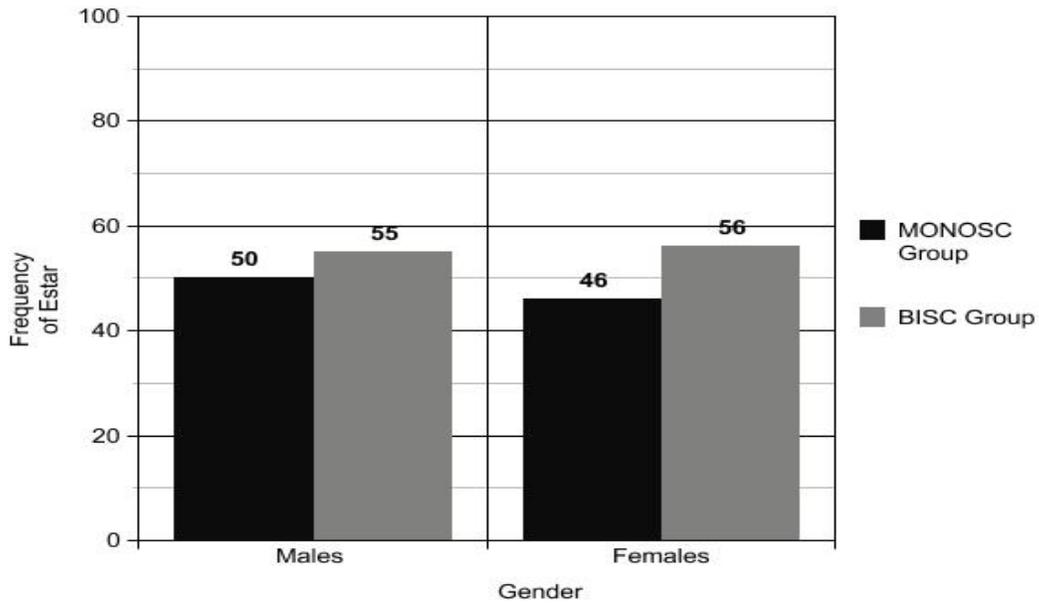


Figure 5-2. Frequency of *estar* according to gender

Figure 5.3 Frequency of Estar According to Number of Years of Formal Instruction in Spanish

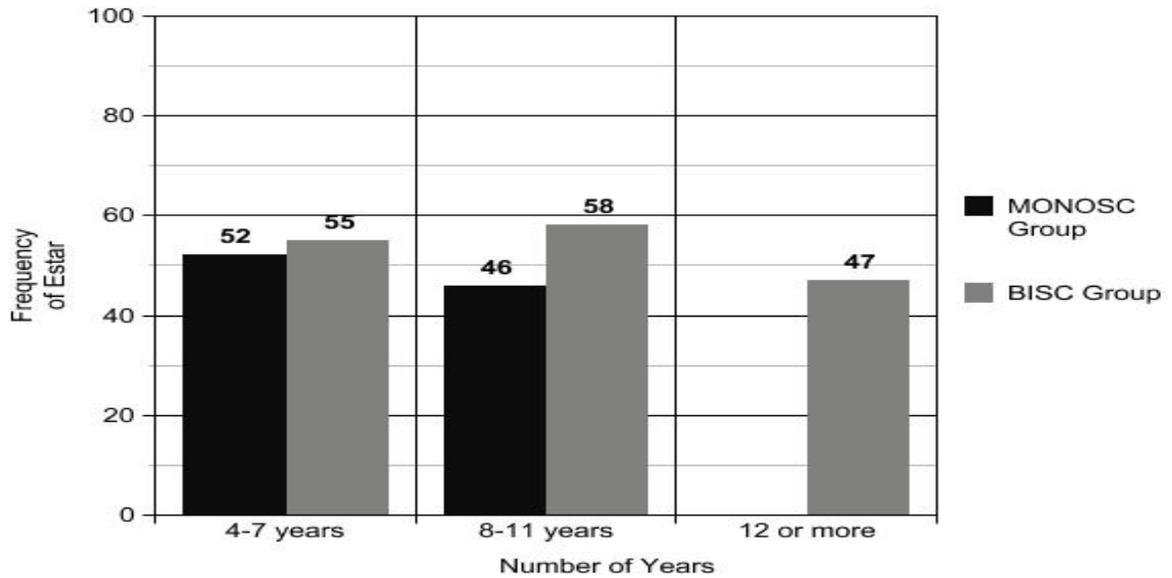


Figure 5-3. Frequency of *estar* according to number of years of formal instruction in Spanish

Figure 5.4 Frequency of Estar According to age of Formal Instruction in Spanish

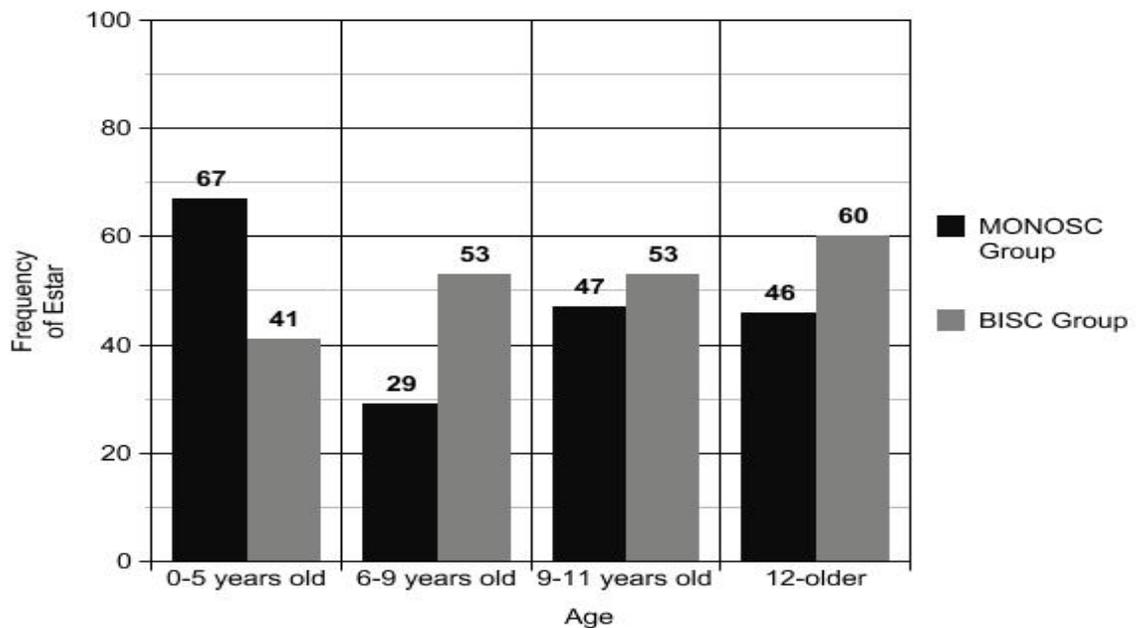


Figure 5-4. Frequency of *estar* according to age of formal instruction in Spanish

Figure 5.5 Frequency of *estar* According to Time Abroad

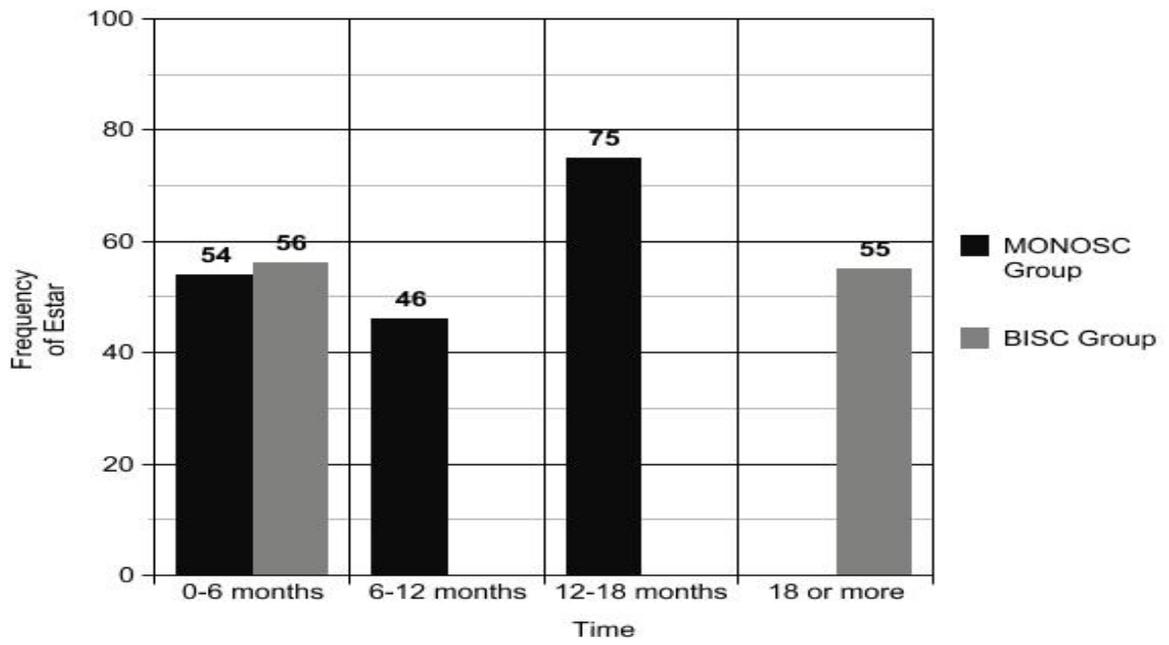


Figure 5-5. Frequency of *estar* according to time abroad

Figure 5.6 Frequency of Estar According to Proficiency level in Spanish

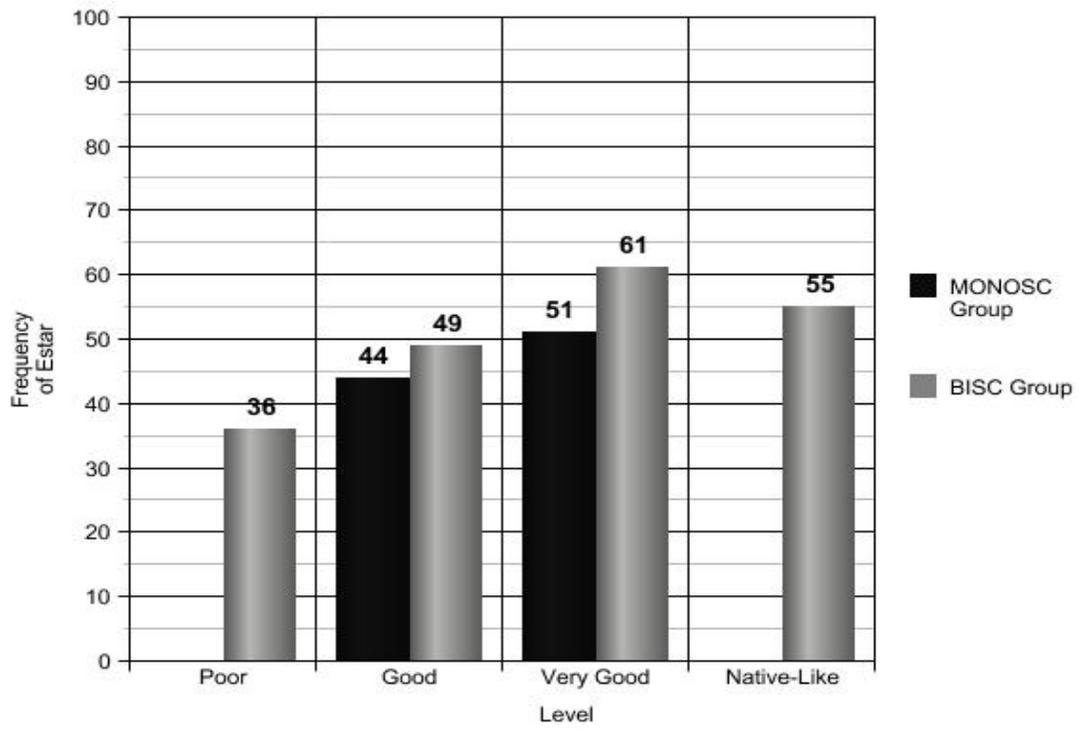


Figure 5-6. Frequency of *estar* according to proficiency level in Spanish

APPENDIX A
INFORMED CONSENT

Variation in Copula Choice: A comparative analysis on advanced second language learners from two speech communities

Purpose of this project: In this dissertation research project I will gather data through the storytelling of American second language learners who are learning Spanish in Gainesville, Florida.

Your role in this project: You will be asked to complete a demographics questionnaire. In addition, you are to create narrative from a children's book that will be provided to you. I will tape record your mini soap opera.

Time required: Each session should take about 30 to 40 minutes.

Compensation: There will be no monetary compensation.

Confidentiality: Your actual name will not be used in any published data and I will only use the information you wish to be included in the project.

Voluntary participation: Your participation in the project is completely voluntary and you have the right to withdraw at any time without consequence. You do not have to answer any question you do not wish to answer.

Risks/Benefits: There are no direct benefits or risks for your participation in this project.

Who to contact with questions: Dorian Dorado, Graduate Student, Department of Spanish and Portuguese Studies, 322 Yon Hall, Gainesville, FL 32611, phone: 352-256-2716, ddorado@ufl.edu

Dr. Gillian Lord, Associate Professor, 235 Dauer Hall, Gainesville, FL 32611, phone: 352.392.2016, glord@ufl.edu

Who to contact about your rights to participate in project: UFIRB Office, Box 112250, University of Florida, Gainesville, FL 32611-2250; ph 392-0433.

Agreement: I have read the project described and I voluntarily agree to participate. I have received a copy of this description.

Participant: _____ Date: _____

Principal Investigator: _____ Date: _____

APPENDIX B
DEMOGRAPHICS

1. Gender _____
2. Age _____
3. Country of birth _____
4. What is your native language? 1) English 2) Spanish 3) Other(s) _____
5. Ethnicity (Cuban, Mexican, White, African-American etc.)

6. Where does your family live?

7. What language(s) do you speak at home? 1) English 2) Spanish 3) Other (s) _____
8. How would you describe language use in your community?

9. How old were you when you were first started formal classes in Spanish?

10. What is your major? _____
11. What is your minor? _____
12. What Spanish classes have you taken? _____
13. Were any of your instructors' native speakers of Spanish? _____ From Where?

14. Have you lived or studied abroad? _____ Where? _____ How long?

15. What is the highest level of education your father has completed?

16. What is the highest level of education your mother has completed?

17. In the boxes below, rate your language ability in each of the languages that you know. Use the following ratings: 0) Poor, 1) Good, 2) Very good 3) Native/native like. How many years (if any) have you studied this language in a formal school setting?

Language	Listening	Speaking	Reading	Writing	# of yrs of study
English					
Spanish					
Other					
Other					

APPENDIX C
THE LANGUAGE CONTACT PROFILE

(MODIFIED FROM FREED, SEGALOWITZ, AND HALTER, 2004)

Mark the strength of your agreement with each statement (7 being “strongly agree” and 1 being “strongly disagree”).

1. I speak to my friends in Spanish almost always.
1 2 3 4 5 6 7
2. I speak Spanish outside class almost always.
1 2 3 4 5 6 7
3. My significant other is a native Spanish speaker.
1 2 3 4 5 6 7
4. I have dated a native Spanish speaker.
1 2 3 4 5 6 7
5. I almost always speak Spanish at home.
1 2 3 4 5 6 7
6. I speak Spanish to family members outside my home.
1 2 3 4 5 6 7
7. I almost always speak Spanish at work.
1 2 3 4 5 6 7
8. I speak more Spanish to women.
1 2 3 4 5 6 7
9. I speak more Spanish to men.
1 2 3 4 5 6 7
10. I almost always watch Spanish television.
1 2 3 4 5 6 7
11. I almost always listen to Spanish radio.
1 2 3 4 5 6 7
12. I almost always listen to Spanish music.
1 2 3 4 5 6 7
13. I write Spanish letters/essays/papers.
1 2 3 4 5 6 7
14. I almost always read Spanish books.
1 2 3 4 5 6 7

- 15.** I almost always read Spanish magazines.
1 2 3 4 5 6 7
- 16.** I often vacation in a Spanish speaking country
1 2 3 4 5 6 7
- 17.** Most businesses in my community are bilingual.
1 2 3 4 5 6 7
- 18.** Most people in my community do not speak English.
1 2 3 4 5 6 7

APPENDIX D INSTRUCTIONS FOR NARRATIVES

This story is about a girl name Rosie. Rosie was adopted as a child. You will be provided a series of pictures to create her story. Please provide a detailed description of each character in the story (physical, emotional, mental and so forth). Also describe location, place, family, friends and so forth.

Guidelines to follow:

1. Must speak 20 minutes
2. Must only speak Spanish
3. Must include specific details of each picture provided
4. Must create a narrative

LIST OF REFERENCES

- Aguilar-Sanchez, J. (2009). *Syntactic Variation: The Case of Copula Choice in Limón, Costa Rica*. Unpublished doctoral dissertation, Indiana University.
- Andrews, G. (1985). *Ser y estar + locativos en Español*. *Hispania*, 68(1), 90-91.
- Bayley, R. (1996). Competing constraints on variation in speech of adult Chinese learners of English. In R. Bayley & D.R Preston (Eds.), *Second Language Acquisition and Linguistic Variation* (pp. 97-120). Amsterdam: Benjamins.
- Bayley, R. (2002). The quantitative. In J.K. Chambers, P. Trudgill and N. Schilling-Estes (Eds.), *The Handbook of Language Variation and Change* (pp. 117-141). Malden, MA: Blackwell Publishers.
- Bayley, R. (2007). *Second language acquisition: a variationist approach*. In Bayley & C. Lucas (Eds.), *Sociolinguistic Variation: Theories, Methods, and Applications* (pp. 133-144). Cambridge: Cambridge University Press.
- Bayley, R., & Regan, B. (2004). Introduction: The acquisition of sociolinguistic competence. *Journal of Sociolinguistics* 8(3), 323-338.
- Bolinger, D. (1975). *Syntactic blends and other matters*. *Language*, 37, 366-381.
- Bonvillain, N. (1993). *Language, culture, and communication*. Englewood Cliffs, N.J.: Prentice Hall.
- Boretti de Macchia, S. and Ferrer de Gregoret, M. C. (1984). *El español hablado en Rosario: Diminutivos*. Cuadernos de Literatura 3. Chaco, Argentina: Universidad Nacional de Nordeste: Instituto de Letras.
- Brecht, R. D., & Robinson, J. L. (1995). On the value of formal instruction in study abroad. In B.F. Freed (Ed.), *Second Language Acquisition in Study Abroad Context* (pp. 318-334). Amsterdam: Benjamins.
- Briscoe, G. (1995). *The Acquisition of Ser and Estar by Non-Native Speakers of Spanish*. Unpublished doctoral dissertation, University of Pennsylvania.
- Bruhn de Garavito, J. & Valenzuela, E. (2006). *The status of ser and estar in late and early bilingual L2 Spanish*. *Selected Proceedings of the 7th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages*, 100-109.
- Carlson, N. (2004). *My Family is Forever*. New York: Viking Press.
- Cortés-Torres, M. (2004). *Ser o estar? La variación lingüística y social de estar más adjetivo en el español de Cuernavaca, México*. *Hispania*, 87(4), 788-795.

- Costello, R. B. (Ed). *The American Heritage College Dictionary* (3rd ed.). New York: Houghton Mifflin Company.
- Crespo, L. (1946). *Los verbos ser y estar explicados por un nativo*. *Hispania*, 29(1), 45-55.
- Dewaele, J. M. (2004). *Retention or omission of the 'ne' in advanced French IL: The variable effect of extralinguistic factors*. *Journal of Sociolinguistics*, 8, 433-450.
- Dewaele, J.M, & Regan, V. (2002). *Maîtriser la norme sociolinguistique en interlangue française: le cas de l'omission variable de 'ne'*. *Journal of French Language Studies*, 12, 123-148
- Díaz-Campos, M., & Geeslin, K. L. (forthcoming). *Copula use in the Spanish of Venezuela: Is the pattern indicative of stable variation or an ongoing change?*
- Donato, R. (2000). *Sociocultural contributions to understanding the foreign and second language classroom*. In J.P. Lantolf (Ed.), *Sociocultural Theory and Second Language Learning* (pp. 27-50). Oxford: Oxford University Press.
- Dorian, N. (1982). *Defining the speech community to include its working margins*. In S. Romaine (Ed.), *Sociolinguistic Variation in Speech Communities* (pp. 26-33). London: Edward Arnold.
- Ferguson, C. A. (1959). *Diglossia*. *Word*, 15, 325-340.
- Ferguson, C. A. (1966). *National sociolinguistic profile formulas*. In W. Bright (Ed.), *Sociolinguistics* (pp. 309-315). The Hague: Mouton.
- Fishman, J. A. (1971). *The sociology of language: An interdisciplinary social science approach to language in society*. In J. A. Fishman (Ed.), *Advances in the Sociology of Language* (pp. 217-404). The Hague: Mouton.
- Freed, B., Dewey, D., Segalowitz, N., & Halter, R. (2004). *The Language Contact Profile*. (349-256). Cambridge University Press.
- Gass, S. & Selinker, L. (2001). *Second Language Acquisition: An Introductory Course*. New Jersey: Lawrence Erlbaum Associates.
- Geeslin, K. (2000). *A new approach to the SLA of copula choice*. *Spanish Applied Linguistics at the Turn of the Millennium: Papers from the 1999 Conference on the L1 & L2 Acquisition of Spanish and Portuguese* (pp. 50-66). Somerville, MA: Cascadilla Press.
- Geeslin, K. (2001). *Linguistic simplification: Past, present and future links to second language acquisition*. *Selected Proceedings of the Second Language Research Forum*, 271-291.

- Geeslin, K. (2001). *Changing norms, moving targets and the SLA of copula choice*. *Spanish applied linguistics*, 5, 29-55.
- Geeslin, K. (2002). *The acquisition of Spanish copula choice and its relationship to language change*. *Studies in Second Language Acquisition*, 24(3), 419-450.
- Geeslin, K. (2003). *A comparison of copula choice: native Spanish speakers and advanced learners*. *Language Learning*, 53(4), 703-764.
- Geeslin, K., & Guijarro-Fuentes, P. (2005). *The acquisition of copula choice in instructed Spanish: The role of individual characteristics*. In D. Eddington (Ed.), *Selected Proceedings of the 6th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages* (pp. 66-77). Somerville, MA: Cascadia Proceedings Project.
- Geeslin, K., & Guijarro-Fuentes, P. (2006). *Second language acquisition of variable structures in Spanish by Portuguese speakers*. *Language Learning*, 56(1), 1-50.
- Guijarro-Fuentes, P., & Geeslin, K. (2006). *A longitudinal study of copula choice: Following development in variable structures*. *Selected Proceedings of the 9th Hispanic Linguistics Symposium*, 144-156.
- Guijarro-Fuentes, P., & Geeslin, K. (2006). *Copula choice in the Spanish of Galicia: The effects of bilingualism on language*. *Spanish in Context*, 3(1), 63-83.
- Gumperz, J. (1968). *The speech community*. In *International Encyclopedia of the Social Sciences*. New York: Macmillan 381-386.
- Guntermann, G. (1992). *An analysis of interlanguage development over time: Part II, ser and estar*. *Hispania*, 75(5), 1294-1303.
- Gutierrez, M. J. (1992). *The extension of estar: A linguistic change in progress in the Spanish of Morelia, Mexico*. *Hispanic Linguistics*, 5(1-2), 109-141.
- Gutierrez, M. J. (2003). *Simplification and innovation in US Spanish*. *Multilingua*, 22(2), 169-184.
- Guy, G. (1993). *The quantitative analysis of linguistic variation*. In D. R. Preston (Ed.), *American dialect research* (pp.223-249). Amsterdam: John Benjamins.
- Hatch, E., & Lazaraton, A. (1991). *The Research Manual: Design and Statistics for Applied Linguistics*. Boston, MA: Heinle & Heinle.
- Hill, J. H. (1985). *Is a sociolinguistics possible?* *Society for Comparative Study of Society and History*, (pp. 461-471).
- Hymes, D. H. (1972). *Models of the interaction of language and social life*. In J. J. Gumperz and

- D. H. Hymes (Eds.), *Directions in Sociolinguistics* (pp. 35-71). New York: Holt, Rinehart, and Winston.
- Jacobson, Adam (2010). *Hispanic Market Overview*. Miami Beach: The Jake Adams Editorial Services and Research Consultancy.
- King, D. L., & Suñer, M. (1999). *Gramática Española: Análisis y Práctica*. New York: McGraw-Hill.
- Labov, W. (1984). *Field methods of the project on linguistic change and variation*. In J. Baugh & J. Scherer (Eds.), *Language in Use: Readings in Sociolinguistic* (pp. 28-53). Englewood Cliffs, NJ: Prentice Hall.
- Labov, W. (1972). *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Larsen-Freeman, D., & Long, M. H. (1991). *An Introduction to Second Language Acquisition Research*. New York: Longman.
- Lenneberg, E. H. (1967). *Biological Foundations of Language*. New York: Wiley.
- Lipski, J. M. (2008). *Varieties of Spanish in the United States*. Washington, D.C: Georgetown University Press.
- Long, M. (2005). *Problems with supposed counter-evidence to the critical period hypothesis*. *International Review of Applied Linguistics in Language Teaching*, 43(4), 287-317.
- Lucas, C. (2007). *Variation and modality*. In Bayley & C. Lucas (Eds.), *Sociolinguistic Variation: Theories, Methods, and Applications* (pp. 145-164). Cambridge: Cambridge University Press.
- Luján, M. (1981). *The Spanish copulas as aspectual indicators*. *Lingua*, 54, 2(3), 165-209.
- Mackey, A., & Gass, M. S. (2005). *Second Language Research: Methodology and Design*. Mahwah, NJ: Lawrence Erlbaum.
- Maienborn, C. (2005). *A discourse-based account of Spanish ser/estar*. *Linguistics*, 43, 1(395), 155-180.
- Marshall, J. (2004). *Language Change and Sociolinguistics: Rethinking Social Networks*. New York: Palgrave MacMillan.
- Mason, K. (1990). *Ser versus estar: A mnemonic device that puts estar in its P.L.A.C.E.* *Hispania*, 73(2), 506-507.

- Milardo, R. M. (1988). *Families and social networks: an overview of theory and methodology*. In R. M. Milardo (Ed.), *Families and Social Networks* (pp. 13-47). Newbury Park, Ca: Sage.
- Milroy, L. (1982). *Social network and linguistic focusing*. In S. Romaine (Ed.), *Sociolinguistic Variation in Speech Communities* (pp. 141-166). London: Edward Arnold.
- Milroy, L. (1999). *Women as innovators and norm-creators: The sociolinguistics of dialect leveling in a northern English city*. In S. Wertheim, A. C. Bailey and M. Corston-Oliver (Eds.), *Engendering Communication: Proceedings of the Fifth Berkeley Women and Language Conference* (pp. 361-376). Berkeley, CA: Berkely Women and Language Group.
- Milroy, L. (2002). *Social networks*. In J.K. Chambers, P. Trudgill and N. Schilling-Estes (Eds.), *The Handbook of Language Variation and Change* (pp. 549-575). Malden, MA: Blackwell Publishers.
- Milroy, L. & Gordon, M. (2003). *Sociolinguistics: Method and Interpretation*. Malden, MA: Blackwell Publishing.
- Montrul, S. (2009). *Knowledge of tense-aspect and mood in Spanish heritage speakers*. *International Journal of Bilingualism*, 13(2), 239-269.
- Mougeon, R., & Dewaele, J. (2004). *Preface to variation in the interlanguage of advanced second language learners*. *IRAL*, 42(4), 295-301.
- Mougeon, R., & Rehner, K. (2001). *Variation in the spoken French of Ontario French immersion students: The case of juste vs seulement vs rien que*. *Modern Language Journal*, 85(3), 398-415.
- Moyer, A. (1999). *Ultimate attainment in L2 phonology: The critical factors of age, motivation and instruction*. *Studies in Second Language Acquisition*, 21(1), 81-108.
- Naro, A. (1981). *The social and structural dimensions of a syntactic change*. *Language* 57, 63-98.
- Navas Sánchez-Élez, M. V. (1997). *Factores lingüísticos y extralingüísticos que determinan la alternancia de las variantes de -s/ and un dialecto luso-español, el barranqueño*. *Revistade Filología Románica*, 14, 391-410.
- Patrick, P. T. (2002). *The speech community*. In J.K. Chambers, P. Trudgill and N. Schilling-Estes (Eds.), *The Handbook of Language Variation and Change* (pp. 574-597). Malden, MA: Blackwell Publishers.
- Paolillo, J. C. (2002). *Analyzing Linguistic Variation: Statistical Models and Methods*. Stanford: CSLI Publications.

- Poplack, S. (1993). *Variation theory and language contact*. In D. Preston (Ed.), *American Dialect Research* (pp. 251-286). Amsterdam: Benjamin.
- Poplack, S. & Tagliamonte, S. (2001). *African American English in the Diaspora*. Malden, MA: Blackwell Publishers.
- Prada de Pérez, A. (2009). *Subject Expression in Minorcan Spanish: Consequences of Contact with Catalan*. Unpublished doctoral dissertation, University of Pennsylvania.
- Preston, D. (1996). *Variationist linguistics and second language acquisition*. In R. Bayley & D. Preston (Eds.), *Second Language Acquisition and Linguistic Variation* (pp. 1-45). Amsterdam: Benjamins.
- Querido, A. M. (1976). *The semantics of copulative constructions in Portuguese*. In M. Lujan & F. Hensey (Eds.), *Current Studies in Romance Linguistics* (pp. 343-366). Washington, D.C: Georgetown University Press
- Ramirez-Gelpi, A. (1995). *The Acquisition of Ser and Estar Among Adult Native English Speakers Learning Spanish as a Second Language*. Unpublished doctoral dissertation, University of Southern California.
- Regan, V. (1996). *Variation in French interlanguage: A longitudinal study of sociolinguistic competence*. In R. Bayley & D.R Preston (Eds.), *Second Language Acquisition and Linguistic Variation* (pp. 177-202). Amsterdam: Benjamins.
- Rehner, K., Mougeon, R., & Nadasdi, T. (2003). *The learning of sociolinguistic variation by advanced FSL learners: The case of nous versus on in immersion French*. *Studies in Second Language Acquisition*, 25, 127-156.
- Roldán, M. (1974). *Toward a semantic characterization of "ser" and "estar"*. *Hispania*, 57(1), 68-75.
- Romaine, S. (1982). *What is a speech community?* In S. Romaine (Ed.), *Sociolinguistic Variation in Speech Communities* (pp. 13-24). London: Edward Arnold.
- Ryan, J. M., & Lafford, B. A. (1992). *Acquisition of lexical meaning in a natural environment: "ser" and "estar" and the Granada experience*. *Hispania*, 75(3), 714-722.
- Salaberry, R. (1999). *The development of past tense verbal morphology in classroom L2 Spanish*. *Applied Linguistics*, 20(2), 151-178.
- Salzar, M. (2007). *Está muy diferente a como era antes: Ser and Estar + Adjective in New Mexico Spanish*. In Potowski & R. Cameron (Eds.), *Spanish in contact: Policy, Social, and Linguistic Inquiries* (pp. 343-353). Amsterdam: Benjamins.

- Sax, K. (1999). *Acquisition of stylistic variation by American learners of French: The case of ne- deletion in negation*. Paper presented at *New Ways of Analyzing Variation 28*, University of Toronto and York University.
- Schilling-Estes, N. (2007). *Sociolinguistic fieldwork*. In Bayley & C. Lucas (Eds.), *Sociolinguistic Variation: Theories, Methods, and Applications* (pp. 165-189). Cambridge: Cambridge University Press.
- Schmitt, C., Holtheuer, C., Miller, K. (2004). *Acquisition of copulas ser and estar in Spanish: learning lexico-semantics, syntax and discourse*. *Proceedings Supplement of the 28th Boston University Conference on Language Development*. Online: <http://www.bu.edu/linguistics/APPLIED/BUCLD/proc.html>.
- Silva-Corvalán, C. (1986). *Bilingualism and language change: The extension of estar in Los Angeles Spanish*. *Language*, 62(3), 587-608.
- Silva-Corvalán, C. (1994). *Language Contact and Change: Spanish in Los Angeles*. Oxford: Clarendon Press.
- Stewart, W. A. (1962). *An outline of linguistic typology for studying multilingualism*, In F. Rice (Ed.), *Study of the Role of Second Languages in Asia, Africa and Latin America* (pp. 15-25), Washington, DC: Center for Applied Linguistics.
- Stewart, M. (1999). *The Spanish Language Today*. New York: Routledge.
- Tagliamonte, S. A. (2006). *Analyzing Sociolinguistic Variation*. Cambridge: Cambridge University Press.
- Tagliamonte, S. A. (2007). *Quantitative analysis*. In Bayley & C. Lucas (Eds.), *Sociolinguistic Variation: Theories, Methods, and Applications* (pp. 190-214). Cambridge: Cambridge University Press.
- Temple Adger, C. & Christian, D. (2007). *Sociolinguistic variation and education*. In Bayley & C. Lucas (Eds.), *Sociolinguistic Variation: Theories, Methods, and Applications* (pp. 237-253). Cambridge: Cambridge University Press.
- Valdés, G. (2000). *Spanish for native speakers: AATSP professional development series handbook for teachers K-16 (1)*. New York, NY: Harcourt College Publishers.
- Van Patten, B. (1985). *The acquisition of ser and estar by adult learners of Spanish: A preliminary investigation of transitional stages of competence*. *Hispania*, 68(2), 399-406.
- Van Patten, B. (1987). *The acquisition of ser and estar: Accounting for development patterns*. In B. Van Patten, T. Dvorak & J. Lee (Eds.), *Foreign Language Learning: A Research Perspective* (pp. 61-75). New York: Newbury House.

- Van Patten, B. (2010). *Some verbs are more perfect than others: Why learners have difficulty with ser and estar and what it means for instruction*. *Hispania* 93(1), 29-38.
- Weinreich, U. (1953). *Languages in Contact: Findings and Problems*. Publications of the Linguistic Circle of New York.
- Whitney, W. D. (1979). *The life and Growth of Language: An outline of Linguistic Science*. (Original work published 1875)
- Watt, D. J. L. & Milroy, L. (1999). *Patterns of variation and change in three Tyneside vowels: Is this dialect leveling?* In P. Foulkes and G. J. Docherty (Eds.), *Urban Voices* (pp. 25-46). London: Arnold.
- Milroy, L. (2002). *Social networks*. In J.K. Chambers, P. Trudgill and N. Schilling-Estes (Eds.), *The Handbook of Language Variation and Change* (pp. 549-575). Malden, MA: Blackwell Publishers.
- Valdivieso, H. & Magaña, J. (1991). *Variación fonética de /s/ en el habla espontánea*. R. L. A. *Revista de Lingüística Teórica y Aplicada*, 29, 97-113.
- Woolsey, D. (2006). *Controlling contexts and clarifying intent: instrument design and the study of the SLA of estar with adjectives*. *Selected Proceedings of the 7th Conference on the Acquisition of Spanish and Portuguese as First and Second Languages*, 86-99.
- Woolsey, D. (2006). *Development of learner use of 'estar+adjective' in contexts of comparison within and individual frame of reference*. *Selected Proceedings of the 9th Hispanic Linguistic symposium*, 180-191.
- Zayas-Bazán, E., Bacon, M. S., & García, M. D. (2006). *Conexiones: Comunicación y Cultura* (3rd ed.), New Jersey: Pearson Prentice Hall.

BIOGRAPHICAL SKETCH

Dorian Dorado was born in Mexico City, Mexico, and has lived in the United States most of her life. Consequently, she grew up speaking both English and Spanish. Because of her bilingual background, she has always been interested in the linguistic processes that are found in speakers particularly in L2 learners and heritage speakers; as a result she studied Hispanic Linguistics at the University of Florida, completing her Ph.D. in August 2011.