

CONTACT-INDUCED CHANGES IN WORD ORDER AND INTONATION IN THE
SPANISH OF NEW YORK CITY BILINGUALS

By

CAROLINA BARRERA-TOBÓN

A dissertation submitted to the Graduate Faculty in Hispanic and Luso-Brazilian Literatures and
Languages in partial fulfillment of the requirements for the degree of Doctor of Philosophy,
The City University of New York
2013

© 2013
CAROLINA BARRERA-TOBÓN
All Rights Reserved

This manuscript has been read and accepted for the Graduate Faculty in
Hispanic and Luso-Brazilian Literatures and Languages in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

Ricardo Otheguy

Date

Chair of Examining Committee

José del Valle

Date

Executive Officer

Laura Callahan

Marcel den Dikken

Eva Fernández

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

CONTACT-INDUCED CHANGES IN WORD ORDER AND INTONATION IN THE
SPANISH OF NEW YORK CITY BILINGUALS

By

CAROLINA BARRERA-TOBÓN

Advisor: Professor Ricardo Otheguy

This dissertation is a variationist sociolinguistic analysis of the variable word order and prosody of copular constructions (*Nicolás es **feliz*** versus ***Feliz** es Nicolás, Es Nicolás **feliz**, Es **feliz** Nicolás*, ‘Nicolas is **happy**’) in the Spanish of first- and second-generation Spanish-English bilinguals in New York City (henceforth NYC). The data used for the study come from a spoken corpus of Spanish in NYC based on 140 sociolinguistic interviews (details of the corpus will be presented in Chapter Three). This dissertation addresses the question of whether second-generation bilinguals have a less flexible word order in Spanish as a result of their increased use of, and contact with, English, where a more fixed order prevails.

We will show that the informants in the present study, like their peers in Los Angeles and other parts of the US, exhibit a more rigid word order compared to their first-generation peers. We have established that this increase in rigidity of word order among the second-generation can be attributed in large part to their increased use of and contact with English. The studies mentioned above have interpreted their results to mean that these speakers are losing or have lost the discourse pragmatic constraints that govern word order. However, the data here show that the first- and second-generation speakers in the present study share many of the same conditioning variables and constraints for word order, although these variables appear to account

for a smaller amount of variance among the second-generation. In this way, we have established that the second-generation is not losing the discourse pragmatic constraints that govern word order, but that they are differently sensitive to these constraints. In fact, we show that second-generation speakers are very capable of communicating the pragmatic functions that the first-generation speakers do using word order because they maintain the prosodic details of their first-generation counterparts. In other words, the second-generation communicates these functions in ways that are slightly different from the first-generation, relying more on prosodic resources than syntactic ones. Furthermore, the data indicate that their prosodic patterns are not modeled after the prosody of English. In general terms we show that the second-generation does not have a different grammar from their first-generation counterparts, as is claimed by other researchers. Instead we show that these speakers favor certain first-generation strategies over others.

Acknowledgements

Several academics have contributed to the development of this dissertation and to my growth as a scholar. I would first like to thank the examining committee: Laura, Eva, and Marcel. Your criticisms of the manuscript have enhanced the value of the work, and I have learned and benefited greatly from your comments. I would also like to thank professor Andrew Lynch, whose Introduction to Spanish Linguistics class first introduced me to the wonderful world of linguistics. To professors Kim Potowski and Richard Cameron, although our tenure together was brief, thank you for broadening my sociolinguistic horizons. To professor Elaine Klein, thank you for encouraging me to explore other avenues within the field of linguistics.

As a student, several groups of people provided me with encouragement and support, both emotional and financial. The first group is my academic family, which includes the professors and students in the Second Language Acquisition Lab, the Research Institute for the Language in Urban Society, the Columbia School Linguistic Society, Writing at Queens, and the Doctoral Student Council. Many of my colleagues, through shared academic struggles and accomplishments, have earned the special label of ‘friend’ especially Joss Ruggles, Lorena Hernández-Ramírez, Luana Ferreira, Rocío Raña-Risso, Rachel Varra, Christen Madsen, Lucia Pozzan, Shawn Rice, Berenice Darwich, Laura Villa, Carissa Halston, Xuan-Nga Cam, Efraín Cardenas, Regina Castro, Marissel Hernández, Noelia Domínguez-Ramos, Marta Albalá, Ingrid Heidrick, Mary Carolan, Ana Martín-Sagredo, Yahilín Rodríguez, Ainhoa Sáinz-Gorbea, Shannon Millikin, Greg Keating, Brad Hoot, Naomi Lapidus Shin, Vítor Meirinho, Jesse Merandy, Michelle Nacimiento, Janine Matts, Paola Evangelista, Teresa and Edmund O’Neill, Katie Hawkland, Jill Jegerski, Ji Young Shim, Dorothy Staub, Rafael Torruella, Ayana Smythe, Marcos Wasem, and many many more whom I am sure I forgot to name. The final group is my

CLACLS familia. Tere, Laura, Debora, Vikki, Lawrence, and Marcela you were my family away from my family, and I love and miss you all dearly. Laird, your unwavering support and confidence in me were unmatched. You gave me the space, both physical and mental, to develop as a scholar and the advice to mature as a person.

I would also like to thank my non-academic family. First, I would like to thank my parents Martha and Guillermo and my grandparents Marta and Fabio for their guidance and unlimited love and support. Mom, you always wanted me to become a ‘doctor’ one way or another. I would also like to thank my brother Fuad for his support and affection. Thank you to my aunts, uncles, and cousins who stopped asking me when I would finish school several years back. Thank you to my in-laws, especially Michelle for helping me plan a wedding so that I could focus on writing. This project would never have been completed without the unconditional support and love of my husband Jordan. Thank you for believing in me when I didn’t believe in myself and for following me to Chicago. I’ll follow you to the ends of the earth. This project would have been completed much sooner if it weren’t for the most important person in my life, my unparalleled blessing, my son Nicolás. You are my biggest and greatest achievement and sharing this accomplishment with you makes it so much more meaningful. Eraina and Martha Lucía, thank you for keeping him while I attended to other ‘less important’ things.

Finally and most notably, I wish to express my deepest gratitude to Professor Ricardo Otheguy. In the acknowledgements of your dissertation you described your director as, “an inspiring instructor, insightful reader, extraordinarily generous with her time and attention, a relentless critic, and a loyal supporter. In one word: she was a teacher.” You have come to embody all of those qualities for your students, especially me. You are an inspiration and a pleasure to work with. Thank you for everything and for making this possible.

Table of Contents

1. Chapter 1 – Preliminaries	1
1. Introduction	1
2. Research questions	2
3. Hypotheses and preview of the results	3
4. Overview of the study	4
5. Relevance of the study	4
2. Chapter 2 - Background and motivation	6
1. Introduction	6
2. Spanish in New York City	6
3. Language contact phenomena	8
4. Variables that affect word order	9
4.1 Pragmatic function	10
4.1.1 Highlighted adjective	10
4.1.2 Subject topic contrast	11
4.1.3 Contrary to expectation	11
4.2 Constituent length	11
4.3 Informational status	12
5. Prosody	14
6. Motivation	16
3. Chapter 3 – Methodology	18
1. Introduction to the study	18
2. Corpus	19

2.1	Interview	19
2.2	Questionnaire	19
2.3	Participants	20
2.3.1	Criteria for selection of informants	21
2.3.2	Composition of the sample	23
2.3.2.1	Age	24
2.3.2.2	Age of arrival in NYC and years lived in NYC	24
2.3.2.3	Social class and education	24
2.3.2.4	English skills	25
2.3.2.5	Spanish use and Spanish skills	26
3.	Envelope of variation	28
4.	Variables and data treatment	33
4.1	Outcome variables	34
4.1.1	Word order	34
4.1.2	Primary stress placement	35
4.2	External (socio-demographic and language use) predictor variables	36
4.3	Internal (linguistic) predictor variables	40
4.3.1	Affirmative or negative utterance	42
4.3.2	Presence of adverb in the predicate	43
4.3.3	Type of copular verb	43
4.3.4	Presence of English elements	44
4.3.5	Syntactic properties of the subject NP	44
4.3.6	Properties of the verb	46

4.3.7	Discourse pragmatics	47
4.4	Summary of variables	51
4.5	Summary of data treatment and tokens	53
5	Statistical methodology	54
5.1	Token-level analysis	54
5.2	Participant-level analysis	55
5.3	Regressions	56
4.	Chapter 4 – External variables	57
1.	Introduction	57
2.	External (socio-demographic and language use) variables predicting word order	57
2.1	Distribution of the dependent variable WOR in the corpus as a whole and by region and generation.	60
2.2	Age of arrival in NYC and years lived in NYC	61
2.3	Regional and national origin	63
2.4	Gender	65
2.5	Age	66
2.6	Social class and education	66
2.7	English skills	67
2.8	Spanish skills	68
2.9	Spanish use	69
3.	Multiple linear regressions	71
3.1	Multiple linear regressions for the entire sample	74
3.2	Multiple linear regressions for the regional sub-samples.	76

3.3 Multiple linear regressions for the generational sub-samples.	78
4. Summary and discussion	80
4.1 Overall results	80
4.2 Results and discussion by region	85
4.3 Results and discussion by generation	86
5. Chapter 5 –Internal variables	91
1. Introduction	91
2. Internal (linguistic) variables predicting word order	91
2.1 Distribution of the dependent variable word order in the corpus as a whole and by region and generation.	93
2.2 Affirmative or negative utterance	94
2.3 Presence of adverb in the predicate	94
2.4 Type of copular verb	95
2.5 Presence of English elements	95
2.6 Syntactic properties of the subject NP	95
2.7 Length and type of adjectival phrase	99
2.8 Number in third person verbs	100
2.9 Tense-Mood-Aspect of the verb	101
2.10 Discourse pragmatics: Informational status of the subject NP	102
2.11 Discourse pragmatics: Function of the utterance	104
2.11.1 Word order expressing the pragmatically neutral function of conveying information	108
2.11.2 Word order expressing the pragmatic function of highlighted adjective	109

2.11.3	Word order expressing the pragmatic function of contrastive subject	110
2.11.4	Word order expressing the pragmatic functions of contrary to expectation and contrastive verb	111
3.	Logistic regressions	112
4.	Summary and discussion	118
4.1	Overall results	118
4.2	Results and discussion by region	125
4.3	Results and discussion by generation	125
6.	Chapter 6 –The internal suprasegmental variable prosody	132
1.	Introduction	132
2.	Pragmatically neutral utterances that convey information only	136
3.	Utterances with the pragmatic function of highlighted adjective	138
4.	Utterances with the pragmatic function of contrastive subject	140
5.	Utterances with the pragmatic function of contrary to expectation	140
6.	Utterances with the pragmatic function of contrastive verb	140
7.	Summary and discussion	141
7.	Chapter 7 –Adverbial predicates	148
1.	Introduction	148
2.	External variables predicting word order for utterances with adverbial predicates	151
2.1	Distribution of the dependent variable WOR(Adv) in the corpus as a whole and by region and generation.	153
2.2	Gender	155
2.3	Language with friends	156

2.4 Language with siblings	157	
2.5 Spanish in reading	158	
3. Internal variables predicting word order for utterances with adverbial predicates		158
3.1 Affirmative or negative utterance	159	
3.2 Type of copula	159	
3.3 English elements	160	
3.4 Syntactic properties of the subject NP	160	
3.5 Length and type of adjectival phrase	160	
3.6 Number ending in third person verbs/TMA of the verb		161
3.7 Discourse pragmatics: informational status of the subject NP		161
4. Summary and discussion	161	
8. Chapter 8 – Conclusions	164	
1. Introduction	164	
2. Summary of conclusions	164	
3. Relevance, avenues for future research, and limitations		168
9. Appendix	171	
10. References	177	

List of Tables

Table 3.1: Participants by National Origin	22
Table 3.2 : Gender Distribution of Participants by National Origin and Generation	23
Table 3.3: Cross-Tabulation of Social Class by Generation	25
Table 3.4: Cross-Tabulation of Educational Attainment by Generation	25
Table 3.5: Cross-Tabulation of English Skills by Generation	26
Table 3.6: Cross-Tabulation of Spanish Skills by Generation	26
Table 3.7: Cross-Tabulation of Spanish Use by Generation	27
Table 4.1: ANOVA, Canonical WOR by Generation	61
Table 4.2: ANOVA, Canonical WOR by Region	61
Table 4.3: ANOVA, Canonical WOR by Age of Arrival, Highlanders	62
Table 4.4: ANOVA, Canonical WOR by Age of Arrival, Caribbeans	63
Table 4.5: ANOVA, Canonical WOR by Region, LARNC	63
Table 4.6: ANOVA, Canonical WOR by Region, NYR	64
Table 4.7: Differences in Canonical WOR by National Origin and Generation	65
Table 4.8: Differences in Canonical WOR by Region and Generation	65
Table 4.9: ANOVA, Canonical WOR by Class, LARNC	66
Table 4.10: ANOVA, Canonical WOR by Class, NYR	67
Table 4.11: ANOVA, Canonical WOR by English Skills	67
Table 4.12: ANOVA, Canonical WOR by English Skills, Highlanders	68
Table 4.13: ANOVA, Canonical WOR by English Skills, Caribbeans	68
Table 4.14: ANOVA, Canonical WOR by Spanish Skills, LARNC	69
Table 4.15: ANOVA, Canonical WOR by Spanish Skills, NYR	69

Table 4.16: ANOVA, Canonical WOR by Language with Siblings	70
Table 4.17: ANOVA, Canonical WOR by Language with Siblings, Highlanders	70
Table 4.18: ANOVA, Canonical WOR by Language with Siblings, Caribbeans	71
Table 4.19: Multiple Regression for Canonical WOR, Whole Sample	74
Table 4.20: Multiple Regression for Canonical WOR, Whole Sample	74
Table 4.21: Multiple Regression for Canonical WOR, Whole Sample	74
Table 4.22: Multiple Regression for Canonical WOR, Whole Sample	75
Table 4.23: Multiple Regression for Canonical WOR, Highlanders	76
Table 4.24: Multiple Regression for Canonical WOR, Caribbeans	76
Table 4.25: Multiple Regression for Canonical WOR, Highlanders	76
Table 4.26: Multiple Regression for Canonical WOR, Caribbeans	76
Table 4.27: Multiple Regression for Canonical WOR, Highlanders	77
Table 4.28: Multiple Regression for Canonical WOR, Caribbeans	77
Table 4.29: Multiple Regression for Canonical WOR, Highlanders	77
Table 4.30: Multiple Regression for Canonical WOR, Caribbeans	77
Table 4.31: Multiple Regression for Canonical WOR, LARNC	78
Table 4.32: Multiple Regression for Canonical WOR, NYR	78
Table 4.33: Multiple Regression for Canonical WOR, LARNC	78
Table 4.34: Multiple Regression for Canonical WOR, NYR	78
Table 4.35: Multiple Regression for Canonical WOR, LARNC	79
Table 4.36: Multiple Regression for Canonical WOR, NYR	79
Table 5.1: Word Order Occurrence	93
Table 5.2: Cross-Tabulation of Word Order by Generation	94

Table 5.3: Cross-Tabulation of Word Order by Region	94
Table 5.4: Cross-Tabulation of Word Order by Type of Subject NP	97
Table 5.5: ANOVA, Mean words by NP Type	98
Table 5.6: Cross-Tabulation of Word Order by Type of Subject NP, LARNC	99
Table 5.7: Cross-Tabulation of Word Order by Type of Subject NP, NYR	99
Table 5.8: Cross-Tabulation of Word Order by Number Ending of the Verb	101
Table 5.9: Cross-Tabulation of Word Order by TMA of the Verb	102
Table 5.10: Cross-Tabulation of Word Order by Informational Status of Subject NP Using a Two-level Informational Status Variable	103
Table 5.11: Cross-Tabulation of Word Order by Informational Status of Subject NP, LARNC, Using a Two-level Informational Status Variable	103
Table 5.12: Cross-Tabulation of Word Order by Informational Status of Subject NP, NYR, Using a Two-level Informational Status Variable	104
Table 5.13: Cross-Tabulation of Word Order by Informational Status of Subject NP, Highlanders Using a Two-level Informational Status Variable	104
Table 5.14: Cross-Tabulation of Word Order by Informational Status of Subject NP, Caribbeans Using a Two-level Informational Status Variable	104
Table 5.15: Cross-Tabulation of Word Order by Pragmatic Function	105
Table 5.16: Cross-Tabulation of Word Order by Pragmatic Function	106
Table 5.17: Cross-Tabulation of Word Order by Pragmatic Function, LARNC	107
Table 5.18: Cross-Tabulation of Word Order by Pragmatic Function, NYR	107
Table 5.19: Cross-Tabulation of Word Order by Pragmatic Function, Highlanders	107
Table 5.20: Cross-Tabulation of Word Order by Pragmatic Function, Caribbeans	108

Table 5.21: Cross-Tabulation of Word Order by Generation for Pragmatically Neutral Utterances	109
Table 5.22: Cross-Tabulation of Word Order by Region for Pragmatically Neutral Utterances	109
Table 5.23: Cross-Tabulation of Word Order by Generation for Utterances with a Highlighted Adjective	110
Table 5.24: Cross-Tabulation of Word Order by Region for Utterances with a Highlighted Adjective	110
Table 5.25: Cross-Tabulation of Word Order by Generation for Utterances with a Contrastive Subject	111
Table 5.26: Cross-Tabulation of Word Order by Region for Utterances with a Contrastive Subject	111
Table 5.27: Variable Hierarchy for Variables Predicting Marked Word Order	114
Table 5.28: Factor Hierarchy for Internal Variable Factors Predicting Marked Word Order	114
Table 5.29: Variable Hierarchies by Region for Internal Variables Predicting Marked Word Order	115
Table 5.30: Factor Hierarchies by Region for Internal Variable Factors Predicting Marked Word Order	116
Table 5.31: Variable Hierarchies by Generation for Internal Variables Predicting Marked Word Order	117
Table 6.1: Cross-Tabulation of Prosody and Word Order by Generation for Utterances With the Pragmatically Neutral Function of Conveying Information	137

Table 6.2: Cross-Tabulation of Prosody and Word Order by Region for Utterances With the Pragmatically Neutral Function of Conveying Information	138
Table 6.3: Cross-Tabulation of Prosody and Word Order by Generation for Utterances With the Pragmatic Function of Highlighted Adjective	139
Table 6.4: Cross-Tabulation of Prosody and Word Order by Region for Utterances With the Pragmatic Function of Highlighted Adjective	139
Table 7.1: ANOVA, Canonical WOR(Adv) by Generation	154
Table 7.2: ANOVA, Canonical WOR(Adv) by Region	155
Table 7.3: ANOVA, Canonical WOR(Adv) by Gender, Highlanders	155
Table 7.4: ANOVA, Canonical WOR(Adv) by Gender, Caribbeans	156
Table 7.5: ANOVA, Canonical WOR(Adv) by Language with Friends	156
Table 7.6: ANOVA, Canonical WOR(Adv) by Language with Friends, Highlanders	157
Table 7.7: ANOVA, Canonical WOR(Adv) by Language with Siblings, LARNC	158
Table 7.8: ANOVA, Canonical WOR(Adv) by Spanish in Reading, Caribbeans	158

CHAPTER 1

PRELIMINARIES

1. Introduction

The present dissertation is a variationist sociolinguistic analysis of the variable word order and prosody of copular constructions in the Spanish of first- and second-generation Spanish-English bilinguals in New York City (henceforth NYC). The copular constructions studied here for word order and prosody (marked with bold) are those involving a subject, a copula, and an adjectival complement (*Nicolás es **feliz*** versus ***Feliz** es Nicolás, Es Nicolás **feliz**, Es **feliz** Nicolás*, ‘Nicolas is **happy**’). We also study, in a separate chapter, the order and the prosody of constructions of this type where the predicate is an adverb rather than an adjective (*Nicolás está aquí* ‘Nicolas is here’). The data used for the study come from a spoken corpus of Spanish in NYC based on 140 sociolinguistic interviews (details of the corpus will be presented in Chapter Three).

This dissertation addresses the question of whether second-generation bilinguals have a less flexible word order in Spanish as a result of their increased use of, and contact with, English, where a more fixed order prevails. A more rigid word order would suggest that these speakers are missing a resource that is used in non-contact varieties of Spanish, and that as a result they have a decreased ability to communicate the wide breadth of pragmatic functions that is expressed by their first-generation counterparts using different word orders. As detailed in the research questions below, we aim to determine whether these speakers are simply losing these expressive devices, replacing them with other mechanisms, or choosing some devices over others. Finally, we examine whether the other resources with which these second-generation speakers may be communicating meaning in Spanish are modeled in some way after English.

2. Research Questions

Below is the list of the research questions for the present study.

Question A: What are the external variables that have predictive statistical effect on word order in the corpus? Do any of these external variables produce categorical effects (that is, effects where the presence of a predictor produces a 100 percent to zero result)?

Question B: What differences, if any, are there between the two geographic regions in the way that the external variables influence word order, either quantitatively or categorically?

Question C: What differences, if any, are there between the two immigrant generations in the way that the external variables influence word order, either quantitatively or categorically?

Question D: What are the internal variables that have predictive statistical effect on word order in the corpus? Do any of these internal variables produce categorical effects (that is, effects where the presence of a predictor produces a 100 percent to zero result)?

Question E: What differences, if any, are there between the two geographic regions in the way that the internal variables influence word order, either quantitatively or categorically?

Question F: What differences, if any, are there between the two immigrant generations in the way that the internal variables influence word order, either quantitatively or categorically?

Question G: Is there a significant relationship between variable word order and variable prosody in our corpus?

Question H: What differences, if any, are there between the two geographic regions in word order and prosody, either quantitatively or categorically?

Question I: What differences, if any, are there between the two immigrant generations in word order and prosody, either quantitatively or categorically?

Question J: If there are prosodic differences across the generations, are these modeled after the prosody of English?

3. Hypotheses and preview of the results

Several studies on Spanish in the United States (henceforth US) have demonstrated that second-generation bilinguals have a more rigid word order than their first-generation counterparts (Silva-Corvalán, 1994; Zapata, Sánchez, & Toribio, 2005), but none of these studies have used naturalistic data to examine prosody nor have they examined the Spanish spoken in NYC. We will show in the present study that second-generation speakers in NYC, like their peers in Los Angeles and other parts of the US, do exhibit a more rigid word order compared to their first-generation peers. And we will support the proposal that this increase in rigidity of word order among second-generation New Yorkers can be attributed in large part to their increased use of and contact with English.

The studies mentioned above have interpreted their results as indicating that second-generation speakers are losing or have lost the discourse-pragmatic constraints that govern word order. However, we show in the present study that first- and second-generation speakers in our sample share many of the same conditioning variables and constraints for word order. We will show here that what distinguishes the two apparent-time generations of our sample is not the conditioning variables but the fact that the variables appear to account for smaller amounts of variance among speakers of the second-generation. In this way, we have established that our second-generation is not losing the discourse-pragmatic constraints that govern word order, but that they are differently sensitive to these constraints. In fact, we show that second-generation speakers are very capable of communicating the pragmatic functions that the first-generation speakers do using word order because they maintain the prosodic details of their first-generation

counterparts. In other words, the second-generation communicates these functions in ways that are slightly different from the first-generation, relying more on prosodic resources than syntactic ones. Furthermore, our data indicate that these prosodic patterns that survive across the generations and help fulfill the functions carried by word order in the first generation are not modeled after the prosody of English. In general terms, we show that the second-generation does not have a very different grammar from that of their first-generation counterparts, as is claimed by other researchers. Instead these speakers favor certain first-generation strategies over others.

4. Overview of the study

In Chapter Two we present the background and motivation for the present study. The methodology used for the study is presented in Chapter Three. In Chapter Four we examine the external variables that affect word order in Spanish and address research questions A, B, and C. Chapter Five is dedicated to the internal variables that affect word order with a focus on research questions D, E, and F. Chapter Six focuses on the analysis of prosody, while Chapter Seven focuses on the analysis of adverbial predicates, which we found to be different from other types of predicates. Finally, in Chapter Eight we summarize the results and aim to provide some general conclusions and interpretations of the data that were presented in Chapters Four, Five, Six, and Seven.

5. Relevance of the study

This study contributes to the growing body of literature on Spanish in the US and in NYC. By carrying out studies like this we can better understand the sociolinguistic profile of Latinos in the US as well as linguistic contact phenomena in general. Most importantly, the study draws attention to the fact that diachronic developments that result from linguistic contact do not always result in loss, and it highlights the importance of exploring alternative explanations

before drawing conclusions about the incompleteness of the Spanish spoken in contact situations, as well as other situations of language contact and change.

CHAPTER 2

BACKGROUND AND MOTIVATION

1. Introduction

This chapter provides details on the Latino population of NYC and gives background information on variable word order and prosody in Spanish and English, thus setting the stage and providing the motivation for the present dissertation study. We begin with a broad overview of Latinos in NYC and the language contact situation, including the outcomes of language contact on variable word order in Spanish. We then turn to a discussion of Spanish word order patterns and the constraints that probabilistically condition word order followed by a discussion of prosody.

2. Spanish in New York City

According to the 2010 Census, Latinos make up almost a third of the population of NYC (29.1 percent). Since the 1950's and 60's the majority of Latinos in NYC were Puerto Ricans; they still make up a third of the City's Latino population (31 percent). Dominicans make up the next largest group (25 percent), followed by Mexicans (14 percent), Ecuadorians (nine percent), Colombians (four percent), and Cubans (two percent). In the past 20 years, the Puerto Rican population rate has decreased by almost one percent, while the rates for the other groups have increased, most notably that of the Mexican population, which has grown by almost 10 percent in the last 10 years and is projected to surpass the Puerto Rican population by 2021 (Bergad, 2011). Our data represent these six Latino national groups. These groups can also be grouped into two major regional groups, the Caribbean and the Latin American Mainland, or Highlanders, which represent one of the major ways that dialectologists typically divide the Spanish spoken in Latin America. The Colombians, Ecuadorians, and Mexicans make up the Highland group, while

the Cubans, Dominicans, and Puerto Ricans make up the Caribbean group. A little over half of the NYC Latino population corresponds to the Caribbean group (58 percent), while 42 percent corresponds to the Highlander group.

Although the Puerto Rican population has been in NYC for several decades—their migration peaked in the 1950's and 1960's—Latinos from the other national groups, such as Colombians, whose migration peaked in the 1980's and 90's, have also been in NYC for decades. Recently, three national groups have had a large increase in migration to the City; more Dominicans arrived in NYC between 2000 and 2010 than any other national group, and since 2000 the Ecuadorian and Mexican migration rates have increased more than any other national group (Bergad, 2011). Additionally, more and more these groups are living in closer and closer proximity, and the linguistic enclaves which were once dominated by one group or another are becoming more and more diverse.

Another major change in the Latino population is in the transformation of the nativity of the City's Latinos. Although a large majority of the Latino population growth is fueled by immigration, in 2010--the first time since World War II--the absolute number of foreign-born Latinos in the NYC has declined. This means that the Latino population growth in the City is being fueled more and more by the increase in domestic-born Latinos (Bergad, 2011). In fact, the domestic-born Latino population grew four percentage points since the last census.

The percentage of Latinos who reported speaking English exclusively, well, or very well did not change since the last two censuses; it has remained constant at around 76 percent. There was a small decrease of six percentage points in the number of Latinos who reported speaking predominantly Spanish at home (from 88 to 82 percent). The fact that a large majority of Latinos reported speaking English exclusively, well, or very well and also reported speaking Spanish at

home indicates that bilingualism is a prominent characteristic among the City's Latinos. This large degree of bilingualism gives rise to several contact linguistic phenomena. Due to this large degree of bilingualism and the richness in dialectal varieties, it presents an ideal setting for a study on language and dialectal contact (Silva-Corvalán, 1995; Silva-Corvalán, 2001).

3. Language contact phenomena

The robust patterns of immigration from Latin America to NYC and to the US in general have prompted a growing interest within the field of contact linguistics in the study of Spanish in the US. Several of these studies have focused on one of the features of interest for the present study, word order. One of the studies is a sociolinguistic study of 50 Mexican-American bilinguals in Los Angeles in which the author found evidence for what she calls *obligatory SVX order* among the second- and third-generation speakers in her corpus, in comparison to the pre-contact lects (popular forms of Spanish in Mexico), in which word order is dependent on many variables, including the discourse-pragmatic function of the utterance (Silva-Corvalán, 1994). The author argues that this is an example of indirect transfer from English, which has a relatively fixed Subject-Verb-Object (henceforth SVO) order. That is, changes in a feature that is present in the pre-contact lect. Specifically, the author found an increased rate of preverbal subject NP's and subject personal pronouns as the speaker's contact with English increased. The author maintains that the increase in SVX order is a "consequence of processes of loss of semantic-pragmatic constraints on preverbal subject placement" (p. 144) that can be attributed to contact with the bilingual's dominant language.

In an experimental study on the production and interpretation of sentences with intransitive verbs by 24 heritage speakers of Spanish, Zapata, Sanchez, and Toribio (2005) also found evidence of convergence toward English word order. The second- and third-generation

bilingual participants in their study preferred SV order in Spanish in contexts where the preference in the pre-contact lect is for VS order. In other words, the second- and third-generation speakers produced word orders that were infelicitous compared to their first-generation counterparts. The authors attribute the increased incidence of SV order to convergence with English and a reduction in the syntactic options that allow the speakers to communicate discourse-semantic information.

Both of these studies suggest that word order variability in the Spanish of second- and third generation bilinguals is susceptible to external influence, and is in fact less variable when compared to the pre-contact lects. They also indicate that second- and third-generation bilinguals are losing the constraints that govern word order in Spanish. In the section below we examine several of the variables that are known to affect word order in the pre-contact lects of Spanish.

4. Variables that affect word order

Traditionally, both Spanish and English have been classified as SVO languages, that is, the basic, or canonical, word order for pragmatically neutral finite declarative sentences with transitive verbs is Subject-Verb-Object. For copular verbs (the verbs *ser* ‘to be’, *estar* ‘to be’, and *parecer* ‘to seem’) the canonical word order is Subject-Copula-Complement (Hawkins, 1983; Tavaniers, 2005; Zagona, 2002). Ocampo (1995a) calls this word order *informational word order* because it is pragmatically neutral. This means that the most common word order is a preverbal subject and a postverbal object or adjectival complement, as in (1) and (2).

(1) Jordan pinta la casa.
S V O
Jordan paints the house.’

(2) Fuad es guapo.
S Cop Adj
‘Fuad is handsome.’

Departure from canonical word order is a function of numerous variables including discourse pragmatic variables, processing variables, and structural variables (Bentivoglio & Sedano, 2001; Bentivoglio, 2003; Givón, 1993; Nava, 2007; Ocampo, 1995a, 1995b, 1995c; Quirk, Greenbaum, Leech, & Svartik, 1972; Tavaniers, 2005; Zagona, 2002). In the sections that follow we will individually discuss several of the variables that are known to affect word order and are relevant to our study.

4.1. Pragmatic function.

Departures from canonical word order are widely recognized as involving pragmatic functions other than the plain conveying of information (Bentivoglio & Sedano, 2001; Bentivoglio, 2003; Givón, 1993; Nava, 2007; Ocampo, 1995a, 1995b, 1995c; Quirk et al. 1972; Tavaniers, 2005; Zagona, 2002). Although this is true for both Spanish and English, the examples and explanations in the sub-sections below are limited to Spanish word order in copular constructions because these are the focus of our study.

In his study of copular verbs in Rio Platense Spanish, Ocampo (2002) identified four of the most common pragmatic functions in addition to the pragmatically neutral function of conveying information. Fernández Leborans (1999) considers the following departures from canonical word order as marked word orders because of their relatively low frequency of occurrence.

4.1.1. Highlighted adjective.

In highlighted adjective constructions the adjective is the focus of the utterance and typically receives primary stress. The subject is typically given or implied. The most common word order for this type of construction is Copula-Adjective-Subject as in (3) below¹. In this example, two speakers are discussing a common friend that one of them ran into recently after

¹ Examples three, four, and five are taken from Ocampo (2002).

not having seen him for a long time. The dialogue following this utterance goes onto describe the friend's physical characteristics.

- (3) Está igualito el tipo
Cop Adj S
Is same the guy.
'The guy is the same.'

4.1.2. Subject topic contrast.

When a constituent is the focus of contrast, in this case the subject, this means that it is placed in contrast to several possible alternatives. When this occurs in Spanish, the constituent that is the focus of contrast appears in first position, as in (4), and also receives contrastive stress (Ocampo, 1995b). In the example below, the speaker opposes the *ranas* 'frogs' to the *sapo* 'toad'. Although the word order for this example is canonical, the stress (which we will discuss in section 5) is contrastive and thus marked.

- (4) El sapo se queda, pero las ranas son tremendas.
S Cop Adj
The toad REFL stays, but the frogs are terrible.
'The toad stays, but the frogs are terrible.'

4.1.3. Contrary to expectation.

In Spanish, speakers may signal that something is not expected by use of inversion, that is, the appearance of the verb before the subject, as in (5) with primary stress typically falling on the adjective, which is the element of surprise and the most salient in this example. Here, the speaker is describing a homeless man that was so incredibly dirty that his skin was literally black. The speaker expresses surprise at the fact that his skin was black because he was so dirty.

- (5) Negro estaba el tipo.
Adj Cop S
Black was the guy.
'The guy was all black.'

4.2. Constituent length

Constituent length and complexity, or *weight*, as it is sometimes called, are also known to affect the word order of constituents in an utterance. The longer or more complex the constituent, the more likely it is to appear in post-verbal position. In example (8) below, the subject NP *las condiciones en que se encuentran* ‘the conditions in which they find themselves’ is in final position. This example illustrates what is known as *heavy noun phrase shift*.

- (8) Son tristes las condiciones en que se encuentran. [269C]²
Cop Adj S
Are sad the conditions in which themselves find.
‘The conditions in which they find themselves are sad.’

This concept, originally attributed to Behaghel (1909), was coined as *end-weight* (Quirk et al., 1972). This tendency is not a language-specific phenomenon. In fact, it has been argued that postponing longer and more complex constituents to utterance final position facilitates processing, and is thus not a language-specific phenomenon (Arnold, Losogco, Wasow & Ginstrom, 2000; Bentivoglio, 2003; Bolinger, 1955). In fact, even in languages with relatively fixed word order, like English, we find this tendency (Arnold et al., 2000).

4.3. Informational status

Another factor that affects the order of constituents in a sentence in Spanish, and also in English, is the given/new informational status of the subject noun phrase (henceforth NP) (Bentivoglio & Sedano, 2001; Bentivoglio, 2003; Ocampo, 1995a, 1995b, 1995c; Quirk et al. 1972; Tavaniers, 2005; Zagana, 2002). NP’s whose subjects are given or presupposed, in other words, those that are stated previously in the discourse or those shared or known by both interlocutors, usually appear in first position, while new information tends to occur in utterance final position (Arnold et al., 2000; Bentivoglio, 2003; Bolinger, 1955, Prince, 1981). In the

² Numbers and letters following examples are examples from the corpus; the number corresponds to the participant number and the letter corresponds to the origin of the participant (C = Colombian, U = Cuban, D = Dominican, E = Ecuadorian, M = Mexican, P = Puerto Rican).

example below, (9), taken from our corpus, the speaker is recommending a Mexican restaurant to the interviewer, and she mentions that the former Mexican President had been to that restaurant during his visit to NYC. Since the two participants were already speaking about the restaurant, this is given information. The new information, the fact that the then president of Mexico went there, appears in sentence final position.

(9) Vino el presidente Fox. [305M]
Came the president Fox.
V S
'President Fox came.'

As we stated previously, the tendency to put new information in final position is not a language specific trend (Arnold et al., 2000; Bentivoglio, 2003; Bolinger, 1955), and is also common in English. However, there are very few circumstances under which we find a postverbal subject in English³. For example, Heavy NP Shift is not possible in English. English systematically requires the structural subject position (SpecTP) to be filled. When there is not another constituent that fulfils the Extended Projection Principle, the subject must appear in that position. It is impossible to both first raise the subject to SpecTP and extrapose it via Heavy NP Shift because the trace that is subsequently left in SpecTP cannot be properly governed (Haegeman, 1994). For Spanish, the situation is quite different. Since Spanish is a pro-drop language, the structural subject position can remain empty, which allows a much broader range of postverbal subject placement possibilities (Zagona, 2002).

As can be seen in the previous examples, there are many similarities regarding the outcomes of these constraints on word order across these two languages, however, there are also important differences regarding how word order is affected by differences in pragmatic function,

³ In intransitive constructions, English deviates from SV order more liberally than in transitive constructions. For example, in intransitive constructions we find copular and locative inversion constructions and *there*-sentences. In transitive constructions we find deviations from SVO order in locative inversion constructions when a non-subject constituent is raised to the structural subject position, (*On this wall hung a picture of U.S. Grant*) as well as in quotative inversion constructions (*"Wow!" said John*), and in the case of *wh*-fronting of the object which produces O(Aux)SV order.

constituent length, and informational status. With few exceptions English has rigid SVO order, relying more on other linguistic resources, such as intonation and stress to convey discourse-pragmatic functions, while Spanish word order tends to be more flexible and sensitive to changes in discourse-pragmatic function and informational status (Nava, 2007; Otheguy, Rodríguez-Bachiller & Canals, 2004). This last point highlights another important difference between Spanish and English, their prosody. Although word order in both English and Spanish varies to account for the need to express a variety of pragmatic functions, the resulting variability in word order differs within each language. English word order is considerably more rigid than Spanish word order, and word order and prosody interact in different ways in both languages (Nava, 2007).

5. Prosody

Prosody refers to the intonation and rhythm of speech, and is based on the pitch, duration, and amplitude of the phonetic segment of an utterance (Cutler, Dahan, & van Donselaar, 1997). The prosodic feature of interest for the present proposal is *primary stress placement*, also called *pitch accent*, at the utterance level; in other words the most prosodically prominent element within an utterance. English and Spanish prosody are similar in a number of ways. Both Spanish and English are considered stress-accent languages, i.e., they signal accent with the phonetic cues of pitch, duration, and intensity (Beckman, 1986). Also, in both languages the pitch accent contour can signal pragmatic information such as informational status, contrast, and focus.

Despite these similarities, each language uses prosody in different ways. For example, Vallduví (1991) considers Spanish, like Italian and French, to be a non-plastic language in terms of its prosody. This means that the pitch accent is relatively fixed, and speakers have to use syntactic cues, such as word order, to indicate the discourse-pragmatic functions of the

constituents. Vallduví considers English, similar to German and Dutch, a plastic language. Unlike Spanish and other non-plastic languages, English pitch accent is much more flexible and falls on the contextually important constituents, since the speaker cannot as easily highlight important information using syntactic cues (Vallduví, 1991).

Another important prosodic difference between Spanish and English is the way in which focus is marked in an utterance, a distinction that separates boundary languages from edge languages. Boundary languages mark focus by inserting a prosodic phrase boundary next to the focused constituent, while edge languages mark focus using a change in word order. English is considered a boundary language, that is, in order to mark focus, speakers insert a prosodic phrase boundary next to the element that is the focus of the utterance (Büring, 2008). For example, in (10), which is said out of the blue, the focus is the word ‘coat’. However, in (11), which is said in response to the question ‘why are you staring at my coat?’, the focus is the word ‘fire’.

(10) Your **coat’s** on fire.

[your coat’s on fire] F
 (*) IP
 (*) pP

(11) Your coat’s on **fire**.

[your coat’s on fire] F
 (*) IP
 (* (*) pP ← insertion

Spanish, however, is considered an edge language. This means that focus is marked using a change in word order, placing the focus in a peripheral position at the end of the utterance. For example, in response to the question *¿Quién compró el periódico ayer?* ‘Who bought the newspaper yesterday?’, Spanish speakers tend to respond as in (12), by moving the subject *Martha* which is the focus of the utterance, to utterance final position.

- (12) Ayer el periódico lo compró **Martha**
Adv DO V S
Yesterday the newspaper bought **Martha**
'Yesterday **Martha** bought the newspaper.'

The alternative, (13), with canonical SVO order and pitch accent placed on the subject NP *Martha*, is considered infelicitous in Spanish.

- (13) **Martha** compró ayer el periódico.
S V Adv DO
Martha bought yesterday the newspaper.
'**Martha** bought the newspaper yesterday'

According to Büring (2008), we would predict that English speakers would respond to this same question using prosody cues, as in (14) or (15), and not by using syntactic resources to place the subject NP in sentence final position, as in (16).

- (14) Yesterday, **Martha** bought the newspaper.
(15) **Martha** bought the newspaper yesterday.
(16) *Yesterday, bought the newspaper **Martha**.

6. Motivation

The fact that discourse categories are realized using both syntactic and prosodic resources in both languages, albeit differently, is of special import because it presents a perfect opportunity for a study on language contact precisely because we are dealing with two languages that have a feature that is similar, but not quite the same. Previous studies on language contact suggest that bilingual speakers tend to search for parallels across two languages and often converge on those features that are shared or common to both languages, such as SVO order. (Silva-Corvalán, 1994). If, in fact, heritage speakers of Spanish in the US do show a preference toward fixed word order, one may ask whether this is attributable to the fact that these speakers are losing the discourse-pragmatic constraints that govern word order or are differently sensitive to these

constraints in that they operate to differing degrees or are operable with different grammatical outcomes. If this is the case, the question arises whether these outcomes are modeled around the prosodic patterns of English.

CHAPTER 3

METHODOLOGY

1. Introduction to the study

This chapter presents the methodology and the materials used to answer the research questions presented in Chapter One and repeated here below.

Question A: What are the external variables that have predictive statistical effect on word order in the corpus? Do any of these external variables produce categorical effects (that is, effects where the presence of a predictor produces a 100 percent to zero result)?

Question B: What differences, if any, are there between the two geographic regions in the way that the external variables influence word order, either quantitatively or categorically?

Question C: What differences, if any, are there between the two immigrant generations in the way that the external variables influence word order, either quantitatively or categorically?

Question D: What are the internal variables that have predictive statistical effect on word order in the corpus? Do any of these internal variables produce categorical effects (that is, effects where the presence of a predictor produces a 100 percent to zero result)?

Question E: What differences, if any, are there between the two geographic regions in the way that the internal variables influence word order, either quantitatively or categorically?

Question F: What differences, if any, are there between the two immigrant generations in the way that the internal variables influence word order, either quantitatively or categorically?

Question G: Is there a significant relationship between variable word order and variable prosody in our corpus?

Question H: What differences, if any, are there between the two geographic regions in word order and prosody, either quantitatively or categorically?

Question I: What differences, if any, are there between the two immigrant generations in word order and prosody, either quantitatively or categorically?

Question J: If there are prosodic differences across the generations, are these modeled after the prosody of English?

Section 2 describes the corpus used for the study, including a detailed description of the participants, the questionnaire, and the oral interviews. Section 3 focuses on the envelope of variation, while section 4 describes the variables and the data treatment. Preliminary results of these variables are also presented in this section. The final section, section 5, describes the statistical methodology used to determine the relationship between the predictor variables described below and word order, as well as how we investigate the association between word order and prosody.

2. Corpus

The data used for this study were taken from the Otheguy – Zentella corpus⁴ (Otheguy, Zentella, & Livert, 2007). The corpus is comprised of 140 taped and transcribed sociolinguist interviews conducted with a sample of NYC Latinos between 2000 and 2005. The corpus is stratified by several social variables including gender, national origin, regional origin, age of arrival, years in NYC, social class, years of education, as well as linguistic variables such as self-reported English skills and amount of Spanish use.

2.1. Interview.

Each interview lasted approximately one hour and was conducted in most cases by an interviewer whose Latin American origin matched that of the interviewee. For example, if a participant was a Colombian he or she was interviewed by a fellow Colombian. The topics were

⁴ I want to thank professors Ricardo Otheguy and Ana Celia Zentella for allowing me to use their corpus, created at the Graduate Center using funds from the National Science Foundation (grant 0004133), for my dissertation.

open and the interviewer and interviewee's shared origin often provided for natural conversations about the homeland, the immigration process, and the immigrant experience in NYC. In case the conversation flagged, interviewers had several prompts to elicit more speech from the interviewee. These questions were based in part on Labov's (1972) sociolinguistic interview technique and were intended to elicit natural speech. They included questions about the participant's first day in NYC, a time when the participant's life was in danger, or how the participant met his or her partner, among others. Some participants were more willing to speak than others. In some, the interviews were characterized by back-and-forth dialogue between the interviewee and the interviewer. But many others were more open and their interviews are characterized by long fluid narratives. The interviewers conducted the interviews entirely in Spanish, but there were many instances of loan words and code-switching into English as well as long narratives in English by some of the participants.

2.2. Questionnaire.

Each participant also completed a questionnaire. The questionnaire was given orally at the end of the interview in order to gather background information and data on language practices and language attitudes. Some participants were also asked to do a picture-naming task as part of the questionnaire. The questionnaire used can be found in Appendix A.

2.3. Participants.

Of the 140 participants in the corpus, 65 are of special interest to the present study. These 65 participants were chosen to represent two apparent-time generations⁵. Of these, 39 are considered Latin American Raised Newcomers (henceforth LARNC), that is, they arrived in

⁵ This is not a longitudinal study. The apparent-time hypothesis allows linguists to test two groups of speakers at the same time (a cross-sectional study), and make predictions about language change over time using those two populations. In other words, if we find linguistic variation occurring across different generations of the same population we can conclude that this is indicative of a change in progress.

NYC after the age of 16 and have spent less than six years in the City. Following the convention in linguistic and sociological research, this group is considered the first generation. The other 26 participants are classified as New York Raised (henceforth NYR); they arrived in NYC before the age of three or were born to first-generation Hispanic immigrants, and were raised in the City. This group is considered the second generation.

The participants belong to six different national groups: Colombians, Cubans, Dominicans, Ecuadorians, Mexicans, and Puerto Ricans. These six groups correspond to the six largest Latino national groups in the City. These six national groups can be arranged into two major regional groups, the Caribbean and the Latin American Mainland, or Highlanders, which represent one of the major ways that dialectologists typically divide the Spanish spoken in Latin America. The Colombian, Ecuadorian, and Mexican participants make up the Highland group, while the Cuban, Dominican, and Puerto Rican participants make up the Caribbean group.⁶

2.3.1. Criteria for selection of informants.

A total of 50 participants, 25 LARNC and 25 NYR, were selected from the Otheguy-Zentella corpus for inclusion in the study based on their sociodemographic traits. The traits, described below in detail, will be used as external predictor variables in the analysis. Given that the pool of NYR participants was smaller than the LARNC pool, all the NYR participants were initially included in the study, but one participant (086P) did not produce any tokens that fell inside the envelope of variation (see below) and was thus excluded.

The 25 LARNC participants were selected by matching the demographic characteristics of the 25 NYR participants as closely as possible, based on the initial available pool of 36. These variables included national origin, age, and gender. For example, although there were eight

⁶ Although several participants in the corpus were from the coastal areas of the three Highland countries, these participants were not included in the present study.

Ecuadorian LARNC in the corpus, only four were included in the study because there were only four NYR Ecuadorian participants in the corpus. Because of the limited number of NYR participants from each national group, not all the national groups were represented equally; but still, each country represented between 12 and 20 percent of the participants included in the study. However, the regional groups were almost equally represented: of the 50 participants, 26 belonged to the Highland group and 24 belonged to the Caribbean group (see Table 3.1).

Table 3.1		
<i>Participants by National Origin</i>		
National Origin	N	Pct
Colombia	10	20
Cuba	8	16
Dominican Republic	6	12
Ecuador	8	16
Mexico	8	16
Puerto Rico	10	20
Total	50	100

Although the number of LARNC included for each national group was determined by the number of NYR participants in the larger corpus, the individual LARNC participants chosen to represent that national group in the present study were determined by two of the participants' socio-demographic characteristics, namely gender and age. For each national group, each LARNC included in the study was matched as closely as possible with a NYR participant of that same nationality with regard to both gender and age. For example, in the larger corpus there were eight Ecuadorian LARNC and four Ecuadorian NYR participants. The NYR Ecuadorians included three females aged 18, 29, and 34 and one male aged 23. The LARNC Ecuadorians included four females aged 18, 25, 37, and 52 and four males aged 19, 24, 30, and 34. Since there was only one Ecuadorian male NYR, we only selected one Ecuadorian male LARNC for

inclusion in the present study. Since the Ecuadorian male NYR was 23 years old at the time of the interview, we selected the Ecuadorian male LARNC that was 24 years old at the time of the interview. The other three Ecuadorian LARNC males were not included in the present analysis. As for the females, we excluded the female aged 52 since there were no middle-aged females among the Ecuadorian NYR females. This resulted in the inclusion of three female NYR Ecuadorians and three female LARNC Ecuadorians in the corpus as well as one male NYR Ecuadorian and one male LARNC Ecuadorian. The average age of the female LARNC Ecuadorians was 26.6 and the average age of the female NYR Ecuadorians was 27. The Ecuadorian LARNC and NYR men differed in age by one year.

This method did not yield completely balanced groups, as sometimes matching was impossible. For example, there were no Puerto Rican LARNC females in the corpus, and as a result only Puerto Rican LARNC males were included in the study. Consequently, among the LARNC, 15 were male and 10 were female, and among the NYR, 13 were male and 12 were female. Table 3.2 compares the gender distribution for each nationality of LARNC and NYR participants.

National origin	N	LARNC		NYR	
		Males	Females	Males	Female
Colombian	10	2	3	3	2
Cuban	8	2	2	3	1
Dominican	6	2	1	1	2
Ecuadorian	8	1	3	1	3
Mexican	8	3	1	2	2
Puerto Rican	10	5	0	3	2
Total	50	15	10	13	12

2.3.2. Composition of the sample.

2.3.2.1. *Age.*

Given that the LARNC were selected by matching the age and gender as closely as possible to those of the NYR participants, it is not surprising that the average age across the two generational groups did not differ significantly. The average age of LARNC at the time of the interview was 28.4 years and for the NYR participants it was 29.6 years.

2.3.2.2. *Age of arrival in NYC and years lived in NYC.*

As we stated in section 2.3 above, the generational cohorts were created based on the number of years spent in NYC and age of arrival to the City. As a result, the years in NYC for the NYR, by definition, are the same, or nearly the same, as their age since they have no real age of arrival since they were born in NYC. Thus, the average age of arrival in NYC among NYR participants was 0.8 years. Nine of the NYR participants were born in the City, while more than half of the participants (13) arrived at the age of one. The average age of arrival to NYC among the LARNC was 25.6 years and the range was from 16 years to 37 years. Similarly, the LARNC had spent an average of 2.8 years in the City, while the NYR participants had spent an average of 28.7 years in the City.

2.3.2.3. *Social class and education.*

Given that social class and education are considered important predictors of language variation, the two groups were also balanced with regard to these two characteristics. Tables 3.3 and 3.4 illustrate that the participants are evenly divided with respect to self-identified social class and relatively balanced regarding educational attainment⁷.

⁷ The N value for some of the tables is not 50 because some participants did not provide this information.

Table 3.3			
<i>Cross-Tabulation of Social Class by Generation</i>			
Generation	N	Working Class	Middle Class
LARNC	24	11 (54%)	13 (46%)
NYR	24	11 (54%)	13 (46%)
Total	48	22 (54%)	26 (46%)

Table 3.4				
<i>Cross-Tabulation of Educational Attainment by Generation</i>				
Generation	N	Secondary or less	College	Graduate
LARNC	25	6 (24%)	13 (52%)	6 (24%)
NYR	25	7 (28%)	16 (64%)	2 (8%)
Total	50	13 (26%)	29 (58%)	8 (16%)

2.3.2.4. *English skills.*

Since our research questions relate to the influence of English on word order and prosody in Spanish, the mastery of English is of special interest to the study. In addition to the socio-demographic characteristics, the participants were also asked a series of questions to obtain information about their language use in English. This information is self reported by the participants in the questionnaire section of the interview and did not involve a task to measure proficiency. The interviewer asked the participants in Spanish how they would rate their English ability and gave them four options: poor, passable, good, and excellent. Given that by definition the two groups differ with regard to the time spent in NYC as well as their age of arrival to the City, it is no surprise that their English skills also differ. The majority of the LARNC reported their English skills to be ‘passable’ while the majority of the NYR reported their English skills to be ‘excellent’. Table 3.5 illustrates the different levels of self-reported English ability in this sample.

Table 3.5					
<i>Cross-Tabulation of English Skills by Generation</i>					
Generation	N	English Skills			
		Poor	Passable	Good	Excellent
LARNC	25	5 (20%)	12 (48%)	4 (16%)	4 (16%)
NYR	25	0 (0%)	1 (4%)	4 (16%)	20 (80%)
Total	50	5 (10%)	13 (26%)	8 (16%)	24 (48%)

2.3.2.5. Spanish use and Spanish skills.

Similar to their self-reported English skills, participants were also asked to report their Spanish skills using the same scale described above. As expected there is a significant difference across the groups with regard to this variable. Not surprisingly, the results for Spanish skills are almost the inverse of the results for English skills: while the majority of the LARNC describe their Spanish as ‘excellent’, the majority of NYR participants describe their Spanish as ‘good’. Table 3.6 illustrates the different levels of self-reported Spanish ability in this sample.

Table 3.6					
<i>Cross-Tabulation of Spanish Skills by Generation</i>					
Generation	N	Spanish Skills			
		Poor	Passable	Good	Excellent
LARNC	24	0 (0%)	1 (4%)	6 (25%)	17 (71%)
NYR	25	2 (8%)	8 (32%)	12 (48%)	3 (12%)
Total	50	2 (4%)	9 (18%)	18 (37%)	20 (41%)

In addition to Spanish skills, participants also reported on the use of Spanish in their everyday lives. First, the participants were asked to report their degree of Spanish use in general giving them four options: none, low, mid, and high. Similar to Spanish skills, the majority of LARNC reported a high degree of Spanish use in general, while the majority of NYR

participants reported low or no use of Spanish in general. The degree of Spanish use across the two generations is reported in Table 3.7.

Table 3.7					
<i>Cross-Tabulation of Spanish Use by Generation</i>					
Generation	N	None	Degree of Spanish Use		
			Low	Mid	High
LARNC	25	0 (0%)	6 (24%)	8 (32%)	11 (44%)
NYR	25	9 (36%)	10 (40%)	5 (20%)	1 (4%)
Total	50	9 (18%)	16 (32%)	13 (26%)	12 (24%)

The interviewers also asked the participants a series of questions about their use of Spanish and English with several people in their lives including their parents, siblings, children, significant other, friends, boss, classmates, and workmates. Participants had to report whether they spoke to these people in Spanish, in English, or in both languages. The majority of both LARNC and NYR participants reported using Spanish with their father and mother. The majority of LARNC reported using Spanish with their siblings, their children, and their significant other while the NYR participants reported using English with these people, but only the difference in language use with siblings and the language use with their significant other was statistically significant across the two generations. Additionally, the majority of both LARNC and NYR participants reported using English with their boss and both languages with their friends. Finally, the majority of LARNC reported using both Spanish and English with their colleagues at work or school, while the majority of NYR participants reported using English in these domains.

The interviewer also asked the participants to report how much Spanish they used or heard in specific contexts such as at home, at school, in social activities, in reading, in listening to the radio, in watching TV. Participants were given two options for this section: a lot, or a little.

Both groups reported using a lot of Spanish at home and in social activities, and using only a little Spanish at school, in listening to the radio, and in watching TV. The only domain in which the two groups differed significantly was in their use of Spanish in reading; the majority of LARNC reported using a lot of Spanish during reading, while the majority of NYR participants reported using only a little Spanish during reading.

Finally participants were asked how much Spanish they used with Spanish speakers from other countries, from their region (Highland or Caribbean), and from their own country. There were no significant differences across the groups. The majority of speakers reported mid and low exposure to Spanish speakers from other countries, mid and high exposure to Spanish speakers from their own region, and mid exposure to speakers of Spanish from their own country.

3. Envelope of variation

The transcriptions of each of the interviews as well as the taped interviews were used to identify utterances with an overt subject, a copular verb, and an attributive complement (*Nicolás es feliz* ‘Nicolas is happy’). The criteria described below were used to determine whether each utterance of this general type found in the corpus fell inside the envelope of variation and was thus to be included in the analysis. Utterances in both the oral interview and in the questionnaire part of the interview were included. Following Ocampo (1995a; 1995b; 1995c), only main finite declarative clauses were included, given that word order in questions and in subordinate clauses appears to be governed by different variables than the ones studied here. Similarly, pronominal subject nouns were not included in the study, given that pronominal subjects and objects also appear to be governed by different constraints (Ocampo, 1995a; Swan, 2005) in both English and Spanish. Utterances with an overt subject, a copular verb, and a nominal predicate (*Nicolás es el presidente* ‘Nicolas is the president’) also did not fall inside the envelope of variation. Initially

both adjectival (*Nicolás es feliz* ‘Nicolas is happy’) and adverbial complements (*Nicolás está aquí* ‘Nicolas is here’) were included, however these were separated and treated differently in the analysis because the adverbial complements behaved differently from the adjectival complements⁸. If the utterance included any other constituents it was not included in the analysis.

Although copular verbs in subordinate clauses following *que* ‘that’, as in (1), were not included in the analysis, there were clauses, as in (2) and (3), that followed *que* which were not considered subordinate clauses and were included because they show main-clause properties in their internal syntax. Similarly, clauses that followed *es que* were not considered subordinate clauses either and were included in the analysis.

- (1) Yo creo que han sido dominicanos y puertorriqueños, ¿no? [305M]
 I think that has been Dominicans and Puerto Ricans, no?
 ‘I think that it has been Dominicans and Puerto Ricans, no?’
- (2) Pues no fijate que fue rápido la ubicación [305M]
 Well no listen that was fast the placement.
 ‘Well no listen, the placement was fast.’
- (3) Y aparte que la comida es buena. [351M]
 And besides that, the food is good.
 ‘And besides that, the food is good.’

Although utterances with subject pronouns were not included, subjects could be simple lexical items (*Mamá*, ‘Mom’) or longer noun phrases (*el trabajo que tienes*, ‘the job you have’) as in (4). Similarly, clausal and infinitival subjects that appeared with a copular verb and an adjectival predicate, like in (5) and (6), were also included.

- (4) Es demasiado el trabajo que tienes. [305M]⁹
 Copula Adj Sub
 Is a lot the work that you have.
 ‘The work you have is a lot.’

⁸ This will be explained further in Chapter Seven.

⁹ Literal and free translations are given for single utterances. Also, when the order of constituents is the focus of the example, the constituent type is also provided. When the context of the utterance is provided, only free translations are given.

- (5) Es fácil comunicarse. [384E]
 Copula Adj Sub
 Is easy communicating.
 ‘Communicating is easy.’
- (6) Es importante que sepa hablar dos idiomas. [330D]
 Copula Adj Sub
 Is important that know to speak two languages.
 ‘Knowing how to speak to languages is important.’

There were several utterances that appeared to fall inside the envelope of variation because they contained a nominal NP, a copular verb, and an adjectival complement as in (7).

- (7) Es una ciudad increíble. [305M]
 Is a city increíble.
 ‘It is an incredible city.’ (referring to NYC)

However, clauses of this type do not enter inside the envelope of variation because although there is a copular verb, an adjective, and an NP, the NP *una ciudad* ‘a city’ is not the subject of the utterance. Instead, the subject is a non-overt or *covert* pronoun that was dropped and refers to a referent previously mentioned in the discourse (Zagona, 2002, p. 25). Similarly, utterances with non-referential subjects, which correspond to English ‘it’ or ‘there’ (Zagona, 2002, p. 31), as in (10), did not fall inside the envelope of variation.

Since Spanish is a pro-drop language, it is sometimes difficult to identify covert pronouns. This can be seen in examples (8) and (10) below, where the appearance of the definite article *las* ‘the’, instead of the indefinite *unas* ‘some’, as well as the context helps determine whether the subject of the verb *han sido* ‘have been’ is *las situaciones* ‘the situations’, a covert pronoun, or a non-referential subject. In example (8), which is underlined and presented in its context in (9), the definite article *las* and the context allow us to determine that the NP *las situaciones* is in fact the subject of the verb *han sido*. In example (10), which is also underlined and presented in its

context in (11), both the context and the lack of the definite article *las* suggest that the verb *han sido* has a non-referential subject.

- (8) Han sido muchas las situaciones [305M]

Copula Adj Sub

Have been many the situations.

‘The situations have been many.’

- (9) Bueno, han sido muchas las situaciones no, pero principalmente es este yo tengo un problema que más o menos aproximadamente tengo año con el este, empecé a tener dolores de estómago, fui a México hace un año y medio.

‘Well the situations have been many, no, but mostly it’s, uh, I have this problem that more or less I have had for a year, uh, I began having stomach pains, I went to Mexico a year and a half ago.’

- (10) Han sido muchas situaciones [305M]

Copula Adj Noun

Have been many situations.

‘There have been many situations.’

- (11) Bueno yo creo, no han sido una no, han sido muchas situaciones no, el hecho de tener que que venir a un país diferente, a un, una sociedad diferente, lenguaje diferente, o sea es un mundo totalmente diferente a a donde yo yo vivía.

‘Well I think, there hasn’t been one no, there have been many situations, no, the fact of having to come to a different country, to a, a different society, different language, in other words, it’s a completely different world to to where I used to live.’

These criteria, however, did not always work and there were cases, like in (12), for which it was especially difficult to determine whether the NP was the subject of the sentence or whether it was a covert subject.

- (12) Durante los años que crecí siempre estuve metido en el baile mexicano, ah ... y cuando regresaba al Bronx me ponía a bailar la música puertorriqueña, y la música, me enamoré de la música dominicana por ahí en 1956, porque vino la primera ola de gente dominicana y con eso vino música de un señor que se llamaba Valladares, Ángel de Gloria, y era el merengue tumbao, y muy diferente al merengue de ahora. [370M]

‘While I was growing up I was always into Mexican dance, uh...and when I would return to the Bronx I would dance to Puerto Rican music, and the music, I fell in love with Dominican music around 1965, because the first wave of Dominicans arrived and with them came the music of a man who’s name is Vallares, Ángel de Gloria, and it was a graceful merengue/the merengue was graceful, and very different from the merengue today.’

Cases in which the copular verb appeared with a noun phrase and a predicate demonstrative pronoun, as in (13), were also not included in the analysis.

- (13) Es ese el problema. [305M]
 Is that the problema
 ‘That is the problem.’

The reason that nominal or pronominal predicates did not fall inside the envelope of variation is precisely because of the difficulty in determining the grammatical subject of the verb when presented with two NP’s in the contact lects of Spanish. Unlike most cases in English, where the grammatical subject of the copular verb is determined mostly by word order (‘Obama is the president’ versus ‘The president is Obama’), in Spanish when a copular verb occurs with two NP’s, it is not only word order that allows us to determine which of the NP’s is the grammatical subject, but the intonation along with the word order and context. For example, in (14) the copular verb occurs with two NP’s: *mi mejor amiga* ‘my best friend’ and *la madrina* ‘the godmother’.

- (14) Es la madrina mi mejor amiga.
 Copula NP1 NP2
 Is the godmother my best friend.
 ‘My best friend is the godmother.’

In order to determine whether the first NP *la madrina* or the second NP *mi mejor amiga* is the grammatical subject of the utterance, we would need to rely on the context, the intonation, the information structure, and the pragmatic function of the utterance. One of our hypotheses (see Chapter One), however, is that the way in which speakers exploit word order and prosody to communicate the pragmatic function may differ across generations. Specifically, we predict that

that the NYR participants, as a result of their increased contact with English, have a grammatical system that exploits word order and prosody to signal pragmatic function in a different way than the LARNC would. If this is the case, then the method above would not be a reliable way to determine the grammatical subject of the utterance for these types of constructions among the NYR. For these reasons, constructions with a nominal or pronominal predicate were not included in the analysis.

Following the criteria outlined in this section, 424 utterances were identified as falling inside the envelope of variation. Another 58 utterances that appeared with adverbial predicates (*Nicolás está aquí* ‘Nicolas is here’) and met all of the other criteria were also coded and included in the study, although their analysis differed somewhat from the tokens with predicate adjectives (*Nicolás es feliz* ‘Nicolas is happy’). The tokens with adverbial predicates will be discussed further in Chapter Seven.

4. Variables and data treatment

In the section that follows we describe the variables that entered into the analysis. Specifically, we discuss how the external (socio-demographic and language use) and internal (linguistic) predictor variables affect word order as an outcome variable, and then address the association between word order and prosody. We will begin with a brief description of our original outcome variable, word order, followed by an explanation of how this variable is associated with our second outcome variable, prosody. We then proceed to detail each predictor variable that will enter into the analysis with our original outcome variable, word order. For each variable we explain how the tokens were coded, whether any tokens were excluded from coding, and why. At the end of the section there is a summary with all the variables as well as a summary of the data treatment.

4.1. Outcome variables.

Recall that word order variability is conditioned by a number of linguistic and social variables (Bentivoglio & Sedano, 2001; Bentivoglio, 2003; Givón, 1993; Nava, 2007; Ocampo, 1995a, 1995b, 1995c; Tavaniers, 2005; Zagona, 2002) and that this conditioning is subject to change in situations of language contact (Klee & Lynch, 2009; Montrul, 2004; Silva-Corvalán, 1994; Toribio, 2004; Zapata, Sánchez & Toribio, 2005). Following this line of reasoning we first try to determine whether there are any relationships between our predictor variables, both internal and external, and word order among our participants, and whether these relationships differ across generations. We also know that word order is related to prosody (Büring, 2008; Nava, 2007; Vallduví, 1991; Zagona, 2002). Thus, assuming we do find reduced word order variability among the NYR participants, we need to determine whether this reduced word order variability occurs with a concomitant change in variability in prosody or whether, instead, the NYR maintain the prosody of the previous generation. In this way for the follow up analysis, we will examine the relationship between our original outcome variable, word order, and prosody, our second outcome variable.

4.1.1. *Word order.*

All of the tokens that fell inside the envelope of variation were coded for the order of the constituents of the utterance, or word order. This is our first outcome variable. For constructions that fall inside the envelope of variation, those with a nominal subject, a copula, and an adjectival complement, there are theoretically six possible orders which appear below. These six possible word orders represent the six factors for this variable.

Dependent (outcome) variable: Word order

Subject-Copula-Adjective *Nicolás es feliz* ‘Nicolas is happy’

Copula-Adjective-Subject *Es feliz Nicolás* ‘lit. Is happy Nicolas,’ ‘Nicolas is happy’

Adjective-Copula-Subject *Feliz es Nicolás* ‘lit. Happy is Nicolas,’ ‘Nicolas is happy’

Copula-Subject-Adjective *Es Nicolás feliz* ‘lit. Is Nicolas happy,’ ‘Nicolas is happy’

Subject-Adjective-Copula *Nicolás feliz es* ‘lit. Nicolas happy is,’ ‘Nicolas is happy’

Adjective-Subject-Copula *Feliz Nicolás es* ‘lit. Happy Nicolas is,’ ‘Nicolas is happy’

Only four of the possible six word orders occurred in the corpus. The last two orders did not occur. For ease of exposition, we refer to the first word order as *canonical* and to the other five as *marked*. Seventy-five percent of the utterances in the corpus had canonical word order. The most frequent marked order was Copula-Adjective-Subject order (*Es feliz Nicolás* ‘lit: Is happy Nicolas,’ ‘Nicolas is happy’); twenty-three percent of the utterances in the corpus were of this type. One percent of the utterances were in Adjective-Copula-Subject order (*Feliz es Nicolás* ‘lit. Happy is Nicolas,’ ‘Nicolas is happy’) and only one utterance (0.2 percent) was in Copula-Subject-Adjective order (*Es Nicolás feliz* ‘lit. Is Nicolas happy,’ ‘Nicolas is happy’). Our data are consistent with a study done on Argentine Spanish by Ocampo (2002).

Although the first part of our analysis focuses on word order as the outcome variable and its relationships with the numerous predictor variables (described in detail below), the second part of the analysis examines the relationship between word order and our second outcome variable, primary stress placement or prosody.

4.1.2. Primary stress placement.

Our second outcome variable, prosody or primary stress placement, was only analyzed among some of the tokens in the corpus for the reasons described below. Only tokens with an

overt nominal subject (not a clausal-infinitival subject) and an adjectival complement (not a predicate adverb) were considered eligible for the prosodic analysis. These criteria were selected to yield utterances mostly consisting of one intonational phrase so that the primary stress placement of the utterance could be more easily determined. Of these 352 eligible tokens, however, 93 were excluded for various reasons. Twenty-nine of the tokens were excluded because the audio file for the participant was either incomplete (some part of it was missing) or missing altogether, thus making it impossible to listen to the prosody. An additional nine of the tokens were excluded because either the subject or the predicate contained some sort of list, thus producing list intonation. Finally, an additional 54 tokens were excluded because the subject NP was too long to do a reliable prosodic analysis of the utterance's primary stress placement.

The primary stress placement for the eligible tokens was determined in two ways. Using the audio file and transcript we created a textgrid to accompany each token using Praat¹⁰. Using the textgrid in Praat to view the pitch, intensity, and duration of each token as well as the audio file we determined the primary stress placement. In cases where the primary stress placement was unclear or difficult to hear the author asked another Spanish native speaker to analyze the token using the textgrid in Praat and the audio file.

4.2. External (socio-demographic and language use) predictor variables.

The different socio-demographic characteristics of the sample described above were carefully chosen so as to become external variables. That is, many of the criteria used for selection of the participants are also used as external predictor variables in the analysis. Below is a list of the socio-demographic and language use predictor variables. Each variable is underlined. If the variable is nominal, its factors are listed below it.

¹⁰ Praat is an acoustic analysis software program developed by Paul Boersma and David Weenick at the University of Amsterdam's Institute of Phonetic Sciences.

National origin

Colombian

Cuban

Dominican

Ecuadorian

Mexican

Puerto Rican

Regional origin

Highland

Caribbean

Age

Gender

Male

Female

Generation (derived from age of arrival in NYC and years lived in NYC)

LARNC

NYR

Social class

Low

Middle

Education

Elementary

Secondary

College

Graduate

English skills

Poor

Passable

Good

Excellent

Spanish skills

Poor

Passable

Good

Excellent

For the following variables the factors were None, Low, Mid, or High.

Spanish use

Spanish in domains

Spanish with speakers from own country

Spanish with other groups

Spanish with speakers from own group

For the following variables the factors were English, Spanish, or Both

Language with father

Language with mother

Language with siblings

Language with children

Language with friends

Language with boss

Language with significant other

For the following variables the factors were A lot or A little

TV in Spanish

Radio in Spanish

Reading in Spanish

Spanish use in social contexts

Spanish at home

Spanish at school

There are many reasons we elected to include these socio-demographic and language use characteristics as predictor variables. Several of the socio-demographic characteristics were selected specifically in order to address the research questions. These characteristics include national and regional origin, as well as age of arrival in NYC and years lived in the City, which were used to assign the participants to the two the generational and regional cohorts. Similarly, since our research questions relate to the influence of English on word order in Spanish, both self-reported English and Spanish skills and Spanish use are of special interest to the study and enter into the analysis as predictor variables. The rest of the socio-demographic characteristics listed above were also included as predictor variables as they have been shown to relate to

linguistic behavior in significant ways. Social class and education were included as variables as several studies have demonstrated a significant relationship between language variability in Spanish and speakers' level of income, occupation, education, neighborhood, and type of residence (Alba, 1988; Cedgren, 1978; Lafford, 1986; Samper-Padilla, 1990). Additionally, we have also included age as a socio-demographic predictor variable since several studies have found that certain linguistic variables are stratified by age (Alba, 1988; Lafford, 1986; Martínez & Moya, 2000; Samper-Padilla, 1990). Gender was also included among the socio-demographic predictor variables given the many documented quantitative and categorical differences in men and women's speech (Martínez & Moya, 2000; Navas Sánchez-Elez, 1997; Rissel, 1989; Valdivieso & Magaña, 1988).

4.3. Internal (linguistic) predictor variables.

Besides the demographic and language-use variables described above, each of the phrases that fell inside the envelope of variation was also coded for a number of linguistic variables in order to determine the relationship, if any, between internal linguistic variables and the outcome variables (word order and prosody). Below is a list of the internal linguistic predictor variables. Each variable is underlined. If the variable is nominal, its factors are listed below it.

Presence of adverb in the predicate¹¹

Adverb is present

Adverb is absent

Affirmative or negative utterance

Affirmative

Negative (contains the adverb *no*, 'no')

¹¹ For example, *La casa es muy grande* 'The house is very big'.

Type of copular verb

Ser, 'To be'

Estar, 'To be'

Parecer, 'To seem'

Presence of English elements

English elements present

English elements absent

Syntactic properties of subject NP

Simple subject NP

Complex subject NP

Clausal-Infinitival subject NP

Length of adjectival phrase in words

Type of adjectival phrase

Simple adjectival phrase

Complex adjectival phrase

Number in third-person verbs

Singular

Plural

TMA of the verb

Present indicative

Preterit indicative

Imperfect indicative

Present perfect indicative

Conditional

Discourse pragmatics: Informational status of the subject NP

Given

New

Implied

Discourse pragmatics: Function of the utterance

Conveying information (pragmatically neutral)

Highlighted adjective (pragmatically marked)

Contrastive subject (pragmatically marked)

Contrary to expectation (pragmatically marked)

Contrastive verb (pragmatically marked)

Below, the choice for the inclusion of each of these internal variables, as well as the data treatment, will be discussed.

4.3.1. Affirmative or negative utterance

The extant literature suggests that the use of adverbs in an utterance can affect word order (Ocampo, 1995b). As a result, each utterance was coded for the appearance of adverbs in the predicate. There were two types of adverbs that were coded for in the corpus. The first, is the adverb *no* ‘no’, so that each phrase was coded for whether it was affirmative or negative. Phrases like (15) were coded as affirmative while phrases like (16) were coded as negative. Some sentences were considered ambiguous because the utterance was affirmative but followed by a pause and then a negative interrogative such as (17).

(15) Nadie es hispano. [401P]
Nobody is Hispanic.
‘Nobody is Hispanic.’

- (16) No es nadie perfecto. [233U]
No is nobody perfect.
'Nobody is perfect.'
- (17) Es incierta mi situación, no? [305M].
Is uncertain my situation, no?
'My situation is uncertain, no?'

The overwhelming majority of the utterances in the envelope of variation were affirmative (93 percent). Only a small percentage (six percent) of the phrases were negative and there were four occurrences (one percent) like (17) above in which the utterance is affirmative but followed by a pause and then *no?*

4.3.2. Presence of adverb in the predicate.

Some of the tokens in the envelope of variation also occurred with an adverb that modified the predicate adjective as in (18). Again, since the extant literature suggests that adverbs can affect word order, whether an adverb modified the predicate adjective was also a predictor variable.

- (18) Mis padres son muy buenos. [201U]
My parents are very good.
'My parents are very good.'

Seventy-five percent of the utterances in the corpus did not occur with an adverb modifying the predicate adjective.

4.3.3. Type of copular verb.

The utterances in the corpus appeared with three different copulas: *ser* 'to be', *estar* 'to be', and *parecer* 'to seem'. Although it is not clear whether the use of one copula versus another can affect word order, which of these verbs appeared in the utterance was also one of the linguistic variables. Seventy-five percent of the copular verbs in the corpus were *ser*, as in (19),

and twenty-four percent of the copular verbs were *estar*, as in (20). There were two instances of *parecer* in the corpus that fell inside the envelope of variation, as in (21).

(19) Algunas palabras son diferentes. [181C]

Some words are different.

‘Some words are different.’

(20) El cheque estaba firmado. [198P]

The check was signed.

‘The check was signed.’

(21) El otro parece Colombiano. [021C]

The other seems Colombian.

‘The other seems Colombian.’

4.3.4. Presence of English elements.

Since there is a known relationship between code-switching and the syntax and prosody of an utterance (Durán Urrea, 2009), whether there were English elements in the utterance was also one of the linguistic variables. Although 96 percent of the utterances in the corpus did not contain English elements such as code-switches or borrowings, there were some 11 utterances that fell inside the envelope of variation and contained English elements such as in (22).

(22) El campus es impresionante. [325E]

The campus is impressive

‘The campus is impressive.’

4.3.5. Syntactic properties of the subject NP.

As mentioned in section 3, utterances in the envelope of variation could have subjects that are simple lexical items, longer noun phrases, clausal phrases, or infinitival phrases. The complexity and length of constituents is known to affect the order of constituents (Arnold et al., 2000). Thus, the type of subject, as well as the number of words that make up the subject NP, were also coded as variables. A subject was considered simple in case it consisted of a one word common noun, as in (23), or a one word common noun modified by a determiner or a possessive,

as in (24), or a combination of a one word noun or a one word adjective with or without determiners, as in (25), or a multiple word proper noun. A subject is complex when the noun is modified by anything other than a single adjective or a determiner. A noun modified by a prepositional phrase is complex, as in (26). A noun modified by a relative clause is complex, as in (27). The clausal and infinitival subjects mentioned above, in (5) and (6), were included but coded as clausal-infinitival subjects.

- (23) Es increíble Manhattan. [305M]
Is incredible Manhattan.
'Manhattan is incredible.'
- (24) La gente es agradable. [308M]
The people are nice.
'The people are nice.'
- (25) Esos grupos viejos son buenos. [201U]
Those group old are good.
'Those old groups are good.'
- (26) La hermana de su novio está desaparecida. [325E]
The sister of her boyfriend is disappeared.
'Her boyfriend's sister is disappeared.'
- (27) El español que habla el colombiano es muy propio. [263C]
The Spanish that speaks the Colombian is very proper.
'The Spanish that a Colombian speaks is very proper.'

The majority of subjects were simple (73 percent). There were similar amounts of infinitival and complex subjects (12 percent and 11 percent of all subjects respectively) and a small amount of clausal subjects (four percent). The minimum number of words in the subject NP was one, while the maximum was 43. The average number of words in the NP was 3.2 ($SD = 3.5$)

Similarly, each token was coded for the type and length of predicate, whether simple or complex. A predicate was considered simple if it contained a single adjective or an adjective modified by an adverb, as in (28). A predicate was considered complex if the adjective was

modified by anything other than a single adverb or if there was more than one adjective. An adjective modified by a prepositional phrase is considered complex, as in (29). An adjective modified by a relative clause is considered complex, as in (30). The majority of all utterances (95 percent) had simple predicates.

(28) El español que habla el colombiano es muy propio. [263C]
The Spanish that speaks the Colombian is very proper.
'The Spanish that a Colombian speaks is very proper.'

(29) El español es igual para todo el mundo. [92P]
The Spanish is the same for everyone.
'Spanish is the same for everyone.'

(30) El college es muy distinto a lo que se llama el high school. [206U]
The college is very different to what is called the high school.
'College is very different from what is called high school.'

The average number of words in the attributive complement was 1.4 (SD = 0.9, upper limit = 9, lower limit = 1).

4.3.6. Properties of the verb.

Some of the properties of the verb were also coded as variables including the verb's person and number ending and its tense-mood-aspect (TMA). Seventy-seven percent of the verbs in the corpus were third person singular verbs *es/está* 'is'. One verb was not conjugated at all and four verbs (one percent) were first person plural *somos* '(we) are'. This last result was somewhat unexpected as verbs that occurred with pronouns were excluded from the study. These first person plural verbs, however, appeared with a nominal subject (31) and were thus included in the envelope of variation.

(31) Los latinos somos bien cariñosos. [384E]
The Latinos are (1pp) very loving.
'We Latinos are very loving.'

The rest of the verbs in the corpus (22 percent) were third person plural verbs. With regard to the TMA of the verbs, 75 percent of the verbs were in the present indicative, followed by the preterit indicative (19 percent), and imperfect indicative (four percent). Three verbs were present perfect indicative (one percent), two were conditional (one percent), and one verb, mentioned above, was not conjugated. There were no verbs in the subjunctive.

4.3.7. Discourse pragmatics.

All of the tokens were coded for the informational status, or the newness of the subject NP given that the newness of constituents can affect the word order of those constituents in an utterance (Arnold et al., 2000). We adopted a modified version of Prince's (1981) typology in order to assign informational status to the subject NP. Prince's typology has seven different categories of subject: situationally evoked, evoked, inferable, unused, brand-new anchored, and brand new. In the current study, the first two information statuses (situationally evoked and evoked) are considered 'given'. Anytime the referent of a Subject NP is explicitly mentioned in the discourse or is brought up situationally, it is considered 'given'. Prince's last three statuses (brand-new, brand-new anchored and unused) are considered 'new'. Anytime the referent of a Subject NP was mentioned for the first time in the discourse and was not inferable from other information in the discourse (e.g. 'my father' can be inferred from 'my parents') it was considered 'new'. The remaining status, inferable, refers to any subject NP whose referent could be inferred by the interviewer or researcher from information that was previously mentioned in the discourse. Following this method, the majority of the subject NP's in the corpus are regarded as given (58 percent), followed by new subjects (29 percent), and finally inferable subjects (13 percent).

In addition to coding for the information status of the subject, all of the tokens, except those with an adverb predicate, were coded for their pragmatic function in the discourse¹². Following Ocampo (2002) we coded the 424 tokens with a predicate adjective for four pragmatic functions: conveying information, highlighted adjective, subject topic contrast, and contrary to expectation. An utterance was coded as ‘conveying information’ if it was pragmatically neutral. Sixty-nine percent of the utterances in the corpus fell into this category. The second most common pragmatic function in the corpus was highlighted adjective (13 percent). In these types of constructions the adjective is the focus of the utterance and typically receives primary stress. The subject is typically given or implied. In (32) below, the interviewee is describing how her husband, then boyfriend, proposed to her in front of her entire family.

- (32) Entonces yo llamé a mi papá y lo mismo de siempre que estaba bravísimo, y yo digo: “Pa’ you know, todavía no hemos ido a cenar”, “¿Qué? todavía no han ido a cenar, qué son estas horas de ustedes estar saliendo”, so, todo estaba planeado perfecto, igualito como él actúa. Así mismo actuó, y entonces yo brava ahí, y entré a la casa y entonces cuando entré estaba todo el mundo ahí, como él la hizo. La casa estaba repleta. [331D]

‘So I called my father and same as always, he was very upset, and I said, “Pa you know, we haven’t gone to eat yet”, “You haven’t gone to eat yet? It’s too late for you two to be out”, so everything was planned perfectly, even how he was acting. He acted that way, and I was there all upset, and I walked in the house, and when I walked in I saw everyone there, like he had planned. The house was full.’

In this case the adjectival predicate receives primary stress and is the focus of the utterance.

Secondary stress falls on the subject NP *la casa* ‘the house’, which is topical and given.

¹² Tokens with predicate adverbs were not coded for pragmatic function for several reasons. First, adverbs behave differently than adjectives in these types of constructions (Ocampo, 1995b). As a result, the tokens with adverb predicates would have to be treated separately and there were simply not enough data points to perform a statistically significant analysis. Only 33 of the 50 participants produced a token of this type, and many of these participants only produced one or two of these tokens, resulting in a total of 58 tokens of this type in the corpus. Further analysis of this type of construction (Subject-Copula-Adverb) in addition to constructions with non-copular verbs could be fruitful but is not part of the present study.

The third most common pragmatic function in the corpus was subject topic contrast (five percent). A subject was considered contrastive when it was placed in opposition to one or several possible alternatives in the discourse. In example (33) below, the interviewee is responding to a question regarding why he did not travel to Cuba after his graduation. He explains that his grandmother did not have any family there. He then describes how two or three of his grandmother's siblings died of smallpox and how she came to the US with a married sister, the one sibling that was left. He then places his grandmother in opposition to her married sister. The contrastive constituent, *mi abuela*, appears in first position and receives primary stress.

- (33) Porque no, tú sabes, ya cuando mi abuela vino, no tenía familia.
Eh... lo... los hermanos de ella y eso, murieron de unas... ¿Cómo se dice?,
viruela? Se murieron dos o tres y después los que quedaban, ella vino con una
hermana casada, mi abuela estaba soltera, y como no quedaba más nadie allá,
pues más nunca ellas quisieron ir allá. [005U]

'Because no, you know, when my grandmother came she didn't have any family left. Uh, ...her... siblings and that, died of some...How do you say? Smallpox? Two or three of them died and then the ones that were left, she came with a married sister, my grandmother was single and since no one else was left there, well they never wanted to go there.'

There was only one token that was considered contrary to expectation given the context of the transcription, but unfortunately the audio tape for this participant was incomplete and the prosody of the utterance could not be analyzed.

We added a fifth pragmatic function, contrastive verb, in addition to the four that Ocampo (2002) included in his analysis of copular verbs. The need for this pragmatic function was not expected because in Spanish the copular verb is almost always unstressed. Its lack of stress is associated with its status as semantically weak and bleached (Fernández Leborans, 1999; López García, 1996). This differs significantly from English in which the copula can easily be stressed to signal contradiction or correction of a previous utterance. As a result, when the

pragmatic function of the utterance requires the copula to be stressed, Spanish speakers insert the word *sí* ‘yes’ to carry the stress of the copula. However, one token in the corpus had a stressed copula and in this case the insertion of *sí* would not have conveyed the same pragmatic function as placing the primary stress on the copula. In this example, the copula was considered contrastive because it was put in opposition with another copula. In the example below (34) the interviewer had asked the interviewee about her relationship with her boyfriend. The interviewee begins using the present indicative copula *está* in her response and then she corrects and uses the imperfect indicative *estaba*. She puts the imperfect indicative copula in opposition to the present indicative copula to highlight the fact that things were ok, but they no longer are because of a comment her boyfriend made about her mother. The contrastive element, the copula *estaba*, is placed first and receives primary stress.

(34) Está todo...estaba todo bien hasta que me dijo eso que me...como me dejó a mí media.. media descontrolada, media...no sé. [371E]

‘Everything is...everything was good until he said that to me which left me...like...left me somewhat out of control, somewhat...I don’t know.’

One important caveat about the difficulty in coding for pragmatic function must be discussed. There were 93 tokens that were eligible but not coded for prosody, but were coded for both pragmatic variables. Determining the pragmatic function of these 93 utterances was especially difficult for a number of reasons. First, for tokens with missing audio it was impossible to hear the prosody of the utterance and thus only the context of the transcript was available for us to interpret the pragmatic function of the utterance. Since primary stress placement is related to pragmatic function, this posed a serious difficulty. Fortunately, this was the case for only 29 of the 93 tokens, which make up a small part of the total number of tokens (< 8 percent). Secondly, for those tokens not eligible for the prosodic analysis because of the

length of the NP or adjectival phrase, determining the pragmatic function was difficult as different constituents in the utterance could serve different pragmatic functions. In the utterance in example (36), the speaker is describing the differences between university and high school. She stresses the words *college* in contrast to *high school*, but she also stresses and highlights the adjective *muy* ‘very’.

- (35) El college es muy distinto a lo que se llama el high school. [206U]
The college is very different to what is called the high school.
‘College is very different to what is called high school.’

In cases such as this, we chose the pragmatic function that best represented the entire utterance. Of course, for these cases, as with all the other cases, it is our judgment that determined the pragmatic function of the utterance, a methodological weakness of this type of analysis.

4.4. Summary of variables.

Below is summary of the variables included in the study separated by dependent and independent as well as internal (linguistic) as well as external (socio-demographic and language use) variables. There were two internal outcome variables in the study:

- Word order
- Prosody (primary stress placement)

There were 27 external predictor variables.

- National origin
- Regional origin
- Gender
- Age
- Generation
- Social class

- Education
- English skills
- Spanish skills
- Spanish use
- Spanish in domains
- Spanish with speakers from own country
- Spanish with other groups
- Spanish with speakers from own group
- Language with father
- Language with mother
- Language with siblings
- Language with children
- Language with friends
- Language with boss
- Language with significant other
- TV in Spanish
- Radio in Spanish
- Reading in Spanish
- Spanish used in social contexts
- Spanish at home
- Spanish at school

Finally, there were 12 internal predictor variables

- Presence of adverb in the predicate

- Negative or affirmative utterance
- Type of copular verb
- Presence of English elements
- Syntactic properties of subject NP
- Length of subject NP
- Length of adjectival phrase
- Syntactic properties of adjective phrase
- Number in third-person verbs
- TMA of the verb
- Discourse pragmatics: Informational status of the subject NP
- Discourse pragmatics: Function of the utterance

4.5. Summary of data treatment and tokens.

A total of 482 tokens were coded for all the external predictor variables, and all of the linguistic predictor variables (except for the pragmatic function of the utterance) as well as the word order outcome variable. Of these, 53 were utterances that contained an adverbial predicate. The other 424 tokens had an adjectival predicate and were also coded for the pragmatic function of the utterance. Of these 424 tokens, 259 were coded for prosody. This yielded the following types of tokens:

- 259 tokens (Nominal Sub-Copula-Predicate Adjective) coded for all predictor and outcome variables
- 93 tokens (Nominal Sub-Copula-Predicate Adjective) coded for all predictor and outcome variables except prosody

- 72 tokens (Clausal-Infinitival Subjects-Copula-Predicate Adjective) coded for all predictor and outcome variables except prosody
- 58 (Nominal Sub-Copula-Predicate Adverb) coded for word order, all external predictor variables, and all internal predictor variables except for pragmatic function of the utterance.

In the section below we discuss how these variables were analyzed.

5. Statistical methodology

Statistical analyses using the Statistical Package for the Social Sciences (SPSS) were performed at both the token-level and the participant-level. Additionally, a series of logistic regressions using first participant-level variables (variables specific to the participant) and then token-level variables (variables specific to the utterance) were also performed. For the token-level analysis, given that the outcome variables were both nominal variables, and that the majority of the predictor variables were also nominal, the majority of predictive statistical analyses involved Cross-Tabulations with a Pearson's Chi-Square statistic. For the participant-level analysis we used canonical word order rate (the number of times the participant used canonical word order divided by the number of tokens that participant produced) as the outcome variable. Thus the majority of our statistical analyses at the participant level involved correlations and ANOVAs. Below the statistical analyses performed at the participant and the token-level will be discussed in further detail.

5.1. Token-level analysis.

Cross-Tabulations were performed at the token-level to determine the joint distribution of several of the outcome and predictor variables listed above. A Pearson's Chi-Square statistic as an index of association was calculated to determine the dependency of an outcome and predictive

variable in each of the contingency tables. All of the linguistic predictor variables described in Section 3.2 were analyzed with the outcome variable word order to see the relationship between the variables, if any. Cross-Tabulations involving the second outcome variable, primary stress placement, were done within each pragmatic function category. In other words, for each pragmatic function listed above we looked at the primary stress placement along with the other predictor variables. Two of the five pragmatic functions (contrary to expectation and contrastive verb) consisted of only one token and thus were not analyzed using this method.

One possible criticism of this type of analysis is that it violates one of the assumptions of the Chi-Square test, namely the independence of the data at the token level. Since each participant produced more than one token, the data at the token-level could be interpreted as a repeated measures design. If this were the case, each participant could contribute to more than one cell of the contingency table and thus we could not perform the Chi-Square test. However, given that the data at the token-level was specific to the utterance, and not the speaker of that utterance, it can be argued that the data are in fact independent. Each token expresses different pragmatic functions and is coded for several linguistic predictor variables that are independent of the participant (e.g. whether the token contains *ser* versus *estar* is dependent on the content of the utterance and not on the participant). This means that the contingency tables at the token-level included only variables specific to each independent utterance. In this way, it can be argued that no one entity contributes to more than one cell and our treatment of the data is valid.

5.2. Participant-level analysis.

In order to avoid the problems explained above with the repeated measures design, a second analysis was performed at the participant level for the socio-demographic predictor variables and the language use variables. As we mentioned previously, the word order variable at

the participant level was a continuous variable, canonical word order rate, which we calculated using the number of tokens the participant produced in canonical word order divided by the total number of tokens that each participant produced. If a participant used only canonical order then he or she was coded as having a canonical word order rate of 1.0. If a participant produced two tokens in canonical word order out of a total of five tokens, then that participant was coded as having a canonical word order rate of .40. This new outcome variable (canonical word order rate) was input into a ANOVAs and correlations along with most of the socio-demographic predictor variables, including Spanish skills, Spanish use, English skills, age (by groups), social class, education, age of arrival, years in the US, as well as generation, which is a variable that is derived from the last two variables, and has two levels: LARNC and NYR.

5.3. Regressions.

Finally, multinomial logistic regressions were performed at the token level using only the variables that were significantly related to word order using the Chi-Square tests, and multiple linear regressions were performed at the participant level using only the variables that were significantly related to canonical WOR using the ANOVAs or correlations. These regressions, performed in the vast majority of sociolinguistic studies (Eddington, 2010), are used to predict the outcome of word order based on a number of token-level and participant-level predictor variables. The results for the external predictor variables and the word order rate variable are presented in Chapter Four, while the results for the internal predictor variables and word order are presented in Chapter Five. Chapter Six focuses on the results for the prosodic analyses, and Chapter Seven focuses exclusively on the results for utterances with an adverbial predicate, instead of an adjectival predicate.

CHAPTER 4

EXTERNAL VARIABLES

1. Introduction

This chapter presents the analysis of the external predictor variables described in Chapter Three. The analysis aims at answering the research questions outlined in Chapter One and repeated here below. Only the research questions addressed in the chapter are listed.

Question A: What are the external variables that have predictive statistical effect on word order in the corpus? Do any of these external variables produce categorical effects (that is, effects where the presence of a predictor produces a 100 percent to zero result)?

Question B: What differences, if any, are there between the two geographic regions in the way that the external variables influence word order, either quantitatively or categorically?

Question C: What differences, if any, are there between the two immigrant generations in the way that the external variables influence word order, either quantitatively or categorically?

In section 2 below we present the results for all the external predictor variables and their relationship with word order. The results for the second outcome variable, prosody, are presented in the Chapter Six. Throughout the analysis of external variables in the present chapter we separate the results out by generation and region. The results of the regression analyses are presented in section 3. A summary and discussion of the results appears in section 4.

2. External (socio-demographic and language use) variables predicting word order

Below we discuss the relationships between the outcome variable word order and the internal predictor variables. The word order outcome variable considered here covers utterances found in the corpus consisting of (a) a simple, complex, infinitival, or clausal subject, (b) a

copular verb and (c) an adjectival predicate. The results for utterances with adverbial predicates will be discussed separately in Chapter Seven.

Recall that for our token-level analysis, that is for the analysis involving internal predictor variables the outcome variable word order has four factors or levels, as outlined below.

Dependent (outcome) variable: Word order

- Subject-Copula-Adjective *Nicolás es feliz* ‘Nicolas is happy’
- Copula-Adjective-Subject *Es feliz Nicolás* ‘lit: Is happy Nicolas,’ ‘Nicolas is happy’
- Adjective-Copula-Subject *Feliz es Nicolás* ‘lit. Happy is Nicolas,’ ‘Nicolas is happy’
- Copula-Subject-Adjective *Es Nicolás feliz* ‘lit. Is Nicolas happy,’ ‘Nicolas is happy’

The first order, Subject-Copula-Adjective, is considered in the present study the canonical word order whereas the other orders are considered the marked orders. For the current analysis, however, we use a continuous word order variable, canonical word order rate (henceforth WOR), for each participant, based on the number of times each participant produced the canonical word order out of his or her total tokens.

The WOR dependent or outcome variable was studied with respect to the independent or predictor variables outlined below.

Independent (predictor) variables initially investigated

- Generation (derived from age of arrival in NYC and years lived in NYC)
- National origin
- Regional origin
- Gender
- Age
- Social class

- Education
- English skills
- Spanish use
- Language with father
- Language with mother
- Language with siblings
- Language with children
- Language with friends
- Language with boss
- Language with significant other
- Language with speakers from own country
- Language with other groups
- Language with speakers from own group
- TV in Spanish
- Radio in Spanish
- Reading in Spanish
- Spanish in use social contexts
- Spanish at home
- Spanish in domains
- Spanish skills

The initial exploration of these predictors showed that many of them bear an association with WOR that is statistically negligible or too inconsistent or unreliable to be revealing of any pattern worth reporting. As a result, this chapter, while devoting a subsection to providing some

details regarding each of the variables above, concentrates primarily on the discussion of the following predictors, which did yield significant results.

Independent (predictor) variables significantly associated with word order

- Generation
- Age of arrival in New York City
- Years lived in New York City
- National origin
- Regional origin
- Class
- English skills
- Spanish skills
- Language with siblings

2.1. Distribution of the dependent variable WOR in the corpus as a whole and by region and generation.

Given our predictions, laid out in Chapters One and Two, we expect to see a higher canonical WOR among NYR. This would suggest that they have less word order variability compared to their first generation counterparts, the LARNC. We would argue that this decreased variability in word order is due, in part, to the NYR participants' increased use of and contact with the English language.

Overall, the average canonical WOR was 75 percent ($SD = 23.9$). One participant never used canonical word order (WOR of 0) while 13 participants used canonical word order only (WOR of 1.0). This may, however, be a result of the few tokens produced by these participants. We did find significant differences in WOR across generations, while the WOR across regions

approached, but did not reach, significance. Tables 4.1 and show the breakdown of WOR by generation and region.

Table 4.1		
<i>ANOVA, Canonical WOR by Generation</i>		
Generation	N speakers	Canonical WOR
LARNC	25	64
NYR	25	87
	50	
$F = 14.169$	$p < .01$	

Table 4.2		
<i>ANOVA, Canonical WOR by Region</i>		
Region	N speakers	Canonical WOR
Highlands	26	71
Caribbean	24	81
	50	
$F = 2.308$	$p = .135$	

The tables show that the LARNC had a lower canonical WOR than the NYR by 23 percentage points. Similarly, the Highlanders had a lower canonical WOR than the Caribbeans by ten percentage points. In the sections below we will present the results for each of the external predictor variables for the entire corpus as well as by generation and region. The order of presentation of the predictor variables follows the order in which they were discussed in the previous chapter. We will postpone discussion and interpretation of our results until section 4.

2.2. Age of arrival in NYC and years lived in NYC.

For the present analysis we had two versions of these variables: a continuous variable and a nominal variable. In other words, for each participant we had a figure for his or her age of arrival and number of years spent in NYC. This is our continuous variable. We also used these

figures to establish two generational groups, which we used for our nominal variable. As a result, the ANOVA for canonical WOR by Age of Arrival or Time Spent in NYC would be identical to the ANOVA for canonical WOR by generation. As a result, we only ran correlations with the continuous variables to confirm that the generational separation as we have it was not arbitrary.

Given that age of arrival in NYC and years lived in the City were variables that were used to calculate our generation variable (which we reported above), it is no surprise that we found a significant correlation between the continuous variable age of arrival and canonical WOR, $r(48) = -.50, p < .01$, as well as between the continuous variable time in NYC and canonical WOR, $r(48) = .44, p < .01$.

We did not find the same results across both regional groups. Among the Highlanders we found a significant correlation between the age of arrival and canonical WOR, $r(24) = -.75, p < .01$, as well as between time in NYC and canonical WOR, $r(22) = .63, p < .01$. Similarly, we found significant differences in age of arrival and years in NYC for the Highlanders, but not among the Caribbeans. The results for the ANOVAs for each regional group are found in Tables 4.3 and 4.4. As we can see, among Highlanders, those participants who arrived after the age of 16 (the LARNC) had a lower canonical WOR than those who arrived before age 3 (the NYR) by 36 percentage points. Among the Caribbeans, this difference was only eight percentage points

Table 4.3		
<i>ANOVA, Canonical WOR by Age of Arrival, Highlanders</i>		
Age of Arrival	N speakers	Canonical WOR
After age 16	13	53
Before age 3	13	89
	26	
$F = 27.71$	$p < .01$	

Table 4.4		
<i>ANOVA, Canonical WOR by Age of Arrival, Caribbeans</i>		
Age of Arrival	N speakers	Canonical WOR
After age 16	12	77
Before age 3	12	85
	24	
$F = .82$	$p = .374$	

The ANOVAs for canonical WOR by Time Spent in NYR for each regional group are identical to Tables 4.3 and 4.4.

2.3. Regional and national origin.

Although we established (in section 2.1) that the differences in WOR across the two regional groups approached significance, we also wanted to determine whether there were any regional differences across the generations or whether there were any differences in WOR across the nationalities.

Similar to our results in section 2.2 we found that there was a significant difference in WOR among Highlanders and Caribbeans for LARNC, but not for NYR. The difference across the regional groups for the LARNC was 24 percentage points while for the NYR it was four percentage points. These results appear below in Tables 4.5 and 4.6.

Table 4.5		
<i>ANOVA, Canonical WOR by Region, LARNC</i>		
Region	N speakers	Canonical WOR
Highlands	13	53
Caribbean	12	77
	25	
$F = 6.122$	$p < .05$	

Table 4.6		
<i>ANOVA, Canonical WOR by Region, NYR</i>		
Region	N speakers	Canonical WOR
Highlands	13	89
Caribbean	12	85
	25	
<i>F</i> = .461	<i>p</i> =.504	

At the national level we found that the average WOR for Colombians, Ecuadorians, and Mexicans (the Highlanders) was lower than the average WOR for Cubans, Dominicans, and Puerto Ricans, but the ANOVA revealed no significant differences across the nationalities mostly likely because there were not enough data. Dominicans had the highest WOR ($M=84.3$, $SD= 21.2$) while Ecuadorians had the lowest WOR ($M=68.0$, $SD=22.5$). We found an increased use of canonical word order across generations among all of the nationalities except for the Cuban participants, which decreased in canonical WOR by one percentage point. The difference from one generational cohort to the other is most marked, and significant, among the Mexican participants. The Mexican LARNC had the lowest canonical WOR among all the groups ($M=47.4$ $SD= 31$) whereas all four Mexican NYR participants used canonical word order only. This latter result, however, may also be attributable to low frequencies especially considering that two of the four NYR Mexican participants produced no more than three tokens each and thus an accident of the sample. We also found significant differences across the generations for Ecuadorians and Colombians (the Highlanders) but not among Dominicans, Puerto Ricans, and Cubans (the Caribbean), although they follow the general trend of an increased use of canonical word order only among the NYR participants. Table 4.7 below summarizes the ANOVAs for generational differences by National Origin.

Table 4.7				
<i>Differences in Canonical WOR by National Origin and Generation</i>				
National Origin	LARNC WOR	NYR WOR	Difference in WOR	p
Mexican	47	100	53	*
Ecuadorian	54	82	28	a
Colombian	56	85	29	**
Puerto Rican	72	87	15	
Dominican	80	89	9	
Cuban	81	79	-2	

As we can see in the table, the three national groups with the largest change in canonical WOR across generations are the three groups that correspond to the Highlander group (Colombians, Ecuadorians, and Mexicans). In Table 4.8 below we have collapsed the national groups into the regional groups and summarized the ANOVAs for generational differences by Region.

Table 4.8				
<i>Differences in Canonical WOR by Region and Generation</i>				
Region	LARNC WOR	NYR WOR	Difference in WOR	p
Highlanders	53	89	36	*
Caribbeans	77	85	8	

2.4. Gender.

Besides exposure, regional, and national origin, all of which we predicted would yield significant differences across the two generational cohorts, we wanted to see whether there were any other differences across the two groups with regard to the rest of the socio-demographic variables discussed in the previous chapter. The first socio-demographic variable we looked at in this way is gender. Recall that there are slightly fewer females than males in our corpus (22 versus 28). When looking at both exposure groups together, we found no significant difference in terms of canonical WOR. When we compare males and females across the generational and regional groups we find the same results. Although females had a higher canonical WOR overall, these results were not significant.

2.5. Age.

The next socio-demographic variable we looked at was age. Since age is not a nominal variable we ran correlations to determine whether there was a significant relationship to canonical WOR. We found no significant relationship for the entire corpus or for each generational and regional group.

2.6. Social Class and education.

Given that social class and education are considered important predictors of language variation, these were the next socio-demographic variables we examined. Recall that the two groups were balanced with regard to these two variables. When we look at all 50 participants combined we found no significant differences across the social class groups with regard to word order. However, when we broke the groups down by generation, we did find some class differences. Specifically, among LARNC, we found no significant differences between the two classes for canonical WOR, but among the NYR the WOR across the classes approached significance. These results can be seen in Tables 4.9 and 4.10.

Table 4.9		
<i>ANOVA, Canonical WOR by Class, LARNC</i>		
Class	N speakers	Canonical WOR
Middle	13	60
Working	11	68
	24	
$F = .417$	$p = .525$	

Table 4.10		
<i>ANOVA, Canonical WOR by Class, NYR</i>		
Class	N speakers	Canonical WOR
Middle	13	82
Working	11	93
	24	
$F = 3.951$	$p = .059$	

With regard to education we found no significant results overall or by generation or region.

2.7. English skills.

The only linguistic variables that were coded and analyzed at the participant-level were the language use and skill variables. Since our research questions relate to the influence of English on word order in Spanish, both self-reported English and Spanish skills and Spanish use are of special interest to the study. Recall that all language use and skill variables are self-reported

Given our prediction that increased use of and contact with the English language is related to an increased canonical WOR, it is not surprising that we found significant differences with regard to English skills. Those participants that reported their English to be poor or passable had a lower canonical WOR by 13 percentage points than those who reported their English to be good or excellent. These results can be seen in Table 4.11.

Table 4.11		
<i>ANOVA, Canonical WOR by English Skills</i>		
English skills	N speakers	Canonical WOR
Poor/Passable	18	67
Good/Excellent	32	80
	50	
$F = 4.111$	$p < .05$	

Among the Highlanders we found the same trend, but among the Caribbeans we did not. These data can be seen in Tables 4.12 and 4.13.

Table 4.12		
<i>ANOVA, Canonical WOR by English Skills, Highlanders</i>		
English skills	N speakers	Canonical WOR
Poor/Passable	9	53
Good/Excellent	17	80
	26	
$F = 8.675$	$p < .01$	

Table 4.13		
<i>ANOVA, Canonical WOR by English Skills, Caribbeans</i>		
English skills	N speakers	Canonical WOR
Poor/Passable	9	80
Good/Excellent	15	81
	24	
$F = .016$	$p = .900$	

Across generations we found no significant differences in WOR by English skills.

2.8. Spanish skills.

Given our hypotheses we would expect the opposite relationship for English skills and WOR than for Spanish skills and word order. However, we did not find any significant differences for word order by Spanish skills overall. Across regions we also did not find significant differences in Spanish skills. When we analyzed the data by generation, we did find significant differences in WOR among the NYR, but not among the LARNC. These results can be seen in Tables 4.14 and 4.15.

Table 4.14		
<i>ANOVA, Canonical WOR by Spanish Skills, LARNC</i>		
Spanish Skills	N speakers	Canonical WOR
Poor/Passable	1	88
Good/Excellent	23	64
	24	
<i>F</i> = .727	<i>p</i> =.403	

Table 4.15		
<i>ANOVA, Canonical WOR by Spanish Skills, NYR</i>		
Spanish Skills	N speakers	Canonical WOR
Poor/Passable	10	80
Good/Excellent	15	91
	25	
<i>F</i> = 5.126	<i>p</i> < .05	

Among the NYR, the ten participants who reported poor or passable Spanish skills had a higher canonical WOR than those who reported good or excellent Spanish skills by eleven percentage points. Although the difference in percentage points across Spanish skills for the LARNC is greater (24 percentage points), the fact that only one participant reported poor or passable Spanish skills is known to affect the significance of these results.

2.9. Spanish use.

Given our predictions, we would expect increased Spanish use to be associated a lower rate of canonical WOR among all participants. Although the data do follow that general trend, that is, the greater the use of Spanish, the lower the canonical WOR, out of the 18 Spanish use variables only one, language with siblings, approached significance. These results can be seen Table 4.16 below.

Table 4.146		
<i>ANOVA, Canonical WOR by Language with Siblings</i>		
Language	N speakers	Canonical WOR
English	12	83
Both	12	80
Spanish	20	65
	44	
$F = 2.508$	$p = .094$	

When we look at the data by generation, these differences disappear in part, we believe, because the generations reflect different trends in language use with siblings. In other words, most LARNC use Spanish with their siblings, while most NYR use English with their siblings. In fact, none of the LARNC reported using English with their siblings. As a result, the data by generation do not yield any significant differences.

When we examine the data by region, however, we do find significant differences among the Highlanders but not among the Caribbeans. Among the Highlanders, those who reported using English with their siblings had a higher canonical WOR than those who reported using Spanish with their siblings by 32 percentage points. Among the Caribbeans, this difference drops to three percentage points. These results can be seen in Tables 4.17 and 4.18.

Table 4.17		
<i>ANOVA, Canonical WOR by Language with Siblings, Highlanders</i>		
Language	N speakers	Canonical WOR
English	4	87
Both	8	78
Spanish	11	55
	23	
$F = 4.015$	$p < .05$	

Table 4.18		
<i>ANOVA, Canonical WOR by Language with Siblings, Caribbeans</i>		
Language	N speakers	Canonical WOR
English	8	81
Both	4	83
Spanish	9	78
	44	
$F = .071$	$p = .932$	

In the following chapter, the relationship between the internal predictor variables and word order at the token level will be reported. The second outcome variable, prosody, will also be discussed in Chapter Six. In the next section below we will look more explicitly at differences across the two generational and regional cohorts using a different statistical test, a logistic regression.

3. Multiple linear regressions

Our analyses in this chapter so far have relied exclusively on ANOVAs or correlations test for significance. However, these types of analyses only allow us to explore the relationship between word order and one predictor variable at a time because they consider the relationship between each predictor variable and the outcome variable separately. In order to examine the effect of several predictor variables upon one outcome variable we can perform a multiple regression, which considers the effect of all of the predictor variables simultaneously.

Additionally, a multiple regression will examine which of the predictor variables accounts for the most variance, thus indicating which predictor variable has the strongest effect on the outcome variable, canonical WOR. Using the results from the multiple regressions we can establish a hierarchy of predictor variables based on their effect on our outcome variable, and thus ensure a more accurate assessment of their relationship.

We performed 11 multiple linear regressions including four regressions for the sample as a whole as well as four regression for each generation and three regressions for the regions. Only variables that were known to have a significant relationship (or approaching significance) with canonical WOR were included in the regressions¹³. These variables are listed below:

- Generation
- Regional origin
- National origin
- Class
- English skills
- Spanish skills
- Language with siblings

In order for multiple regressions to be reliable, the variables included in the regression must be independent, that is, there cannot be a relationship between them either statistically or abstractly. If the predictor variables in a regression are not independent, the variables overlap, yielding a less effective and less reliable regression. This overlap is known as *multicollinearity* (Field, 2009, p. 257) and should be avoided. Thus, we would not include National Origin and Regional Origin in the same regression since these two variables are related conceptually. Similarly, we would also not include English Skills and Generation in the same regression, not because these variables overlap conceptually, but because English linguistic proficiency increases with a younger age of arrival and time in NYC. For similar reasons we did not include Generation and Spanish Skills or Generation and Language with Siblings in the same regression model.

¹³ Variables included in the logistic regression were selected based on the ANOVAs and the correlations in this chapter.

A linear regression calculates the amount of variance accounted for by the variables in the regression, expressed as an R^2 ('R square') value. Included in the output of the regression is an ANOVA which tests whether the model as a whole is statistically significant. The regression also calculates the relative weight or strength of association between each independent variable and the dependent variable using a Standardized Beta (β_1). The β_1 value indicates "the number of standard deviation changes that the outcome will change as a result of one standard deviation change in the predictor" (Field, 2009, p. 274). As a result, the higher the β_1 value, the greater influence that predictor has on the outcome variable, and thus we can use the β_1 values to establish a hierarchy of predictor variables. In addition to allowing us to calculate the weight of each predictor variable in the model, the β_1 , which can be either positive or negative, also allows us to understand the direction of the relationship between each of the predictor variables and the outcome variable. For example, a negative β_1 value for Language with Siblings, which was coded (1 = English, 2 = Both, and 3 = Spanish), would indicate that an increase in reported use of Spanish with siblings would yield a decrease in canonical WOR. In the same way, a positive β_1 value for English Skills (which was coded as 1 = Poor, 2 = Passable, 3 = Good, and 4 = Excellent) would indicate that an increase in reported English skills would correspond to an increase in canonical WOR. Finally, the regression also calculates a p value for each β_1 , which indicates the statistical significance of each individual variable when it is taken into account simultaneously with all the other variables in the model.

In the bottom left hand corner of each table below is the R^2 value, followed by one or two asterisks or an ^a ($p < .01$, $p < .05$ $p < .10$ respectively) depending on whether the model as a whole is statistically significant. For each variable in the tables there is also a β_1 value followed

by one or two asterisks or an ^a depending on whether the individual variable retained statistical significance in the model (at $p < .01$, $p < .05$ $p < .10$ respectively).

In the following section we analyze the regression results for the entire sample. In sections 3.2 and 3.3 we analyze each of the regional and generational sub-samples.

3.1. Multiple linear regressions for the entire sample.

The variables Region and Class appear in each of the first four regressions along with Generation, English Skills, Language with Siblings, and Spanish Skills alternatively. We will postpone discussion and interpretation of our results until section 4¹⁴.

Table 4.19		
<i>Multiple Regression for Canonical WOR, Whole Sample</i>		
Variable	β_1	p
Generation	.490	**
Region	.219	^a
Class	.185	
$R^2 = .321^{**}$		N = 48

Table 4.20		
<i>Multiple Regression for Canonical WOR, Whole Sample</i>		
Variable	β_1	p
English Skills	.420	**
Class	.339	*
Region	.227	^a
$R^2 = .234^{**}$		N = 48

Table 4.21		
<i>Multiple Regression for Canonical WOR, Whole Sample</i>		
Variable	β_1	p
Language with Siblings	-.332	*
Class	.202	
Region	.183	
$R^2 = .189^*$		N = 42

¹⁴ In the tables below, a double asterisk indicates $p < .01$, a single asterisk indicates $p < .05$, and a superscript a indicates $p < .10$.

Table 4.22		
<i>Multiple Regression for Canonical WOR, Whole Sample</i>		
Variable	β_1	<i>p</i>
Region	.241	^a
Class	.207	
Spanish Skills	-.162	
$R^2 = .136^a$		N = 47

The R^2 values in Tables 4.19-4.22 indicate several things. First, they indicate that we can have a great deal of confidence in most of our models, except the model presented in Table 4.22. Secondly it indicates that the three variables presented in Table 4.19 together account for almost a third of the variance in canonical WOR, more than the three variables presented in Tables 4.20-4.22.

The column labeled β_1 indicates both the strength of the relationship between each predictor variable and the outcome variable, permitting us to rank each of the variables in the model, as well as the direction of the relationship between the predictor variable and the outcome variable. In the first table we find Generation at the top of the list. In the next two tables we find English Skills and Language Spoken with Siblings at the top of the list. Recall that we predicted that these variables would be related to Generation and thus lead to multicollinearity if we included them in the same model. Spanish Skills is at the bottom of the final Table. The direction of the results aligns with those from the bivariate analyses: an increase in reported English skills corresponds to an increase in canonical WOR; a decrease in reported Spanish skills and Spanish spoken with siblings corresponds to an increase in canonical WOR. Similarly, increased canonical WOR is associated with the Caribbeans (coded as 1 = Highlander and 2 = Caribbean) the working class (coded as 1 = High, 2 = Middle, and 3 = Working), and the NYR (coded as 1 = LARNC and 2 = NYR).

The p value in the right hand column indicates the significance of each individual variable when it is taken into account simultaneously with the other variables in the model. Our results indicate that no one variable is statistically significant in all of the models, although two of the three variables ‘related’ to Generation are significant in three of the models. Region is significant in one model, and approaches significance in two of the models. Class is only significant in one of the models, and as a result we can only have limited confidence in its results.

3.2. Multiple linear regressions for the regional sub-samples.

In this section we analyze some of the same variables that were considered above for the whole sample regressions, but we look at each regional sub-sample separately. For each regional group we include National Origin and Class in each of the four regressions along with Generation, English Skills, Language with Siblings, and Spanish Skills alternatively. The tables on the left present the results for the Highlanders and those on the right present the results for the Caribbeans.

Table 4.23		
<i>Multiple Regression for Canonical WOR, Highlanders</i>		
Variable	β_1	p
Generation	.723	**
National Origin	.041	
Class	.030	
$R^2 = .538^{**}$		N = 26

Table 4.24		
<i>Multiple Regression for Canonical WOR, Caribbeans</i>		
Variable	β_1	p
Generation	.281	
Class	.210	
National Origin	-.080	
$R^2 = .086$		N = 22

Table 4.25		
<i>Multiple Regression for Canonical WOR, Highlanders</i>		
Variable	β_1	p
English Skills	.589	**
Class	.339	*
National Origin	-.113	
$R^2 = .621^*$		N = 26

Table 4.26		
<i>Multiple Regression for Canonical WOR, Caribbeans</i>		
Variable	β_1	p
Class	.207	
English Skills	.156	
National Origin	-.067	
$R^2 = .031$		N = 22

Table 4.27		
<i>Multiple Regression for Canonical WOR, Highlanders</i>		
Variable	β_1	<i>p</i>
Language with Siblings	-.531	*
Class	.183	
National Origin	.067	
$R^2 = .311^a$		N = 23

Table 4.28		
<i>Multiple Regression for Canonical WOR, Caribbeans</i>		
Variable	β_1	<i>p</i>
Class	.162	
Language with Siblings	-.123	
National Origin	.006	
$R^2 = .029$		N = 19

Table 4.29		
<i>Multiple Regression for Canonical WOR, Highlanders</i>		
Variable	β_1	<i>p</i>
Spanish Skills	-.418	^a
National Origin	.187	
Class	.113	
$R^2 = .199$		N = 26

Table 4.30		
<i>Multiple Regression for Canonical WOR, Caribbeans</i>		
Variable	β_1	<i>p</i>
National Origin	-.229	
Class	.223	
Spanish Skills	-.018	
$R^2 = .088$		N = 21

In general, the results from Tables 4.23-4.30 indicate that the regional sub-samples vary in significance and the amount of variance accounted for by the variables in the models. The R^2 values in the tables indicate several things. First, they indicate that we can have very little confidence in our models for the Caribbeans as none of the models accounted for more than nine percent of the variance in canonical WOR, and none of the models were overall significant. For the Highlanders the model that included Spanish Skills was not significant at all, while the model that included Language with Siblings approached significance and accounted for 31 percent of the variance in canonical WOR. The models that included English Skills and Generation were both statistically significant and accounted for 62 and 54 percent of the variance respectively.

Based on the column labeled β_1 we can make a few descriptive remarks about our variables. First, National Origin did not survive in any of the regressions for either Caribbeans or Highlanders. In fact, none of the variables in any of the models were significant for the Caribbeans. Among the Highlanders, all of the variables related to Generation were either significant or approached significance and in each of their respective models they were at the top

of the variable hierarchy. As in the tables for the entire sample, Class was significant in only one of the models, and as a result we can only have limited confidence in its results. The directions of the relationships for the statistically significant variables were also as expected based on the results from the bivariate analyses.

3.3. Multiple linear regressions for the generational sub-samples.

In this section we analyze some of the same variables that were considered above for the whole sample regressions, but we look at each generational sub-sample separately. As we did above with Region, we excluded Generation from the models because it is no longer relevant when analyzing each generational sub-sample individually. For each regression below we include Class and Region as well as Language with Siblings, English Skills, and Spanish Skill alternatively. The tables on the left present the results for the LARNC and those on the right present the results for the NYR.

Table 4.31		
<i>Multiple Regression for Canonical WOR, LARNC</i>		
Variable	β_1	<i>p</i>
Region	.457	*
Class	-.038	
English Skill	-.038	
R ² = .202		N = 24

Table 4.32		
<i>Multiple Regression for Canonical WOR, NYR</i>		
Variable	β_1	<i>p</i>
Class	.429	^a
English Skills	.156	
Region	.011	
R ² = .176		N = 24

Table 4.33		
<i>Multiple Regression for Canonical WOR, LARNC</i>		
Variable	β_1	<i>p</i>
Region	.433	*
Language with Siblings	-.077	
Class	-.025	
R ² = .193		N = 23

Table 4.34		
<i>Multiple Regression for Canonical WOR, NYR</i>		
Variable	β_1	<i>p</i>
Class	.392	
Language with Siblings	.154	
Region	.018	
R ² = .227		N = 19

Table 4.35		
<i>Multiple Regression for Canonical WOR, LARNC</i>		
Variable	β_1	<i>p</i>
Region	.508	*
Spanish Skills	-.130	
Class	-.034	
$R^2 = .252$		N = 23

Table 4.36		
<i>Multiple Regression for Canonical WOR, NYR</i>		
Variable	β_1	<i>p</i>
Spanish Skills	.400	*
Class	.378	^a
Region	.101	
$R^2 = .307^*$		N = 24

In general, the results from Tables 4.31-4.36 indicate that the generational sub-samples vary in significance and the amount of variance accounted for by the variables in the models. The R^2 values in the tables indicate several things. First, although our models accounted for 18-31 percent of the variance in canonical word order rate, we can have very little confidence in these data as only one model reached significance for the NYR and none of the models for the LARNC did.

At the individual variable level, based on the column labeled β_1 , we can make a few descriptive remarks about our variables. Among the LARNC the only variable that reached significance in any and all of the models was Region. For the NYR this variable never reached significance. Spanish Skills was the only variable that reached significance among the NYR and Class approached significance in two of the three models. Language with Siblings and English Skills did not survive in any of the regressions for either LARNC or NYR. The directions of the relationships for the statistically significant variables for the LARNC were as expected based on the results from the bivariate analyses. For the NYR, however, we found that an increase in reported Spanish skills corresponds to an increase in canonical WOR, which is in the opposite direction of the relationship we found for the sample as a whole.

We attempted to separate each regional group by generation as well in order to compare LARNC Caribbean participants to NYR Caribbean participants and LARNC Highlanders to NYR Highlanders, but none of the regression models yielded significant results at either the

overall model or individual variable levels. We believe that the reason our results did not yield significance is because of the low frequencies as the N for our models was between nine and twelve.

4. Summary and discussion

In order to address our research questions, we begin this section with a discussion and summary of the overall results (research question a) for the entire corpus. We then examine our results by region (research question b) and by generation (research question c).

4.1. Overall results.

Based on the canonical WOR for the entire corpus (75 percent), we can conclude that the most common word order for all the participants combined as well as for each of the regional and generational sub-samples is canonical word order. These results are consistent with those studies on Spanish word order with copular constructions (Ocampo, 2002).

Using a series of bivariate analyses (including correlations and ANOVAs) we determined that there were no significant differences in canonical WOR for some of our external predictor variables. These variables are listed below:

Predictor variables that did not appear to have a predictive statistical effect on Canonical WOR

- Gender
- Age
- Education
- Spanish use
- Language with father
- Language with mother
- Language with children

- Language with friends
- Language with boss
- Language with significant other
- Language with speakers from own country
- Language with other groups
- Language with speakers from own group
- TV in Spanish
- Radio in Spanish
- Reading in Spanish
- Spanish use in social contexts
- Spanish at home
- Spanish in domains

The lack of significance for the first three variables (Gender, Age, and Education) aligns with Silva-Corvalán's (2001) generalization that external or social variables have minimal impact on the variability of syntactic or morpho-syntactic features (133). Although we had expected that the other 16 variables would yield significant results, given our expectations that an increased use of Spanish with a variety of speakers and in various contexts would correspond to a decreased use of canonical WOR, we did not find any of the language use variables above to be significant. The reason the questions that elicited these data were included in the questionnaire was to provide a less direct, but more reliable assessment of the participants' English and Spanish skills. Although we had specific questions addressing both Spanish and English skills, oftentimes the problem with self-reported proficiency is that participants suffer from linguistic insecurity and rate themselves on correctness norms instead of on competency

and fluency (Otheguy & Zentella, 2012, p. 93). As a result, their perceived proficiency may not align with their actual linguistic proficiency. The purpose, then, of the above language use variables was to better judge the speakers' competence and fluency based on how much English or Spanish they used with their close relatives and colleagues (mother, father, spouse, siblings, children, boss, friends) and in familiar contexts (home, school, and social). In this way, we expected participants who claimed to use English more with these people and in these contexts to have higher proficiency in English. Our expectations were correct. As a result, it is especially surprising that so few of the language use variables achieved significance (only one) even though English Skills did. However, it is important to note that although the p value for the majority of the language-use variables did not approach significance, almost all of variables followed the expected pattern of a decrease in canonical WOR corresponding to an increase in Spanish use.

There were several variables that did have a predictive statistical effect on canonical WOR and which will be the focus of our discussion. These are listed below.

Independent (predictor) variables significantly associated with word order

- Generation
- Age of arrival in NYC
- Years lived in NYC
- National origin
- Regional origin
- Class
- English skills
- Spanish skills
- Language with siblings

We will leave the first four variables for our discussion in section 4.2 and discussion of Regional Origin for both sections 4.2 and 4.3. Class and Spanish Skills will be discussed in section 4.3 as well as they were only significant for one of the generational sub-samples. The last three variables will be discussed below as well as in sections 4.2 and 4.3.

English Skills. Since our research questions relate to the influence of English on word order in Spanish, the mastery of English was of special interest to the study. In our bivariate analyses we found that participants that reported their English to be poor or passable had a lower canonical WOR by 13 percentage points than those who reported their English to be good or excellent. In our multivariate analyses English Skills was a significant predictor of canonical WOR, the highest predictor in its regression model with an increase in reported English skills corresponding to a higher canonical WOR. Given our predictions, it is not surprising that we find a relationship between an increase in reported English skills and increased canonical WOR. We believe that it is due to these participants' increased contact with and strong command of English, a language with a rigid SVO order, that they have an increased canonical WOR in Spanish. We argue that this phenomenon is an example of simplification (Silva-Corvalán, 1994) in that we find a reduction in the use of alternative or competing forms among those speakers with increased language contact. In other words, both forms (canonical word order and marked word order) exist in the language of the LARNC, before the start of simplification. The process of simplification involves a reduction in the use of one of the forms, thus we would expect that NYR to have an increased canonical WOR as a result of the decreased use of marked word order, which is precisely what we find in our data. It has been suggested by several researchers that one of the consequences of this reduction in forms for variable word order is the concomitant reduction in communication of meaning. This occurs, they argue, as a result of the reduction in

discourse strategies to communicate the variety of pragmatic functions that first generation bilinguals do using word order (Silva-Corvalán, 1994; Zapata, Sánchez & Toribio, 2005). We maintain, however, that in order to arrive at such conclusions we need to determine two things: first, we would need to compare the discourse pragmatic and internal constraints that govern word order across the two generations; and secondly we would need to determine whether there is another resource with which these speakers maintain meaning. That is, do they compensate for the loss of the discourse-pragmatic constraints that usually govern word order in the pre-contact lects by using other resources, such as prosody? The answer to the first question will be addressed in Chapter Five, while the answer to the second question will be addressed in Chapter Six.

Language with Siblings. One of our most curious findings is that out of all the variables on language use with close relatives (mother, father, spouse, siblings, children), the only variable that approached significance in the bivariate analyses was Language Spoken with Siblings. Recall that this variable was also a significant predictor of canonical WOR in one of the regression models for the entire sample. We did not expect to find significant results for Language Spoken with Mother or Language Spoken with Father considering that none of the LARNC reported speaking English with their parents and most of the NYR also reported speaking Spanish with their parents. In other words, because speaking the heritage language with your parents is quite common, we would not expect those variables to be predictive of language contact phenomenon such as canonical WOR. The variable Language with Children, which would be highly suggestive of heritage language maintenance, especially among the second generation, did not reach significance in our sample probably because of the low frequencies, that is, because many of our participants did not have children. Only six LARNC

and ten NYR responded to this question. A similar argument can be made for Language Spoken with Spouse or Significant Other. More than half of our LARNC did not respond to this question (13) while only two-thirds of NYR responded to this question (10). Among our participants it was much more common to have a sibling than to have a child or be in a relationship (only six missing cases for the entire corpus). Language Spoken with Siblings is also different from the other variables (although similar to Language Spoken with Children) in that it may be predictive of language contact phenomenon. For example, birth order has been found to be an important predictor of Spanish language maintenance and use in the US. Children who are first-born tend to develop higher levels of proficiency in the heritage language than their second-born or third-born siblings (Lambert & Taylor, 1996; Zentella, 1997). This has been attributed, in part, to the fact that second and third-born children are exposed to English, through their first-born sibling, much earlier than those first-born children are exposed to the dominant language through entry into the education system or daycare. We can infer, then, that those participants who reported speaking only Spanish with their siblings have a higher proficiency in Spanish than those who reported speaking to their siblings in English or in both languages. Based on this assumption, our data align with the idea that increased proficiency and use of Spanish would correspond to a lower canonical WOR.

4.2. Results and discussion by region.

We found several differences between regional groups with regard canonical WOR. Recall that overall, Highlanders used canonical WOR more often than Caribbeans by ten percentage points, though this difference was not significant ($p = .135$). Similarly, in the multivariate analyses we found Region approached significance in three of the four regression

models for the entire sample. These data would have us believe that there are several differences across the two regional sub-samples.

English Skills and Language Spoken with Siblings. As we stated previously, we predicted and confirmed for the entire sample that increased command of English corresponds to an increase in canonical WOR in Spanish. Similarly, an increased use of Spanish with your siblings corresponds to a decrease in canonical WOR. Both of these patterns are also true among our Highlanders. Those that reported poor or passable English skills had a canonical WOR 27 percentage points lower than those that reported good or excellent English skills. Highlanders that spoke Spanish with their siblings had a lower canonical WOR by 32 percentage points than those that spoken English with their siblings. However, we did not find these same patterns among the Caribbeans. We found almost no difference (one percent) for reported English skills and a small difference (three percent) for Language Spoken with Siblings. Our multivariate analyses confirm these results: neither of these variables was significant in any of the models for the Caribbeans while both variables were significant in the models for the Highlanders. Additionally, we found that English skills (along with Class) significantly accounted for 62 percent of the variance in canonical WOR among Highlanders, but only three percent (and not significantly) among Caribbeans. These differences across regions may have to do with the Caribbean participants' overall higher canonical WOR or their increased contact with English, as well as their unique political ties to the United States, especially among the Puerto Ricans (Lynch & Klee, 2009, p 200; Zentella, 2000, p. 139).

4.3. Results and discussion by generation.

We also found several differences between generational groups with regard to canonical WOR. Since the two generational sub-samples were created using the continuous variables Age

of Arrival in NYC and Time Spent in NYC we will focus our discussion on the Generation variable given that the results for all three variables would be identical. For the entire corpus as well as for the Highlanders we found a significant difference in canonical WOR across generations. Specifically, the NYR participants had a higher canonical WOR than the LARNC by 23 percentage points. However, it appears that the Highlanders were driving this generational difference. NYR Highlanders had a higher canonical WOR than the LARNC Highlanders by 36 percentage points, while the difference across generations for the Caribbeans was only eight percentage points. This generational change in the Highlanders is even more pronounced when we look at each specific nationality.

National and Regional Origin. We found an increased use of canonical WOR across the generations among all the nationalities except Cubans. Only the generational differences for the Highlander nationalities (Colombian, Ecuadorian, Mexican) were significant. The difference from one generational cohort to the other is most marked among the Mexican participants. However, the latter result could be due to the fact that some of the NYR Mexican participants produced relatively few tokens.

The results from our bivariate analyses for Region were confirmed in the multivariate analyses. Recall that Region was the only variable that was significant in all three of the regression models for LARNC, but was not significant in any of the models for the NYR. Based on these data we can make several arguments. First, we can attribute the slight increase in canonical WOR, albeit small, to the NYR's increased contact with English. The same case can be made for the Highlander group. However, we can also argue that the Highlander group appears to be *Caribbeanizing* their canonical WOR. That is, by increasing their canonical WOR, the Highlanders are assimilating to the Caribbean lects with regard to this feature. It is not

uncommon for a minority group to assimilate to the lexical or phonological features of the linguistic majority (Zentella, 1990). Considering that Puerto Ricans, Dominicans, and Cubans combine to make up almost two-thirds of the Latinos in NYC (Bergad, 2011), it would not be unusual for the Highlander immigrants, who compared to the Caribbeans in NYC are relatively new immigrants, to assimilate to other types of linguistic features, such as a more fixed SVO order, especially when that feature is also shared by the dominant language, English. In other words, although it appears that we have clear evidence of language contact, we may also be able to argue that, among Highlanders, we have a case of dialectal contact too.

Class. We included Social Class and Education as variables given that several studies have demonstrated a significant relationship between language variability in Spanish and speakers' level of income, occupation, education, neighborhood, and type of residence (Alba, 1988; Cedgren, 1978; Lafford, 1986; Samper-Padilla, 1990). We did not find any differences in canonical WOR across the two classes for the entire sample or for the Newcomers; among the NYR we found that the working class participants had a higher canonical WOR, and this difference approached significance. In the multivariate analyses one of our regression models labeled Class as a significant variable for the entire sample as well as for the Highlanders. Among the LARNC it did not survive any of the regressions, while for the NYR it approached significance in two models. Based on the combination of both our multivariate and bivariate results we can argue that Class is an important predictor of canonical WOR among the NYR, although it may not be among the LARNC or the entire sample. There are several possible explanations for these findings. First, it is well documented that the upper classes tend to maintain the standard features of a language (Alba, 1988; Cedgren, 1978; Lafford, 1986; Samper-Padilla, 1990). In our case that would mean that the middle class participants would tend

to use the marked forms more often than the working class participants. That is precisely the tendency we find for both generational sub-samples, although this difference is not significant among the LARNC. It is possible, however, that the working class LARNC may also be limited in their access to English and English language education. If this is the case, then, the working class LARNC, by virtue of not having as much access to English and English language education as the middle class LARNC, would not have a significantly increased canonical WOR compared to their middle class counterparts. This idea is supported by the fact that we found a significant correlation between English Skills and Class among the LARNC ($r(23) = -.650, p < .01$), but not among the NYR. That is, being working class is associated with lower English skills among the LARNC, but not among the NYR.

Spanish Skills. As we saw previously, an increase in reported English skills corresponds to an increase in canonical WOR. Conversely, we would expect that an increase in reported Spanish skills would correspond to a decrease in canonical WOR. However, we did not find this pattern among our entire corpus. When we separated out the participants by generation, we found this pattern among the NYR, but not among the LARNC. Similarly, in our multivariate analyses, Spanish Skills was the only variable that significantly accounted for variance in canonical WOR among the NYR, but was not significant among the LARNC. There are several reasons why we would not expect Spanish Skills to be predictive of canonical WOR among the LARNC. First, only one LARNC participant rated his Spanish as poor or passable. Secondly, the fact that not all the LARNC, whose native language is Spanish and whose primary and secondary education were exclusively in Spanish, rated their Spanish skills as excellent indicates that these speakers' self-reported proficiency is based on correctness norms and not fluency or competence, demonstrating their linguistic insecurity (Otheguy & Zentella, 2012, p. 93). As a

result, it is not surprising that we did not find a relationship between Spanish skills and canonical WOR among LARNC. Among NYR, however, we did find a relationship between these two variables: Those who reported poor or passable Spanish skills had a higher canonical WOR than those who reported good or excellent skills by 11 percentage points. We believe that a weak command of Spanish, along with increased contact with English yields an increased canonical WOR in Spanish among the NYR. These data support the argument that increased canonical WOR among NYR participants is an example of simplification (Silva-Corvalán, 1994), or a reduction in the use of alternative or competing forms among those speakers with increased language contact. Recall that this reduction of forms in word order variability has been interpreted to mean a consequent loss of communicative resources among the second generation (Silva-Corvalán, 1994; Zapata, Sánchez & Toribio, 2005). In the next chapter we examine the effects of the internal predictor variables on word order in order to compare the discourse pragmatic and internal constraints that govern word order across the two generations. These data, along with the data on prosody in Chapter Six will allow us to determine whether such changes are taking place among the second-generation speakers in our sample.

CHAPTER 5

INTERNAL VARIABLES

1. Introduction

This chapter presents the analysis of the internal segmental predictor variables described in Chapter Three. The analysis aims at answering the research questions outlined in Chapter One and repeated here below. Only the research questions addressed in the chapter are listed.

Question D: What are the internal variables that have predictive statistical effect on word order in the corpus? Do any of these internal variables produce categorical effects (that is, effects where the presence of a predictor produces a 100 percent to zero result)?

Question E: What differences, if any, are there between the two geographic regions in the way that the internal variables influence word order, either quantitatively or categorically?

Question F: What differences, if any, are there between the two immigrant generations in the way that the internal variables influence word order, either quantitatively or categorically?

In section 2 below we present the results for all the internal predictor variables and their relationship with word order. The results for the second outcome variable, prosody, are presented in the next chapter. Throughout the analysis of internal variables in the present chapter we separate the results out by generation and region. The results of the regression analyses are presented in section 3. A summary and discussion of the results appears in section 4.

2. Internal (linguistic) variables predicting word order

Below we discuss the relationships between the outcome variable word order and the internal predictor variables. The word order outcome variable considered here covers utterances found in the corpus consisting of (a) a simple, complex, infinitival, or clausal subject, (b) a

copular verb and (c) an adjectival predicate. The results for utterances with adverbial predicates will be discussed separately in Chapter Seven.

Recall that the outcome variable word order has four factors or levels, as outlined below.

Dependent (outcome) variable: Word order

- Subject-Copula-Adjective *Nicolás es feliz* ‘Nicolas is happy’
- Copula-Adjective-Subject *Es feliz Nicolás* ‘lit. Is happy Nicolas,’ ‘Nicolas is happy’
- Adjective-Copula-Subject *Feliz es Nicolás* ‘lit. Happy is Nicolas,’ ‘Nicolas is happy’
- Copula-Subject-Adjective *Es Nicolás feliz* ‘lit. Is Nicolas Happy,’ ‘Nicolas is happy’

The first order, Subject-Copula-Adjective, is considered in the present study the canonical word order whereas the other orders are considered the marked orders.

The word order dependent or outcome variable was studied with respect to the independent or predictor variables outlined below.

Independent (predictor) variables initially investigated

- Presence of adverb in the predicate
- Negative or affirmative utterance
- Type of copular verb
- Presence of English elements
- Syntactic properties of subject NP
- Length and type of adjectival phrase
- Number in third-person verbs
- TMA of the verb
- Discourse pragmatics: Informational status of the subject NP
- Discourse pragmatics: Function of the utterance

The initial exploration of these predictors showed that many of them bear an association with word order that is statistically negligible or too inconsistent or unreliable to be revealing of any pattern worth reporting. As a result, this chapter, while devoting a subsection to providing some details regarding each of the variables above, concentrates primarily on the discussion of the following predictors, which did yield significant results.

Independent (predictor) variables significantly associated with word order

- Syntactic properties of subject NP
- Number in third-person verbs
- Discourse pragmatics: Function of the utterance
- Discourse pragmatics: Informational status of the subject NP

2.1. Distribution of the dependent variable word order in the corpus as a whole and by region and generation.

As we saw in section 3.3.1 of the previous chapter, 75 percent of the utterances produced by all the participants had canonical word order. Specific results regarding word order at the token level can be seen in Table 5.1.

Table 5.1	
<i>Word Order Occurrence</i>	
Word Order	Tokens
Canonical Word Order	317 (75%)
Marked Orders	107 (25%)
Total	424 (100%)

Tables 5.2 and 5.3 below show the breakdown for word order by generation and region using the two-level variable.

Table 5.2			
<i>Cross-Tabulation of Word Order by Generation</i>			
Generation	N	Word Order	
		Canonical order	Marked orders
LARNC	233	153 (66%)	80 (34%)
NYR	191	164 (86%)	27 (14%)
Total	424	317 (75%)	107 (25%)

Table 5.3			
<i>Cross-Tabulation of Word Order by Region</i>			
Region	N	Word Order	
		Canonical order	Marked orders
Highlanders	239	162 (68%)	77 (32%)
Caribbeans	185	155 (86%)	30 (14%)
Total	424	317 (75%)	107 (25%)

The tables show that the LARNC produced a total of 233 tokens while the NYR participants produced 191 tokens. Similarly, the Highlanders produced 239 while Caribbean participants produced 185 utterances.

In the sections below, we will present tables with the two-level variable unless specified. We will postpone discussion and interpretation of our results until section 4. The order of presentation of the predictor variables follows the order in which they were discussed in Chapter Three.

2.2. Affirmative or negative utterance.

Whether the adverb *no* occurred in the utterance did not significantly affect word order in the whole corpus or across regional or generational groups.

2.3. Presence of adverb in the predicate.

We found no significant differences in word order for utterances with or without an adverb in the predicate. We also did not find any significant differences across the two generational cohorts with regard to this variable. With regard to region, there were no significant results of interest to report. The presence of an adverb, by and large, does not have an impact on word order.

2.4. Type of copular verb.

The type of copula in the utterance, that is, whether the copular verb in the utterance was either *ser* ‘to be’, *estar* ‘to be’, or *parecer* ‘to seem’, did not appear to have any significant relationship with word order. Three quarters of the utterances that appeared with either *ser* ‘to be’ or *estar* ‘to be’ had canonical order. Utterances with the verb *parecer* ‘to seem’ occurred with canonical order only but this cannot be generalized since this verb only occurred twice in the corpus.

2.5. Presence of English elements.

As expected, utterances with English elements had an overall higher rate of canonical order than those without English elements. However, this result, while showing the predicted trend, was not statistically significant. When we separated out the data by generation we found that the impact of English elements on word order approached significance only among LARNC.

2.6. Syntactic properties of the subject NP.

Given that the length and syntactic structure of a constituent are known to affect its position in the utterance we sorted the subjects by the type of NP in order to determine whether this variable was significantly related to word order. Recall from Chapter Three that a subject was considered simple as long as it consisted of a one word common noun, as in (1), or a one word common noun modified by a determiner or a possessive, as in (2), or a combination of a

one word noun or a one word adjective with or without determiners, as in (3), or a multiple word proper noun. A subject is complex when the noun is modified by anything other than a single adjective or a determiner. A noun modified by a prepositional phrase is complex, as in (4). A noun modified by a relative clause is complex, as in (5). Clausal and infinitival subjects that appeared with a copular verb and an adjectival predicate, like in (6) and (7), were considered clausal-infinitival.

- (1) Es increíble Manhattan. [305M]
Is incredible Manhattan.
'Manhattan is incredible.'
- (2) La gente es agradable. [308M]
The people are nice.
'The people are nice.'
- (3) Esos grupos viejos son buenos. [201U]
Those group old are good.
'Those old groups are good.'
- (4) La hermana de su novio está desaparecida. [325E]
The sister of her boyfriend is disappeared.
'Her boyfriend's sister is disappeared.'
- (5) El español que habla el colombiano es muy propio. [263C]
The Spanish that speaks the Colombian is very proper.
'The Spanish that a Colombian speaks is very proper.'
- (6) Es fácil comunicarse. [384E]
Is easy communicating.
'Communicating is easy'
- (7) Es importante, que sepa hablar dos idiomas. [330D]
Is important that know to speak two languages.
'Knowing how to speak to languages is important.'

Our data indicate that the syntactic type of subject NP was significantly related to word order. Additionally, among both generational and regional cohorts we also found a statistically significant relationship between this variable and word order. For this analysis, in order to avoid

frequency issues, we combined the clausal and infinitival subjects into one variant. The results for the entire corpus can be seen in Table 5.4 below.

Table 5.4			
<i>Cross-Tabulation of Word Order by Type of Subject NP</i>			
Type of Subject NP	N	Word Order	
		Canonical order	Marked orders
Simple	301	267 (89%)	34 (11%)
Complex	51	41 (80%)	10 (20%)
Clausal-Infinitival	72	9 (12%)	63 (88%)
Total	424	317 (75%)	107 (25%)
$\chi^2 = 179.802$ $df=2$, $p < .01$			

Table 5.4 shows a clear pattern; the use of canonical word is highest in sentences with syntactically simple subjects, it is somewhat lower in those with complex subjects, and much lower still in sentences with clausal-infinitival subjects.

This finding requires further exploration in terms of the connection between syntactic complexity of the subject and simple length of the subject measured in number of words. In section 4.2 in Chapter Two we discussed the relationship between the length of constituents in an utterance and word order. It would be useful, then, to determine whether there is a relationship between the length of a subject NP and the syntactic type of subject NP, since we would expect clausal-infinitival and complex subjects to be longer than simple subjects and we would also expect longer constituents to be utterance final. To this end, we calculated the average number of words for each type of subject NP (simple, complex, and clausal-infinitival). Table 5.5 below shows that simple subjects had the fewest words on average, while clausal-infinitival subjects had the greatest number of words. The overall differences in numbers of words across NP types are significant, and follow a pattern that parallels Table 5.4 above (though we note that the 6.08 to 6.94 difference between complex and clausal-infinitival subjects, which follows the trend, is

not itself significant).

Table 5.5		
<i>ANOVA, Mean words by NP Type</i>		
NP Type	N tokens	Average Length in Words
Simple	301	1.98
Complex	51	6.08
Infinitival/Clausal	72	6.94
	424	
$F = 95.524$	$p < .01$	

It appears, then, that the longer the subject NP, the more likely it is to appear in post-verbal position (marked order).

This three-way correspondence between increased syntactic complexity of the subject, increased length of the subject, and increased marked order of words continues to hold up, though not as neatly, when we separate the data by generation. Although the syntactic type of subject NP was significantly related to word order among both generational groups, the patterns observed across the two groups differed. Among LARNC we found an increased incidence of marked order as the complexity of the subject NP increased, but among NYR we did not find the exact same pattern; although NYR participants used more canonical word order with utterances with simple NPs and more marked order with utterances with clausal-infinitival NPs, they used canonical word order exclusively in utterances with a complex subject NP. These differences across the generations can be seen in Tables 5.6 and 5.7 below.

Table 5.6			
<i>Cross-Tabulation of Word Order by Type of Subject NP, LARNC</i>			
Type of Subject NP	N	Word Order	
		Canonical order	Marked orders
Simple	142	122 (86%)	20 (14%)
Complex	34	24 (71%)	10 (29%)
Clausal-Infinitival	57	7 (12%)	50 (88%)
Total	233	153 (66%)	80 (34%)
$\chi^2 = 98.243$ $df=2$, $p < .01$			

Table 5.7			
<i>Cross-Tabulation of Word Order by Type of Subject NP, NYR</i>			
Type of Subject NP	N	Word Order	
		Canonical order	Marked orders
Simple	159	145 (91%)	14 (9%)
Complex	17	17 (100%)	0 (0%)
Clausal-Infinitival	15	2 (13%)	13 (87%)
Total	191	164 (86%)	27 (14%)
$\chi^2 = 71.534$ $df=2$, $p < .01$			

With some exceptions, then, we find across both regional groups the same significant relationship that we found overall, namely an increased incidence of marked order as the syntactic complexity of the subject NP increases, a complexity that itself correlates highly with the simpler measure of length, measured in number of words.

2.7. Length and type of adjectival phrase.

Each token was also coded for the type of predicate, whether simple or complex. Recall from Chapter Three that a predicate was considered simple if it contained a single adjective or an adjective modified by an adverb, as in (8). A predicate was considered complex if the adjective was modified by anything other than a single adverb or if there was more than one adjective, as in (9). Finally, an adjective modified by a relative clause is considered complex, as in (10).

- (8) El español que habla el colombiano es muy propio. [263C]
The Spanish that speaks the Colombian is very proper.
'The Spanish that a Colombian speaks is very proper.'
- (9) El español es igual para todo el mundo. [92P]
The Spanish is the same for everyone.
'Spanish is the same for everyone.'
- (10) El college es muy distinto a lo que se llama el high school. [206U]
The college is very different to what is called the high school.
'College is very different from what is called high school.'

Based on our findings in the previous section, as well as our discussion of constituent length in Chapter Two, we would expect to find the same relationship between adjective type and word order as between NP type and word order. Thus, we would expect complex adjectives to appear in final utterance position, both because we found that more complex NP subjects are utterance final in greater proportions than simple ones and because longer constituents tend to appear utterance final, and syntactically complex items tend to be longer. Specifically, we would expect complex adjectives to occur in utterances with canonical word order, since the adjective is in final position, or in one of the marked orders with the adjective in final position. However, there were no significant differences in word order for utterances with complex or simple predicate adjectives. There was also no significant relationship between word order and adjective type among the two generational cohorts or the regional groups. Our results in this and the previous section indicate that in our corpus complexity (and length) are determinants of word order for subject NPs but not for predicate adjectives.

2.8. Number in third-person verbs.

As a way to delimit the scope of our study, we limited our analysis to third-person verbs. In these verbs, we studied the influence of verb number on word order in the corpus. Table 5.8 below shows that there is a significant relationship between word order and whether the number

ending of the verb is singular or plural. Very few utterances with third person plural verbs had marked order, whereas almost a third of the utterances with third person singular verbs had marked word order.

Table 5.8			
<i>Cross-Tabulation of Word Order by Number Ending of the Verb</i>			
Verb	N	Word Order	
		Canonical Order	Marked Orders
3sg	326	229 (70%)	97 (30%)
3pl	95	85 (90%)	9 (10%)
Total	421	317 (75%)	107 (25%)
$\chi^2 = 13.352$ $df=1$, $p < .01$			

These same results were also found among the LARNC and Highlanders, that is, utterances with plural verbs produced by LARNC or by Highlanders appeared more often in canonical word order compared to utterances with singular verbs and this difference is significant. Among NYR participants, the same pattern is found, with statistically near significant results ($p < .08$). Among Caribbean speakers, however, the difference was not significant. It is important to point out that utterances with first person plural verbs, which we excluded from the contingency tables, also had canonical word order. It appears, then, that utterances with plural verbs favor canonical word order more so than utterances with singular verbs.

2.9. Tense-Mood-Aspect of the verb.

Another verb property whose relationship with word order was examined was the tense-mood-aspect (TMA) of the verb. Recall from Chapter Three that the verbs in the corpus occurred in the present indicative, preterit indicative, imperfect indicative, present perfect indicative, and conditional. In order to avoid these frequency issues the verb that was not conjugated was excluded from this analysis.

Table 5.9			
<i>Cross-Tabulation of Word Order by TMA of the Verb</i>			
TMA	N	Word Order	
		Canonical order	Marked orders
Preterit Indicative	73	60 (82%)	13 (18%)
Present Indicative	322	241 (75%)	81 (25%)
Present Perfect Indicative	3	2 (67%)	1 (33%)
Imperfect Indicative	22	12 (55%)	10 (45%)
Conditional	3	1 (33%)	2 (67%)
Total	423	316 (75%)	107 (25%)
$\chi^2 = 9.720$ $df=4$, $p < .05$			

Based on the results above, it appears that utterances with verbs in the conditional tend to favor marked word order whereas utterances with indicative tend to favor canonical word order. We found the same trend among the LARNC participants. Because several of the tenses have so few tokens, however, these results are not very reliable. Among the we did not find that the verb's TMA was not related to word order among the NYR as there was a preference for canonical word order regardless of the TMA of the verb. Additionally, NYR participants did not produce any tokens with verbs in the conditional or present perfect.

2.10. Discourse pragmatics: Informational status of the subject NP.

In addition to the syntactic properties of the subject NP discussed in section 2.5 above, we also examined the subject's discourse-pragmatic properties. Specifically, we examined whether the givenness of the subject NP was related to the word order of the constituents of the utterance. Recall that informational status in the corpus had three levels: new, given, and implied. However, with three factors, the results were not revealing and proved difficult to interpret. Instead, we collapsed this three-level factor into two levels (given versus new subjects) based on the fact that the majority of the extant literature follows this format. Following Ocampo (2002), who considers inferred subjects as new referents, we collapsed implied subjects into the new

subjects group. In doing so our results, for the entire sample as well as for each generation, align better with the extant literature, which indicates that new information tends to occur in utterance final position (Arnold et al., 2000; Bentivoglio, 2003; Bolinger, 1955). Table 5.10 shows significant results for the whole sample and Table 5.12 shows significant results for the NYR, while Table 5.11 shows near-significant results for the LARNC.

Table 5.10			
<i>Cross-Tabulation of Word Order by Informational Status of Subject NP Using a Two-level Informational Status Variable</i>			
Informational Status	N	Word Order	
		Canonical order	Marked orders
New or implied	173	113 (65%)	60 (35%)
Given	251	204 (81%)	47 (19%)
Total	424	317 (75%)	107 (25%)
$\chi^2 = 13.821$ $df=1$, $p < .01$			

Table 5.11			
<i>Cross-Tabulation of Word Order by Informational Status of Subject NP, LARNC, Using a Two-level Informational Status Variable</i>			
Informational Status	N	Word Order	
		Canonical order	Marked orders
New or implied	117	70 (60%)	47 (40%)
Given	116	83 (72%)	33 (28%)
Total	233	153 (66%)	80 (34%)
$\chi^2 = 3.550$ $df=1$, $p < .06$			

Table 5.12			
<i>Cross-Tabulation of Word Order by Informational Status of Subject NP, NYR, Using a Two-level Informational Status Variable</i>			
Informational Status	N	Word Order	
		Canonical order	Marked orders
New or implied	56	36 (77%)	13 (23%)
Given	135	121 (90%)	14 (10%)
Total	191	164 (86%)	27 (14%)
$\chi^2 = 5.380$ $df=1$, $p < .02$			

We found the same trend for each of the regional groups; results are significant for the Highlanders and near-significant for the Caribbeans (see Tables 5.13 and 5.14).

Table 5.13			
<i>Cross-Tabulation of Word Order by Informational Status of Subject NP, Highlanders Using a Two-level Informational Status Variable</i>			
Informational Status	N	Word Order	
		Canonical order	Marked orders
New or implied	100	56 (56%)	44 (44%)
Given	139	106 (76%)	33 (24%)
Total	239	162 (68%)	77 (32%)
$\chi^2 = 10.931$ $df=1$, $p < .01$			

Table 5.14			
<i>Cross-Tabulation of Word Order by Informational Status of Subject NP, Caribbeans Using a Two-level Informational Status Variable</i>			
Informational Status	N	Word Order	
		Canonical order	Marked orders
New or implied	73	57 (78.%)	16 (22%)
Given	112	98 (88%)	14 (12%)
Total	185	155 (84%)	30 (16%)
$\chi^2 = 2.885$ $df=1$, $p < .09$			

2.11 Discourse pragmatics: Function of the utterance.

The second discourse pragmatic predictor variable we examined was the pragmatic function of the utterance. Of the 424 tokens coded for pragmatic function, 78 percent (331 tokens) of the utterances were pragmatically neutral. Highlighted adjective was the second most common pragmatic function with 65 tokens followed by subject topic contrast with 26 tokens. There was only one token for each of the last two pragmatic functions of contrary to expectation and contrastive verb.

We found clear differences in word order across the pragmatic functions. While pragmatically neutral utterances and those that conveyed the pragmatic function of subject topic contrast tended to occur in canonical word order, utterances that communicated the pragmatic function of highlighted adjective, contrary to expectation, and contrastive verb tended to occur in marked word orders. These results can be seen in Table 5.15.

Table 5.15			
<i>Cross-Tabulation of Word Order by Pragmatic Function</i>			
Pragmatic Function	N	Word Order	
		Canonical order	Marked orders
Pragmatically Neutral	331	270 (82%)	61 (18%)
Subject Topic Contrast	26	21 (81%)	5 (19%)
Highlighted Adjective	65	26 (40%)	39 (60%)
Contrary to Expectation	1	0 (0%)	1 (100%)
Contrastive Verb	1	0 (0%)	1 (100%)
Total	424	317 (75%)	107 (25%)
$\chi^2 = 58.947$ $df=4$, $p < .01$			

For both generational groups and regional groups Pragmatically Neutral was also the most common pragmatic function. However, there were a few differences across generations and regional groups. There were not many differences in the frequencies of pragmatic functions used, but the NYR did not produce any utterances that communicated the pragmatic function of

contrary to expectation or contrastive verb. Similarly, the Highlanders did not produce any utterances that communicated the pragmatic function of contrary to expectation, while the Caribbean participants did not produce any utterances that communicated the pragmatic function of contrastive verb. We believe that these differences are due to accidental gaps based on the relative infrequency of these types of tokens.

Although we coded for these five pragmatic functions these results are difficult to interpret and not revealing of a strong relationship to word order. In order to gain insights into the connection between pragmatic function and word order we chose to use a simpler distinction, a two-variant variable for pragmatic function. The first factor is pragmatically neutral which includes utterances that convey information while the other factor, pragmatically marked, includes utterances that convey the pragmatic functions of highlighted adjective, subject topic contrast, contrary to expectation and contrastive verb. In doing so we find that pragmatically neutral utterances tend to appear in canonical word order significantly more than utterances that communicate pragmatically marked functions. The results for word order by pragmatic function for the entire corpus appear in Table 5.16 below.

Table 5.16			
<i>Cross-Tabulation of Word Order by Pragmatic Function</i>			
Pragmatic Function	N	Word Order	
		Canonical order	Marked Orders
Pragmatically Neutral	331	270 (82%)	61 (18%)
Pragmatically Marked	93	47 (51%)	46 (49%)
Total	424	317 (75%)	107 (25%)
$\chi^2 = 35.432$ $df=1$, $p < .01$			

We found the same significant relationship between pragmatic function and word order in both generational and regional groups. That is, among all four groups, differences in pragmatic function were related to differences in word order (see Tables 5.17, 5.18, 5.19, and 5.20).

Table 5.17			
<i>Cross-Tabulation of Word Order by Pragmatic Function, LARNC</i>			
Pragmatic Function	N	Word Order	
		Canonical order	Marked Orders
Pragmatically Neutral	178	127 (71%)	51 (29%)
Pragmatically Marked	55	26 (47%)	29 (53%)
Total	233	153 (66%)	80 (34%)
$\chi^2 = 9.761$ $df=1$, $p < .01$			

Table 5.18			
<i>Cross-Tabulation of Word Order by Pragmatic Function, NYR</i>			
Pragmatic Function	N	Word Order	
		Canonical order	Marked Orders
Pragmatically Neutral	153	143 (94%)	10 (6%)
Pragmatically Marked	38	21 (55%)	17 (45%)
Total	191	164 (86%)	27 (14%)
$\chi^2 = 33.518$ $df=1$, $p < .01$			

Table 5.19			
<i>Cross-Tabulation of Word Order by Pragmatic Function, Highlanders</i>			
Pragmatic Function	N	Word Order	
		Canonical order	Marked Orders
Pragmatically Neutral	187	142 (76%)	45 (24%)
Pragmatically Marked	52	20 (38%)	32 (62%)
Total	239	162 (68%)	77 (32%)
$\chi^2 = 26.164$ $df=1$, $p < .01$			

Table 5.20			
<i>Cross-Tabulation of Word Order by Pragmatic Function, Caribbeans</i>			
Pragmatic Function	N	Word Order	
		Canonical order	Marked Orders
Pragmatically Neutral	144	128 (89%)	16 (11%)
Pragmatically Marked	41	27 (66%)	14 (34%)
Total	185	155 (84%)	30 (16%)
$\chi^2 = 12.464$ $df=1$, $p < .01$			

The data in the tables above indicate that there are frequency differences across both the generations and the regions. Considering Tables 5.19 and 5.20, both generations favor canonical word order, but the preference for canonical word order in pragmatically neutral sentences is much stronger among the NYR than among the LARNC. Considering Tables 5.21 and 5.22, we see that both regions favor canonical word order, but the preference for canonical order for pragmatically neutral function is much stronger among Caribbeans.

In order to better understand the frequency differences we see in the tables above we need to look specifically at how each of these pragmatic functions was communicated. In other words, we cannot simply examine the relationship between word order and pragmatic function by looking at all the tokens in the corpus together in the way we have for the other internal and external the predictor variables. Instead we need to look at word order variability by immigrant generation and geographic region for each pragmatic function. In this way, the variable word order of a pragmatically neutral utterance is analyzed separately from the variable word order of an utterance that communicates one of the four pragmatically marked functions. In the next four sections below we will discuss word order variability for each pragmatic function in the corpus.

2.11.1 Word order expressing the pragmatically neutral function of conveying information.

Following Ocampo (2002) we would expect utterances that are pragmatically neutral to appear in the canonical word order. As we saw in the tables above, this was the case for utterances produced by both regional and generational groups. Tables 5.21 and 5.22 show the word orders for pragmatically neutral utterances by generation and by region.

Table 5.21			
<i>Cross-Tabulation of Word Order by Generation for Pragmatically Neutral Utterances</i>			
Generation	N	Word Order	
		Canonical Order	Marked orders
LARNC	178	127 (71%)	51 (29%)
NYR	153	141 (94%)	10 (6%)
Total	331	270 (82%)	61 (18%)

Table 5.22			
<i>Cross-Tabulation of Word Order by Region for Pragmatically Neutral Utterances</i>			
Region	N	Word Order	
		Canonical Order	Marked orders
Highlanders	187	142 (76%)	45 (24%)
Caribbeans	144	128 (89%)	16 (11%)
Total	331	270 (82%)	61 (18%)

There are a few quantitative differences across both the regional groups and the generational groups. In pragmatically neutral utterances, whose overall tendency is toward canonical word order, the NYR participants used this order more often than the LARNC, and the Caribbean participants more often than the Highlanders. Although we cannot subject these differences to a significance test, it is clear that the Caribbean participants, and the NYR in particular, had an overwhelming preference for canonical word order when the utterance was pragmatically neutral.

2.11.2 Word order expressing the pragmatic function of Highlighted Adjective.

An adjective is considered highlighted when it is the focus of the utterance. In the corpus, there were 66 occurrences of utterances that communicated the pragmatic function of highlighted adjective. The results for word order by generation and region appear in Tables 5.23 and 5.24 below.

Table 5.23			
<i>Cross-Tabulation of Word Order by Generation for Utterances with a Highlighted Adjective</i>			
Generation	N	Word Order	
		Canonical Order	Marked orders
LARNC	40	16 (40%)	24 (60%)
NYR	26	10 (38%)	16 (62%)
Total	66	26 (39%)	40 (61%)

Table 5.24			
<i>Cross-Tabulation of Word Order by Region for Utterances with a Highlighted Adjective</i>			
Region	N	Word Order	
		Canonical Order	Marked orders
Highlanders	40	11 (28%)	28 (72%)
Caribbeans	26	15 (56%)	12 (44%)
Total	66	26 (39%)	40 (61%)

Both generational cohorts and the Highlanders preferred marked orders to convey this pragmatic function. The Caribbean participants, however, did not prefer marked orders for this pragmatic function. Although fewer utterances appeared in canonical word order for this function than for pragmatically neutral utterances, the Caribbean participants were the only group that still used more canonical word order than marked orders for this pragmatic function.

2.11.3 Word order expressing the pragmatic function of Contrastive Subject.

In the corpus there were 26 occurrences of an utterance with a contrastive subject. Recall that a subject is considered contrastive when it is presented in opposition to one or several

possible alternatives in the discourse. The word orders and their frequencies by region and generation appear in Tables 5.25 and 5.26 below.

Table 5.25			
<i>Cross-Tabulation of Word Order by Generation for Utterances with a Contrastive Subject</i>			
Generation	N	Word Order	
		Canonical Order	Marked orders
LARNC	14	10 (71%)	4 (29%)
NYR	12	11 (92%)	1 (8%)
Total	26	21 (81%)	5 (19%)

Table 5.26			
<i>Cross-Tabulation of Word Order by Region for Utterances with a Contrastive Subject</i>			
Region	N	Word Order	
		Canonical Order	Marked orders
Highlander	12	9 (75%)	3 (25%)
Caribbean	14	12 (86%)	2 (14%)
Total	26	21 (81%)	5 (19%)

Although we had few data points for this pragmatic function, we can see that the NYR and Caribbean participants favored canonical word order more than did the LARNC and the Highlanders. It is important to note, though, that these results are not entirely reliable, as we have so little data to draw upon. This is also the case for the next two pragmatic functions.

2.11.4 Word order expressing the pragmatic functions of Contrary to Expectation and Contrastive Verb.

Only one case of an utterance with the pragmatic function of contrary to expectation occurred in the corpus. The token, produced by a Caribbean LARNC, occurred in the marked word order Adjective-Copula-Subject (*Feliz es Nicolás* ‘lit. Happy is Nicolas,’ ‘Nicolas is happy’). There was only one token with the pragmatic function of contrastive verb and this token

was produced by a Highlander LARNC. This token occurred with the marked word order Copula-Subject-Adjective (*Es Nicolás feliz* ‘lit. Is Nicolas happy,’ ‘Nicolas is happy’). Since these two tokens were produced by LARNC, and the NYR participants did not produce any tokens of these types, we have no way of comparing the two generational cohorts. In the next section below we will look more explicitly at differences across the two generational and regional cohorts using a different statistical test, a logistic regression.

3 Logistic regressions

Our analyses in this chapter so far have relied exclusively on contingency tables or Cross-Tabulations with a Chi-Square test for significance. However, Cross-Tabulations only allow us to explore the relationship between word order and one predictor variable at a time. To examine the effect of several predictor variables upon one nominal outcome variable we can perform a logistic regression. The logistic regression predicts the outcome of a nominal dependent variable using several nominal or continuous predictor variables and determines which of them account for the most variance. Like we did in the previous chapter, we can use the results from the logistic regression to establish a hierarchy of predictor variables based on their effect on word order.

We performed five logistic regressions. The first regression included the tokens from all the participants. For the second regression we separated the participants by generational cohorts and for the third by regional cohorts, in order to determine whether the factors that condition word order were similar across the groups. For the fourth and fifth regressions we separated the participants by both generation and region. We only included in the logistic regression those variables that based on the Chi-Square tests in the previous sections of this chapter were known

to have a bivariate significant relationship with word order¹⁵. For all of the regressions we used the two-level word order dependent variable (that is, canonical versus marked order).

Additionally we used the two-level pragmatic function variable (pragmatically neutral versus pragmatically marked functions); the two-level subject givenness variable (given subject versus new or implied subject); and the two-level number ending of the verb variable (singular verbs versus plural verbs)¹⁶.

A logistic regression calculates the amount of variance accounted for by the variables in the regression, expressed as an R^2 ('R square') value. The regression also calculates the relative weight or strength of association between each independent variable and the dependent variable using a Wald Value. The variable with the highest Wald value accounts for the largest amount of word order variance when compared to all the other variables entered in the regression. A second value, the Exponential B value, also allows us to determine both the weight and the direction of the relationship between each of the factors of the independent variables and the dependent variable. For the present study the variables are coded so that a factor with an Exponential B value greater than one favors marked word order and a factor with an Exponential B value lower than one favors canonical word order. In this way we can create two hierarchies, a variable hierarchy based on Wald values and a factor hierarchy based on Exponential B values.

The variable hierarchy for the first logistic regression including all 424 tokens can be found in Table 5.27. Following convention, significant results are marked with asterisks; non-significant results are not marked¹⁷.

¹⁵ Variables included in the logistic regression were selected based on the Chi-Square tests done in section 2 of this chapter. TMA of the verb was not included in the regression analysis because of frequency issues with three out of its five factors. Four variables were thus included: Number ending of the verb, type of subject NP, pragmatic function, and subject givenness.

¹⁶ Because pragmatic function and number ending of the verb had one or two factors with a low frequency we chose to use the two or three-level versions of these variables as we did in the previous sections which collapse the factors into a fewer number of factors, thus eliminating the frequency issue. For the subject givenness variable, we used the two-level variable following Ocampo (2002).

¹⁷ In the tables below, a double asterisk indicates $p < .01$, a single asterisk indicates $p < .05$, and a superscript a indicates $p < .10$.

Table 5.27		
<i>Variable Hierarchy for Variables Predicting Marked Word Order</i>		
Order	Variable	Wald Value
1	Type of Subject NP	95.081**
2	Pragmatic Function	51.538**
3	Subject Givenness	1.061
4	Verb Number Ending	0.337
R ² = .606		N = 424

Table 5.30 shows that only two variables significantly account for 60 percent of the variance ($R^2 = .606$) in word order for all the participants: type of subject NP and pragmatic function. The type of subject NP accounted for more variance than the pragmatic function.

As mentioned earlier, the variable hierarchy in Table 5.27 above tells us about the importance of the type of subject NP and the type of pragmatic function in predicting word order; but it does not tell us which type of NP or which type of function favors the marked orders. That is left for the factor or constraint hierarchy, given in Table 5.28 below, and which, as did the variable hierarchy, covers every token in the corpus.

Table 5.28		
<i>Factor Hierarchy for Internal Variable Factors Predicting Marked Word Order</i>		
Order	Factor	Exp (B)
1	Clausal-Infinitival Subject NP	18.259**
2	Pragmatically Marked	3.734**
3	New or Implied Subject	1.187
4	Singular Verb	1.132
5	Plural Verb	0.883
6	Given Subject	0.842
7	Complex Subject NP	0.323**
8	Pragmatically Neutral	0.268**
9	Simple Subject NP	0.169**
		N = 424

Recall that an Exponential B value greater than one favors marked order whereas a value lower than one favors canonical order. We can see from our results above that an utterance with a clausal or infinitival subject NP significantly favors marked order whereas one with a complex or simple subject NP significantly disfavors it, i.e., favors canonical word order. These results are consistent with those from our bivariate results. Similarly, a pragmatically neutral utterance significantly favors canonical word order whereas a pragmatically marked utterance significantly favors marked word order, which also aligns with our bivariate results. The other factors, although not significant, align with the results from our bivariate analyses. Utterances with a singular verb or a new or implied subject favored no-canonical word order, while those with a given subject or a plural verb favored canonical word order.

A logistic regression allows us to compare the variable and factor hierarchies across two groups in order to determine whether the conditioning of a dependent variable is similar across the two groups. To do so we created variable and factor hierarchies for each generation and regional group. The variable hierarchies for each regional group are in Table 5.29 below.

Table 5.29					
<i>Variable Hierarchies by Region for Internal Variables Predicting Marked Word Order</i>					
Highlanders			Caribbeans		
Order	Variable	Wald Value	Order	Variable	Wald Value
1	Type of Subject NP	58.179**	1	Type of Subject NP	30.475**
2	Pragmatic Function	37.460**	2	Pragmatic Function	14.484**
3	Verb Number Ending	1.451	3	Verb Number Ending	0.311
4	Subject Givenness	1.433	4	Subject Givenness	0.041
R ² = .637		N = 239	R ² = .556		N = 185

Based on the results in Table 5.29 we see that both regional groups are the same. For both regional groups the type of subject NP and the pragmatic function of the utterance were the two strongest predictors of word order with the type of subject NP being the strongest for both groups.

The other two variables were not significant predictors of word order. The variables account for roughly similar amounts of total variance in word order in the two regions, about 63 percent of word order variance in the Highlanders and 55 percent in the Caribbeans. In order to compare how the two groups behave in more detail we turn to the factor hierarchies in Table 5.30 below.

Highlanders			Caribbean		
Order	Factor	Exp (B)	Order	Factor	Exp (B)
1	Clausal-Infinitival Subject NP	15.844**	1	Clausal-Infinitival Subject NP	25.699**
2	Pragmatically Marked	4.408**	2	Pragmatically Marked	3.348**
3	Singular Verb	1.415	3	Plural Verb	1.206
4	New or Implied Subject	1.289	4	Given Subject	1.063
5	Given Subject	0.776	5	New or Implied Subject	0.941
6	Plural Verb	0.707	6	Singular Verb	0.829
7	Complex Subject NP	0.289**	7	Complex Subject NP	0.380*
8	Pragmatically Neutral	0.277**	8	Pragmatically Neutral	0.299**
9	Simple Subject NP	0.219**	9	Simple Subject NP	0.102**
N = 239			N = 185		

Although there were some differences across the two regional groups, there were more similarities than differences and, in particular, both groups are the same with regard to the significant factors. For both groups the following factors significantly favored marked word order: clausal-infinitival subject NP and pragmatically marked functions. Similarly, an utterance with a complex or simple NP or a pragmatically neutral utterance significantly favored canonical word order for both regional groups.

There were also a few differences, all in the non-significant factors. An utterance with a singular verb or a new or implied subject favored marked word order among Highlanders but favored canonical word order among Caribbean participants. Whereas an utterance with a plural

verb or a given subject favored canonical word order among Highlanders but favored marked word order among Caribbean participants.

Although the similarities and differences described above provide us specifics regarding word order variability across the two regional groups, these groups include speakers from both generational cohorts. That is, the Highlander group includes LARNC Highlanders and NYR Highlanders while the Caribbean group includes LARNC Caribbean participants as well as NYR Caribbean participants. We saw in sections 2.1 and 2.2 in Chapter Four that generation is also significantly related to word order. It would follow, then, that we perform a logistic regression comparing the two generational cohorts as well. Below in Table 5.31 we find the variable hierarchies for the two generational cohorts.

Table 5.31					
<i>Variable Hierarchies by Generation for Internal Variables Predicting Marked Word Order</i>					
LARNC			NYR		
Order	Variable	Wald Value	Order	Variable	Wald Value
1	Type of Subject NP	62.913**	1	Type of Subject NP	24.860**
2	Pragmatic Function	23.943**	2	Pragmatic Function	21.771**
3	Subject Givenness	1.407	3	Subject Givenness	0.007
4	Verb Number Ending	0.505	4	Verb Number Ending	0.006
R ² = .523		N = 233	R ² = .166		N = 191

Table 5.31 illustrates clear similarities in the variables that condition word order across the two generations. For both groups the pragmatic function of the utterance as well as the type of subject NP significantly conditioned word order. There were two differences of import. First, the type of subject NP, which was the strongest predictor variable among both groups, is a much stronger predictor among LARNC than among NYR, with a much higher (more than twice as big) Wald Value. Secondly, we see a large difference in the variance accounted for in the two

regions. The variables account for 52 percent of total variance in word order for the LARNC but only 17 percent among the NYR.

The reader may now expect factor hierarchies comparing the two regional groups. But when we computed these factor hierarchies we obtained Exponential B values that were inconsistent and difficult to interpret. The Exponential B value for Clausal-Infinitival NP was unusually high and may not be accurate, and the statistical program returned significant results for the factors related to the third-ranked variable, but not the first and second ranked variables, which is also a suspect result. We obtained the same outcome when we calculated the constraint and variable hierarchies comparing the generations within their own regions, i.e., Highlander LARNC to Highlander NYR participants and Caribbean LARNC to Caribbean NYR participants. We concluded that the data do not provide us enough material with which to provide a multivariate answer especially for the factors of the syntactic properties of the subject NP variable. As a result, we have chosen not to include the variable and constraint hierarchies comparing the generations within their own regions or the factor hierarchies comparing the two generations. Instead we will draw upon the five reliable hierarchies and our bivariate results to address our research questions.

4 Summary and discussion

In order to address our research questions, we begin this section with a discussion and summary of the overall results (research question d) for the entire corpus. We then examine our results by region (research question e) and by generation (research question f).

4.1 Overall results.

The most common word order for the entire corpus as well as for each of the regional and generational groups was the canonical word order, followed by Copula-Adjective-Subject (*Es*

feliz Nicolás ‘lit: Is happy Nicolas,’ ‘Nicolas is happy’). These results are similar to those from other studies on Spanish word order in utterances with copular verbs (Ocampo, 2002).

We first examined our data using Chi-Square tests (bivariate analyses) to determine whether there were significant relationships between the linguistic predictor variables and word order (research question d). We found that several of our potential internal predictor variables were in fact not predictors, that is, were not significantly related to word order. These variables are listed below.

Independent (predictor) variables that did not have a predictive statistical effect on word order

- Presence of adverb in the predicate
- Affirmative or negative utterance
- Type of copular verb
- Length and type of adjectival phrase
- Presence of English elements

The first four of these non-significant variables will not be discussed here. We address the last variable because the results approached significance for one of the generational groups.

Presence of English elements. We had expected differences in utterances with or without English elements since there is a known relationship between code-switching and the syntax and prosody of an utterance (Durán Urrea, 2009). We also know that intrasentential code-switching is not arbitrary, but is constrained by syntactic principles (Belazi, Rubin & Toribio, 1994; Poplack, 1980). We had posited that utterances with English elements would occur more frequently in canonical word order due to English’s more fixed word order. In section 2.4 we saw that this expectation was fulfilled, that is, utterances that had English elements indeed had an overall higher rate of canonical word order than those without English elements. This pattern

occurred in both regional and generational groups, although it only approached significance among the LARNC. We believe that the reason these results were not significant was because very few of the utterances in the corpus included English elements (only 11). If our corpus were larger, or if more utterances with English elements had fallen inside the envelope of variation, it is very probable that this result would have been significant among the NYR, Caribbean, Highlanders and for the entire corpus.

There were two other variables that were significantly related to word order but which we will not discuss at length:

- TMA of the verb
- Number in third-person verbs

TMA of the verb. We found that verbs in the indicative mood appeared more frequently in canonical word order while those in the conditional appeared most often in marked word order. However, we were unable to compare this variable across generations because the NYR did not produce tokens in the conditional yielding several zero-values in the tables. Moreover, this bivariate result did not survive the multivariate analysis. That is, this variable was discarded as non-significant when we entered it into the regression along with the other variables. For this these reasons, our discussion of this variable below will be brief.

Number in third person verbs. Similarly, number in third-person verbs was a property of the verb that was significantly related to word order in the bivariate analysis. We found that utterances with verbs that were plural appeared more frequently in canonical word order than those with singular verbs. But as was the case with TMA of the verb, these results were not confirmed by our regression analyses. And, again as in the case of TMA, we found no

generational differences and few regional differences with regard to this variable. For these reasons, this variable too is discussed only briefly here.

There is two possible explanation for why the favorings involving verb TMA and Verb Number are as we find them and not in the opposite direction. First, as we saw above, we found an increased use of canonical word order among the Caribbean participants. As we will explain below, we also find an increased use of subject personal pronouns among this population (Otheguy & Zentella, 2012; Raña-Risso, 2013). According to Morales (1989, 1997, 1999) this abundance of subject pronouns among the Caribbeans favors a ‘fixing’ of the canonical word order in these lects. If there were a relationship between the rate of overt subject personal pronouns and the rate of canonical word order, we would expect to find more canonical word order among verbs that also favor overt subject personal pronouns. However, we find somewhat contradictory results. In their comprehensive study on subject personal pronouns in the Spanish in NYC (using the same corpus that we used for the present study) Otheguy and Zentella (2012) found that singular verbs and verbs in the imperfect significantly favored overt subject pronoun expression (pp. 163-164). Additionally, other studies on Spanish personal pronouns indicate that there is an increased use of overt pronouns with verb tenses that are morphologically ambiguous such as the conditional or the imperfect (Lipski, 1996). This is related to the phonological erosion of verb endings among these types of verbs, a fact that makes verbs lack person identification. Although we would expect this to lead to a tendency to say the subject first, whether this subject is nominal or pronominal, for functional-communicative reasons, we actually find the opposite results. Ambiguous verbs, like those in the conditional, occur more often in marked word order than those verbs that are not ambiguous, like preterit or present indicative verbs. Although this observation is somewhat tenuous, it appears that verb properties

that are associated with null subject pronouns are also associated with canonical word order. Secondly, our data for singular verbs may be skewed by the fact that all of the utterances with clausal-infinitival subjects had singular verbs and an overwhelming amount of these (88 percent) appeared in marked word order.

There were three variables that did have a predictive statistical effect on word order and which will be the focus of our discussion. These are listed below.

Independent (predictor) variables significantly associated with word order

- Discourse pragmatics: Informational status of the subject NP
- Syntactic properties of subject NP
- Discourse pragmatics: Function of the utterance

Discourse pragmatics: Informational status of subject NP. The discourse properties of the subject NP related to informational status proved to be fruitful variables in our analysis. For the entire corpus as well as for each generational and regional group we found clear differences in word order when the subject was given versus when the subject was new or implied. Although utterances with new or implied subjects in our corpus appeared most often in canonical word order, they did so less often than utterances with given subjects. In this way, our results align with the extant literature in that new subjects tend to appear more often in final position than given subjects in order to facilitate processing (Arnold et al., 2000; Bentivoglio, 2003; Bolinger, 1955). However, our results regarding subject givenness as a predictor of word order were not confirmed by our regression analysis. This variable was not a significant predictor of word order among any of the regional or generational groups, nor for the entire corpus. Although the tendency to put new information in final position is not a language specific trend (Arnold et al., 2000; Bentivoglio, 2003; Bolinger, 1955), and despite our bivariate results being significant for

this variable, it appears that this variable and its factors do not significantly account for word order variability in our corpus.

Syntactic properties of the subject NP. The syntactic type of Subject NP also proved to be an important variable for predicting word order. For the entire corpus as well as for each regional and generational group we found that utterances with simple subjects appeared more often in canonical word order than utterances with clausal-infinitival subjects, which appeared most often in marked word orders. This finding, as we saw, needs to be considered in light of the existing correlation between syntactic complexity and constituent length. Since we found a relationship between NP type and NP length we can also interpret these results to mean that constituent length as a predictor of word order is operative among the participants in our corpus. That is, the longer the subject NP, the more likely it is to appear in post-verbal position (marked order). It would be strange to find that this variable was not a significant predictor of word order considering that the tendency for longer constituents to appear in utterance final position is also not a language specific phenomenon. In fact, it has been argued that postponing longer constituents to utterance final position facilitates processing and is thus not a language specific phenomenon (Arnold et al., 2000; Bentivoglio, 2003; Bentivoglio, 2003; Bolinger, 1955). However, we also find that complex subjects, which do not differ significantly in length from clausal-infinitival subjects, appear more often in canonical word order than in marked word order. In fact, clausal-infinitival subjects tend to appear in marked word order even when these are that are only one word (e.g. *cantar* ‘to sing’ or ‘singing’). Thus it would be more useful to focus our discussion on the relationship between NP type and word order instead of the discussion of the relationship between NP length and word order. For example, although there is a nine percent point decrease in canonical word order compared to simple subjects, the

difference between complex subjects and clausal-infinitival subjects is much greater (68 percentage points). It is clear, then, that subject complexity is an independent predictor of word order, and that there is more than just a length difference between complex subjects and clausal-infinitival subjects. In fact, the large difference in canonical word order across among clausal-infinitival subjects and complex and simple subjects may be attributable to a category effect. Specifically, it is possible that the infinitival subjects are not nominal¹⁸. If so, there is a clear opposition between CP and NP subjects, the former occurring in the marked word order significantly more often than the latter, regardless of length. These results are also confirmed by our multivariate analyses. The type of subject NP was the most important variable in our regression analyses. Overall, utterances with clausal-infinitival subjects favored marked word order while those with simple or complex subjects favored canonical word order. For the entire corpus as well as for both generational and regional groups this variable and its factors accounted for the most word order variability in the regressions.

Discourse pragmatics: Function of the utterance. Our final variable, pragmatic function of the utterance, was also significant in both the bivariate and multivariate analyses. In general, pragmatically neutral utterances favored canonical word order while utterances that communicated the pragmatically marked functions favored the marked word orders. This result is consistent with our expectations that speakers would exploit word order resources to signal a variety of pragmatic functions (Ocampo, 2002), i.e. produce marked word order, for utterances that were not pragmatically neutral.

Although many of our results for these variables were similar when we looked at the regions and generations separately, we did find some important differences across the

¹⁸ Following this argument we would add that they are always contained in CPs, hence not a category of 'N'.

generations and regions. In the following sections we discuss and compare our results across generations and regions in order to address our two final research questions (e and f).

4.2 Results and discussion by region.

We found several quantitative differences between regional groups with regard to word order. Recall that for the corpus as a whole we found that both the NP's informational status and the syntactic type of NP were significantly related to word order. Across regional groups the Caribbean participants used canonical word order with given subjects more often than Highlander participants by 12 percentage points. This increased use of canonical word order among the Caribbean participants was not limited to this context. In general, Caribbean participants produced more utterances with canonical word order than the Highlanders by 18 percentage points. This difference is expected and well documented in the extant literature. The preference for canonical word order in the Caribbean lects has been attributed in part to their increased use of overt subject pronouns, a finding that is also confirmed for this corpus (Raña-Risso, 2013). According to Morales (1989, 1997, 1999) the abundance of the subject pronouns favors a 'fixing' of the canonical word order in these lects. This argument, however, would contradict our previous argument—for the increase in canonical word order with plural verbs and conditional verbs—that verbs that tend to occur with more overt subject personal pronouns favor marked word order. Another possible explanation for the increased preference for canonical word order is this group's increased contact with English, as well as their unique political ties to the United States, especially among the Puerto Ricans (Lynch & Klee, 2009, p 200; Zentella, 2000, p. 139).

4.3 Results and discussion by generation.

We also found several differences between generational groups with regard to word order. There were three categorical differences (i.e., differences of 100 percent to zero) across the groups. First, the LARNC Caribbeans produced three different marked word orders whereas the other groups only produced two. Specifically, a LARNC Caribbean participant produced an utterance with Adjective-Copula-Subject order (*Feliz es Nicolás* ‘lit. Happy is Nicolas,’ ‘Nicolas is happy’), while none of the other groups did.

The second categorical difference is that the NYR participants did not produce any tokens for two of the five pragmatic functions (contrastive verb and contrary to expectation). This last observation would appear to support the literature that maintains that second generation speakers, like our NYR, have a decreased ability to communicate the wide breadth of pragmatic functions that is expressed by their first generation counterparts. It is important to point out that although no NYR participants produced an utterance with these two pragmatic functions, we cannot interpret this to mean that NYR participants cannot produce them at all. We believe that the reason for this is because of these tokens’ relative infrequency in the speech of the population. Out of 424 utterances only one utterance in the corpus had the pragmatic functions of contrary to expectation or contrastive verb. Given that 25 participants and dozens of hours of discourse produced only one token of each type, we can conclude that these types of utterances are very rare. We thus attribute both of these categorical differences to frequency issues. It is very possible that the NYR participants would have produced an utterance with these two pragmatic functions if they had produced more discourse, or if we had had interviewed more participants. It is also possible that the discourse contexts simply never called for these token types.

Finally, a third categorical difference is that for pragmatically neutral utterances, a NYR Caribbean participant produced a marked word order that was not found among the NYR

Highlanders nor among any of the LARNC. Although the lack of this type of token among LARNC participants may again be a result of low overall frequency, it is also possible that the NYR are beginning to produce novel orders for this utterance type. If this were the case, however, it would be difficult to argue that this word order (Copula-Subject-Adjective *Es Nicolás feliz* ‘lit. Is Nicolas Happy,’ ‘Nicolas is happy’) was modeled after English word order, given that the canonical word order for English utterances of this type is Subject-Copula-Adjective. For these reasons we attribute this categorical difference to frequency issues as well, and do not interpret this difference, as well as the two previously mentioned, as actual categorical differences or significant changes across generations.

There were also several quantitative differences across the groups (research questions e and f). Recall that for the corpus as a whole we found that both the NP’s informational status and the type of NP were significantly related to word order. For the entire corpus as well as for each regional group and the LARNC we found a decreased use of canonical word order as the complexity of the subject NP increases. Although we found that simple subjects favored canonical word order while clausal-infinitival subjects favored marked word order among the NYR, we also found that they used canonical word order categorically for complex subjects. Although we know that constituent length is only somewhat related to word order, we calculated the number of words for complex subjects for both LARNC and NYR and ran an ANOVA to determine if there was a significant difference across the groups that might account for the categoricity among the NYR. We found no significant differences ($F = 1.288, p = .261$) in word order across the two generations. Since our factor hierarchies across the generations yielded unreliable results, we are unable to confirm this result using a multivariate analysis. We do know, however, that for the variable hierarchy this variable was the strongest predictor of word order

for both generations. It is possible that this result is due to the relative low frequency of data for complex subjects versus simple subjects (17 tokens versus 159 tokens), although we did not find the same issue for the clausal-infinitival utterances, which also had low frequency (15 tokens).

Across generations we also found that NYR participants used marked word orders with new or implied subjects less frequently than the LARNC by 17 percentage points. This increased use of canonical word order among the NYR was not limited to this context. In general, the NYR produced more utterances with canonical word order than the LARNC by 20.2 percentage points. This difference is expected and well documented in the extant literature. The word order differences across our generations are of special interest because they appear to support the findings of several studies on Spanish in the US that suggest that second and third-generation bilinguals are differently sensitive to the discourse-pragmatic constraints that govern word order in Spanish, and by being differently sensitive to these constraints, may have lost the ability to communicate a variety of meanings using word order (Silva-Corvalán, 1994; Zapata, Sánchez, & Toribio, 2005). The differences across our generations are similar to those reported in other studies on word order in US Spanish (Silva-Corvalán, 1994; Zapata, Sánchez & Toribio, 2005). Recall that Silva-Corvalán (1994), in her sociolinguistic study of 50 Mexican-American bilinguals in Los Angeles, found evidence for what she calls ‘obligatory SVX order’ among the second and third-generation speakers in her corpus, in comparison to the pre-contact lects (popular forms of Spanish in Mexico), in which word order is dependent on many variables, including the discourse-pragmatic function of the utterance. Specifically, the author found an increased rate of preverbal subject NP’s and subject pronouns as the speakers’ contact with English increased. The author argues that this increase in SVX order reflects the “consequence of processes of loss of semantic-pragmatic constraints on preverbal subject placement” (p. 144) that

can be attributed to contact with the bilingual's dominant language, English, which has a more rigid SVO order. Similarly, in their experimental study on the production and interpretation of sentences with ergative and unergative verbs by 24 heritage speakers of Spanish, Zapata, Sanchez, and Toribio (2005) found evidence of convergence toward English word order. Specifically, their second and third-generation bilingual participants preferred SV order in Spanish for unaccusative subjects in a full focus context although the preference in the pre-contact lect is for VS order. In other words, the heritage speakers produced word orders that were not target-like compared to their first generation counterparts. The authors attribute the increased incidence of SV order to convergence with English and a reduction in the syntactic options that allow the speakers to communicate discourse-semantic information.

One way to interpret our NYR's increased use of canonical word order is what Silva-Corvalán calls 'simplification', which is one of the many linguistic phenomena have been documented in US Spanish US (Silva Corvalán, 1994, 1995, 2001). Silva-Corvalán (1994) defines simplification as:

The higher frequency in the use of a form X in context Y (i. e., generalization) at the expense of a form Z, usually in competition with and semantically closely related to X, where both X and Z existed in the language prior to the start of simplification. Thus, X is an expanding form while Z is a *shrinking/contracting* form. If simplification reaches completion, its final outcome is reduction or loss of forms and elimination of alternatives, i. e., a *simplified* system with fewer forms and possibly, though not necessarily, loss of meanings (p. 3).

According to this definition, we would interpret canonical word order as form X in any one of our pragmatic functions, context Y, and the marked word orders as form Z. Considering that our

LARNC use both forms in all the contexts (pragmatic contexts) we can argue that both forms existed in the language before the start of simplification—and the contact situation. As the process of simplification unfolds we would expect to find among the second generation, our NYR, an increased use of canonical word order (expansion of form X) with a decreased use of marked word order (shrinking of form Z). This is, in fact, precisely what we find in our data.

It is important to highlight an important point that Silva-Corvalán makes in her definition of simplification, which is the idea that an increase in form X at the expense of form Z does not necessarily imply a loss of meanings. As we saw above, however, the extant literature, including Silva-Corvalán, has arrived at another interpretation arguing that second and third-generation Spanish-English bilinguals in the US, by being differently sensitive to the discourse-pragmatic constraints that govern word order in Spanish, have lost the ability to communicate these meanings (Silva-Corvalán, 1994; Zapata, Sánchez & Toribio, 2005). In other words these researchers argue that these bilinguals have fewer discourse strategies to communicate the variety of pragmatic functions that first generation bilinguals do using word order, thus resulting in a loss of communicative resources. We believe, however, that in order to make such an assertion we would first need to determine two things: first, we would need to compare the discourse pragmatic and internal constraints that govern word order across the two generations; and secondly we would need to determine whether there is another resource with which these speakers can maintain meaning. That is, do they compensate for the loss of the discourse-pragmatic constraints that usually govern word order in the pre-contact lects by using other resources, such as prosody? In the next chapter we will address the latter question. So far, we have demonstrated that there are few differences in the discourse-pragmatic constraints that govern word order. Below we will address the former question using the bivariate and

multivariate results for the variable pragmatic function of the utterance presented earlier in this chapter.

One of our strongest arguments against the claim that the second-generation Spanish-English bilinguals in our corpus (NYR) have lost the ability to communicate certain meanings through the loss of the discourse-pragmatic constraints that govern word order is our data for the variables and factors related to the pragmatic function of the utterance. In general, pragmatically neutral utterances favored canonical word order while utterances that communicated the pragmatically marked functions favored marked word order. This result is similar to several studies on word order and pragmatic function (Ocampo, 1995a, 1995b, 1995c, 2002, 2005). The most noteworthy argument against claims that second and third-generation bilinguals have fewer expressive devices to communicate pragmatic intent is that we found the same results across our generations (as well as across regions). The results from our bivariate analyses were also confirmed by our multivariate analyses. For the entire corpus as well as for each generational and regional group the pragmatic function of the utterance was the second strongest predictor of word order. Additionally, among both generational groups, the order of the constraint and factor hierarchies was identical for this variable, and the factors related to the pragmatic function of the utterance were identical in the direction of their favorings. These data appear to refute the idea that our NYR are losing the capacity to communicate the wide breadth of pragmatic functions that their first generation counterparts communicate using word order variability. The importance of the similarities in constraints across the generations is not only illustrative of structural continuities across the two generations. More important is the fact that one of the two constraints that remains viable in the second generation is precisely the one related to the pragmatic function of the utterance.

However, there was an important difference between the two generations' hierarchies. When we look at the R Square value in Table 5.31, which specifies the amount of variance accounted for by each regression, we see that among the LARNC, the two variables (type of Subject NP and pragmatic function of the utterance) significantly account for 52 percent of the variance in the regression, whereas for the NYR the same two variables only account for 17 percent of the variance. This means that the first-generation predictors are no longer as important among the NYR, and that there are very likely to be other variables accounting for the variance among this group. Most importantly, the large difference in the R Square value across the two generations indicates that there are other variables that account for a larger amount of variance than NP type and pragmatic function of the utterance, the two variables which account for the majority of the variance among our LARNC. As a result, although we maintain that our NYR participants are very much capable of communicating a wide variety of pragmatic functions using word order resources similar to those that their first generation counterparts use, we should note that there appear to be other variables that account for word order variability among our NYR that we have not yet explored. We can conclude, then that our NYR are in fact differently sensitive to the discourse-pragmatic constraints that govern word order in Spanish compared to the LARNC, but we have yet to find evidence that by being differently sensitive to these constraints they have lost the ability to communicate certain meanings. It is possible that the variables are being used to communicate the meanings that the LARNC communicate using word order. One possible variable that may account for some of this variance, and which can also be used to signal discourse-pragmatic information, is prosody. In the next chapter we examine variable prosody for the entire corpus as well as how each regional and generational group uses prosody to communicate each pragmatic function.

CHAPTER 6

THE INTERNAL SUPRASEGMENTAL VARIABLE PROSODY

1. Introduction

This chapter presents the analysis of the internal suprasegmental outcome variable, prosody, which was described in Chapter Three. The previous two chapters focused on examining the relationship between, on the one hand, the internal and external predictor variables and, on the other, the word order outcome variable. As we saw in Chapter Two, we know that there is a relationship between word order and prosody (Büring, 2008; Nava, 2007; Vallduví, 1991; Zagona, 2002). Throughout this chapter, then, we will attempt to determine the specifics of that relationship. Specifically, we aim to determine how prosody is used, with or without word order, to communicate the various pragmatic functions, and whether there are any differences across the immigrant generations or geographic regions with regard to this variable.

The prosody outcome variable that, together with word order, is the focus of this chapter contains five factors or levels. Each factor combines a particular word order with a particular stress placement. That is, the factors of this variable consist of order-prosody packages. We list the factors below. In each factor, the word that receives primary stress is in bold, both on this list and on the tables that follow.

Variable: Prosody

- Subject-Copula-**Adjective**, *Nicolás es feliz* ‘Nicolas is **happy**’
- **Subject**-Copula-Adjective, ***Nicolás** es feliz* ‘**Nicolas** is happy’
- Copula-**Adjective**-Subject, *Es **feliz** Nicolás* ‘lit: Is **happy** Nicolas,’ ‘Nicolas is **happy**’

- Copula-Subject-**Adjective**, *Es Nicolás feliz* ‘lit. Is Nicolas **happy**,’ ‘Nicolas is **happy**’
- **Copula**-Subject-Adjective, *Es Nicolás feliz* ‘lit. **Is** Nicolas happy,’ ‘Nicolas **is** happy’

In this chapter we also examine and follow up on several of the results from our previous chapters. Our data in Chapters Four and Five indicate that there is word order variability across both generations, but that the NYR use canonical word order more frequently than their first generation counterparts. That is, the NYR have less word order variability than the LARNC. We hope to determine whether this reduced word order variability among NYR participants occurs with a concomitant increase of variability in prosody. That is, given that our NYR have less flexible word order in Spanish, it would seem to follow that these speakers are missing one of the resources that are used in non-contact varieties for communicative purposes. As a result, one might think that these bilinguals are simply not expressing or not able to express these communicative objectives, but the point has to be investigated whether they are using other mechanisms, be these completely new mechanisms or relying more on the other resources that are used in the non-contact varieties for these purposes. That is, do they compensate for the loss of these discourse-pragmatic constraints by using other resources, such as prosody? If this is the case, then we can argue that these second generation speakers do not have fewer expressive devices, as some have argued (Silva-Corvalán, 1994; Zapata, Sánchez & Toribio, 2004), but different expressive devices than their first generation counterparts.

Additionally, given that we found variability in word order in Chapter Six, we would also expect to find variability in the order-prosody packages. If we do find variability in the order-prosody packages, especially among the LARNC, this would indicate several things. First, it would highlight the fact that prosody in non-contact forms of Spanish is more flexible than

originally thought (Gabriel, 2007, 2010; Hoot, 2012; Leal-Méndez & Shea, 2012; Mutendam, 2009). The consensus in the extant literature is that Spanish prosody is relatively invariable and fixed, and for this reason Spanish is considered a non-plastic language (Vallduví, 1999). Also, it would indicate that order-prosody variability among the NYR is not necessarily attributable to their increased contact with English. If the NYR modeled their word order and prosody after English we would expect to find very little variability in word order and a lot of variability in prosody. Based on our results in Chapters Four and Five we know that the first expectation is not true. Although we found reduced word order variability among NYR, we did not find categorical use of canonical word order. Our analysis of prosody in this chapter will allow us to determine whether any generational differences in prosody are modeled after English.

In addition to addressing these issues, the analysis of these two variables (prosody and word order) aims at answering the research questions outlined in Chapter One and repeated here below. Only those research questions addressed in the chapter are listed.

Question G: Is there a significant relationship between variable word order and variable prosody in our corpus?

Question H: What differences, if any, are there between the two geographic regions in word order and prosody, either quantitatively or categorically?

Question I: What differences, if any, are there between the two immigrant generations in word order and prosody, either quantitatively or categorically?

Question J: If there are prosodic differences across the generations, are these modeled after the prosody of English?

There are several notable differences between the analyses presented in this chapter and those presented in the previous chapters. First, only 258 of the 424 tokens were coded for

prosody.¹⁹ As a result, we do not have enough data to examine two of the five pragmatic functions (contrastive verb and contrary to expectation). Second, the dependent variable for this chapter, as we see above, does not encompass a single phenomenon such as word order, but embodies order-prosody packages where the order phenomenon and the stress phenomenon are combined. Additionally, since variable prosody and variable word order are used to express pragmatic functions,²⁰ we cannot simply examine the relationship between the two by looking at all the tokens in the corpus together, as we've done in previous chapters. Instead we look at word order and prosodic variability by immigrant generation and geographic region separately for each pragmatic function. In this way, the variable word order and prosody of an utterance that is pragmatically neutral is analyzed separately from the variable word order and prosody of the utterances that communicate each of the four other pragmatic functions. Finally, we note that the term *canonical* has an additional usage in this chapter. Until now, we have spoken of canonical word order. In this chapter, we will also speak of canonical prosody, i.e., canonical stress placement. We will postpone discussion and interpretation of our results until section 7.

2. Pragmatically neutral utterances that convey information only

The word order and stress patterns in the first of the factors listed above are considered canonical for pragmatically neutral utterances. Of the 331 pragmatically neutral utterances described in section 2.10.1 in Chapter Five, 206 were coded for prosody. Only two out of these had marked word order (Copula-Adjective-Subject *Es feliz Nicolás* 'lit: Is happy Nicolas,' 'Nicolás is happy'). The primary stress placement by generation for the 206 utterances is

¹⁹ The reasons for this are explained in section 4.1.2 in Chapter Three.

²⁰ The reasons for this organization are presented in more detail in Chapters Three and Five.

presented in Table 6.1 below²¹. Only three of the five factors for variable prosody were used for this pragmatic function.

Table 6.1				
<i>Cross-Tabulation of Prosody and Word Order by Generation for Utterances With the Pragmatically Neutral Function of Conveying Information</i>				
Word Order and Primary Stress Placement				
Generation	N	Sub Cop Adj	Sub Cop Adj	Cop Adj Sub
LARNC	94	54 (57%)	39 (42%)	1 (1%)
NYR	112	73 (65%)	38 (34%)	1 (1%)
Total	206	127 (62%)	77 (37%)	2 (1%)

While there was a difference of eight percentage points across the generations, we could not use a Chi-Square test to confirm these results because we cannot subject token-level data to a Chi-Square analysis at the participant-level as it would violate the assumption of the independence of the data. As a result, there is no way of knowing whether the differences in this contingency table are statistically significant. However, it appears that the two generational cohorts behave similarly with regard to primary stress placement and word order when the pragmatic function of the utterance is neutral, i.e., conveys information only. Most LARNC and NYR prefer canonical word order and canonical stress to communicate this pragmatic function, although the NYR prefer the canonical word order with canonical prosody slightly more so than the LARNC. Each generation's second and third word order-prosody choices are also similar: the LARNC and NYR use canonical word order with marked stress (stress on the subject) for 42 percent and 34 percent of the utterances respectively, and use marked word order and marked prosody for one percent of the utterances.

²¹ In the tables that present prosody the constituent upon which the primary stress placement falls is also in bold. For example, in Table 7.1 the first prosodic variant is Sub Cop **Adj** which indicates first, that the order of the constituents was Subject-Copula-Adjective and secondly that the adjective in that utterance received primary stress placement.

We find even fewer differences across the regional groups, that is, the choices of order-prosody packages made by Highlanders and Caribbeans to convey pragmatically neutral information are nearly the same (Table 6.2).

Table 6.2				
<i>Cross-Tabulation of Prosody and Word Order by Region for Utterances With the Pragmatically Neutral Function of Conveying Information</i>				
Word Order and Primary Stress Placement				
Region	N	Sub Cop Adj	Sub Cop Adj	Cop Adj Sub
Highlands	105	64 (61%)	40 (38%)	1 (1%)
Caribbean	101	63 (62%)	37 (37%)	1 (1%)
Total	206	127 (62%)	77 (37%)	2 (1%)

Across the different regions it appears that there are no differences in primary stress placement and word order for utterances that are pragmatically neutral. In the sections that follow we will examine whether the two generational and regional cohorts also behave similarly in their primary stress placement when the utterances are not pragmatically neutral.

3. Utterances with the pragmatic function of Highlighted Adjective

Of the 65 highlighted adjective tokens, 41 tokens were coded for primary stress placement. When the adjective in an utterance is highlighted, the canonical order-prosody pattern involves placing the adjective in utterance final position and stressing it. As a result, we can have canonical prosody (primary stress placement on the final constituent) in utterances with marked word order (where the order is not Subject-Copula-Adjective). This is the case, for example, in utterances with the pattern Copula-Subject-**Adjective** (*Es Nicolás **feliz*** ‘lit. Is Nicolas **happy**,’ ‘Nicolas is **happy**’). Alternatively, Canonical prosody can also occur in utterances with canonical word order such as Subject-Copula-**Adjective**’ (*Nicolás es **feliz*** ‘Nicolas is **happy**’). In Tables 6.3 and 6.4 below we present the primary stress placement and word order by generation and region. Again, the participants only used three of the five factors listed above,

although one of these three was different from the three factors that were used to communicate the pragmatically neutral function of conveying information.

Table 6.3				
<i>Cross-Tabulation of Prosody and Word Order by Generation for Utterances With the Pragmatic Function of Highlighted Adjective</i>				
Word Order and Primary Stress Placement				
Generation	N	Sub Cop Adj	Cop Sub Adj	Cop Adj Sub
LARNC	21	6 (29%)	0 (0%)	15 (71%)
NYR	20	7 (35%)	3 (15%)	10 (50%)
Total	41	13 (32%)	3 (7%)	25 (61%)

Table 6.4				
<i>Cross-Tabulation of Prosody and Word Order by Region for Utterances With the Pragmatic Function of Highlighted Adjective</i>				
Word Order and Primary Stress Placement				
Region	N	Sub Cop Adj	Cop Sub Adj	Cop Adj Sub
Highlands	22	2 (9%)	2 (9%)	18 (82%)
Caribbean	29	11 (58%)	1 (5%)	7 (37%)
Total	41	13 (32%)	3 (7%)	25 (61%)

It is important to note that although Table 6.3 shows that LARNC did not produce any tokens with the order-prosody package Copula-Subject-**Adjective** (*Es Nicolás **feliz*** ‘lit. Is Nicolas **happy**’ ‘Nicolas is **happy**), this is true only for tokens that were coded for prosody. Recall from Table 5.25 in Chapter Five that the LARNC produced two tokens with this word order. These tokens were not included in Table 6.3, however, because they were not coded for prosody. As a result, we cannot assume that this is a categorical difference across the two generational cohorts with regard to this word order and stress order-prosody pattern. We simply do not have enough data to make such conclusions. However, there are several interesting observations that we can make based on the data in Tables 6.3 and 6.4. First, all groups except the Caribbean speakers preferred marked word order and prosody (Copula-**Adjective**-Subject *Es **feliz** Nicolás* ‘lit. Is

happy Nicolas,’ ‘Nicolas is **happy**’) while the Caribbean participants preferred canonical word order and prosody for this pragmatic function. Secondly, without exception both regional and generational groups placed the primary stress on the adjective regardless of the adjective’s position in the utterance. Again, while we cannot test for statistical significance, we can see that the two generational groups appear to behave similarly with only a six percent difference in preference for canonical word order, while the two regional groups behave quite differently.

4. Utterances with the pragmatic function of Contrastive Subject

Of the 26 tokens with the pragmatic function of contrastive subject in the corpus, only ten were coded for prosody. For this pragmatic function, canonical word order with the primary stress placement on the subject would be considered the canonical stress pattern since the subject is contrastive. All ten of the tokens coded for prosody appeared in this way (**Subject-Copula-Adjective** *Nicolás es feliz* ‘**Nicolas** is happy’). Although these results may not be very reliable as we have so little data to draw upon, we can make two general observations. First, there was a categorical favoring of canonical word order among both generational and regional cohorts for this pragmatic function. Secondly, among all the tokens with a contrastive subject the primary stress placement always fell on that constituent. We can easily say, then, there are no differences across generations or regions for this pragmatic function.

5. Utterances with the pragmatic function of Contrary to Expectation

The one token that communicated contrary to expectation was not coded for prosody so there are no data regarding primary stress placement for this pragmatic function.

6. Utterances with the pragmatic function of Contrastive Verb

As with utterances that conveyed the pragmatic function of contrary to expectation, there was only one token that communicated the pragmatic function of contrastive verb, but this token

was coded for prosody. Recall that this pragmatic function was created precisely for this token because in Spanish the copular verb is almost always unstressed (Fernández Leborans, 1999; López García, 1996). Typically instead of stressing the copula, Spanish speakers insert the word *sí* ‘yes’ to carry the stress of the copula. However, in this case the insertion of *sí* would not have conveyed the same pragmatic function as placing the primary stress on the copula. In this example, the copula was considered contrastive because it was put into opposition with another copula that immediately preceded it. Thus, this was the only token in the entire corpus with primary stress placement that fell on the copula. The prosody and order for this token was **Copula-Subject-Adjective** (*Es Nicolás feliz* ‘lit. **Is** Nicolas happy,’ ‘Nicolas **is** happy’). Since this is a contrastive utterance, this stress pattern is considered the canonical stress pattern for this pragmatic function. We have no way of comparing the two generational or regional cohorts with regard to this type of token since there was only one token produced and it was produced by a Highlander LARNC.

7. Summary and discussion

Although we do not have sufficient data to draw any conclusions for two of the five pragmatic functions (utterances conveying the pragmatic functions of contrastive verb and contrary to expectation), we can answer our research questions with some confidence for the other three pragmatic functions. First, we did find a relationship between variable word order and variable prosody (research question g). For pragmatically neutral utterances we found a strong tendency toward canonical word order and canonical prosody with the primary stress on the final constituent (the adjective). This was true for all regional and generational groups. This finding is consistent with the extant literature (Bolinger, 1954; Bolinger, 1955; Buring & Gutiérrez-Bravo,

2001; Contreras, 1978, Costa, 2001; Domínguez, 2004; Gutiérrez-Bravo, 2002; Gutiérrez-Bravo, 2008; Samek-Lodovici, 2001; Ocampo, 2002; Vallduví, 1991; Zubizarreta, 1998).

However, it is of special import that this order-prosody package was not categorical among speakers of the first generation, and that LARNC's produced pragmatically neutral utterances in which the primary stress placement was not in utterance final position. This finding indicates that prosody in non-contact forms of Spanish is more flexible than originally thought (Gabriel, 2007; Gabriel, 2010; Hoot, 2012; Leal-Méndez & Shea, 2012; Muntendam, 2009). To be sure, for most non-contrastive utterances the constituent in utterance final position is stressed. But our results indicate that this is not always the case, even for our LARNC, who have had little contact with English. Consequently, we can affirm that prosody is a variable phenomenon in both generations, and that the mere fact of variability in the NYR is not necessarily due to their increased contact with English.

The only pragmatic function for which we did not find a clear relationship between prosody and word order was for utterances with a highlighted adjective. All the participants preferred the marked Copula-**Adjective**-Subject order-prosody package, except the Caribbean participants who preferred canonical word order with primary stress on the adjective.

Regarding our next research questions (research questions h and i) it appears as though there are no categorical differences and few quantitative differences across the two generational and the regional cohorts with regard to prosody for three of the five functions. There was only one major quantitative difference across the regions; Caribbean participants preferred canonical word order with canonical prosody for utterances conveying the pragmatic function of highlighted adjective while the other participants preferred marked word order and marked prosody. Recall from Chapter Five that the preference for canonical word order in the Caribbean

lects has been attributed in part to their increased use of overt subject pronouns, which favors a fixing of the canonical word order (Morales 1989, 1997, 1999), and in part to their increased contact with English, especially among the Puerto Ricans (Lynch & Klee, 2009, p 200; Zentella, 2000, p. 139). What is interesting here is that even though both of these explanations would seem to apply to the NYR participants as well, the results were not as expected. Despite the NYR's increased overt pronoun rate—a finding that has been confirmed in various analyses of this corpus (Otheguy & Zentella, 2012; Raña-Risso, 2013)—and their increased contact with English, this group did not behave similarly to the Caribbean participants in this respect.

For utterances that convey the pragmatic function of contrastive subject we found no differences across the generational or regional groups. For those utterances that were pragmatically neutral, we found no differences regionally but did find a generational difference, namely an increased use of canonical word order and canonical prosody among NYR.

This was also the case across the generations for utterances with the pragmatic function of highlighted adjective. Although we cannot make any broad generalizations due to the fact that these data were not tested for significance, and that there were few data to draw upon, we can make a few observations. Recall from Chapter Two that Spanish speakers can use a number of resources including syntactic resources, such as word order, and prosodic resources, such as primary stress placement, together as well as independently to communicate their pragmatic intent. We saw in Chapters Four and Five that the NYR rely less on the syntactic resources by producing fewer utterances with marked word orders. We posited in Chapter Six that it was possible that the NYR were relying less on syntactic resources such as word order to communicate their pragmatic intent. If this were the case we would expect the NYR to rely more on other resources, such as prosodic resources. Some of our data, however limited, support this

expectation. For utterances with a highlighted adjective, the NYR prefer to use prosody only in 35 percent of their utterances, while the LARNC do so in 29 percent of theirs. This difference, albeit small, supports the idea that the NYR rely more on prosodic resources to communicate this particular pragmatic intent (highlighted adjective) than their first generation counterparts, the LARNC. When we look at the results for the other two pragmatic functions, conveying information and contrastive subject, we find that the two generational groups prefer to use prosodic and syntactic resources at similar rates. Thus we can conclude that when all pragmatic functions are taken together, the NYR rely on the prosodic resource more than or as much as the LARNC.

In order to address whether the prosodic differences across the generations are modeled after the prosody of English, our final research question (research question j), we need to focus exclusively on those instances where we found generational differences in prosody. Let us return, then, to utterances with a highlighted adjective where, as we noted above, the NYR rely more than the LARNC on prosodic resources than on syntactic and prosodic resources combined. In order to address our question we need to determine whether the increased use of prosody-only resources to communicate this pragmatic function appears to be modeled after the prosody of English. The data in Table 6.3 suggest that NYR are using an order-prosody package that the LARNC do not use. However, this pattern does not appear to be modeled after the prosody of English since it has final constituent stress and marked word order, both of which are more common in Spanish than in English (Vallduví, 1999). The one word order and stress pattern combination that could appear to be modeled after English is Subject-Copula-**Adjective** (*Nicolás es **feliz*** ‘Nicolas is **happy**’). However, this order-prosody package is found and used by the LARNC, albeit slightly less. Surprisingly, the most common order among both generational

groups for this pragmatic function is Copula-Adjective-Subject (*Es feliz Nicolás* ‘lit: Is **happy** Nicolas,’ ‘Nicolas is **happy**’), which is somewhat unexpected due to the stress shift. In short, the increased use of prosody-only resources to communicate this pragmatic function among the NYR is quite real, but does not appear to be modeled after the prosody of English.

There are two other arguments that support the claim that the NYR’s prosody is not modeled after English in general. If the NYR were modeling their Spanish prosody after English prosody, we would expect to find a preference among NYR for canonical word order with marked prosody for each of the pragmatic functions. That is, we would expect the NYR to use mostly prosodic resources, to the detriment of word order ones, to express the pragmatic functions of the utterances. We would also expect the NYR to employ novel stress patterns in their Spanish prosody. But this was not the case; with one exception, which we argue is not modeled after English, all of the stress patterns used by the NYR were also used by the LARNC.

This adherence on the part of the NYR to general Spanish prosodic patterns is especially true in contexts where English-like prosody would be infelicitous in the Spanish of the first generation. For example, as we have mentioned, in Spanish the copular verb is almost categorically unstressed due to its status as a semantically weak and bleached constituent (Fernández Leborans, 1999; López García, 1996). In English, on the other hand, and as we have also mentioned, the copula can easily be stressed in order to communicate correction or contradiction of a previous utterance, whereas in Spanish, speakers must insert the word *sí* ‘yes’ before the verb with primary stress falling on this constituent. If NYR participants were modeling their Spanish prosody after English prosody, we would expect them to stress the copula (the way they would in English) to communicate correction or contradiction. Although the tokens for this pragmatic function were not included in our original analysis because they

contained the extra constituent *sí*, a total of seven utterances of this type occurred in the corpus; two were produced by NYR and five were produced by LARNC. All of them contained a stressed *sí*. That is, none of them, including neither of the two by the NYR, followed the English stress pattern, with stress on the copula, which would be infelicitous in Spanish.

Based on the results presented in this chapter we can make several arguments. First, that the relationship between word order and prosody is very similar across the two generations. There were no categorical and few quantitative differences in prosody across the two generational cohorts. Secondly, that the NYR have not introduced novel stress patterns in their Spanish for the types of utterances that we studied.

Finally, since we provided evidence to demonstrate that prosody in Spanish is a variable phenomenon, we can argue that both generational groups have several linguistic resources at their disposal to communicate pragmatic functions. It is not true that the LARNC always use syntax, or changes in word order, to communicate pragmatic functions. Instead, they can use word order, prosody, and often both to communicate these functions. In short, the NYR employ various resources to communicate the pragmatic function of the utterance, but it appears that they prefer to use their prosodic resources, instead of the syntactic resources, more so than the LARNC. (Although this preference may be attributed to their increased contact with English, this possibility will be discussed more in depth in Chapter Seven.) In this way, the NYR do not necessarily have fewer expressive devices compared to their first generation counterparts, as has been argued by the extant literature (Silva-Corvalán, 1994; Zapata, Sánchez & Toribio, 2004). Instead, we argue that they have similar expressive devices, which they use in ways that are quantitatively different from those of their first generation counterparts.

Recall from our multivariate analyses in Chapter Five that two variables, type of subject NP and pragmatic function of the utterance, significantly accounted for only 17 percent of the variance of the regression among NYR, compared to 52 percent among the LARNC. We said there that this meant that other variables that were not considered in the regression could be accounting for some of the remaining variance among the NYR. We posited that one of these variables could be prosody. Here we have demonstrated the increased reliance of the NYR on prosody, suggesting that this variable is indeed covering some of the variance that we saw was not covered by word order. However, given the limited number of tokens on which this analysis is based, it is important to recognize that our conclusions regarding prosody should be regarded as preliminary until more data are gathered. This points to a way of deepening the present research, developing additional data that would allow us to test for statistical significance and to test directly the matter of increased variance accounted for by prosody among the NYR.

Thus far in our analysis we have focused exclusively on utterances with adjectival predicates. However, we found several utterances with overt subjects, copular verbs, and attributive complements that had adverbs instead of adjectives in the predicate. Although we first discarded these tokens because it appeared that they behaved differently from those tokens with an adjectival predicate, we believe that the differences between these two types of tokens speak to some of our research questions. In the next chapter we examine the relationship between word order and several of our predictor variables for tokens with adverbial, instead of adjectival, predicates.

CHAPTER 7

ADVERBIAL PREDICATES

1. Introduction

This chapter presents the analysis of the internal segmental predictor variables and the external predictor variables described in Chapter Three as they relate to utterances with adverbial predicates, instead of utterances with adjectival predicates, which have been the focus of the analysis for the past three chapters. The analysis aims at answering the research questions outlined in Chapter One and repeated here below. Only the research questions addressed in the chapter are listed.

Question A: What are the external variables that have predictive statistical effect on word order in the corpus? Do any of these external variables produce categorical effects (that is, effects where the presence of a predictor produces a 100 percent to zero result)?

Question B: What differences, if any, are there between the two geographic regions in the way that the external variables influence word order, either quantitatively or categorically?

Question C: What differences, if any, are there between the two immigrant generations in the way that the external variables influence word order, either quantitatively or categorically?

Question D: What are the internal variables that have a predictive statistical effect on word order in the corpus? Do any of these internal variables produce categorical effects (that is, effects where the presence of a predictor produces a 100 percent to zero result)?

Question E: What differences, if any, are there between the two geographic regions in the way that the internal variables influence word order, either quantitatively or categorically?

Question F: What differences, if any, are there between the two immigrant generations in the way that the internal variables influence word order, either quantitatively or categorically?

Although answers to these research questions have been provided for utterances with adjectival predicates, we wanted to determine whether the answers to these questions are the same for utterances with adverbial predicates. Recall from Chapter Three that tokens with adverbial predicates (*Nicolás está aquí*, ‘Nicolas is here’) were collected and included originally in our analysis. However, after preliminary analyses revealed that these types of utterances behaved differently from those with adjectival predicates we chose to analyze them separately from the adjectival predicates. Specifically, including the adverbial predicate utterances in our analysis of the adjectival predicate utterances would increase word order variability, thus confounding the data and leading us to conclude that there is more variability than there actually is.

Our inclination that adverbial predicates and adjectival predicates are different is not just based on our preliminary results that adverbial predicate utterances have greater word order variability. Ocampo (2002) treats adverbs and adjectives separately in his studies on word order in Spanish. Also, in many contemporary syntactic treatments adverbs like *aquí* ‘here’, *ahí* ‘there’ and *allá* ‘over there’ share the external syntactic distribution of prepositional phrase predicates (Haegeman, 1994). Additionally, we know that both Spanish and English speakers may signal pragmatic functions by use of *inversion*, that is, the appearance of the verb before the subject. Inversion to VS order in English may occur with an adverb in a copular construction as in (1) (Birner, 1994). This type of inversion does not typically occur with a pronominal subject (Quirk et al. 1972; Swan, 2005), as in (2).

(1) Here is the doorman.
Adv Cop S

(2) *Here is he.
Adv Cop S

Inversion is also common in Spanish with adverbial predicates, as in (3).

- (3) Aquí está el portero.
Adv Cop S
'Here is the doorman.'

This type of locative inversion, especially in copular constructions, is used in English commonly, especially in conversational speech. Due to the relative frequency with which adverbial predicates can occur in marked order in English, we believe that marked order in Spanish will also be more frequent among utterances with adverbial predicates than those with adjectival predicates. That is, we predict that utterances with adverbial predicates will occur less frequently in canonical word order than utterances with adjectival predicates. We predict that we will find more word order variability among utterances with adverbial predicates because utterances with adverbial predicates occur in marked order more often than utterances with adjectival predicates in English. As a result, the tendency toward SVX is not as salient for these types of structures in English, and thus the influence toward SVX in Spanish due to language contact will also not be as strong.

Recall that these utterances were not coded for prosody or for pragmatic function so we cannot address whether the word order variability for adverbial predicate structures in our sample is related to the communication of pragmatic function through word order or prosody. However, these tokens were coded for the socio-demographic, language use, and the other predictor variables including informational status of the subject NP. As a result, we can examine whether any of these variables, some of which were predictors of word order variability for utterances with adjectival predicates, are predictive of word order variability for utterances with adverbial predicates. The focus of this chapter, then, is to determine the differences between the two types of tokens (adjectival predicates versus adverbial predicates).

In section 2 below we present the results for all the external predictor variables and their

relationship with word order, as well as the differences between the results for adjectival and adverbial predicates. The results for all the internal predictor variables and their relationship with word order, as well as the differences between the results for adjectival and adverbial predicates is presented in section 3. A summary and discussion of the results appears in section 4.

2. External variables predicting word order for utterances with adverbial predicates

Below we discuss the relationships between the outcome canonical word order rate for adverbs (henceforth WOR(Adv)) and the internal predictor variables. The word order outcome variable considered here covers utterances found in the corpus consisting of (a) a simple, complex, infinitival, or clausal subject, (b) a copular verb and (c) an adverbial predicate.

Recall that for our token-level analysis, that is for the analysis involving internal predictor variables the outcome variable word order has four factors or levels, as outlined below.

Dependent (outcome) variable: Word order

- Subject-Copula-Adverb *Nicolás está aquí* ‘Nicolas is here’
- Copula-Adverb-Subject *Está aquí Nicolás* ‘lit: Is here Nicolas,’ ‘Nicolas is here’
- Adverb-Copula-Subject *Aquí está Nicolás* ‘Here is Nicolas’
- Copula-Subject-Adverb *Está Nicolás aquí* ‘lit. Is Nicolas here,’ ‘Nicolas is here’

The first order, Subject/Copula/Adverb, is considered in the present study the canonical word order whereas the other orders are considered the marked orders. For the current analysis, however, we use a continuous word order variable, canonical WOR(Adv), for each participant, based on the number of times each participant produced the canonical word order out of his or her total tokens.

The word order rate dependent or outcome variable was studied with respect to the independent or predictor variables outlined below.

Independent (predictor) variables initially investigated

- Generation (derived from age of arrival in NYC and years lived in NYC)
- National origin
- Regional origin
- Gender
- Age
- Social class
- Education
- English skills
- Spanish use
- Language with father
- Language with mother
- Language with siblings
- Language with children
- Language with friends
- Language with boss
- Language with significant other
- Language with speakers from own country
- Language with other groups
- Language with speakers from own group
- TV in Spanish
- Radio in Spanish
- Reading in Spanish

- Spanish use in social settings
- Spanish at home
- Spanish in domains
- Spanish skills

The initial exploration of these predictors showed that many of them bear an association with WOR(Adv) that is statistically negligible or too inconsistent or unreliable to be revealing of any pattern worth reporting. This is in part due to the few data we have to draw upon. As a result, this chapter, while devoting a subsection to providing some details regarding each of the variables above, concentrates primarily on the discussion of the following predictors, which did yield significant results. Also, given that Region and Generation were two very predictive variables in the previous three chapters, we will discuss these variables below, even though they were not significant for this analysis.

Independent (predictor) variables significantly associated with word order

- Language with Friends
- Language with Siblings
- Spanish in Reading
- Gender

2.1. Distribution of the dependent variable WOR(Adv) in the corpus as a whole and by region and generation.

Given our predictions, laid out in Chapters One and Two, as well as our results in Chapters Four and Five, we would expect higher canonical WOR(Adv) among NYR. This would indicate that they have less word order variability compared to the LARNC, which we attribute, in part, to the NYR participants' increased use of and contact with the English language.

However, we expect utterances with adverbial predicates to behave slightly different from those with adjectival predicates because those with adverbial predicates have more parallel structures in the contact language.

Overall, there were many fewer tokens with adverbial predicates; only 58 tokens occurred in the corpus. Additionally, not all speakers produced tokens of this type. Many speakers only produced one token—meaning that their values for canonical WOR(Adv) were either 100 or 0, which we believe may have skewed the results—and 17 speakers produced zero tokens. As a result, only 33 participants entered into the analysis for the external variables. For these reason, our conclusions are only preliminary. Another consequence of the low frequencies is that many of the analyses did not reach significance. Despite these limitations, we feel that the differences between adverbial and adjectival predicates highlight an important point in our broader discussion.

The average canonical WOR(Adv) for utterances with adverbial predicates was 70 percent, which is slightly lower than the canonical WOR(Adv) for adjectival predicate utterances (75 percent). Unlike our results for utterances with adjectival predicates, we did not find significant differences in canonical WOR(Adv) across generations or regions for adverbial predicates. Tables 7.1 and 7.2 show the breakdown of WOR(Adv) by Generation and Region.

Table 7.1		
<i>ANOVA, Canonical WOR(Adv) by Generation</i>		
Generation	N speakers	Canonical WOR(Adv)
LARNC	17	78
NYR	16	62
	33	
$F = 1.274$	$p = .268$	

Table 7.2		
<i>ANOVA, Canonical WOR(Adv) by Region</i>		
Region	N speakers	Canonical WOR(Adv)
Highlanders	14	67
Caribbeans	19	72
	33	
$F = .122$	$p = .729$	

When we compare canonical WOR(Adv) and WOR for adjectival predicates we do not find any significant differences for the two token types for the entire sample or across regions. However, when we compare generations we find that there is a significant difference in canonical WOR(Adv) among the NYR participants, $t(15) = -2.187, p < .05$.

2.2. Gender.

For the corpus as a whole we found no significant difference in terms of canonical WOR(Adv) by gender. We find the same results when we compare males and females by generation. Across region, however, we did find a significant difference in canonical WOR(Adv) between males and females. Among Highlanders the females had a canonical WOR(Adv) 58 percentage points higher than the males. Recall that for adjectival predicates women also had a higher Canonical WOR(Adv) than men. Among the Caribbeans, the males had a higher canonical WOR(Adv) by two percentage points. These data can be seen in Tables 7.3 and 7.4

Table 7.3		
<i>ANOVA, Canonical WOR(Adv) by Gender, Highlanders</i>		
Gender	N speakers	Canonical WOR(Adv)
Male	5	30
Female	9	88
	14	
$F = 11.884$	$p < .05$	

Table 7.4		
<i>ANOVA, Canonical WOR(Adv) by Gender, Caribbeans</i>		
Gender	N speakers	Canonical WOR(Adv)
Male	13	73
Female	6	71
	14	
$F = .001$	$p = .917$	

2.3. Language with friends.

Three of the language use variables yielded significant results for adverbial predicates. The first, Language with Friends, was significant for the entire corpus as well as for the Highlanders. The direction of the results, however, was not as we outlined above in our expectations. The participants who used more Spanish with their friends had higher canonical WOR(Adv) than those that used both languages or English. The results for the entire corpus as well as for the Highlanders can be seen in Tables 7.5 and 7.6 respectively.

Table 7.5		
<i>ANOVA, Canonical WOR(Adv) by Language with Friends</i>		
Language	N speakers	Canonical WOR(Adv)
English	9	44
Both	16	75
Spanish	7	95
	32	
$F = 3.713$	$p < .05$	

Table 7.6		
<i>ANOVA, Canonical WOR(Adv) by Language with Friends, Highlanders</i>		
Language	N speakers	Canonical WOR(Adv)
English	2	0
Both	8	78
Spanish	3	89
	32	
$F = 5.680$	$p < .05$	

Although the results for the Caribbean participants followed this direction, the results were not significant.

2.4. Language with siblings.

The second language use variable, Language with Siblings, was not significant, but approached significance. Recall that this variable was the only language use variable that was significant for the adjectival predicates, and participants who reported speaking Spanish with their siblings tended to have a lower canonical WOR(Adv) than those who reported speaking English or both languages with their siblings. This variable was not significant for the corpus overall, for either regional sub-sample, or for the NYR; it was only significant among the LARNC. As we saw above, the direction of the results was not as we outlined above in our expectations. Participants who spoke Spanish with their siblings had a higher canonical WOR(Adv) than those that spoke both languages with their siblings. No LARNC reported speaking English only with their siblings. These data can be found in Table 7.7.

Table 7.7		
<i>ANOVA, Canonical WOR(Adv) by Language with Siblings, LARNC</i>		
Language	N speakers	Canonical WOR(Adv)
Both	4	50
Spanish	12	85
	16	
$F = 3.383$	$p = .087$	

2.5. Spanish in reading.

The final language use variable that was significant was Spanish Reading. This variable was not significant for the corpus overall, for either generational sub-sample, or for the Highlanders; it was only significant among the Caribbeans. As with the other language use variable the direction of the results was not as we outlined above in our expectations. The participants who reported reading a lot in Spanish had a higher canonical WOR(Adv) than those that reported reading a little in Spanish. These results appear in Table 7.8

Table 7.8		
<i>ANOVA, Canonical WOR(Adv) by Spanish in Reading, Caribbeans</i>		
Spanish in Reading	N speakers	Canonical WOR(Adv)
A lot	11	61
A little	7	100
	32	
$F = 5.320$	$p < .05$	

In section 3, below, we turn to our analysis of the internal predictor variables.

3. Internal variables predicting word order for utterances with adverbial predicates

In this section, instead of using the continuous variable Canonical WOR, we use the nominal variable. For each token, then, the variable Word Order has two factors: canonical word

order or marked word order. This variable was studied with respect to the independent or predictor variables outlined below.

Independent (predictor) variables initially investigated

- Affirmative or negative utterance
- Type of copular verb
- Presence of English elements
- Syntactic properties of subject NP
- Length and type of adjectival phrase
- Number in third-person verbs
- TMA of the verb
- Discourse pragmatics: Informational status of the subject NP
- Discourse pragmatics: Function of the utterance

The initial exploration of these predictors showed that all of them bear an association with word order that is statistically negligible or too inconsistent or unreliable to be revealing of any pattern worth reporting. Additionally, we found no significant differences in word order across generations or regions. As a result, in each corresponding section we will focus on pointing out any notable differences or similarities to the adjectival predicate results.

3.1. Affirmative or negative utterance.

There were no significant differences in word order between utterances that included the adverb *no* and those that did not. However, this result may have to do with frequency issues as only seven of the tokens were negative. All of the negative tokens had canonical word order.

3.2. Type of copula.

There were also no significant differences in word order for type of copula. Seventy-four percent of the utterances with adverbial predicates used the verb *estar*. Recall from section 2.3 of Chapter Five that the opposite was true for utterances with an adjectival predicate: seventy-five percent of tokens with an adjectival predicate had the verb *ser*. This is more evidence in favor of the argument to treat these two constructions separately.

3.3. English elements.

There was not enough data to perform an analysis for English elements with only two tokens among the 58 adverbial predicate tokens. One of the tokens appeared in canonical word order and the other in marked word order.

3.4. Syntactic properties of the subject NP.

More evidence to support our choice to treat these constructions separately is that utterances with adverbial predicates did not appear with infinitival or clausal subjects at all in the corpus, while 17 percent of adjectival predicate constructions did. This is not surprising considering that clauses express propositions, and these do not tend to be localized the way the locative adverbs like *aquí* ‘here’ do. There were no significant differences with regard to type of subject NP for the adverbial predicate tokens: about 72 percent of utterances with either a complex or simple NP appeared in canonical word order. The average length of the subject NP for utterances with adjectival predicates was 3.32 while utterances with adverbial predicates had a subject with an average of 2.84 words.

3.5. Length and type of adjectival phrase.

Similarly, utterances with adverbial predicates only had simple predicates while five percent of adjectival predicates were complex. The average length of the adjectival predicates

was 1.5 while the average length of the adverbial predicates was 1.0. There were no complex predicates in the corpus for this type of token.

3.6. Number ending in third person verbs/TMA of the verb.

There were also no significant differences regarding the number ending of the verb and word order. Just like utterances with adjectival predicates, utterances with plural verbs occurred more often in canonical word order than utterances with singular verbs, although the difference was very small. There were also no significant differences with regard to the TMA of the verb. Utterances with adverbial predicates only occurred with present indicative and preterite indicative verbs in the corpus. Both types of verbs occurred most often with canonical word order (70 percent and 78 percent accordingly).

3.7. Discourse pragmatics: Informational status of the subject NP.

Unlike utterances with adjectival predicates, the informational status of the subject was not significantly related to word order for adverbial predicates. All three levels of informational status (given, new, and implied) occurred most often with canonical word order.

4. Summary and discussion

The previous sections examined the variables that were significantly related to word order for utterances with adverbial predicates. Only four of the external variables reached or approached significance and for three of these variables, our results did not coincide with our expectations that increased Spanish use would correspond to a lower Canonical WOR(Adv). Although this may be due to the fact that there were many fewer tokens with adverbial predicates and that many speakers only produced one token, which may have skewed the results, it also supports the idea that these two token types are inherently different. In addition to the directional difference, we found many more differences than similarities between utterances with

an adjectival predicate and those with an adverbial predicate. The only commonality between these two types of tokens is that women appear to have a higher canonical word order rate for both types of constructions. One of the most important differences was the lack of significant differences in word order across generation or region for utterances with an adverbial predicate. Similarly, when we compared canonical WOR for adverbial predicates and adjectival predicates we found no differences two token types among the entire sample, across regions, or among the Highlanders, but we did find significant differences among the NYR participants.

At the token level we found no significant relationship between any of the predictor variables and word order for the entire sample as well as by region or generation. Although the lack of results may have to do with the low frequencies in the analyses, the differences between these two types of tokens support our choice to analyze these two types of utterances separately.

We believe that the differences outlined above highlight several important points. First, they indicate that these two token types are structurally different given that in many contemporary syntactic treatments adverbs like *aquí* ‘here’, *ahí* ‘there’, and *allá* ‘over there’ share the external syntactic distribution of prepositional phrase predicates (Haegeman, 1994). For this reason it is important that researchers looking at Spanish word order consider different types of structures separately and not overgeneralize findings from one structure type to another. Also, it is possible that one of the reasons we do not find an significant increase in canonical WOR(Adv) among the NYR (and we do find a significant difference in WOR(Adv) across the two token types for NYR only) is that adverbial predicate tokens are not subject to simplification processes (outlined in Chapters Four and Five) that the adjectival predicate tokens are subject to. We believe adverbial predicates are somewhat immune to these processes because the parallel construction in the contact language (utterances with Subject/Copula/Adverb in English) behaves

similar in Spanish. In the following chapter we will discuss the notion of parallel constructions more in depth and how we believe this concept can be used to explain why two superficially similar constructions behave so differently in our corpus.

CHAPTER 8

CONCLUSIONS

1. Introduction

In this chapter, we present a summary of the major results and conclusions of the dissertation. We also attempt to contextualize these results in the wider field of contact linguistics, and present several avenues for future research.

2. Summary of conclusions

Our study yielded results that showed differences between two generational groups of Spanish speakers in New York (the first generation of Latin American raised newcomers, or LARNC, and the second generation of New York Raised speakers, or NYR). We discuss each in turn.

Generational differences. We discuss our generational results by separating findings related to word order rates (WOR) from findings related to conditioning factors. With regard to rates, our results indicate that, as we had expected, the NYR have a more rigid word order than the LARNC. Our data also indicate that these apparent-time generational differences can be attributed in large part to the NYR's increased use of and contact with English. Overall in our sample, exposure and mastery of English were related to greater use of canonical word order while exposure and mastery of Spanish were related to greater use of marked word order. Our results showed clear correlations between higher self-reported English skills and higher canonical WORs. We also found a correlation between, on the one hand, greater Spanish skills and reported use of Spanish with siblings and, on the other hand, a lower canonical WOR.

These results align with experimental studies (Zapata, Sánchez & Toribio, 2005) and sociolinguistic studies (Silva-Corvalán, 1994) that also report an increased use of canonical word

order among second- and third-generation bilinguals in the US. Both of these studies attribute their findings to indirect transfer from, or convergence with, English word order patterns, an interpretation with which we concur and that is supported by our correlational findings. Additionally, we concur with Silva-Corvalán in seeing the greater rigidity of word order in the second generation as an example of simplification, in that it involves a reduction in the use of alternative or competing forms (canonical versus marked word orders), and an overgeneralization of the use of canonical word order.

With regard to conditioning factors, we found that although the variables that condition word order in the two generational groups are mostly the same (Type of Subject NP, Pragmatic Function, Subject Givenness, and Verb Number Ending), these variables do not account for the same amount of word order variance in the two generations (variance-accounted-for as measured by R-square coefficients in the regressions). The internal variables of our study account for three times more word-order variance among the LARNC than among the NYR (they account for 52 percent of the variance among the LARNC but only for 17 percent of the among the NYR). For the external variables, the variance accounted for in the two generations is about the same, but the pattern of significant, near-significant, and non-significant variables is somewhat different.

We have suggested that our results correspond, but only to some extent, to what Silva-Corvalán calls *syntactic permeability*, which she defines as the increased use of a parallel structure (in our case, canonical word order) and the loss of semantic-pragmatic constraints. Although we concur with the first part of this description, our results do not support the second part, as we have not been able to confirm the loss of constraints among the NYR. Instead, we have shown that the NYR are differently sensitive to the variables and constraints that govern word order among the LARNC. We maintain that despite the constraint differences across the

generations, the NYR are very much capable of communicating the pragmatic functions that their first-generation counterparts communicate using word order. We maintain this position for two reasons. First, because the variable constituted by the pragmatic function of the utterance, along with its constraints, is still a significant predictor of word order among the NYR. Second, we believe that the NYR are still capable of communicating the pragmatic functions for which the first generation relies mostly on word order because there are among the NYR other variables that account for the variance among the NYR, such as prosody.

Before we turn to our discussion of prosody, we would like to focus briefly on Silva-Corvalán's (1994) the notion of parallel structures. The idea that speakers would converge on strategies that are shared or common to both languages (Silva-Corvalán, 1994) is supported in our data by two separate analyses, the first of which (the greater use of canonical WOR among the LARNC) we have just described. The second analysis that supports this idea is our examination of copular utterances with adverbial predicates. Here we found that for constructions whose English word order and Spanish word order are more similar—English adverbial predicates are more flexible in their order than adjectival predicates—we do not find a significant increase in canonical WOR across the generations (nor do we find significant differences between the regions). We interpret these results to mean that that adverbial predicates are not equally subject to the processes of simplification and overgeneralization that the adjectival predicates are subject to, given that the parallel construction in the contact language behaves similarly in non-contact Spanish. Even if the bilinguals did overgeneralize English word order to Spanish utterances with adverbial predicates, this process would not be very salient since adverbial predicates in Spanish behave similar to English adverbial predicates anyway. However, given that we had very few tokens for this analysis, we recognize the limitations of the

data and our conclusions regarding adverbial predicates are preliminary until more data are gathered.

Returning to prosody, we found that, as the NYR exhibited a more rigid word order, they concomitantly employed prosodic resources to express discourse-pragmatic functions with greater frequency than the LARNC. Importantly, we also found that Spanish order-prosody packages are much more variable in the LARNC than the literature on Spanish prosody might lead one to believe. Therefore, we did not interpret the variability in prosody that we found among the NYR as an innovation, much less a sign of reduced expressive capacity. The differences in prosody across the generations are quantitative, not qualitative; the NYR are not doing something novel (or infelicitous) with their prosody, just more of what the LARNC are already doing. Finally, we have shown that prosodic differences among the NYR are not in the direction of English.

Regional differences. Between the regional groups, we also found differences in the use of canonical word order. In general the Caribbean participants had a higher canonical WOR than their Highlander counterparts. This difference was significant among the LARNC, but disappears among the NYR. Our explanation for these regional differences is bipartite. First, these regional differences in canonical word order in the construction under study here can be folded under more general dialectal differences, which we know exist for word order with other types of constructions, such as with subject personal pronouns (Lipski, 1996; Raña-Risso, 2013). We can also attribute the Caribbeans' increased use of canonical word order to their increased contact with English (Lynch & Klee, 2009; Zentella, 2000). According to the 2010 Census, Caribbeans in New York City (NYC) report speaking English at home nearly three times more than Highlanders do (17.3 percent versus 5.8 percent respectively), and they also report speaking

better English in general. When asked to rate their English skills, 78 percent of Caribbeans reported speaking English exclusively, well, or very well, compared to 63 percent of Highlanders. Given these differences, we argue that we can make the same generalization for the regions that we made across generations: Greater exposure and mastery of English are statistically related to greater use of canonical word order while greater exposure and mastery of Spanish are statistically related to an increase in the use of marked word order. Similar to our results for the two generational sub-samples, we also found that the Caribbeans relied slightly more on prosodic resources than on word order resources. We believe that the parallels between regional and generational findings in our study are an example of what Otheguy (1995) calls *synchronic creativity*, that is, that dialectal differences and changes in non-contact varieties of Spanish are oftentimes no different from the processes that result in innovations in the Spanish in the US and other situations of linguistic contact.

3. Relevance, avenues for future research, and limitations

Our data show that second-generation Spanish-English bilinguals in NYC have not lost the discourse-pragmatic constraints that govern word order, but are instead differently sensitive to these constraints than their first-generation counterparts. As a result, we believe this study contributes to the dialogue regarding which linguistic domains are vulnerable to contact-induced change. Differences in frequency between the pre-contact and contact lects at the discourse-pragmatic and prosodic level are not necessarily evidence of structural change. These differences are, according to Silva-Corvalán (1994), common outcomes in situations of language contact and an attempt to reduce the cognitive load of using two or more linguistic systems. These and many other strategies do not necessarily result in different grammars, but in the abandonment or preference of certain outcomes in favor of others.

Differences in word order can also be explained by appealing to cognitive processes that govern other linguistic phenomena such as language acquisition. Donati and Nespors (2003) maintain that the acquisition of word order is prompted by prosody, so that speakers who experience incomplete acquisition may not fully acquire all of the discourse-pragmatic constraints that govern the use of variable word order. Other researchers have pointed out that the interface between syntax and pragmatics is especially susceptible to cross-linguistic influence in bilingual acquisition (Hulk & Müller, 2000). This is especially true when the two languages share a parallel structure whose discourse-pragmatic constraints partially overlap. In other words, bilingual children may prefer one syntactic construction (Subject-Verb order) over another (Verb-Subject order) in one language (Spanish) due to the fact that the other language (English) is biased toward one of these two possible constructions (Meisel, 2001). Our data, then, may suggest that domains that are vulnerable in bilingual acquisition correspond to those that are vulnerable in language contact situations. This point however has yet to be investigated.

Finally, we believe this dissertation brings a new perspective to the field of contact linguistics and prosody. Although word-order rigidity in the Spanish of second- and third-generation Latinos in the US has been documented by other researchers (Silva-Corvalán 1994; Zapata, Sánchez, & Toribio, 2005), these studies are mostly descriptive and were not preoccupied with looking for functional explanations to account for language change. Additionally, recent technological developments have yielded huge advances for the study of prosody. However, most studies have been restricted to the examination of native prosody and prosody in second language learners. By examining the prosody of heritage language speakers of Spanish, this dissertation presents a novel direction for the study of prosody.

However, much more work needs to be done. For example, we need to analyze the word order and prosody for other types of constructions such as those with pronouns and other types (non-copular) verbs. Future corpus-based studies on prosody and word order may have to be accompanied by experimental studies on heritage language prosody, such as those by Hoot (2012). The present study, and other sociolinguistic corpus based studies like Silva-Corvalán's (1994), may not yield enough data to arrive at reliable conclusions because one cannot control the discourse contexts, nor can you standardize the contexts across participants. Complementing corpus-based prosody studies with experimental data would provide a broader perspective of Spanish heritage language prosody, and help prevent the frequency issues that we encountered throughout our analyses of prosody and adjectival predicates.

APPENDIX
QUESTIONNAIRE

Pronoun Study

CUNY Project on Spanish in New York

Ricardo Otheguy¹ and Ana Celia Zentella²

¹Graduate Center, City University of New York

²University of California / San Diego

This questionnaire for CUNY's Project on the Spanish of New York's Pronoun study is based on a previous questionnaire designed by Ana Celia Zentella for her own research. The abbreviation acz found after most items refers to the number that the item had in Zentella's questionnaire. The abbreviation FG, for factor group, refers to the number of the variable or factor group in the Pronoun study's Coding Manual.

This questionnaire is to be filled out by the investigator on the basis of information provided by the consultant, or on the basis of a prior questionnaire already filled out by the consultant. Use only the consultant's three-digit-plus-letter identification code, not the consultant's name. Make sure to match the identification code with the one used in the interview transcription. A transcription without a matching questionnaire is of no use to the project.

A. (acz 55, FG 14)

País de origen _____

Ciudad o pueblo de origen _____

B. (acz 53, FG 15)

Sexo _____

C. (acz 52, FG 16)

Edad _____

D. (acz 59b, FG 17)

Edad a la que llegó a EE.UU. _____

E. (acz 58, FG 18)

Años en los EE.UU. _____

EE. (acz 57, FG 19)

Clase social a la que ud. pertenece

Alta _____

Media _____

Obrera _____

F. (a: acz 54a; b: acz 54b, FGs 19 and 20)

a. Profesión del informante en EE.UU. _____

(Si es estudiante, profesión del padre y la madre en EEUU

b. Profesión del informante en el país de origen _____

c. Profesión del padre en el país de origen _____

G. (acz 56, 57, FG 21)

a. ¿Dónde se educó? _____

b. Años completados _____

H. (acz 1, 9, FG 22)

a. ¿Habla inglés? Sí _____ No _____

_____ Excelente

_____ Muy bien

_____ Pasable

_____ Pobre

b. ¿Sabe leer y escribir en inglés?

_____ Leer

_____ Escribir

I. (acz 2, 8, FG 23)

a. ¿Habla español? Sí _____ No _____

_____ Excelente

_____ Muy bien

_____ Pasable

_____ Pobre

b. ¿Sabe leer y escribir en español?

_____ Leer

_____ Escribir

J. (acz 3,4,5,6, y 7)

a. ¿Qué idioma aprendió primero?

_____ inglés

_____ español

_____ ambos

b. ¿A qué edad aprendió el otro idioma? _____

c. ¿Cómo y dónde lo aprendió?

_____ escuela

_____ TV

_____ familia

_____ otro

d. ¿Qué idioma habla mejor?

_____ inglés

_____ español

e. ¿Cuál le gusta más?

_____ inglés

_____ español

K. (acz 10, FG 24)

Ponga I, E o A en los espacios correspondientes.

I = inglés

E = español

A = ambos

¿Cuál idioma(s) habla [o hablaba] con su(s):

_____ Papá

_____ Mamá

_____ hermanas/os

_____ hijos menores

_____ hijos mayores

_____ amigos

_____ jefe

_____ compañeros de trabajo

_____ compañeros de escuela

_____ esposa/o o novia/o

L. (acz 11, FG 25)

Ponga P o M en los espacios correspondientes a la cantidad de español que use el informante en las distintas actividades.

P = poco español

M = mucho español

N = nada de español

NA = no aplicable

Cerciorarse de poner P o M solamente cuando el informante de hecho participa de estas actividades. Si alguna de las actividades no son pertinentes al informante porque, por ejemplo, no trabaja, no va a la escuela, etc., no se pone ni P ni M, sino que se pone NA = no aplicable.

¿Cuánto español usa Ud. en:

_____ casa

_____ la escuela

_____ el trabajo

_____ actividades sociales

_____ al leer

_____ al escuchar la radio

_____ al mirar la televisión

M. (acz 17, FG 26)

Ponga una marca debajo de la columna correspondiente.

Use las tres primeras columnas, Mucho, Algo, Poco.

Si el informante ofrece información suplementaria, escríbala en una o dos palabras bajo la columna de Dónde.

¿Cuánto contacto tiene Ud. con los siguientes grupos, y dónde [en la casa, el trabajo, el vecindario]?:

	<u>Mucho</u>	<u>Algo</u>	<u>Poco</u>	<u>Dónde</u>
Colombianos	_____	_____	_____	_____
Ecuatorianos	_____	_____	_____	_____
Mexicanos	_____	_____	_____	_____
Cubanos	_____	_____	_____	_____
Dominicanos	_____	_____	_____	_____
Puertorriqueños	_____	_____	_____	_____

N. Vecindario, ciudad y país donde se realizó la entrevista

REFERENCES

- Alba, O. (1988). Estudio sociolingüístico de la variación de las líquidas finales de palabra en el español cibaño. In R. Hammond & M. Resnick (Eds.), *Studies in Caribbean Spanish Dialectology* (pp. 1–12). Washington DC: Georgetown University Press.
- Arnold, J., Losongco, A., Washow, T., & Ginstrom, R. (2000). Heaviness versus newness: The effects of structural complexity and discourse status on constituent ordering. *Language*, 76, 28–55.
- Beckman, M. (1986). *Stress and non-stress accent*. Dordrecht, Holland/ Riverton, USA: Foris Publications
- Behaghel, O. (1909). Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen*, (25), 110–142.
- Belazi, H., Rubin, E., & Toribio, A. J. (n.d.). Code Switching and X-Bar Theory: The Functional Head Constraint. *Linguistic Inquiry*, 25(2), 221–237.
- Bentivoglio, P. & Sedano, M. (2000). El sujeto liviano: Una restricción de tipo funcional. *Boletín de Filología* (Universidad de Chile), 38.9-22
- Bentivoglio, P. (2003). Orden de palabras en español: Un análisis sintáctico-semántico-pragmático del sujeto. *Lexis*, 1-2, 235-260.
- Bergad, L. (2011). The Latino Population of New York City, 1990 - 2010. *The CUNY Center for Latin American, Caribbean, and Latino Studies' Latino Data Project Report 44*.
- Birner, B. (1994). Pragmatic constraints on the verb in English inversion. *Lingua*, 97, 233-256.
- Büring, D. (2008). Towards a typology of focus realization. In M. Zimmermann & C. Féry (Eds.), *Information structure*. Oxford: Oxford University Press.
- Büring, D., & Gutiérrez-Bravo, R. (2001). Focus-related constituent order without the NSR: A prosody-based crosslinguistic analysis. In S. Bhloscaidh (Ed.), *Syntax at Santa Cruz 3* (pp. 41-58). Santa Cruz: University of California
- Cedgren, H. (1978). En torno a la variación de la /s/ final de sílaba en Panamá: análisis cuantitativo. In H. López Morales (Ed.), *Corrientes actuales en la dialectología del Caribe hispánico* (pp. 80–103). Río Piedras: Universidad de Puerto Rico.
- Contreras, H. (1978). *El orden de palabras en español*. Madrid: Cátedra.
- Costa, João. (2001) Emergence of Unmarked Word Order. In G. Legendre, J. Grimshaw, & J. Vikner (Eds.), *Optimality-Theoretic syntax* (pp. 171-204). Cambridge, MA: MIT Press.

- Cutler, A., Dahan, D., & Van Donselaar, W.A. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, 40 (2), 141-202.
- Domínguez, L. (2004). The effects of phonological cues on the syntax of focus constructions in Spanish. In R. Bok-Bennema (Ed.), *Romance Languages and Linguistic Theory 2002: Selected papers from 'Going Romance', Groningen, 28-30 November 2002* (pp.69-81). Philadelphia: John Benjamins.
- Donati, C. & Nespor, M. (2003). From focus to syntax. *Lingua*, 113, 1119-1142.
- Durán Urrea, E. (2009). The Syntax and Prosody of Codeswitching in New Mexican Spanish-English Discourse. Presented at the 38th Annual Meeting of New Ways of Analyzing Variation (N WAV), University of Ottawa, Ottawa, CAN.
- Eddington, D. (2010). A comparison of two tools for analyzing linguistic data: logistic regression and decision trees. *Italian Journal of Linguistics*, 22(2), 265–286.
- Fernández Leborans, M. J. (1999). La predicación: las oraciones copulativas. In I. Bosque & V. Demonte (Eds.), *Gramática descriptiva de la lengua española* (pp. 2357–2460). Madrid: Espasa Calpe.
- Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London: Sage Publications.
- Gabriel, C. (2007). *Fokus im Spannungsfeld von Phonologie und Syntax. Eine Studie zum Spanischen*. Frankfurt: Vervuert.
- Gabriel, C. (2010). On focus, prosody, and word order in Argentinean Spanish: A Minimalist OT account. *ReVEL* 4:183-222.
- Givón, T. (1993). *English grammar: A function-based introduction*. Amsterdam / Philadelphia: John Benjamins Publishing Co.
- Gutiérrez-Bravo, R. (2002). Focus, word order variation and intonation in Spanish and English: An OT account. In C. Wiltshire & J. Camps (Eds.), *Romance phonology and variation* (pp. 39-53). Amsterdam: John Benjamins.
- Gutiérrez-Bravo, R. (2008). La identificación de los tópicos y los focos. *Nueva Revista de Filología Hispánica* 56: 362-401.
- Hawkins, J. A. (1983). *Word order universals*. New York, NY: Academic Press.
- Haegeman, L. (1994). *Introduction to government and binding theory* (2nd ed.). Oxford: Blackwell.
- Hoot, B. 2012a. Presentational focus in heritage and monolingual Spanish. Ph.D. dissertation, University of Illinois at Chicago.

- Hulk, A., & Müller, N. (2000). Bilingual first language acquisition at the interface between syntax and pragmatics. *Bilingualism: Language and Cognition*, 3(3), 227-244.
- Klee, C. & Lynch, A. (2009). *El español en contacto con otras lenguas*. Washington DC: Georgetown University Press.
- Labov, W. (1972). The isolation of contextual styles. In *Sociolinguistic patterns* (pp. 70–109). Philadelphia, Pennsylvania: University of Pennsylvania Press.
- Lafford, B. (1986). Valor diagnóstico-social del uso de ciertas variantes de /s/ en el español de Cartagena, Colombia. In R. Nuñez-Cedeño, I. Paéz, & J. Guitart (Eds.), *Estudios sobre la fonología del español del Caribe* (pp. 53–75). Caracas: La Casa de Bello.
- Lambert, W., & Taylor, D. (1996). Language in the lives of ethnic minorities: Cuban American families in Miami. *Applied linguistics*, 17, 477–500.
- Leal-Méndez, T., & Shea, C. (2012). L1 Mexican Spanish and L2 Spanish learners at the syntax-discourse interface: P-movement or in-situ prosody? Paper presented at the 42nd Linguistic Symposium on Romance Languages, Cedar City, UT, April 21, 2012.
- Lipski, John. (1996). Patterns of pronominal evolution in Cuban-American bilinguals. Spanish in contact: Issues in bilingualism, ed. by Ana Roca & John B. Jensen. Somerville, MA: Cascadilla Press.
- López García, Á. (1996). *Gramática del español*. Madrid: Arco Libros.
- Martínez, M., & Moya, J. (2000). Reacciones actitudinales hacia la variación dialectal en hablantes granadino. *LEA*, 22(2), 137–160.
- Meisel, J. (2001). From bilingual language acquisition to theories of diachronic change. *Working papers in multilingualism (Series B, Number 30)*. University of Hamburg: Collaborative Research Center on Multilingualism.
- Montrul, S. (2004). Subject and object expression in Spanish heritage speakers: a case of morphosyntactic convergence. *Bilingualism: Language and Cognition* 7, 125-142.
- Morales, A. (1989). Hacia un universal sintáctico del español del Caribe: el orden SVO. *Anuario de Lingüística Hispánica*, 5, 139–152.
- Morales, A. (1997). La hipótesis funcional y la aparición de sujeto pronominal: el español de Puerto Rico. *Hispania*, 80(1), 153–165.

- Morales, A. (1999). Anteposición del sujeto en el español del Caribe. In L. Ortiz (Ed.), *El Caribe hispánico: perspectivas lingüísticas actuales* (pp. 78–99). Madrid: Iberoamericana.
- Muntendam, A. (2009). *Linguistic transfer in Andean Spanish: Syntax or pragmatics?* Ph.D. dissertation, University of Illinois at Urbana-Champaign.
- Nava, E. H. (2007). Word order in bilingual Spanish: Convergence and Intonation Strategy. In J. Holmquist et al. (Eds.), *Selected proceedings of the third workshop on Spanish sociolinguistics*. Somerville, MA: Cascadilla Proceedings Project.
- Navas Sánchez-Elez, M. V. (1997). Factores lingüísticos y extralingüísticos que determinan la alternancia de las variantes de -“s” en un dialecto luso-español, el barranqueño. *Revista de filología románica*, 14(1), 391–410.
- Ocampo, F. (1995a). The word order of two-constituent structures in modern Spanish. In P. Downing. & M. Noonan (Eds.), *Word order in discourse*. Amsterdam/ Philadelphia: John Benjamins.
- Ocampo, F. (1995b). The word order of constructions with a verb, a subject, and a direct object in spoken Spanish. In J. Amastae, G. Goodall, M. Montalbetti & M. Phinney (Eds.), *Contemporary research in romance linguistics*. Amsterdam/ Philadelphia: John Benjamins.
- Ocampo, F. (1995c). Pragmatic factors in word order: Constructions with a verb and an adverb in spoken Spanish. *Probus*. 7.69-88.
- Ocampo, F. (2002). The word order of constructions with ser and estar, a subject NP, and an adjective in spoken Spanish. In J. Lee, K. Geeslin & C. Clemens (Eds.) *Structure, Meaning, and Acquisition in Spanish: Papers from the 4th Hispanic Linguistics Symposium*.
- Ocampo, F. (2005). The Word Order of Constructions with an Intransitive Verb, a Subject, and an Adverb in Spoken Spanish. In D. Eddington (Ed.), *Selected Proceedings of the 7th Hispanic Linguistics Symposium* (pp. 142–158). Somerville, MA: Cascadilla Proceedings Project.
- Otheguy, R. (1995). When contact speakers talk, linguistic theory listens. Meaning as explanation: Advances in linguistic sign theory, ed. by Ellen Contini-Morava & Barbara Sussman Goldberg. Berlin: Mouton de Gruyter.
- Otheguy, R., Rodriguez-Bachiller, B., & Canals, E. (2004). Length of the extra-information phrase as a predictor of word order: A cross language comparison. In E. Contini-Morava, R. Kirsner & B. Rodriguez-Bachiller (Eds.), *Cognitive and communicative approaches to linguistic analysis*. Amsterdam / Philadelphia: John Benjamins Publishing Co.

- Otheguy, R., & Zentella, A. C. (2012). *Spanish in New York: Language Contact, Dialect Leveling, and Structural Continuity*. New York, NY: Oxford University Press.
- Otheguy, R., Zentella, A.C., & Livert, D. (2007). Language and dialect contact in Spanish in New York: Toward the formation of a speech community. *Language*, 83, 770-802.
- Prince, E. (1981). Toward a taxonomy of given-new information. In P. Cole (Ed.), *Radical pragmatics*. New York: Academic Press.
- Poplack, S. (1980). “Sometimes I’ll start a sentence in Spanish y termino en español”: toward a typology of code-switching. *Linguistics*, 18(7/8), 581–618.
- Quirk, R., Greenbaum, S., Leech, G., & Svartik, J. (1972). *A grammar of contemporary English*. Essex: Longman.
- Raña-Risso, R. (2013). *A corpus-based sociolinguistic study of subject pronoun placement in Spanish in New York*. Ph.D. dissertation. The Graduate Center of the City University of New York, New York, NY.
- Rissel, D. (1989). Sex, attitudes, and the assibilation of /r/ among young people in San Luis Potosí, Mexico. *Language variation and change*, 1(3), 269–283.
- Samek-Lodovici, V. (2005). Prosody-syntax interaction in the expression of focus. *Natural Language and Linguistic Theory* 23:687-755.
- Samper Padilla, J. A. (1990). *Estudio sociolingüístico del español de Las Palmas de Gran Canaria*. Las Palmas: La Caja de Canarias.
- Schwegler, A., Kempff, J., & Ameal-Guerra, A. (2010). *Fonética y fonología españolas* (4th ed.). Hoboken, NJ: John Wiley & Sons.
- Silva-Corvalán, C. (1994). *Language contact and change: Spanish in Los Angeles*. Oxford: Oxford University Press.
- Silva-Corvalán, C. (1995). The study of language contact: An overview of the issues. *Spanish in four continents: Studies in language contact and bilingualism*, ed. by Carmen Silva-Corvalán. Washington DC: Georgetown University Press.
- Silva-Corvalán, C. (2001). Lenguas en contacto y bilingüismo – el español en los EE.UU. In *Sociolingüística y pragmática del español*. Washington, DC: Georgetown University Press.
- Swan, M. (2005). *Practical English Usage*. Oxford: Oxford University Press.
- Taverniers, M. (2005). Subjecthood and the notion of instantiation. *Language Sciences*, 27, 651-678.

- Toribio, A. J. (2004). Convergence as an optimization strategy in bilingual speech: Evidence from code-switching. *Bilingualism: Language and Cognition*, 7, 165-173. Vallduví, E. (1991), *The Informational Component*, New York, NY: Garland.
- Trager, G., & Smith, H. (1951). An outline of English structures. In *Studies in Linguistics occasional papers*. Norman, OK: Battenberg Press.
- Valdivieso, H., & Magaña, J. (1988). Variación lingüística: La /s/ implosiva en Concepción. *Revista de Lingüística Teórica y Aplicada*, 26, 91-103.
- Vallduví, E. 1991. The role of plasticity in the association of focus and prominence. In Y. No & M. Libucha (Eds.), *Conference on Linguistics* (pp. 295-306). Columbus, OH: OSU Press.
- Zagona, K. (2002). *The syntax of Spanish*. New York, NY: Cambridge University Press.
- Zapata, G. C., Sánchez, L. & Toribio, J. A. (2005). Contact and contracting Spanish. *International Journal of Bilingualism*, 9, 377-395.
- Zentella, A. C. (1990). Lexical leveling in four New York City Spanish dialects: Linguistic and social factors. *Hispania*, 73, 1094-1105.
- Zentella, A.C. (1997). *Growing up Bilingual: Puerto Rican Children in New York*. Malden, MA: Blackwell Publishers; Ch 9
- Zubizarreta, M. L. (1998). *Prosody, focus, and word order*. Cambridge, MA: MIT Press.