A CORPUS-BASED SOCIOLINGUISTIC STUDY OF SUBJECT PRONOUN PLACEMENT IN SPANISH IN NEW YORK

by

ROCÍO RAÑA RISSO

A dissertation submitted to the Graduate Faculty in Linguistics in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2013

This manuscript has been read and accepted for the
Graduate Faculty in Linguistics in satisfaction of the
dissertation requirement for the degree of Doctor of Philosophy.

Ricardo Otheguy

_____ _____

Date Chair of Examining Committee

Gita Martohardjono

_____ _____

Date Executive Officer

Gita Martohardjono

Virginia Teller

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

A CORPUS-BASED SOCIOLINGUISTIC STUDY OF SUBJECT PRONOUN PLACEMENT IN SPANISH IN NEW YORK

by

ROCÍO RAÑA RISSO

Adviser: Professor Ricardo Otheguy

This dissertation presents a variationist sociolinguistic study of the variable placement of subject personal pronouns before or after verbs in Spanish in New York City (e.g. *ella canta; canta ella*, both 'she sings'). It pursues a line of inquiry that partially replicates recent work by Otheguy & Zentella (2012). The research addresses, on the basis of a different morphosyntactic phenomenon, the same question of whether bilingualism in NYC is giving rise to language contact phenomena under which patterns of English usage are affecting the use of Spanish. In addition, this project expands the original database extracted from the Otheguy – Zentella corpus to include not only finite verbs but also non-finites, and it expands as well the tools of the analysis by introducing additional linguistic predictor variables.

This project will demonstrate that, with respect to the feature under study, there are indeed two distinct groups of newcomers entering the City, those from the Caribbean and those from the Latin American Mainland. It will also show that while continuity with Latin American ways of speaking remains the strongest force shaping Spanish in NYC, changes in the internal

grammar of the speakers begin with slight force in the first generation and continue in the second.

On the basis of extensive internal evidence, this project advances our knowledge of variables conditioning the placement of subject pronouns. It is shown that the influencing variables for Mainlanders coincide for the most part with those for Caribbeans; that the speech of Caribbeans is not as rigidly set on an SVO word order as it has been claimed in the literature; and that the speech of Mainlanders is more set on an SVO order as well, rather than on the required VSO word order claimed by several authors in the literature.

Lastly, the study demonstrates that the linguistic variables constraining pronoun placement in the speech of newcomers begin to change in significance and hierarchy in the first generation with more exposure to the City, and become insignificant for the second generation of speakers.

Acknowledgments

I'd like to thank my parents, Yuyo and Gustavo, for their guidance and unlimited support. They allowed me to try out as many professional paths as I was able to think of. *Gracias por todo el apoyo y las oportunidades, los quiero infinitamente.* I'd also like to thank my brother, Sebastián, for his support and for giving me my adorable bilingual niece and nephew. I wouldn't have been able to finish this long project without the unconditional love and support from Gonzalo, my favorite person in the whole world. And the two best pieces of me, Salvador and Cristóbal. Without you I would have finished way sooner!

I'd also like to thank my academic family. Virginia, thanks for always believing in me. Gita, you're an inspiration, a great role model for women in the program. And Ricardo, I can't thank you enough for your guidance, help, patience, and generosity. I hope we can continue to work together in the future.

I have met great people during the course of this project, who have helped me in different ways, most especially Xuan-Nga. I wish you were in NYC to celebrate with me.

Table of Contents

List of Tables

CHAPTER 1

INTRODUCTION AND BACKGROUND: SPANISH IN NEW YORK AND THE STUDY OF PRONOUN PLACEMENT

1. Introduction

The present work is a variationist sociolinguistic study of the variable placement of subject personal pronouns before or after verbs in Spanish (*ella llegó* 'she arrived,' ~ *llegó ella* 'she arrived'). The study takes its data from a corpus of Spanish as it is spoken in New York City and addresses matters of language contact. Additionally, several of its findings can be generalized to monolingual Spanish settings where this phenomenon has received only limited attention. In this introductory chapter I will provide an overview of the research presented in this dissertation. I will also present demographic information about the Latino community in New York City (NYC), taken from the 1990, 2000, and 2010 Census data[1].

This dissertation pursues a line of inquiry about the use of pronouns in Spanish in New York City (NYC) that partially parallels recent work by Otheguy & Zentella (2012). The research addresses the question of whether bilingualism in NYC is giving rise to language contact phenomena, under which patterns of English usage are affecting the use of Spanish in the City, and whether, in addition, the interaction of Spanish-speakers from different countries and regions is leveling out the dialectal differences that enter the City from Latin America. In the work just cited, these questions have been addressed through the analysis of variation between presence and absence of subject personal pronouns with finite verbs (as in *vino* 'she came' versus *ella vino*, also 'she came'.) In the present

research, a different but related phenomenon is analyzed, namely the placement of occurring pronouns either before or after the verb (as in *ella vino* versus *vino ella*, both 'she came'). In addition, the data base is expanded to include not only finite verbs but also non-finites.

The grammar of Spanish allows subjects to be preverbal, as in (1a), or postverbal, as in (1b). The possibility to alternate between preverbal and postverbal overt subject pronouns is found, to my knowledge, in all speakers of Spanish.

(1) a. *Ella* ***es*** *mayor de edad.* [234U]

'She is of age'

b. ***Estaba*** *yo aquí estudiando.* [317M]

'I was studying here'

At the beginning of this project, it was hypothesized that this alternation would respond to linguistic-internal and speaker-external conditioning variables, such as, for example, the type of clause where the verb is found and the immigrant generation to which the speaker belongs. The novelty of the project lies in the fact that the factors that affect pronoun placement in Spanish have not been fully analyzed either in Latin America or Spain, nor in the contact Spanish of the U.S. In addition to elucidating the role of internal and external conditioning factors on pronoun placement, the project will help to further analyze the role of English, as well as of ways of speaking Spanish other than one's own, in shaping Spanish in the multilingual and multidialectal context of New York.

2. Spanish in New York City

According to the 2010 Census data, the Latino population of NYC constitutes 29.1 percent of the City's total population. There are six leading groups that conform 85 percent of the total Latino population for this year: Puerto Ricans (31 percent), Dominicans (25 percent), Mexicans (14 percent), Ecuadorians (9 percent), Colombians (4 percent), and Cubans (2 percent). In this introduction, all data presented is based on these six groups of Latinos only, which I have termed 'Major Latino Groups'. These numbers show that 58 percent of NYC's Latino population is from the Caribbean and 42 percent from the Latin American Mainland.

In order to explore the language usage trends of speakers in Major Latino Groups in New York City, I looked at the numbers of people who reported speaking Spanish at home according to the 1990, 2000, and 2010 Census.

Table 1.1

Population in major Latino groups 5 years of age and over that speaks Spanish at home per Census year

Census Year	**Percent**
1990	82
2000	79
2010	76

Although the percentage of the total Latino population that reports speaking Spanish at home is slowly decreasing (from 81.5 percent in 1990 to 76.2 percent in 2007), it can be concluded that the Latino population of NYC is still maintaining its heritage language since the decrease (5.3 percent) in the use of Spanish at home in the last 20 years has

been small. However, when I looked at each of the six leading groups individually, I noticed that there are differences among the groups. While use of Spanish at home has remained stable for Dominicans, Ecuadorians, and Colombians, it has significantly decreased for Cubans (from 82 percent in 1990 to 55.6 percent in 2010) and Puerto Ricans (from 79 to 65.4 percent), and increased for Mexicans (from 73 to 75.5 percent). One way of explaining these findings is by looking at the distribution of the population, where it can be noticed that while the population of Mexicans more than tripled in 20 years (from 4 to 14 percent), the population of Cubans was cut in half (from 4 to 2 percent), and the population of Puerto Ricans also decreased significantly (from 58 to 31 percent). Therefore, it seems that an increase in the number of speakers from a particular nationality correlates with maintenance of Spanish in that subgroup, while a decrease in the number of speakers correlates with assimilation to English in that subgroup.

Furthermore, if we classify each group by their place of birth (domestic or abroad) as shown in Table 1.2 below, we see that except for Puerto Ricans, whose population is two-thirds domestic born and one third foreign born, and Cubans who are half domestic born and half foreign born, every body else is the opposite: two-thirds foreign born, one third domestic born, which means that these groups are continuously receiving new immigrants, which helps maintain Spanish alive in the city. The results presented in Table 1.1 also help illustrate the reason why although Spanish is very much alive in the city, the number of speakers is slowly decreasing: the number of foreign-born speakers is also slowly decreasing for every single group as the years go by.

Table 1.2

Distribution of the total population in major Latino groups born domestically or abroad by census year

Major Latino Groups	Place of Birth	1990	2000	2010
Mexican	Domestic Born	34%	32%	42%
	Foreign Born	66%	68%	58%
Ecuadorian	Domestic Born	24%	25%	33%
	Foreign Born	76%	75%	67%
Colombian	Domestic Born	21%	23%	34%
	Foreign Born	79%	77%	66%
Cuban	Domestic Born	28%	42%	55%
	Foreign Born	72%	58%	45%
Puerto Rican	Domestic Born	58%	63%	71%
	Foreign Born	42%	37%	29%
Dominican	Domestic Born	30%	31%	36%
	Foreign Born	70%	69%	64%

I also used census data to investigate whether those who speak Spanish at home also speak English well, or whether use of Spanish at home correlates with lack of skills in English. As illustrated by Table 1.3 below, I found that approximately 70 percent of the total population in major Latino groups that speaks Spanish at home, also reports

speaking English well or very well, 20 percent reports speaking English but not well, and only 10 percent reports not speaking English. Furthermore, this situation has remained stable throughout the last 20 years. There are two interesting comments to make about these results: 1) use of Spanish at home does not seem to negatively affect the development of English language skills in the Latino population; 2) the vast majority of the Latino population appears to be bilingual.

Table 1.3

Level of English Skills among those who speak Spanish at Home ages 5 and over in Major Latino Groups

English skills	**1990**	**2000**	**2010**
Yes, speaks very well	49%	48%	50%
Yes, speaks well	23%	22%	19%
Yes, but not well	19%	20%	22%
Does not speak English	9%	10%	10%

The following conclusions have arisen from the analysis of the 1990, 2000 and 2010 census data:

- The Latino population of NYC is maintaining its heritage language.
- 58 percent of NYC's Latino population is from the Caribbean and 42 percent is from the Mainland.
- The ongoing arrival of Spanish-speaking newcomers helps maintain Spanish alive in the City.
- 70 percent of NYC's Spanish-speaking population also speaks English well.

Spanish dialectal studies have consistently separated the Caribbean and the Latin American Mainland into two separate dialectal zones based on a number of phonological and morphosyntactic features (López-Morales, 1992). The census data presented above shows that the majority of Spanish-speakers in NYC come from one of these two dialectal zones, and that the majority of New York Latinos are bilingual. This validates the research questions set forth above in terms of dialect contact and language contact, and it shows that New York City is an ideal environment to explore these phenomena: a) there is a large group of Spanish-speakers from the Mainland interacting with an even larger group of Spanish-speakers from the Caribbean, giving rise to contact between these two dialects of Spanish; b) 70 percent of NYC's Spanish-speakers report being balanced bilinguals, giving rise to contact between Spanish and English in those speakers, possibly leading to an influence of English on the Spanish spoken in the City.

3. The study of subject pronoun placement in Spanish in New York City

Data for this study of variable pronoun placement has been drawn from the Otheguy-Zentella corpus of Spanish sociolinguistic interviews conducted in New York[2]. Its accompanying database identifies and codes for several internal and external variables, including a coding of all finite verb tokens that participate in the alternation between present and absent pronouns. The present project uses this existing data and the accompanying coding and expands it so as to include new data from the corpus as well as new codings. A list of the socio-demographic and linguistic variables included in the present study is provided in chapter 2.

There were originally two lines of expectations in the present research: one stemming from the dialect leveling hypothesis and one from the language contact

hypothesis. First, I expected my analysis to confirm that there are in fact dialectal differences in the speech of newcomers to NYC from Latin America. The expectation was that there would be more preverbal pronouns in the speech of Caribbeans than in that of Mainlanders (Morales 1997, 1999; Toribio 2000; Ordoñez and Olarrea, 2001). However, I also expected that as these newcomers are exposed to speakers from other parts of Latin America, these initial dialectal differences would, with the passing of time, tend to be leveled out. Second, with the passage of time spent in New York, as Spanish-speakers become more and more exposed to English, I expected that patterns of English usage would affect their usage of Spanish, leading to an increase not only in the rate of occurring pronouns (the pronoun rate), as shown by Otheguy & Zentella (2007, 2012) and Otheguy et al. (2007, 2008), but also an increase in the rate of preverbal pronouns (the preverbal rate). Both lines of expectations have been explored in this dissertation by means of two basic statistical inquiries that characterize variationist sociolinguistic research, namely the occurrence rate for the phenomenon in question and the hierarchies of variables and constraints that are statistically associated with, and predictive of, the phenomenon.

In addition to studying the preverbal rate and its predictive constraints, a running comparison is drawn in the present study between the pronoun rate and the preverbal rate. I will establish in the present research that there is a strong correlation between the rate of occurrence of preverbal and overt pronouns, which means that those individuals who use more preverbal pronouns tend to be the same as those who use more overt pronouns (as first noted in Raña Risso, 2010:106). I will also show that there is a distinction in the placement of overt pronouns between the two regions under consideration, evidenced by

the speech of newcomers, to wit speakers from the Caribbean use more overt and more preverbal pronouns, and speakers from the Mainland use significantly fewer overt and fewer preverbal pronouns. This will demonstrate that the Caribbean and the Mainland constitute two separate dialectal zones for the feature under study, and also that, on this feature, there are two distinct dialects of Spanish interacting in NYC.

In the context of generative parameter theory, I will discuss the connection between the possibility for a language to have null subjects and the possibility of having postverbal subjects. I will investigate whether this theoretical perspective can help to offer an account for the more frequent preverbal order in the speech of Caribbeans than of Mainlanders, given the higher rate of overt pronouns among Caribbeans.

With regard to exposure to the New York environment, the dissertation will show that there is a significant increase in the rates of preverbal pronouns with more years in the City, especially in Caribbean and second-generation speakers. This increase parallels the increase in overt pronouns, a fact already established in the literature. As with the increase in overtness, which has been attributed to the influence of English, I will propose an explanation for the increase in preverbal pronouns in terms of language contact. The speakers' exposure to English, where preverbal rates approach categoricity, influences their use of preverbal subject pronouns in Spanish. However, I will also show, through the use of regression analysis, that despite the changes taking place in pronominal use in NYC, the region of origin continues to be the strongest force shaping pronominal placement in Spanish in New York.

---

[1] The demographic information on this chapter is taken from a report created to fulfill the requirements of a fellowship given by the Center for Latin American, Caribbean & Latino Studies at the Graduate Center, CUNY. The author is grateful to Professor Laird W. Bergad for providing the census data, as well as his assistance and comments.

[2] The Otheguy-Zentella corpus contains 140 interviews conducted in New York in Spanish with speakers who trace their origins to six different Latin American countries (Colombia, Cuba, Dominican Republic, Ecuador, Mexico, Puerto Rico). The participants belong to different immigrant generations and, among those of the first generation, have different numbers of years of residence in New York. The interviews were conducted beginning in the year 2000, by Ricardo Otheguy (City University of New York), Ana Celia Zentella (University of California, San Diego), and their graduate assistants, with support from the City University of New York and the National Science Foundation (NSF grant number BCS 0004133). The author is grateful to Professors Otheguy and Zentella for allowing the use of data from the corpus.

CHAPTER 2

DATA, METHODOLOGY, AND VARIABLES IN THE STUDY

1. Introduction

The previous chapter described the socio-demographic context of New York City, where the research took place**,** and provided a general sketch of the informants in the study as well as a description of the variable linguistic phenomenon under investigation. This chapter describes the origin of the data, and the data itself, as well as the methodology used throughout the investigation. It also lists all dependent and independent variables in the study with their corresponding factors.

2. Data: The Otheguy – Zentella Corpus

The data for this project was extracted from the Otheguy – Zentella corpus (henceforth referred to as 'the corpus' or 'the original corpus') which contains over a million words and consists of 140 transcripts of sociolinguistic interviews conducted in NYC, lasting between 60 and 90 minutes each. The corpus was constructed mostly in the period between 2000 and 2004 at the Graduate Center, CUNY, by a team of researchers from the Research Institute for the Study of Languages in Urban Society (RISLUS)[1]. From these sociolinguistic interviews, separately stratified samples of Colombians, Dominicans, Ecuadorians, Mexicans, Puerto Ricans, and Cubans living in NYC were created. The stratification criteria used to create the sample were, within each national group, self-reports of: (a) Place of birth (Latin America or NYC), (b) age, (c) age of arrival in NYC (for informants born in the city, age of arrival is zero), (d) years of residence in NYC (for informants born in the city, years in NYC is equivalent to their age, with some exceptions to account for intervening years spent back in their parents' country of origin), (e) levels of education, (f) level of English skills, (g) social class, (h) extent of use of Spanish in

general and in specific domains, (i) extent of use of Spanish with Latinos from one's own country, and (j) extent of use of Spanish with Latinos from countries other than one's own. For additional information regarding the data see Otheguy & Zentella (2012).

In the present dissertation, all analyses were actually conducted with 139 informants, instead of 140. One informant was eliminated because his preverbal rate was significantly different from the rest of the informants (he had, by far, the smallest preverbal rate). Upon closer inspection, it turned out that he had spent his life traveling back and forth between Mexico and the U.S., living in each country for short periods of time, which made it harder to fit him into some of the socio-demographic groups investigated in the following chapters.

Previous research using this corpus has focused on finite verbs occurring either with or without pronouns. The present study focuses only on finite verbs that occur with a pronoun, either preverbal or postverbal. For some of the analyses here the entire corpus (minus the aforementioned informant) will be used, as described in the Methodology section below. For other analyses, a portion of the corpus will be used, which will be referred to as the subset corpus.

3. Data: The Subset Corpus

The need for a subset corpus arose because two of the structures investigated here, interrogative sentences and non-finite verbs, had not been fully extracted from, or systematically coded in, the original corpus. These two structures needed investigating because they were consistently mentioned in the literature as playing a role on subject pronoun placement. Therefore, the initial work of this dissertation consisted of selecting a stratified subset of interviews from the original corpus, and then extracting from there all additional verb tokens relevant for the present project that had not been previously entered in the database. For this

subset corpus, a set of 22 interviews containing approximately 200,000 transcribed words were selected. These subset interviews were chosen to represent three levels of exposure, namely participants born or raised in NYC, newcomers, and immigrants, and two regions, namely the Caribbean (Cuba, Puerto Rico, Dominican Republic) and the Latin American Mainland (Mexico, Colombia, Ecuadorians). Since only finite verbs had been extracted from the original corpus, early work in the present project consisted of extracting all non-finite verbs occurring with an overt pre- or post-verbal pronoun. Furthermore, the speech from the research interviewers, which was not part of the original extraction, was also analyzed, since it is naturally in the speech of interviewers where most of the interrogative phrases are found. The speech of interviewers thus provided additional finite and non-finite verbs to be included in the present project's database.

4. Methodology

For the analysis of this dissertation's data I have mostly used the Statistical Package for the Social Sciences (SPSS) as well as small programs I developed using the Perl programming language in order to look for additional linguistic patterns in the corpus[2]. The latter step led to the creation of the aforementioned subset corpus, which is described in more detail in chapter 7.

The analysis of the variable placement of overt subject pronouns described in the first four chapters of this dissertation (chapters 3 through 6), was conducted using a dependent variable that I called the preverbal rate. The preverbal rate is the percentage of finite verb tokens, out of all finite verb tokens occurring with an overt pronoun, that are found with a preverbal subject pronoun. These four chapters are a partial replication of the study conducted by Otheguy & Zentella (2012) for the analysis of the variable use of present and omitted subject pronouns. While the work of those two scholars focused on analyzing the social and linguistic variables that play a role in the speakers' choice between presence and absence of subject pronouns, the

study presented in this dissertation focuses on overt (present) pronouns, and studies the social and linguistic variables motivating speakers to place them either before or after verbs. Therefore, following the structure of the study under replication, statistical analysis is used to explore New York Spanish-speaking communities through the preverbal rate (chapter 3), guided by the language contact hypothesis (chapter 4), the dialect leveling hypothesis (chapter 5), and the possibility of predicting the occurrence of preverbal pronouns based on these findings (chapter 6). For exploring the communities in NYC, ANOVAs were calculated for each of the potential groups. For researching language contact, ANOVAs are used again to analyze the mean rate of preverbal pronouns of the regional, generational, exposure, and language skills groups. For the exploration of the dialect leveling hypothesis, several statistical measures are used to compare exposure and regional groups, as well as ANOVASs. Then, for predicting the occurrence of preverbal pronouns in NYC, the analysis moves from bivariate to multivariate, through the use of multiple regression as it was done in the study under replication.

For the exploration of internal variables playing a role in the variable placement of overt subject pronouns, the investigation moves to the subset corpus. From this subset corpus, which contains 1740 verb tokens, both finite and non-finite verbs occurring with an overt pronoun are used, taken both from the speech of interviewees as well as interviewers. One of the main differences between chapters 3 – 6 and chapter 7 is that two different dependent variables are used. In chapters 3 – 6, I look at a property of informants, encoded in terms of their preverbal rates, and use this rate as the dependent variable. In chapter 7, I look at a property of verb tokens, consisting of whether they are accompanied by a preverbal or a postverbal subject pronoun, and I use Placement as the dependent variable. For this reason, the co-varying independent variables of chapters 3 – 6 are based on personal and demographic properties of individuals (Region,

Exposure, etc.) whereas the co-varying independent variables in chapter 7 are grammatical properties associated with each verb token (Clause, Transitivity, etc.). With regards to the statistical analysis, while the study of social variables in chapters 3 – 6 is conducted at the informant level, the analysis of grammatical variables of chapter 7 is conducted at the item level. Another difference is that the dependent variable Preverbal Rate is a continuous variable, that is, a percentage. The dependent variable Placement in chapter 7 is a categorical variable, that is, the placement of a pronoun before or after a particular verb token. In the bivariate analysis of chapter 7, I use crosstabulations to identify significant variables. Then, for the multivariate analysis, I use logistic regressions to build variable and constraint hierarchies.

5. Variables in the study

The independent predictors in chapters 3 – 6 are twelve socio-demographic or external variables. They are listed below, along with their values or factors.

Dependent variable

Preverbal Rate

Independent variables

Informant's region of origin

Caribbean

Mainland

Informant's national origin

Colombia
Dominican Republic
Ecuador
Mexico
Puerto Rico
Cuba

Informant's age of arrival to the U.S.

0 for informants born in the U.S.

Corresponding number for all others

Informant's number of years in the U.S.

Their age for informants born in the U.S.

Corresponding number for all others

Informant's exposure to NYC

Latin American Raised (LAR)

New York Raised (NYR)

Informant's exposure to NYC

LAR Immigrant Newcomers

LAR Established Immigrants

New York Raised (NYR)

Informant's sex

Female

Male

Informant's socioeconomic status

Lower SES

Higher SES

Informant's level of education

High School or less

College or more

Informant's Spanish Dialect

Reference Spanish

Contact Spanish

Informant's level of Spanish skills

Spanish less than excellent

Spanish excellent

Informant's level of English skills

English less than excellent

English excellent

Details on, and the motivation for, each one of these variables are offered when results are presented in chapters 3-6.

For the exploration of internal variables conducted in Chapter 7 on the basis of the subset corpus, the following dependent and independent variables were identified:

Dependent variable

Pronoun Placement Type (Placement)

Pronoun is preverbal

Pronoun is postverbal

Independent variables

Verb morphology: Verb Finiteness Type (Finiteness)

Pronoun appears with a finite verb

Pronoun appears with a non-finite verb

Sentence structure: Clause Type (Clause)

Pronoun is in a main clause in a simple sentence

Pronoun is in a main clause in complex sentence

Pronoun is in a subordinate clause

Pronoun is in a coordinate clause

Sentence structure: Subordinate Clause Type (Subordinate)

Pronoun is in an Argument/Nominal Clause

Pronoun is in a Relative/Adjectival Clause

Pronoun is in a Temporal Adjunct Clause

Pronoun is in a Gerundive and Participial Adjunct Clause

Pronoun is in an Infinitival Adjunct Clause

Sentence Structure: Relative Clause Type (Relative)

Pronoun is in a Relativization of Direct Object Clause

Pronoun is in a Relativization of Indirect Object Clause

Verbal Syntax: Verb Transitivity Type (Transitivity)

Pronoun appears with a Transitive verb

Pronoun appears with a Intransitive verb

Pronoun appears with a Ditransitive

Sentence Structure: Phrase Type (Phrase)

Pronoun is in a Declarative Sentence

Pronoun is in a Non-declarative Sentence

Sentence Structure: Interrogative Phrase Type (Interrogative)

Pronoun is in a yes/no direct question

Pronoun is in a Wh question

Sentence Structure: Wh Question Type (Wh Question)

Pronoun is in a Bare Wh question

Pronoun is in a Non-bare Wh question

Pronoun is in an indirect question

Lexical Category: Pronoun Type (Pronoun)

Pronoun is *yo*

Pronoun is *tú*

Pronoun is *él, ella*

Pronoun is *nosotros, nosotras*

Pronoun is *vos*

Pronoun is *ellos, ellas*

Pronoun is *usted*

Pronoun is *ustedes*

Pronoun is *uno, una*

Pronoun is *unos, unas*

Pronoun Morphology: Pronoun Gender Type (Gender)

Pronoun is Feminine

Pronoun is Masculine

Gender is Not applicable

Details on and the motivation for each one of these variables are offered when results are presented in chapter 7.

---

[1] This corpus was developed with grants from the National Science Foundation (BCS 0004133), the City University of New York (09-91917), and CUNY's Professional Staff Congress (62666-00-31).

[2] The Perl programs developed to automatically extract patterns from the subset corpus are included in the Appendix.

CHAPTER 3

AN EXPLORATION OF NEW YORK SPANISH-SPEAKING COMMUNITIES THROUGH THE PREVERBAL RATE

1. Introduction

In the Spanish of New Yorkers, there is variability between preverbal and postverbal placement of subject personal pronouns, much as there is in the rest of the Spanish-speaking world. There are no sub-communities in the City whose verbs occur always with a preverbal pronoun, or always with a postverbal pronoun, but the proportions vary considerably between communities. Given this variability, a first investigation into this issue had to involve finding out which are the sub-communities of Latino New Yorkers that can be isolated on the basis of subject personal pronoun placement. To that end, I have availed myself of a measure that I will call the PREVERBAL RATE. The preverbal rate is the percentage of eligible finite verb tokens that appear in constructions with a preverbal subject pronoun in the corpus. There follows a passage from one of the transcripts that will help illustrate the preverbal rate.

Ella lo quería mucho, y entonces después se trajo a Socorro, y entonces **ellos ya constantemente me estaban** informando de que era mejor, que acá **tenía yo** mejor porvenir que en Colombia, y al cabo del tiempo, pues su mamita se murió, ¿no?, y **yo me vine** al lado de los dos muchachos, de Socorro y de Fabio, ¿no?, pues… cuando **ella murió.** Estaba donde su tía Lía, como su tía Lía no podía, mejor dicho, verlos porque se mantenía trabajando, por lo mismo el marido, Luis, se mantenía muy enfermo, porque era alcohólico, ¿no? Ahí fue cuando los muchachos me escribieron, ahí fue cuando me vine, y después con lo… con el tiempo ya **reaccioné yo** y me… me quedé aquí, no quise volver más porque vi que había mejor porvenir aquí que en Colombia. [174C]

She loved him very much, and then she brought Socorro, and then they were constantly telling me that it was better, that here I would have a better future than in Colombia, and after a while, her mother died, right? And I came to be with these guys, with Socorro and Fabio, right? That's… when she died. She was at her aunt Lia's, as her aunt Lia couldn't, I mean, see them because she was always working, and her husband, Luis, was always sick, because he was an alcoholic, right? That's when the guys wrote to me, that's when I came, and afterwards with… with time I reacted and … I stayed here, I never wanted to go back because I realized there were more possibilities here than in Colombia.

The passage above, corresponding to an informant from Colombia, contains six verb tokens, of which four appear with a preverbal pronoun (the other two verbs appear with postverbal pronouns). The preverbal rate for this passage is 4/6, or 67 percent. This is how the preverbal rate is calculated for all eligible verbs in all transcripts. This allows me to group individuals by their national origin, generation, age, etc. and compare their preverbal rates to investigate whether it makes sense to group them under any of these criteria.

In order to establish which groups of individuals have significant differences in their preverbal rates to merit being grouped together, I first ascertain the preverbal rate for all speakers in the corpus (see Table 3.1).

Table 3.1
Percent of verbs found with preverbal subject pronoun

	N verbs	Percent
Preverbal	20269	**95.5**
Postverbal	924	**4.5**
Total	21193	**100**

The 139[1] speakers in the corpus produced 21,193 finite verbs with an overt subject pronoun, of which 20,269 (95.5 percent) appeared with a preverbal subject pronoun, and 924 (4.5 percent) with a postverbal subject pronoun. This table shows that in New York, as in all forms of Spanish, the vast majority of subject personal pronouns are placed before verbs. Although there may be differences by communities of speakers, given the stratified and diverse nature of the corpus of interviews that I am using, it is probably safe to extrapolate our sample findings to our population, and conclude that 95 percent of all

subject personal pronouns with finite verbs in New York appear with a preverbal pronoun.

In the rest of this chapter, and in addition to investigating whether it makes sense to group individuals based on their gender, age, and socioeconomic status, I will establish whether the preverbal rate allows us to group individuals based on their place of origin, as it has been suggested in the literature (Lipski 1994:241, López-Morales 1992:137, Ordoñez & Olarrea 2001, Toribio 2000, Zagona 2002, Goodall 2004, Otheguy & Zentella 2012). This will provide a foundation to explore, in the following chapters, the possibility of language contact and dialectal leveling. Furthermore, given that the corpus used for this project is essentially the same as in Otheguy & Zentella's (2012) investigation of the pronoun rate, I will be able to compare the results based on both measures.

2. The use of preverbal pronouns by basic demographic groups

2.1 Gender, age, and education

ANOVAs were calculated for each of the potential groups. This analysis of the mean rate of preverbal pronouns of each demographic group yields the *F* ratio, which provides an indication of the proportion of variance across the groups to internal variance within each group. In the study of any particular demographic independent variable, an *F* close to 1 indicates that there is no effect of the demographic factor on the placement of pronouns, whereas a large *F* indicates an effect of the demographic factor on pronoun placement. With respect to co-variation between the preverbal rate and the four traits under study in this section, results were not statistically significant for two of the three factors: Gender ($F = 0.66$, $p < .41$), and Age ($F = 0.66$, $p < .57$). The very low *F* values indicate that in these two cases variance between the groups is smaller, or the same, as

variance within the groups, i.e., the groups are not different. With respect to the preverbal rate, no evidence is found that there are sub-communities in Spanish-speaking New York made up of men or women, younger or older speakers. From the evidence provided by our sample, it appears that members of these different groups in the City's Latino population, when taken as a whole, are not distinguishable on the basis of their preverbal rates.

These results are in line with the findings of the study under replication for the same factors with regards to the pronoun rate. However, the two studies differ in their findings with regards to Education. While Otheguy & Zentella (2012) found that the education groups were indistinguishable from each other with regards to the pronoun rate, I found that the level of education of the speakers plays a significant role in their use of preverbal pronouns (see Table 3.2).

Table 3.2

Preverbal rate by education

Education	N Speakers	Preverbal Rate
High school or less	58	**93**
College or more	80	**96**
	138	

F = 6.85
p = 0.01

Table 3.2 shows that the difference in mean rates of preverbal pronouns between the two education groups is statistically significant. The higher the level of education of the speakers, the more preverbal pronouns they have. Given this finding, I will further study the level of education of the speakers when I divide the sample into sub-groups based on the speaker's regional origin, to see if this difference stays valid for all groups.

As an initial explanation, it could be the case that for all or most speakers living in the U.S., those who completed high school and went into college have a high level of proficiency in English and that their grammar of Spanish is affected by the grammar of English. Given that English is almost exclusively an SVO language, then the speakers' high proficiency in English could be affecting their grammar of Spanish, leading to more preverbal subjects in Spanish. If this were indeed the case, then the Education results may be a consequence of the more educated having more English than the less educated, which in turn would mean that the real grouping is not More Educated versus Less Educated, but Higher Proficiency in English versus Lower or No Proficiency in English. This issue is investigated in the following chapter, where I will group speakers by their skills in English and investigate the incidence that this factor has on the preverbal rate.

2.2 Socio-economic status (SES)

An additional analysis of variance on the preverbal rate with respect to the speakers' socio-economic status (SES) was conducted. As it was noted in Otheguy & Zentella (2012) the classification of speakers as working class or middle class is the result of the speaker's self-report. Therefore, it is not entirely reliable and that was the reason for creating a composite SES measure based on the speakers' education and occupation[2]. In the case of Otheguy & Zentella's investigation for the pronoun rate, it was found that this more abstract measure showed that speakers of lower SES had higher pronoun rates than those of higher SES. However, in the present study of preverbal rates with the same speakers, this measure worked in the opposite direction: speakers of lower SES have lower preverbal rates and speakers of higher SES have higher preverbal rates (see Table 3.3).

Table 3.3

Preverbal rate by socioeconomic status

SES	N Speakers	Preverbal Rate
Lower SES	58	**93**
Higher SES	78	**95**
	136	

F = 4.05
p = 0.04

Table 3.3 shows that the group of speakers reporting a higher socio-economic status presented slightly more preverbal pronouns (95 percent) than the group of speakers reporting a lower socio-economic status (93 percent). This difference in mean rates of preverbal pronouns between the two socio-economic groups was statistically significant. The *F* value of 4.05 indicates that variance is more than four times greater across the two SES groups than within each group, and the *p* value below .05 shows that these results are revealing information about the New York Latino population, which will be discussed again below and in later chapters.

3. ***National and regional origin of speakers***

In this section the goal is to investigate whether the geographic origin of the speakers is related to the use of overt pronouns. More specifically, I wanted to know whether, on the basis of the preverbal rate, there are sub-communities of Latino New Yorkers that can be defined by where in Latin America the speakers or their parents lived before coming to New York. The Spanish-speaking New Yorkers in the corpus were divided into six groups by their COUNTRY OF ORIGIN: Colombians, Cubans, Dominicans, Ecuadorians, Mexicans, and Puerto Ricans (Table 3.4a).

Table 3.4a
Preverbal rate by country of origin

	N speakers	Preverbal rate
Puerto Ricans	24	98
Dominicans	24	96
Cubans	24	96
Ecuadorians	24	94
Colombians	21	94
Mexicans	22	89
	139	

F = 8.62
p < .00

Table 3.4a lists in descending order the average preverbal rate for each of the groups of speakers from each country, showing that Puerto Ricans have the highest preverbal rate and Mexicans the lowest. On average, 98 percent of verbs used by Puerto Ricans occur with a preverbal pronoun, whereas only 89 percent of verbs used by Mexicans have a preverbal pronoun. The ANOVA indicates that the overall country difference in mean rates of preverbal pronouns was statistically significant at $p < .001$.

In addition to learning about overall significance, the question arises of which of the differences between adjacent countries is significant. For example, is the Dominican average rate of 96 percent statistically different from the Puerto Rican average rate of 98 percent? Based on a test of pair-wise significance, Table 3.4b marks with an asterisk the intersections between countries that are adjacent to each other in Table 3.4a whose preverbal rate differences are statistically significant, and leaves blank the intersections where the rate differences are not significant.

Table 3.4b
ANOVA, Post-hoc Tukey test of pair-wise significance

	PR	Dom	Cu	Ecu	Col	Mex
Puerto Ricans					*	*
Dominicans						*
Cubans						*
Ecuadorians						*
Colombians	*					*
Mexicans	*	*	*	*	*	*

Asterisks indicate a significant difference in preverbal rate

The data in these tables assists in the decision of whether there is justification for establishing a six-way characterization of Latino New Yorkers based on the association between the preverbal rate and their country of origin. Table 3.4b indicates that a nation by nation grouping is not justified. Rather it indicates that Mexicans are clearly distinct from everybody else and also that Colombians are clearly distinct from Puerto Ricans.

Comparing these results on the preverbal rate to those of the overt rate in the study under replication, we find that our findings go in the same direction but are not as clear cut. Similar to my findings, Otheguy & Zentella (2012) found no justification for a country by country grouping, but were able to justify a two-way distinction between Caribbeans (Cubans, Dominicans, and Puerto Ricans taken together) versus Mainlanders (Colombians, Ecuadorians, and Mexicans taken together). In contrast, in the present study of the preverbal rate, the Caribbean is more integrated, but the Mainland less so, since one of the countries, Ecuador, is, in strictly statistical terms, behaving differently from the other two. Still, given the results in Table 3.4a where the preverbal rate for Ecuadorians is almost the same as that for Colombians and represents a clear drop from the Caribbean rates; given too the well-known existence of substantial numbers of phonological features that do distinguish the three Caribbean countries from most of

Colombia, Ecuador, and Mexico; given also the widespread cultural acceptance of the Caribbean vs Mainland division; and given the findings on the overt rate in the work that I am replicating, I have decided not to pay attention to the small problem of our post-hoc findings on Ecuadorians in Table 3.4b, and will divide the sample into the two conventional regional groups, placing the Ecuadorians within the Mainlanders. That is, when all these considerations are kept in mind, Tables 3.4a and 3.4b support an interpretation that maintains the existence in Spanish in New York of a national continuum of preverbal rate differences, and of a sharper distinction between two regional groups (Caribbeans and Mainlanders), and two countries at the opposite ends of the continuum (Puerto Rico and Mexico).

The appropriateness of groups created on the basis of regional differences for capturing differences of preverbal rate in the Latino population of New York is shown in Table 3.5.

Table 3.5

Preverbal rate by region of origin

	N speakers	Preverbal rate
Caribbean	72	96
Mainlanders	67	92
	139	
F = 22.03	p < .00	

The table shows that our 72 Caribbean speakers use preverbal subject pronouns at a substantially higher rate (96 percent of their verb tokens are found with a preverbal pronoun) than do our 67 Mainlanders (92 percent of their verbs appear with a preverbal pronoun). The table also shows that differences of preverbal rate across the two groups

are 22 times greater than within each group ($F = 22.03$) and that these differences are statistically significant ($p < .001$). These results indicate that the categorization of speakers by region, in part based on Tables 3.4a and 3.4b and in part established deductively, is also inductively justified by the facts of pronominal distribution in the sample. Based on the evidence in Tables 3.4a, 3.4b, and 3.5, we can safely conclude that, for the feature under study, Caribbeans and Mainlanders very likely constitute two sub-communities of Spanish-speakers in New York City.

In order to qualitatively illustrate the differences between the Caribbean and the Mainland, two passages from informants from both regions are presented below. These informants are typical in that their overall preverbal preverbal rates are the same as the average for their respective regions; the Dominican in 1a has an overall personal rate of 96 percent, while the Ecuadorian in 1b has an overall personal rate of 92.

1a. [About learning to milk cows] **yo rebalé** que venía con con una lata de lecho como con quince o diecisei litro(s) de leche, y rebalé y se botaron todas y nosotro(s) "¿Ay y cómo **nosotro(s) vamo(s) a llegar** a la casa que hoy me mata mi mamá" y "Ay cómo lo vamo(s) a hacei cuando llegamo(s), **nosotro(s) le dijimo(s)** a mi mamá lo que había pasao y **ella no no(s) lo creía** y entonce mi papá dijo, "Pero que eso muchachito(s) no se podían mandar hoy a ordenai hoy, ¿Por qué no-no me llama(s)te pa yo ir o **fui(s)te tú**, e-e- eso fue que tú **tú sabe(s)** como e(s) que está el día y eso, hoy no (es)taba de eso muchachito(s) ir a ordeñai." [125D]

I slipped and I was coming with a tin of milk with fifteen or sixteen liters of milk, and I slipped and they were all spilled and we "how are we going to get home, my mom will kill me today" and "What are we going to do when we get there?", we told my mom what had happened and she didn't believe it and then my dad said, "But these boys should not have been sent to milk cows today. Why didn't you call me so that I could go or you go, you know how the day is today, it was not a good day for the boys to go milk."

1b. [About visiting their home country] y **yo he sido** una persona que a mí no me ha gustado.. no me gusta ir a mi país porque **creo yo** que la gente que van a hacer eso, van endeudándose de aquí, a gastar un dinero que no tienen y luego vienen aquí y se encuentran con la deuda que han dejado. Y **yo no he sido**.. no **he participado yo** de esas ideas, **yo he ido** con lo que **yo he tenido**, y **yo nunca he hecho** ostentación de nada. [326E]

And I have been a person who hasn't liked… doesn't like to go to my country because I think those who do that, get in debt here to do that, to spend money that they don't have and then they come back here and they find themselves with the debt that they left behind. And I haven't been… I haven't shared those ideas, I have gone with what I had, and I have never been ostentatious with anything.

4. Regional differences in preverbal rate by the basic demographic groups

It was seen above that, in the whole sample, there are no differences with respect to preverbal rates between groups defined by their gender or age. But when individuals were grouped by their level of education and SES, they do behave differently in the placement of subject pronouns. In this section, I investigate the same background variables separately for Caribbeans and Mainlanders. The results are very revealing about the distinctive regional patterns of overt pronoun use in Spanish in New York.

As it turns out, none of the basic demographic groups show any validity with respect to differences of preverbal rate for Mainlanders, but education and SES are valid for the Caribbeans. The results of ANOVA for the basic demographic groupings among Mainlanders are: Gender ($F = 0.01$, $p < .89$), Age ($F = 1.17$, $p < .31$), Education ($F = 0.45$, $p < .50$), SES ($F = 0.45$, $p < .50$). The values for Caribbeans are: Gender ($F = 1.12$, $p < .29$), Age ($F = .75$, $p < .52$), Education ($F = 4.70$, $p < .03$), SES ($F = 8.13$, $p < .00$). The results show that whereas it is true that men and women, the young and the old, the more or less educated, etc. are not distinguished by preverbal rates in the Mainlander sub-sample, they do use pronouns at different rates in the Caribbean sub-sample. Saying it another way, whereas Mainlanders tend to have uniform preverbal rates throughout the population, Caribbeans are well differentiated by education and SES. Among Mainlanders, none of the *F* values rises above 1.20, and all the significance values are $p > .30$; but among Caribbeans, the *F* values for education and SES are above 4.70, and the *p* values for these variables are all below .05.

When we investigate socio-economic status, the mean rates of preverbal pronouns for the Caribbean and the Mainland are as follows: in the Caribbean, the higher the socio-economic status, the more preverbal pronouns speakers have. This difference between the two groups is statistically significant. It can also be seen that Mainlanders who have a higher socio-economic status do tend to have more preverbal pronouns than those with a lower socio-economic status, but this difference is not statistically significant.

Regarding the impact of education on the preverbal rate, the situation is similar. For Caribbean speakers, the higher the level of education, the more preverbal pronouns. This difference in mean rates of preverbal pronouns between those who attended college or had a higher level of education, and those who finished high school or had a lower level of education is statistically significant. Although again here there is a similar tendency for the Mainland, that is**,** those with a higher level of education have more preverbal pronouns, the ANOVA revealed that this difference between the two levels of education is not statistically significant.

Recapitulating the findings for the Mainland population, it has been found that none of the factors considered (gender, age, socio-economic status, and education) play a role in the placement of overt subject pronouns. On the other hand, two factors do influence placement of subject pronouns in the Caribbean population: socio-economic status and education.

5. From groups to individuals

The present work focuses primarily on the linguistic behavior of groups, and only secondarily on that of individuals. However, group results can be difficult to grasp and

may obscure individual differences, requiring a more individualized display of data. The central discovery in this chapter with regard to the preverbal rate, presented in Table 3.5, is that Spanish in New York is regionally differentiated, with a statistically significant spread of four percentage points between members of the two regional groups and with a much greater difference across than within the groups. This can be more easily visualized in Table 3.6, which lists all 139 consultants, ranked in descending order by their preverbal rate. The first column in the table gives the informant's ranking, starting with the informant with the highest preverbal rate. The second column gives the informant's identification number. The third column gives the informant's preverbal rate. Under the column marked Nationality, the informant's national origin is indicated. The last column lists the informant's region, with Mainlanders in bold to facilitate reading the table.

Table 3.6

Informants by preverbal rate

Ranking	Informant	Rate	Nationality	Region
1	11	100%	Cuba	Caribbean
2	13	100%	Cuba	Caribbean
3	42	100%	Cuba	Caribbean
4	86	100%	Puerto Rico	Caribbean
5	112	100%	Dominican Republic	Caribbean
6	113	100%	Dominican Republic	Caribbean
7	201	100%	Cuba	Caribbean
8	318	100%	Dominican Republic	Caribbean
9	333	100%	Dominican Republic	Caribbean
10	359	100%	Mexico	**Mainland**
11	368	100%	Dominican Republic	Caribbean
12	403	100%	Puerto Rico	Caribbean
13	417	100%	Puerto Rico	Caribbean
14	435	100%	Puerto Rico	Caribbean
15	12	99%	Cuba	Caribbean
16	65	99%	Puerto Rico	Caribbean
17	181	99%	Colombia	**Mainland**
18	206	99%	Cuba	Caribbean

continued

Table 3.6 (continued)

19	220	99%	Puerto Rico	Caribbean
20	310	99%	Colombia	**Mainland**
21	319	99%	Dominican Republic	Caribbean
22	323	99%	Ecuador	**Mainland**
23	331.1	99%	Dominican Republic	Caribbean
24	331.2	99%	Dominican Republic	Caribbean
25	332.1	99%	Dominican Republic	Caribbean
26	334	99%	Ecuador	**Mainland**
27	342	99%	Colombia	**Mainland**
28	376	99%	Cuba	Caribbean
29	413	99%	Puerto Rico	Caribbean
30	416	99%	Puerto Rico	Caribbean
31	423	99%	Puerto Rico	Caribbean
32	428	99%	Puerto Rico	Caribbean
33	5	98%	Cuba	Caribbean
34	92	98%	Puerto Rico	Caribbean
35	153	98%	Puerto Rico	Caribbean
36	183	98%	Cuba	Caribbean
37	305	98%	Mexico	**Mainland**
38	311	98%	Colombia	**Mainland**
39	325	98%	Ecuador	**Mainland**
40	336	98%	Dominican Republic	Caribbean
41	346	98%	Mexico	**Mainland**
42	365	98%	Ecuador	**Mainland**
43	373	98%	Puerto Rico	Caribbean
44	377	98%	Puerto Rico	Caribbean
45	381	98%	Dominican Republic	Caribbean
46	427	98%	Puerto Rico	Caribbean
47	432	98%	Puerto Rico	Caribbean
48	96	97%	Puerto Rico	Caribbean
49	158	97%	Colombia	**Mainland**
50	194	97%	Colombia	**Mainland**
51	208	97%	Colombia	**Mainland**
52	233	97%	Cuba	Caribbean
53	238	97%	Cuba	Caribbean
54	300	97%	Ecuador	**Mainland**
55	302	97%	Dominican Republic	Caribbean
56	329	97%	Dominican Republic	Caribbean
57	339	97%	Mexico	**Mainland**
58	344	97%	Colombia	**Mainland**
59	363	97%	Ecuador	**Mainland**

continued

Table 3.6 (continued)

60	367	97%	Ecuador	**Mainland**
61	369	97%	Dominican Republic	Caribbean
62	372	97%	Cuba	Caribbean
63	374	97%	Dominican Republic	Caribbean
64	379	97%	Colombia	**Mainland**
65	384	97%	Ecuador	**Mainland**
66	422	97%	Puerto Rico	Caribbean
67	434	97%	Puerto Rico	Caribbean
68	2	96%	Cuba	Caribbean
69	24	96%	Colombia	**Mainland**
70	37	96%	Dominican Republic	Caribbean
71	102	96%	Puerto Rico	Caribbean
72	125	96%	Dominican Republic	Caribbean
73	229	96%	Dominican Republic	Caribbean
74	272	96%	Cuba	Caribbean
75	303	96%	Puerto Rico	Caribbean
76	364	96%	Ecuador	**Mainland**
77	366	96%	Ecuador	**Mainland**
78	401	96%	Puerto Rico	Caribbean
79	10	95%	Cuba	Caribbean
80	44	95%	Cuba	Caribbean
81	172	95%	Colombia	**Mainland**
82	180	95%	Colombia	**Mainland**
83	237	95%	Cuba	Caribbean
84	304	95%	Mexico	**Mainland**
85	308	95%	Mexico	**Mainland**
86	309	95%	Ecuador	**Mainland**
87	322	95%	Ecuador	**Mainland**
88	340	95%	Mexico	**Mainland**
89	378	95%	Puerto Rico	Caribbean
90	3.1	94%	Cuba	Caribbean
91	26	94%	Colombia	**Mainland**
92	230	94%	Dominican Republic	Caribbean
93	301.1	94%	Ecuador	**Mainland**
94	324	94%	Ecuador	**Mainland**
95	6	93%	Cuba	Caribbean
96	9	93%	Cuba	Caribbean
97	173	93%	Colombia	**Mainland**
98	198	93%	Puerto Rico	Caribbean
99	203	93%	Cuba	Caribbean
100	228	93%	Dominican Republic	Caribbean

continued

Table 3.6 (continued)

101	258	93%	Colombia	**Mainland**
102	326	93%	Ecuador	**Mainland**
103	328	93%	Ecuador	**Mainland**
104	338	93%	Ecuador	**Mainland**
105	347	93%	Mexico	**Mainland**
106	351	93%	Mexico	**Mainland**
107	271.2	92%	Mexico	**Mainland**
108	313	92%	Ecuador	**Mainland**
109	362	92%	Ecuador	**Mainland**
110	371	92%	Ecuador	**Mainland**
111	234	91%	Cuba	Caribbean
112	312	91%	Ecuador	**Mainland**
113	321	91%	Ecuador	**Mainland**
114	263	90%	Colombia	**Mainland**
115	271.1	90%	Mexico	**Mainland**
116	306	90%	Mexico	**Mainland**
117	8	89%	Cuba	Caribbean
118	38	89%	Colombia	**Mainland**
119	120	89%	Dominican Republic	Caribbean
120	314	89%	Mexico	**Mainland**
121	316	89%	Ecuador	**Mainland**
122	320	89%	Ecuador	**Mainland**
123	349	89%	Mexico	**Mainland**
124	21	88%	Colombia	**Mainland**
125	227	88%	Dominican Republic	Caribbean
126	25	87%	Colombia	**Mainland**
127	269	87%	Colombia	**Mainland**
128	330	87%	Dominican Republic	Caribbean
129	350	87%	Mexico	**Mainland**
130	354	87%	Mexico	**Mainland**
131	317	86%	Mexico	**Mainland**
132	370	85%	Mexico	**Mainland**
133	118	83%	Dominican Republic	Caribbean
134	270	82%	Mexico	**Mainland**
135	7	81%	Cuba	Caribbean
136	174	80%	Colombia	**Mainland**
137	356	79%	Mexico	**Mainland**
138	348	74%	Mexico	**Mainland**
139	352	71%	Mexico	**Mainland**

Puerto Rico	(N = 24, Preverbal rate = 98 %)
Cuba	(N = 24, Preverbal rate = 96 %)
Dominican	(N = 24, Preverbal rate = 96 %)
Ecuador	(N = 24, Preverbal rate = 94 %)
Colombian	(N = 21, Preverbal rate = 94 %)
Mexico	(N = 22, Preverbal rate = 89 %)

In looking at the individual preverbal rates on the table, regional differences can be appreciated. Caribbeans tend to concentrate towards the top of the table, with higher preverbal rates, while Mainlanders tend to concentrate towards the bottom of the table, with lower preverbal rates. When looking at the individuals who occupy the top of the list, there are 14 speakers who registered a preverbal rate of 100 percent, which means that all their overt subject pronouns were placed preverbally. Among them, there is only one informant from the Mainland. In looking towards the bottom of the table, among the 49 speakers who had a preverbal rate below 95 percent (which is the mean rate for the whole sample), there are only 14 informants from the Caribbean (approximately 30 percent of the subset), while the remaining 35 are Mainlanders. Furthermore, if the table is divided by half, the top half of the sample contains 48 Caribbeans but only 22 Mainlanders, whereas the bottom half contains 45 Mainlanders and only 24 Caribbeans. If attention is focused on the column listing preverbal rates for each informant, it can clearly be seen how much of a continuum there is in New York.

While Table 3.6 shows the individual and nationality continuum and the two regional groupings quite clearly, it leaves other questions unanswered. It is not known at this point, for example, why informant 359M who is a Mainlander from Mexico, is among those with the highest possible preverbal rate, nor do I know why informants 007U, 118D and 330D, who are Caribbeans, have such low rates and rankings (they rank 128th, 133rd, and 135th in the sample). Some of these answers will come from the results

of further questions about these participants, including their membership in other groups, as it will be presented in the next chapters.

The point to remember is that while it is true of Spanish in New York that Caribbeans tend to use more preverbal pronouns than Mainlanders, this does not mean that all Caribbeans use more preverbal pronouns than Mainlanders. As it can clearly be seen on Table 3.6, there are Mainlanders who use more preverbal pronouns than Caribbeans. What is shown in Table 3.5 is that the individual Mainland informants who appear toward the top of Table 3.6 where Caribbeans predominate, and the Caribbean speakers who appear toward the bottom where Mainlanders predominate, do not negate the basic statistical fact that the sample can be divided into the two regional groups. In other words, it is true enough that it cannot be predicted, given any two randomly chosen Spanish speakers in New York, one from the Caribbean and one from the Mainland, which one is going to use more preverbal pronouns. But on the basis of Tables 3.5 and 3.6, and on the basis of the careful sampling used to select speakers in the corpus, it can be said that the Caribbean speaker is much more likely to use more preverbal pronouns than the Mainlander. Therefore, I can never be sure that a randomly picked Latino New Yorker from the Mainland will not turn out to be like consultant Number 359M (the Mainlander from Mexico unexpectedly ranked 10th in preverbal rate), or that a randomly picked Caribbean speaker will not be like participant Number 007C (the Caribbean from Cuba unexpectedly ranked 135th), both reversing the predominant statistical pattern. But I can be more confident that our two randomly selected Latinos will have preverbal rates similar to those from their region, i.e., the Caribbean will be like consultant Number 201C (a Cuban ranked 7th) or Number 112D (a Dominican ranked 5th), and the

Mainlander will be closer to Number 263C (a Colombian ranked 114th), or Number 320E (an Ecuadorian ranked 122th).

Table 3.6 is an individualized, and more visually compelling, display of the same statistical information that appears in Tables 3.4 and 3.5, and it highlights the variable nature of preverbal rates within national groups as well as the dominant patterns across regional groupings.

6. Summary and discussion

In the foregoing investigations I aimed to establish which groups of individuals had significant differences in their preverbal rates to merit being grouped together. To that end, six basic socio-demographic factors were considered for all speakers in the corpus, and four of those factors were further investigated separately for speakers from the Caribbean and the Mainland. The results obtained in terms of which factors play a role in speakers' use of preverbal pronouns are presented in Table 3.7 below.

Table 3.7
Basic demographic factors by group

Factors	**Whole Group**	**Caribbeans**	**Mainlanders**
Gender	No	No	No
Age	No	No	No
SES	**Yes**	**Yes**	No
Education	**Yes**	**Yes**	No
Country of Origin	**Yes**	-	-
Region of Origin	**Yes**	-	-

The first observation to make from the results in the previous sections outlined in Table 3.7 above is that there is a clear and statistically significant difference in rates of preverbal pronouns between the two regions from which the majority of New York City

Spanish speakers originate. The claim that the Caribbean and the Mainland constitute two dialectal regions in terms of placement of overt subject pronouns has been made in the literature before, usually in qualitative observations regarding linguistic environments, such as non-finites and interrogatives, that are said to favor preverbal subjects in the Caribbean and null or postverbal subjects in the Mainland (Lipski 1994; Morales 1988, 1989, 1999; Ordoñez & Olarrea 2001; Toribio 2000; Zagona 2002; Goodall 2004). The facts stated above for my corpus confirm this distinction quantitatively, and for a much broader set of environments. This finding regarding the preverbal rate is parallel to the finding regarding the pronoun rate, which has also been shown to be higher in the Caribbean (López-Morales 1992, Otheguy & Zentella 2012).

One can see, then, a clear difference in two aspects of subject pronoun usage between the Caribbean and the Mainland: Caribbeans use more pronouns than Mainlanders, and they place them in preverbal position more than Mainlanders. Thus, I have established that, for at least the feature under study, there are two distinct dialects of Spanish, corresponding to the Caribbean and the Mainland. These two separate dialectal groupings enter the City in the speech of newcomers and provide the baseline for subsequent developments in Spanish in New York.

The dialectal separation between a region that has higher rates of both overt and preverbal pronouns, and another region that shows lower rates in both measures can usefully be considered in the context of parameterization in formal theories of syntax, specifically with regard to claims regarding the null subject parameter (Chomsky 1981). It has been claimed in the generative literature that the null subject parameter is a cluster of properties including the possibility of null subjects and the possibility of postverbal

subjects (Chomsky 1981, Rizzi 1982, Jaeggli and Safir 1989). This means that if a language can have null subjects, it can also have postverbal subjects and that if a language cannot have null subjects, then it cannot freely place their subjects postverbally. Spanish, as is well known, conforms well to this characterization, displaying the traits of a null-subject language; its subjects can be null or overt, and when they are overt, they can appear before or after the verb.

The dialectal findings in this project support the idea of a cluster of properties around the null subject parameter in the two dialectal regions studied. In the Mainland, less overt pronouns are found (i.e. more null subjects), and more postverbal subject pronouns than in the Caribbean. On the other hand, in the Caribbean more overt subject pronouns and more preverbal subject pronouns are found. It seems that the more the Caribbean Spanish dialect leans towards behaving like a non-null-subject language, the less it places subjects in postverbal position. If this linguistic behavior of Caribbean speakers represents a case of change in progress, it would be safe to assume that in the distant future Caribbean speakers would only have overt subjects, and a very limited capacity to place subjects postverbally.

Furthermore, the dialectal separation evidenced by the difference in mean rates of preverbal pronouns between the Caribbean and the Mainland is further confirmed when some of the basic socio-demographic factors are investigated. While in the Caribbean speakers of a higher socio-economic status and higher level of education favor the use of preverbal pronouns, no such role is played by those factors in the Mainland where none of these variables was statistically significant. The findings in use of preverbal pronouns

in the Caribbean suggest that placing subject pronouns before verbs is a valued and approved of practice, favored by the higher socio-economic strata in the region.

The results presented in this chapter shed light on variation in subject pronoun placement witnessed in New York City. I have shown that this variation can be at least partly explained by considering the origin of New York City Spanish-speakers. Another observation to make about Spanish-speaking New Yorkers as a whole is that their speech resembles more the speech of Caribbeans than of Mainlanders. The statistical analysis revealed that none of the four socio-demographic factors considered were significant in the Mainland while two were significant in the Caribbean. Those two were also significant when all Spanish speakers were considered as one group. This indicates, once again, that Spanish-speaking New Yorkers approve of preverbal pronouns and that use of preverbal pronouns is favored by the higher strata of Spanish speakers in the City. There are two possible explanations for this phenomenon: 1) Caribbean ways of speaking Spanish have a preeminent place in New York City, and 2) the speakers' bilingualism is leading to language contact phenomena. Several demographic facts favor the first explanation offered. The Caribbean population in New York City is larger than the Mainlander population (61 percent of Latinos are of Caribbean origin) and they have been in the City in larger numbers for a longer period of time than Mainlanders. Furthermore, the largest group of Caribbeans in the City, Puerto Ricans, is composed of U.S. citizens while the majority of Mainlanders are legal and illegal immigrants. For these reasons, it may be the case that Caribbeans have achieved a higher socio-economic status and a more privileged standing in this society, imposing their speaking characteristics on the rest of the speakers of Spanish. The second explanation which deals

with the possibility that the speech of Mainlanders and Caribbeans is further influenced by English, is considered in the next chapter.

[1] While Otheguy & Zentella's (2012) investigation was performed with a corpus of 140 speakers, a corpus of 139 speakers is used in the research herein described. See chapter 2 for an explanation of the reason for eliminating one speaker.

[2] For an explanation of how the composite measure was created, see Otheguy & Zentella (2012: 71).

CHAPTER 4

THE PREVERBAL RATE AS EVIDENCE OF LANGUAGE CONTACT

1. *Introduction*

In this chapter I investigate whether the speakers' exposure to New York City, which involves contact with monolingual and bilingual speakers of English, is related to an increase in their use of preverbal pronouns, in the same way that it is connected to an increase in the use of pronouns as shown in the study under replication.

Otheguy & Zentella (2012) found support for the hypothesis that English, where subject pronouns are vastly more frequent than in Spanish, is one of the forces motivating an increase in the pronoun rate in the Spanish spoken in New York City. They observed that second generation speakers (those born or raised from early infancy in New York) registered a significantly higher pronoun rate than do first generation speakers (those born in Latin America). They also noted that the speech of bilinguals, including both the Latin American-raised established immigrants and the New York raised, had a significantly higher pronoun rate than the newcomers. Furthermore, they found that established immigrants used more pronouns than newcomers, and that in turn, New York raised speakers used more than established immigrants. Regarding the role of English proficiency, they noticed that Latinos with excellent English proficiency had more pronouns than those with lower proficiency in English. They also noted that all these findings applied to the community studied as a whole, as well as when it was divided into Latinos from Caribbean and Mainland origins.

I aim to establish a similar connection between the immigrant generation of the speakers and their rate of preverbal pronouns, as well as a connection between their level of English proficiency and their rate of preverbal pronouns. To remind, in this investigation the rate of preverbal pronouns or *preverbal rate* is the percentage of overt pronouns that are placed before the verb (as opposed to after the verb). Other related factors such as the time that Latin American raised speakers have been in the U.S. and the extent of use of English and Spanish in their daily lives are also investigated with reference to the preverbal rate.

2. Exposure factors in the whole sample

2.1. Reference Spanish and Contact Spanish

I start by offering a three-way comparison between the three groups of informants established in the study under replication, namely *newcomer immigrants*, consisting of those first generation speakers who arrived at age 17 or older and who have lived in the City for five years or less, *established immigrants*, those first generation speakers who do not qualify as immigrant newcomers because they came younger than 17 and/or have been in the city for over five years, and New York raised speakers (NYR) who were born in New York City or were brought to the City before the age of three. The reasoning behind this comparison is that immigrant newcomers should still speak a very close approximation to the Spanish spoken in Latin America, and therefore are, in the present investigation, properly considered speakers of *reference Spanish* whereas established immigrants and NYRs have been exposed to English for a longer period of time, or in their formative years, or during their whole life, and therefore they are classified here as speakers of *contact Spanish*[1].

In this investigation, contact with English would be evidenced by an increase in preverbal pronouns given that English places its subjects in preverbal position almost categorically whereas Spanish may place them both pre- and post-verbally. Therefore, the expectation is that newcomers will have less preverbal pronouns than the rest of the speakers in the sample. The results are shown in Table 4.1 below.

Table 4.1
ANOVA
Preverbal rate by Reference and Contact Spanish

	N speakers	Preverbal rate
Reference Spanish (newcomers)	39	94
Contact Spanish (all others)	100	95
	139	
F = .36	p < .54	

The table shows that even though the results go in the expected direction (a slightly higher preverbal rate in Contact Spanish than in Reference Spanish), the difference is minimal and not statistically significant. If we compare these preverbal rate results with those obtained in the study of pronoun rates under replication, a difference between the two investigations becomes immediately apparent. Both inquiries resulted in findings going in the expected direction, showing an increase in preverbal pronouns in this case, and in overt pronouns in Otheguy & Zentella (2012). However, in their study, the difference between the two groups was larger and statistically significant: a five percentage-point difference and a p value of < .02. This suggests that the alternation between preverbal/postverbal pronouns in Spanish is probably more subtle, less salient, than the alternation between presence and absence of pronouns. This point will be

addressed again after the investigation regarding the effect of exposure to New York on the preverbal rate.

2.2. Latin American Raised speakers and New York Raised speakers

The study under replication found an increase in the use of overt pronouns from the first immigrant generation of Spanish-speakers living in New York (the Latin American Raised or LAR), and the second immigrant generation of Spanish-speakers (the New York raised or NYR). I expect a similar increase in the use of preverbal pronouns and for the same reasons: (1) The NYR group has been exposed to English from birth or very early in their lives, and (2) The NYR group has been also exposed to the speech of Caribbeans whose Spanish has higher rates of preverbal pronouns than that of Mainlanders. The results are in Table 4.2.

Table 4.2
ANOVA
Preverbal rate by generation

	N speakers	Preverbal rate
LAR	114	94
NYR	25	97
	139	
F = 4.94	P < .02	

As expected, the NYR group has a significantly higher rate of preverbal pronouns than the LAR group. While 94 percent of the verbs in the speech of the first generation appear with a preverbal pronoun, the number increases to 97 percent of verbs with a preverbal pronoun in the speech of the second generation. Variance in the use of preverbal pronouns is almost five times greater (F = 4.94) across the generational groups than

inside each of the groups, and the four percentage-point difference between the LAR and the NYR registers a high level of statistical significance ($p < .02$). These findings parallel those for the overt rate.

2.3 Immigrant Newcomers, Established Immigrants, and the NYR

The next question we raise is whether the process of change is confined to the passage between the immigrant generations or whether instead it begins within the LAR group, among those with more years in the city. To that end, in Table 4.3 the LAR group is divided into immigrant newcomers (the recent arrivals) and the established immigrants (the rest of the LAR); the table also considers the NYRs to see if there is a continuum in the increase of preverbal placement.

Table 4.3
ANOVA
Preverbal rate by exposure

	N speakers	Preverbal rate
LAR immigrant newcomers	39	94
LAR established immigrants	75	94
NYR	25	97
	139	
F = 2.45	P < .09	

Unlike the results of Otheguy & Zentella (2012) for the overt rate, I do not see an increase in the preverbal rate with more exposure to New York City within the LAR group. Clearly, the shift to increased preverbality occurs only across the generations and not within the first generation.

In comparing the two studies, it seems that for the increase in preverbal placement to occur, the increase to more frequent occurrence of pronouns has to take place first. As discussed in the previous chapter, one way to interpret our findings is with reference to

pronominal parametric settings in formal grammars. I speculate that the increase in overt pronouns is affecting the null subject parameter in the speakers' native language. One of the characteristics of the null subject parameter is the ability to place subjects postverbally, i.e. languages which allow subjects to be null also allow subjects to be placed after the verb. Conversely, it seems that if the speakers use fewer null pronouns, they will also use fewer postverbal pronouns. However, given the results above, it seems to be the case that the shift to fewer postverbal pronouns occurs after the shift to fewer null pronouns. In this view of the matter, the placement change may be a consequence of the frequency change, and could be seen as evidence that a parametric change in the speakers grammar has occurred.

3. English proficiency groups

3.1 English Excellent and English less than excellent

We have seen above that although no changes with regards to the placement of subject pronouns seem to occur in the first generation, the second generation, born and/or raised in New York City and bilingual, exhibits a statistically significant increase in the preverbal rate which mirrors English grammar. These results seem to indicate that those who are born or raised with Spanish and English, in a medium where English has a higher status and is prevalent, find their grammar of Spanish affected by their grammar of English. Consequently, I wanted to know whether this is just an effect of the environment and the higher status of one of the languages, or whether it is the case that, independently of where the speakers are born or raised, those who claim to have a high proficiency in English usually exhibit more preverbal pronouns than those who claim to have lower or no proficiency. For the purpose of studying the effect of higher proficiency in English on

the Spanish spoken in New York City, the sample is divided into two groups: those who claim to speak excellent English and everybody else. If excellent English speakers have a higher preverbal rate, the language contact hypothesis would be supported. The results of this investigation are presented in Table 4.4a below. Also, the same division is carried out in a subset of the sample of speakers: the LAR. The goal of this inquiry is to establish whether excellent English skills (or fluent bilingualism) are a driving force behind the changes in Spanish grammar, independently of where the speakers were born and raised, that is independently of generation, exposure, etc. The results of this second inquiry are presented in Table 4.4b below.

Table 4.4a

ANOVA

Preverbal rate by English skills

	N speakers	Preverbal rate
Eng. Less than excellent	94	94
Eng. Excellent	44	96
	138	
F = 8.38	p < .001	

Table 4.4b

ANOVA
Preverbal rate by English skills, LAR only

	N speakers	Preverbal rate
Eng. Less than excellent	89	94
Eng. Excellent	24	96
	113	
F = 4.28	p < .04	

Both investigations yielded the expected results: speakers who claim to have excellent English proficiency also have more preverbal pronouns in Spanish. Variance in the use of preverbal pronouns is more than eight times greater ($F = 8.38$) across the English proficiency groups than inside each of the groups**,** and the two percentage-point difference between those who claim to be highly fluent in English and the rest of the speakers registers a high level of statistical significance ($p < .00$). The same holds true when we subdivide the sample to just contain LAR speakers and we run the same test. Speakers who claimed to have a high proficiency in English also had more preverbal pronouns, and the difference between these two groups was more than four times greater (4.28) and highly significant ($p < .02$). These results suggest very strongly that there is language contact between English and Spanish at work, and that English is one of the factors shaping the Spanish spoken in New York City.

3.2 Language choice with interlocutors and the English proficiency groups

As a way to temper subjectivism in the answers provided by informants with regards to their ability in English, the Otheguy-Zentella corpus questionnaire includes a set of language-choice questions that less directly but more reliably address**es** the informant's English ability, in the sense of competence and fluency. In this questionnaire, informants were asked how much English they used with their five nearest relatives (mother, father, siblings, spouse, children), and how much Spanish they used in the three most familiar domains of action (home, school, social activities). Compared to the English-skills question, the language-choice questions have the advantage that, being more concrete and specific, and much less subject to a prescriptive interpretation, they are more likely to produce accurate answers about the use of English and Spanish. These

more reliable language choice answers serve as a check on the language proficiency answers: the expectation being that informants who say that their English is excellent should also report using English with more interlocutors than informants who are less confident in their English.

Informants were asked each language-choice question about the use of English with specific interlocutors separately. Answers were coded in an ordinal scale, with steps ascending toward higher use of English. The scale had three values: (1) 'I speak to this person in Spanish,' (2) 'I speak to this person in both Spanish and English,' and (3) 'I speak to this person in English.' The results, reported in Otheguy & Zentella (2012: 94), are clearly supportive of this reasoning. As expected, the authors found that Latino informants who say that their English is excellent are still more inclined to use Spanish with their parents than with anyone else, since Spanish, as can be attested through informal observation, is a very frequent choice in parent /child communication in the New York Latino community, even when the child knows a lot of English; thus the correlations between reports of excellent levels of English and English language choice for father and mother are in the predicted direction and statistically significant, but low (in the .20+ range). But with the interlocutors with whom language choice is more revealing of language ability, namely siblings, spouses, and children, the correlations found by Otheguy & Zentella between reported language choice and reported language proficiency range from moderate upwards. The results indicate a clear pattern; informants who say they have high English skills tend to be the same who prefer English, to different degrees, with their closest interlocutors other than their parents.

3.3 Language choice in domains and the English proficiency groups

In the questionnaire that accompanied the corpus, separate questions were asked for language choice at home, in school, and in social settings. Possible answers were coded on a binary scale whose values were: (1) 'In this domain I speak little Spanish' and (2) 'In this domain I speak a lot of Spanish.' As discussed by Otheguy & Zentella, results concerning language-choice in domains cannot be as telling as those just described for interlocutors, for two reasons. First, it is much more difficult to report accurately on the language one speaks in something as broad as a domain, e.g., 'social activities,' than to report on the language one speaks to a specific interlocutor, e.g., one's sister. Second, these questions bear less directly on the issue under investigation, since they ask about use of Spanish, not English. Still, the reasoning behind is that if the self-reports on English proficiency are about actual fluency and competence, they should show an inverse correlation with these answers, that is, the informants who say they know a lot of English should be more likely to say that they use less Spanish in these three domains. The results in Otheguy & Zentella (2012: 95) reveal that informants who say they know a lot of English make less use of Spanish; they speak more English with their closest and most frequent interlocutors and the least Spanish in their most familiar settings. The congruence between the answers to these different types of questions supports the validity of the answers to the English proficiency question, thereby confirming the existence of a sub-community of Latino New Yorkers who have a strong command of, and a clear preference for, English, and who are distinguished by their significantly higher use of preverbal subject pronouns in Spanish.

3.4 Spanish proficiency and the preverbal rate

It was shown above that higher proficiency in English was correlated with more preverbal pronouns in the whole sample and in the LAR group, which is interpreted as evidence of an effect of language contact. The next question that arises is whether those speakers who claim to have lower proficiency in Spanish are more susceptible to the influence of English than speakers who claim to have a high proficiency in Spanish. If that were the case, then speakers with low Spanish proficiency would have more preverbal pronouns and speakers with high Spanish proficiency would have less preverbal pronouns. To investigate this possibility and to parallel the investigation for English proficiency, I looked at the connection between Spanish proficiency and the preverbal rate in the whole sample and in the LAR group as well. Tables 4.5a and 4.5b below show that no clear connection was discovered.

Table 4.5a

ANOVA
Preverbal rate by Spanish skills

	N speakers	Preverbal rate
Spanish less than excellent	81	95
Spanish excellent	57	94
	138	

F = .19 p < .66

Table 4.5b

ANOVA
Preverbal rate by Spanish skills, LAR only

	N speakers	Preverbal rate
Spanish less than excellent	58	94
Spanish excellent	55	94
	113	

F = .02 p < .87

Although in the whole sample there is a one-point difference in the preverbal rate between speakers who have high and low Spanish proficiency, these numbers are not strong enough to yield statistical significance and to conclude that Spanish proficiency is related to the preverbal rate. Furthermore, when we look at the subset of LAR speakers, there is no difference in the preverbal rate between speakers who claim to have low and high Spanish proficiency. Whereas we had seen that reported high proficiency in English could affect the speakers' grammar of Spanish, it does not seem to be the case that reported low proficiency in Spanish makes its grammar more vulnerable to the influence of English[2].

4. Differences of preverbal rate in the regional and SES subsamples

In the preceding investigations for the whole sample, it was discovered that preverbal rates allowed for the grouping of speakers by generation and English proficiency. However, I was not able to clearly distinguish groups based on exposure to New York City nor on whether the speakers were users of Reference or Contact Spanish. Given the disparity between these findings for the preverbal rate and those of the work under replication for the pronoun rate, I have offered the explanation that for changes in preverbal rates to take place, changes in pronoun rates have to happen first, which would be aligned with the generativists' hypothesis of the null subject parameter as encompassing both the possibility of dropping subjects and the possibility of placing subjects after the verb. In this section, I will be looking at the same grouping possibilities after subdividing the sample into the two regional groups whose validity was established in the previous chapter, namely Caribbeans and Mainlanders. I will also investigate

whether socioeconomic status is related to the preverbal rate in informants from these two regions.

4.1 Exposure and English proficiency in Caribbeans

EXPOSURE IN CARIBBEANS. As exposure increases among Caribbeans from immigrant newcomers to established immigrants to NYR, the preverbal rate does show an increase. This positive finding means that, among Caribbeans, the preverbal rate differences associated with exposure differ from the trend of the whole sample (see Table 4.6).

Table 4.6
ANOVA
Preverbal rate by exposure among Caribbeans

	N speakers	Preverbal rate
LAR immigrant newcomers	19	95
LAR established immigrants	40	97
NYR	13	98
	72	
F = 2.49	P < .09	

The results of Table 4.6 above support the language contact hypothesis among Caribbeans, and unlike the results for the whole sample, where we saw that contact was evident only in the second generation, we can see changes taking place in the first generation, with a two-point increase in the preverbal rate from immigrant newcomers to established immigrants. These findings parallel those of Otheguy & Zentella (2012) for the pronoun rate in the same subsample of speakers.

ENGLISH PROFICIENCY IN CARIBBEANS. The differences in preverbal rates of Caribbeans with different English abilities are as expected, and they also move in the

same direction as in the whole sample, yielding strong differences with clear statistical significance.

Table 4.7
ANOVA
Preverbal rate by English skills among Caribbeans

	N speakers	Preverbal rate
Eng. Less than excellent	45	95
Eng. Excellent	26	99
	71	
F = 14.5	p < .001	

Among Caribbeans whose English competence is high, the ability to place subjects postverbally has almost disappeared. Furthermore, the difference between these two groups was more than fourteen times greater (14.5) and highly significant (p < .00**1**). These findings provide strong evidence of language contact between English and Spanish in this subset of speakers. Furthermore, they are aligned with the findings of Otheguy & Zentella (2012) for the pronoun rate.

4.2 Exposure and English proficiency in Mainlanders

EXPOSURE IN MAINLANDERS. Preverbal rate differences for Mainlanders of different exposure groups do not follow the same pattern as for Caribbeans, since there is not an increase within the first generation of speakers.

Table 4.8
ANOVA
Preverbal rate by exposure among Mainlanders

	N speakers	Preverbal rate
LAR immigrant newcomers	20	93
LAR established immigrants	35	91
NYR	12	95
	67	
F = 2.14	P < .12	

As exposure increases from Mainlander immigrant newcomers to established immigrants, the preverbal rate decreases. Then, it increases again with the NYR group to surpass that of LAR immigrant newcomers. The big increase for Mainlanders is inter-generational and not intra-generational. Regarding the decrease in preverbal rates in the LAR established immigrant group, it is interesting to note that 15 out the 35 speakers in this group registered preverbal rates that were much lower than the average rate of the whole group (from 71 to 89) and that they came mostly from Mexico, but there were also two speakers from Colombia and four from Ecuador with preverbal rates in the 71-to-92 range, which is below the mean rate for newcomers from that region, which is 93. It could be the case that LAR established immigrants from the Mainland make an unconscious attempt to differentiate themselves from Caribbean immigrants and English speakers and that this translates into placing more subjects postverbally. This behavior would be similar to the one reported by Labov (1963) in Martha's Vineyard, where the need of a group to assert its unique identity translated into an increase in the use of the phonological patterns typically associated with the speech of local fishermen families. In that study, Labov showed that language (or dialect) contact situations do not necessarily lead to a melting-pot linguistic outcome, but rather to an increase in the differences that distinguish each group.

ENGLISH PROFICIENCY IN MAINLANDERS. The differences in preverbal rate between Mainlanders with different levels of English-proficiency follow the same pattern as those of the whole sample and the Caribbean group. However, there is only one

percentage point difference between Mainland speakers with high and low English skills, and the results lack statistical significance ($F = .40$, $p < .52$).

In subdividing the whole sample into two subsets by region, I discovered that the factors of exposure and English proficiency affect the preverbal rate differently in each region. Whereas in the Caribbean more exposure to the U.S. and more fluency in English clearly affect the placement of pronominal subjects, the same does not hold true in the Mainland. For Mainland speakers, the increase takes place in the second generation but not in the first, and an excellent level of English fluency does not seem to be a clear indicator of increased preverbal placement of pronouns. The fact that each region behaves differently for the two-aforementioned factors with regards to the preverbal rate supports the division of speakers from the two regions into two groups. Furthermore, the behavior of both regions supports the language contact hypothesis, although in one group more strongly than in the other.

4.3 Exposure and English proficiency in lower SES consultants

EXPOSURE IN LOWER SES CONSULTANTS. Differences in preverbal rate associated with exposure follow the expected direction in lower SES speakers. Among lower SES speakers, the preverbal rate of immigrant newcomers, established immigrants, and the NYR increases as predicted by the language contact hypothesis; it is, respectively, 91 percent, 93 percent, and 99 percent ($F = 3.86$, $p < .02$).

ENGLISH PROFICIENCY IN LOWER SES CONSULTANTS. The differences in preverbal rates of speakers of lower SES with different English abilities follow the trend of the whole sample and are as expected according to the language contact hypothesis. Speakers

of lower SES whose English proficiency is low have a 93 percent preverbal rate, whereas those whose English proficiency is high have a 96 percent rate ($F = 3.01$, $p < .08$).

4.4 Exposure and English proficiency in higher SES consultants

EXPOSURE IN HIGHER SES CONSULTANTS. Preverbal rate differences among consultants of higher SES status who belong to different exposure groups are not found at all in the sample ($F = .09$, $p < .91$).

ENGLISH PROFICIENCY IN HIGHER SES CONSULTANTS. The differences in preverbal rate between higher SES consultants with different levels of English proficiency are as predicted by the language contact hypothesis, follow the trend of the whole sample, and produce clear results that are statistically significant. Among consultants of higher SES, those with lower English proficiency have a 94.5 percent preverbal rate, while those with excellent English proficiency have a 97 percent rate ($F = 4.74$, $p < .02$).

The analysis of the preverbal rate in speakers of different socioeconomic status yields results that are aligned with the language contact hypothesis: more exposure and better English skills mean more preverbal pronouns, independently of whether the speakers have a higher or a lower socioeconomic status.

5. Summary and discussion

In this chapter, I investigated whether the placement of overt pronouns with regards to the verb can lend as much support to a language contact hypothesis for Spanish in New York as the occurrence of pronouns did in Otheguy & Zentella (2012). I believe that the answer is affirmative, and that there is a strong influence of English on the placement of overt pronouns by speakers of Spanish in New York City. Next, I summarize the findings that support this view.

A) Even though the difference does not reach levels of statistical significance, speakers of New York Contact Spanish (represented by established LAR immigrants and the NYR) use more preverbal pronouns than do speakers of Reference Spanish (represented by the newcomers).

B) In a statistically significant fashion, second-generation (NYR) speakers as a whole use a higher percent of Spanish pronouns than first-generation (LAR) speakers as a whole.

C) When we subdivide the LAR New York population into immigrant newcomers and established immigrants, using a linguistically sensitive combination of age of arrival and years in New York, we obtain results supportive of language contact as a force in one of the two regions. Even though the established immigrants do not use more preverbal pronouns than the newcomers in the whole sample and in the subset of Mainlanders, they do in the subsample of Caribbeans.

D) Latinos in New York who are more proficient in English use more preverbal pronouns in their Spanish than those who are less proficient. The connection between English proficiency and higher preverbal rates is very clear among Caribbeans, and while it has not been statistically significantly confirmed for Mainlanders, it also has not been strongly disconfirmed.

E) These differences in preverbal rate associated with reference/contact Spanish, generation, and English proficiency groups are true for the City population as a whole, as well as for each of the Caribbean and Mainlander regional sub-communities and for both lower-SES and higher-SES groups, except for one case. The only exception is the case of exposure given that established immigrants from the Mainland seem to decrease their

preverbal rate as a consequence of contact with English and other communities of Spanish speakers who have a higher rate of preverbal pronouns.

[1] Otheguy and Zentella (2012) have termed this division *Reference Lect* and *Bilingual Lect*. Although I am using the same classification, I have decided to call the two groups *Reference Spanish* and *Contact Spanish* respectively.

[2] For an explanation of similar findings with the overt rate, see Otheguy & Zentella (2012: 96-97).

CHAPTER 5

EXPLORING THE DIALECT LEVELING HYPOTHESIS WITH THE PREVERBAL RATE

1. Introduction

This chapter reports an important negative finding regarding the absence of dialectal leveling in the preverbal rate. I address here the question of whether there is evidence of dialect contact between the different regional varieties of Spanish that enter New York City, and whether, as a result of that contact, dialect leveling is taking place and shaping the Spanish speech community in the city as it has been proposed by Otheguy & Zentella (2012). In their study it was found that the differences in *pronoun rates* between newcomers from the Caribbean and newcomers from the Mainland narrowed among Caribbeans and Mainlanders of the second generation. In the present study, I aimed to investigate whether the same had happened to the *preverbal rate*, that is whether the differences in preverbal rate between Caribbeans and Mainlanders (established in chapter 3 diminished in the second generation. Otheguy & Zentella (2012) also gathered sufficient evidence to show that the dialectal leveling process is skewed towards one of the regions: the Caribbeanization of Mainlanders is somewhat stronger than the Mainlanderization of Caribbeans with respect to the use of pronouns. In the present study I aimed to find out whether this can also be said of preverbal pronouns.

The dialect leveling hypothesis encompasses the expectation that the regional differences in preverbal rates that come into the City with the immigrant newcomers will tend to be gradually smoothed out among the established immigrants and the NYR, as speakers weaken their linguistic ties to the pronominally differentiated norms of their Latin American areas of origin and increase their acquaintance with ways of using pronouns other than their own.

Therefore, the analysis of dialectal convergence will be based herein on the already established sub-categorization of the sample in terms of generational and exposure groups, as it was done in the study under comparative replication. For the purpose of researching dialectal convergence, informants were further classified according to how much they use Spanish with Latinos from their own and other countries and calculated how much each individual Caribbean in the sample is in contact with Mainlanders and how much each individual Mainlander is in contact with Caribbeans.

Anticipating my results, the evidence presented briefly in this chapter will show that after considering the same factors as the study being compared, I could not find evidence of dialect leveling with regards to the preverbal rate. For instance, regional differences did not narrow with more exposure to New York City or in the second generation. Furthermore, I was not able to show that the differences were narrower among out-group oriented individuals (speakers from one region claiming to be in touch with speakers from the other region) than in-group oriented individuals (speakers claiming to only be in touch with other speakers from their same region) either.

This is an important negative finding. Because of these results, we now know that the intermingling of Spanish-speakers from different parts of Latin America is deep enough to impact the pronoun rate, but not the preverbal rate. Why this is so will be discussed in the conclusion of this chapter.

2. Differences between the exposure groups

The expectation was that the regional gap in preverbal rate would be narrower among the NYR because their ties to the Latin American regional norms should be weaker, while their ties to what is, by hypothesis, a less pronominally differentiated New York norm should be stronger.

That is, whereas the immigrant newcomer's use of preverbal pronouns should still be either normatively Caribbean (higher use of preverbal pronouns) or normatively Mainlander (lower use of preverbal pronouns), the use of preverbal pronouns by Spanish-speakers born in New York or raised in the City from early infancy should be departing from these patterns. However, the results show that whereas the regions do exhibit an increase in preverbal use of pronouns by the second generation, there is no clear indication or tendency toward dialectal convergence (see Table 5.1a). These results for the preverbal rate are strikingly different from those obtained for the pronoun rate in the study under comparative replication (see Table 5.1b).

Table 5.1a

Preverbal rate: Region and generation

	LAR Newcomers			**NYR**		
	N	**Preverbal**	SD	N	**Preverbal**	SD
Caribbean	19	**95**	5	13	**98**	2
Mainland	20	**93**	4	12	**95**	5
P	p < .19			p < .04		
F	1.75			4.38		
Range	2			3		
d	0.5			0.6		

Table 5.1b (from Otheguy & Zentella 2012)

Pronoun rate: Region and generation

	LAR Newcomers			**NYR**		
	N	**Pro**	SD	N	**Pro**	SD
Caribbean	19	**36**	8	13	**44**	11
Mainland	20	**24**	9	13	**33**	8
P	p < .01			p < .01		
F	19.07			8.04		
Range	12			11		
d	1.33			1.00		

Table 5.1a shows the results of the present study and Table 5.1b shows the results of the study under comparative replication. For readers to clearly see the difference between the two

investigations, we will do the same with subsequent tables in this chapter. The tables present regional differences at each of the two exposure stages, immigrant newcomers on the left and second-generation NYR on the right. The columns marked *N*, *Preverbal*, and *SD* indicate respectively the number of informants, the preverbal rate, and the standard deviation. The rows marked *F*, *Range*, and *d* provide three separate measures of the strength of the regional differences. The F coefficient, which in this table compares variance across the regions to variance within the regions at each exposure stage, was already discussed in chapter 3. The row for Range shows the percentage-point difference between the regions at each of the two exposure stages. And the row marked *d* shows the effect size of the regional difference, again at each of the two exposure stages.

From Table 5.1a it becomes evident that, as opposed to what was discovered for the pronoun rate in the study under comparative replication, a comparison of the three measures does not show consistent narrowing of regional differences among the NYR. The F coefficient is bigger (rather than smaller) in the right-side panel of Table 5.1a than in the left-side panel. A smaller F coefficient would have indicated that Caribbeans and Mainlanders are less distinct at the NYR than at the newcomer stage. The range also increases instead of decreasing: whereas there was a 2-point difference between the regions in the LAR generation, there is a 3-point difference between the regions in the NYR generation. The results of Table 5.1a, then, do not support the notion that there is dialectal convergence occurring in the Spanish-speaking population, even though it does show that changes are occurring in both groups.

Rather than a leveling effect, the results point to a continuation of the Vineyard effect mentioned in the previous chapter, that is, an increase rather than a decrease in group differences: the regional difference is not significant among the newcomers ($p < .19$) but is

indeed significant among the NYR ($p < .04$). These findings seem to suggest that not only there is no leveling between the regions, but there seems to be at least a hint of hyper-differentiation, in that the behavior of speakers from the two regions becomes significantly different in the second generation. Overt preverbal pronouns are predominant (almost categorically) in English grammar. What was seen for the second generation in Otheguy & Zentella (2012) was that both regions increase the use of overt pronouns, and that the increase is bigger in Mainlanders than in Caribbeans, leading to a smoothing out of regional differences. However, what Table 5.1a is showing in the present study is that in the case of preverbal pronouns, Caribbeans are more inclined to follow the English trend whereas Mainlanders seem to be making an effort to not be influenced by Caribbeans or English-speakers, and to place more overt pronouns postverbally. These phenomena of hyper-differentiation could be explained by considering the difference in perception of preverbal pronouns by speakers from each region. Whereas LAR Caribbeans of middle class status have more preverbal pronouns than working class speakers of the same region, LAR Mainlanders of middle class have fewer preverbal pronouns than working class speakers of the same region. This fact indicates that while placing pronouns preverbally may be prestigious in the Caribbean, it does not seem to be the case in the Mainland. Therefore, it could be the case that this perception is carried on into the second generation, leading to the phenomenon of hyper-differentiation that we are witnessing.

3. Differences between informants with in-group versus out-group orientations

3.1 Orientation groups

In the study under comparison, support for dialectal leveling with regard to the pronoun rate also emerged when the sample was partitioned, separating informants with a greater out-group orientation (Caribbeans who say they frequently interact with Mainlanders and

Mainlanders who say they frequently interact with Caribbeans) from those with a more in-group orientation (Caribbeans and Mainlanders who say that they seldom or never interact with speakers of the other region). The authors found that the more in-group oriented consultants preserved the pronoun rate patterns of their reference Spanish more than consultants whose daily life involved frequent conversations with Latinos from the other region (see Table 5.2b). However, I did not have the same results for the preverbal rate, since while Mainlanders with an out-group orientation increased their preverbal rate, Caribbeans with an out-group orientation also increased their preverbal rate (see Table 5.2a). The expectation was that the preverbal rate of outwardly oriented Caribbeans would decrease.

Table 5.2a
Preverbal rate, orientation

	In-group orientation			**Out-group orientation**		
	N	**Pro**	SD	N	**Pro**	SD
Caribbean	46	**96**	4.5	24	**97**	2
Mainland	30	**92**	6	37	**93**	6
P	p < .001			p < .001		
F	11.38			13.10		
Range	4			4		
D	0.66			0.66		

Table 5.2b (from Otheguy & Zentella 2012)

Pronoun rate, orientation

	In-group orientation			**Out-group orientation**		
	N	**Pro**	SD	N	**Pro**	SD
Caribbean	46	**40**	10	24	**38**	10
Mainland	30	**26**	10	38	**29**	10
P	P < .01			p < .02		
F	39.98			11.75		
Range	14			9		
D	1.40			0.90		

None of the measures used in Table 5.2a supports the dialect leveling hypothesis: the range does not change but remains at 4; the ratio of cross-group to within-group regional variance is larger among the out-group oriented informants (F = 13.10) than among the in-group oriented ones (F = 11.38), showing that Caribbeans and Mainlanders represent more distinct groupings among informants with an out-group orientation than among those with an in-group orientation. Likewise, the d coefficient measuring regional effect size is the same for the two groups, thus not providing additional information.

3.2 Cross-orientation groups

In the study being compared, cross-orientation groups were created by assembling the IN-GROUP CARIBBEANS with OUT-GROUP MAINLANDERS, and the OUT-GROUP CARIBBEANS and IN-GROUP MAINLANDERS, to bring together informants from different regions according to the particular orientation that should produce equivalent tendencies of pronoun use. It was found that the group consisting of IN-GROUP CARIBBEANS and OUT-GROUP MAINLANDERS had a significantly higher pronoun rate than the group consisting of all the OUT-GROUP CARIBBEANS and IN-GROUP MAINLANDERS (see Table 5.3b.) Because of the tendency toward accommodation on the part of the out-group members, the authors interpreted that in New York, due to dialectal convergence, the speakers of each region who reach out to those of the other region become, when grouped with the inwardly oriented members of those groups, a statistically significant collectivity of informants based on the pronoun rate. In this study, when we grouped informants in the same way, we had results in the same direction, but not statistically significant (see Table 5.3a).

Table 5.3a
ANOVA
Preverbal rate, informants by cross-orientation group

	N	Pro rate
In group Mainlanders & out group Caribbeans	54	94
In group Caribbeans & out group Mainlanders	84	95
	F = .17	p < .67

Table5. 3b (from Otheguy & Zentella 2012)
ANOVA
Pronoun rate, informants by cross-orientation group

	N	Pro rate
In group Mainlanders & out group Caribbeans	54	31
In group Caribbeans & out group Mainlanders	84	35
	F = 4.28	p < .05

4. Exploring direction of leveling based on orientation and exposure

In the study under comparative replication, the authors showed that the Caribbean-induced upward pull among outwardly oriented Mainlanders was greater than the Mainlander-induced downward push on outwardly oriented Caribbeans. See Table 5.4a and 5.4b for a comparison of the results, taken from the data in Table 5.2a and 5.2b.

Table 5.4a
ANOVA
Preverbal rate, orientation differences by region

	Caribbean		Mainlander	
	N	Pro rate	N	Pro rate
In-region orientation	46	96	30	92
Out-region orientation	24	97	37	93
Range		1		1

Table 5. 4b (from Otheguy & Zentella 2012)

ANOVA

Pronoun rate, orientation differences by region

	Caribbean		Mainlander	
	N	Pro rate	N	Pro rate
In-region orientation	46	40	30	26
Out-region orientation	24	38	38	29
Range		-2		3

As noted earlier, the findings for the overt rate (Table 5.4b) are different than the findings for the preverbal rate (Table 5.4a), given that both cases show the same direction and range between in-region and out-region orientation.

Regarding the role of language contact, which is present in the city and also affects dialect leveling, in the study being compared it was found that everybody's pronoun rates were on the increase in New York, but that the increase was GREATER among Mainlanders than among Caribbeans (see Table 5.5b). However, this was not the case with the preverbal rate.

Table 5.5a

ANOVA

Preverbal rate, exposure differences by region

	Caribbean		Mainlander	
	N	Pro rate	N	Pro rate
Immigrant newcomer	19	95	20	93
NYR	13	98	12	95
Range		3		2

Table 5.5b (from Otheguy & Zentella 2012)

ANOVA

Pronoun rate, exposure differences by region

	Caribbean		Mainlander	
	N	Pro rate	N	Pro rate
Immigrant newcomer	19	36	20	24
NYR	13	44	13	33
Range		8		9

Table 5.5a shows that the increase is greater among Caribbeans than among Mainlanders, which is contrary to what was found in the study being compared in which the exposure increase was greater among Mainlanders than among Caribbeans, even if the difference (one percent) was considered too small to take into account.

5. Summary and discussion

In this chapter I explored the dialect leveling hypothesis proposed by Otheguy & Zentella (2012). In their study it was found that the differences in pronoun rates between newcomers from the Caribbean and newcomers from the Mainland narrowed among Caribbeans and Mainlanders of the second generation. The authors also showed that the dialectal leveling process is skewed towards one of the regions: the Caribbeanization of Mainlanders is somewhat stronger than the Mainlanderization of Caribbeans with respect to the use of overt pronouns.

The evidence presented here showed that, after considering the same factors, but this time with regard to the preverbal rate, no evidence of dialect leveling was found: regional differences did not narrow with more exposure to New York City or in the second generation, and the differences were not narrower among individuals who related often with members of the other region than among those who did not. Therefore, comparing the two studies and results, it has

become apparent that the intermingling of Spanish-speakers from different parts of Latin America is deep enough to impact the pronoun rate, but not the preverbal rate.

One of the possible explanations is that regional differences with respect to the preverbal rate are not as clearly marked as they are for the overt rate (see chapter 3). Rather, preverbal rate national means constitute more of a continuum in which countries have an ascending mean rate that is at most 2 points higher than the preceding country. The only speakers behaving strikingly differently from the rest are of Mexican origin, with a mean preverbal rate which is 5 points lower than the mean rate of the following country going up. Table 5.4a from chapter 3 is copied here for illustration purposes, but re-numbered to follow the numbering of this chapter.

Table 5.6
Preverbal rate by country of origin

	N speakers	Preverbal rate
Puerto Ricans	24	98
Dominicans	24	96
Cubans	24	96
Ecuadorians	24	94
Colombians	21	94
Mexicans	22	89
	139	

F = 8.62
p < .00

In looking at this table again, it would be interesting to see what happened if we grouped all countries but Mexico on one side, and just Mexico on the other, in order to re-explore the dialect leveling hypothesis. This could be the subject of a future investigation.

To conclude, I have mentioned in previous chapters that changes in preverbal rates seem to take place in the grammar after pronoun rate changes occur. This may be part of the reason why no telling results regarding dialect leveling arise of this exploration. A way to overcome this

could be to analyze the speech of second-generation speakers, but then it would be hard to tell whether changes are a consequence of dialect or language contact.

CHAPTER 6

MULTIVARIATE ANALYSIS: PREDICTING PREVERBAL PRONOUNS IN NEW YORK

1. Introduction

It was established in Chapters 3 and 4 that groups of Spanish-speakers who are distinguished in New York by regional origin, immigrant generation, length of exposure to the City, and knowledge of English have preverbal rates that are distinct in statistically significant ways. However, the question of which of these differences between informants offers the most telling picture of the use of preverbal pronouns in the City was not answered.

My aim, when considering the importance of each of these groups to the analysis of the variable placement of preverbal pronouns in New York, is to offer answers to important questions about the use of Spanish in the City. Some of these groupings, such as the regional one, reflect continuity between the Spanish spoken in Latin America and the Spanish spoken in the City; others, such as the exposure one, reflect the new factors affecting Spanish in the U.S. setting. Groups based on generation and exposure mirror the force of language contact; others are connected to only one or the other of these pressures (knowledge of English, for example, references only language contact). One of the most important questions in this project addresses the extent to which pronominal placement variation in Spanish in New York represents continuity with patterns of variability imported from Latin America, and to what extent it represents instead newly emerging speech communities that are being shaped by local cross-linguistic or cross-dialectal influences.

These different groupings of preverbal pronoun users allow me not only to consider the City as a whole, but also to look more narrowly at the regional sub-communities. It was discovered previously that informants from the Caribbean (but not from the Mainland) differ in

their preverbal rate by SES, education, and English proficiency, so now we want to know which of these groupings is most relevant for understanding the preverbal rate, which less, and which least. The present chapter addresses these questions through the use of MULTIPLE REGRESSION ANALYSIS, details of which are provided below. First, I discuss the conversion of the grouping categories into predictor variables, and present preliminary findings based on bivariate correlations. I then turn to analyses based on multivariate regression covering the whole sample, and finally move to a comparative study covering each of the regions separately. Beginning in Section 5, and after the bivariate and multivariate findings have been fully laid out, an interpretation of these findings is offered.

2. The variables defining the grouping criteria and the preverbal rate

Following the approach of Otheguy & Zentella (2012), I now consider the groupings of the previous chapters as variables, naming the variables with a single capitalized word, and re-stating the findings as follows: the variables Education, SES, Region, Generation, English excellent (English for short), and Exposure have been found relevant to an understanding of pronominal placement variation for the whole sample. Additionally, the variables English, SES, and Education are also important for studying the preverbal rate but are mostly applicable to one region. It was already ascertained that the variable Spanish excellent, or Spanish for short, is neither relevant to the study of the whole sample nor the regional sub-samples. Expressing the analysis in these terms, the present chapter is about a DEPENDENT VARIABLE, namely the Preverbal Rate, and about the differential impact upon this outcome variable of the several INDEPENDENT VARIABLES, or predictor variables, that I have just named.

In statistical terms, the dependent variable is a continuous variable consisting of a number, namely each informant's preverbal rate. The independent predictors are either binary

nominal variables containing two levels or factors, or ordinal variables containing two or three factors, as follows:

Region

Caribbean

Mainlander

Generation

LAR

NYR

English

English less than excellent

English excellent

SES

Lower SES

Higher SES

Education

Secondary education (high school) or less

Tertiary education (college or university) or more

Exposure

Immigrant newcomer

Established immigrant

NYR

3. Ranking of independent variables by bivariate correlations

Previously, I focused on whether the preverbal rate placed Caribbeans above or below Mainlanders, the LAR above or below the NYR, etc. Now, I am looking to discover whether, in their capacity to predict the preverbal rate, Region is stronger than Generation, SES stronger than Education or than English, etc. I start with simple correlations between the Preverbal Rate and all the variables that were relevant in previous chapters (Table 6.1). The variables are listed in order of the strength of their correlation with the Preverbal Rate.

Table 6.1
Pearson correlations with Preverbal Rate

	N	R	P
Region	139	-0.37	**
English	137	0.24	**
Education	139	0.20	*
Generation	139	0.19	*
SES	136	0.17	*
Exposure	139	0.14	a

* = p < .05
** = p < .01
a = p < .10

POSITIVE AND NEGATIVE MARKINGS. In Table 6.1, the statement of correlations as positive or negative is simply a matter of coding with no impact on the analysis. Region is consistently marked with a minus sign because the Caribbean region was coded with a 0 and the Mainland region was coded with 1, and it is the Caribbean region that registers the most preverbal pronouns. The negative marking helps indicate towards which region the preverbal rate is skewed.

As expected, the table corroborates the findings of previous chapters. When the whole sample is taken into account, the statistically significant or nearly significant correlations with Preverbal Rate involve the variables Region, English, Generation, SES, Exposure, and

Education.

Table 6.1 provides a preliminary indication of the relative importance of the variables in predicting the preverbal rate. The variable representative of continuity with the Spanish of Latin America, namely Region, is a stronger predictor than the variables that reflect language contact in New York, which are English and Generation, and these in turn are more predictive of the preverbal rate than the basic demographic variable of SES. However, the variable Education, which is a basic demographic variable, comes after Region, and the variable Exposure, indicative of language contact, comes at the very end.

As it was done in Chapter 3, I report the statistical significance value for Exposure even though it is not $p < .05$, in keeping with the goal of avoiding Type I errors (incorrectly attributing sample results to the population) while also avoiding Type II errors (incorrectly failing to attribute sample results to the population). In the regression tables that follow, I also offer cautious interpretations of results that are not only at $p < .05$ or $p < .01$, but also at $p < .10$.

4. A multivariate regression analysis of language contact and language continuity

The bivariate analysis of Table 6.1 above does not provide an entirely reliable answer to the question of the relative strength of the variables in an account of the use of preverbal pronouns because it considers separately the impact of Region on Preverbal Rate, of Generation on Preverbal Rate, etc. Given that the same informant is simultaneously a member of a region, a generation, a socio-economic group, etc., there is the danger that one (or more) of the correlations may be statistically subsumed under another of the correlations. For example, Table 6.1 tells us that Caribbeans use more preverbal pronouns than Mainlanders; and that people who know more English use more preverbal pronouns than those who know less English. Even though this sounds like two separate facts, two things that affect the use of overt pronouns, one

cannot be sure, just by looking at correlations. It could be the case that Caribbeans consistently and throughout the sample knew more English than Mainlanders. If so, then either Region or English would be an irrelevant variable to pronoun placement; one or the other would not be in itself a true predictor of pronoun placement. In other words, under such a hypothetical scenario, either Region would really be responsible for differing rates of pronoun placement among informants, and knowledge of English would play no independent role; or, alternatively, knowledge of English would really be responsible for differing rates of pronoun placement among informants, and Region would play no role. Table 6.1 does not allow me to decide which of these imagined possibilities would be the case; many other problems like this one can lie hidden in bivariate analyses.

These issues can be resolved by applying the algorithm of multiple regression, which stacks up all the independent variables against the dependent variable at once, in order to investigate the unique effect of each predictor. A multiple regression analysis resembles the correlation analysis of Table 6.1 in generating a comparison of independent variables, but differs from it in that all the independent variables are brought into the analysis at the same time, producing a more reliable ranking. In determining the unique contribution of each independent variable, multiple regression analysis does more than uncover predictors that are eliminated or re-ranked. An association between variables that is not significant, or perhaps only marginally so, when studied as a correlation, can turn out to be fully significant when the variable's unique contribution is investigated through regression analysis. Thus, and as noted in the work being replicated in the present study, multiple regression analysis is not only a ranking procedure that creates a more reliable hierarchy of independent variables, but an elimination and incorporation exercise that provides a more accurate assessment of which variables belong in the hierarchy. By

means of regression analysis, we will be able to ascertain, in a manner not possible in bivariate work, the relative importance of the regional and contact forces operating in New York, as well as of pressures related to education and socio-economic status. The regressions, despite containing an element of abstraction that makes them at times difficult to interpret, thus insure that the more accessible findings presented in the previous chapters can be fine-tuned, and verified, thereby encouraging greater confidence in their projection to the New York Latino population.

For multiple regression analysis to work well, independent variables should be independent measures, both conceptually and statistically. When the factors of the independent variables overlap, the regression is less effective. This overlap is referred to by statisticians as 'collinearity' (Newton and Rudestam, 1999:264). Thus it makes perfect sense to include Region and Exposure in the same regression model, since there is no connection between whether an informant is Caribbean or Mainlander and whether he or she was raised in Latin America or in New York. The variables Region and Exposure represent two totally different concepts and, in addition, the methodology used in this project for selecting informants for the sample insures that there are newcomers, established immigrants, and NYR in similar proportions in both the Caribbean and Mainland regional groups.

On the other hand, it would not be as useful to include in the same regression the contact-sensitive variables Generation and Exposure. These variables are related conceptually, since the factors of Exposure overlap with those of Generation. For similar reasons of collinearity, it would not be wise to include in the same regression these two variables along with the variable English. Although a variable of linguistic proficiency is conceptually distinct from one having to do with place of birth, age of arrival, and years in New York, there is a fair amount of overlap in

factor membership, since English proficiency increases with exposure to New York. To avoid the collinearity problem, none of the regressions that follow include any of these variables together. For similar reasons, the variables Education and SES are not included in the same models, given that the variable SES was created from the informants' reports of education and profession.

4.1 Multivariate regressions on preverbal rate for the whole sample of speakers

In the first three regressions that follow, the variables studied are Region and SES, which appear in three regression models alternatively with Generation, Exposure, or English. In the following three regressions, the variables studied are Region and Education, which appear alternatively with Generation, Exposure, or English. I first walk the reader through the regression tables and present results for the whole sample. Then, I perform separate regression analyses for the individual regional groups, followed by an interpretation of the findings.

As in Table 6.1, the results for Region in Tables 6.2 below are marked with a minus sign because informants from the Caribbean, who were coded with a lower number (1 for Caribbeans, 2 for Mainlanders), have a higher preverbal rate.

Table 6.2a
MULTIPLE REGRESSION
Dependent variable: Preverbal rate
Whole Sample

R2 = .20**	N	Standardized Beta	P
Region	139	-0.36	**
Education	138	0.17	*
Generation	137	0.16	*

* = p < .05
** = p < .01

Table 6.2b
MULTIPLE REGRESSION
Dependent variable: Preverbal rate
Whole Sample

R2 = .19**	N	Standardized Beta	P
Region	139	-0.35	**
Education	139	0.19	**
Exposure	137	-0.14	A

** = p < .01
a = p < .08

Table 6.2c
MULTIPLE REGRESSION
Dependent variable: Preverbal rate
Whole Sample

R2 = .20**	N	Standardized Beta	P
Region	139	-0.34	**
English	137	0.16	*
Education	137	0.14	A

** = p < .01
* = p < .05
a = p < .10

Table 6.2d
MULTIPLE REGRESSION
Dependent variable: Preverbal rate
Whole Sample

R2 = .19**	N	Standardized Beta	P
Region	139	-0.37	**
Generation	138	0.17	**
SES	137	0.15	*

* = p < .05
** = p < .01

Table 6.2e
MULTIPLE REGRESSION
Dependent variable: Preverbal rate
Whole Sample

R2 = .18**	N	Standardized Beta	P
Region	139	-0.36	**
SES	138	0.17	**
Exposure	137	0.14	a

** = p < .01
a = p < .08

Table 6.2f
MULTIPLE REGRESSION
Dependent variable: Preverbal rate
Whole Sample

R2 = .20**	N	Standardized Beta	P
Region	139	-0.35	**
English	138	0.19	*
SES	137	0.12	

* = p < .05
** = p < .01

OVERALL SIGNIFICANCE, INDIVIDUAL SIGNIFICANCE, AND VARIANCE. In the upper left hand corner of the regression tables, R-square, the multiple coefficient of determination, indicates whether the particular regression model as a whole is statistically significant and, if it is, how much variance in the dependent variable is accounted for by the independent variables entered into the particular regression model (Newton and Rudestam, 1999:249). The two asterisks next to the R-square figure in Tables 6.2a-f show that one can have a great deal of confidence in these models ($p < .01$). The tables also show that, in all cases, the three variables together account for about a fifth of the variance in preverbal rates among the 139 informants in the sample (R2 is .20, .19, .18).

RANKING VARIABLES BY STANDARDIZED BETA. The column marked Standardized Beta gives information on the ranking of the independent variables with respect to one another (Newton and Rudestam, 1999:267). Region is at the top of the hierarchy in all six models, just as it was in Table 6.1. Exposure is at the bottom in the two regressions where it appears.

STATISTICAL SIGNIFICANCE. The *p* column on the right hand side of the table reveals whether each variable, when taken into account simultaneously with all the others, retains statistical significance. Each variable is marked for its significance value by means of asterisks or letter, or is left blank when $p > .10$.

REGION, ENGLISH, GENERATION, EDUCATION. The regression tables show that the results for Region are clear and one can have full confidence in them: this variable is statistically significant in all six models. The variables English, Generation, Education are also all significant in their respective models. In the case of Education, it is worth mentioning that in the bivariate analysis performed in Chapter 3, it was only significant for informants from the Caribbean. Interestingly here, this variable emerges as more significant than expected for the whole set of informants.

SES. The variable SES is fully significant in Tables 6.2d and 6.2e and close to statistically significant in Table 6.2f. Whereas SES was only meaningful within one of the regions in our bivariate analyses, SES emerges from the regression analysis as a variable that plays a larger role than expected in our understanding of pronouns for the City as a whole.

EXPOSURE. With regard to the variable Exposure, we can have only limited confidence in its results in Tables 6.2b and 6.2e. This variable, which was also only close to statistically significant in Chapter 4 ($p < .09$), barely survives the regression test, which throws more light on the role that Exposure plays in accounting for variance in the use of preverbal pronouns, and on

how we should interpret the exposure to NYC force that underlies it.

4.2 Multivariate regressions on the preverbal rate for the regional sub-samples

In this section I analyze the same variables that were considered above for the whole sample, but taking into consideration each region separately. This time, I eliminate the variable Region, which in the previous section was applicable to the study of the whole sample but is obviously not relevant when each individual region is studied separately. The variables involved in the following regressions, then, are Generation, Education, English, Exposure, and SES.

The tables on the left (6.3a-f) present the results for Caribbeans and those on the right (6.4a-f) present the results for Mainlanders.

Table 6.3a
MULTIPLE REGRESSION
Dependent variable: Preverbal rate
Caribbeans

		Standardized	
R2 = .13**	N	Beta	P
SES	71	0.30	**
Generation	72	0.17	

** = p < .01

Table 6.4a
MULTIPLE REGRESSION
Dependent variable: Preverbal rate
Mainlanders

		Standardized	
R2 = .05	N	Beta	P
Generation	67	0.20	
SES	67	0.07	

Table 6.3b
MULTIPLE REGRESSION
Dependent variable: Preverbal rate
Caribbeans

		Standardized	
R2 = .16**	N	Beta	P
SES	71	0.31	**
Exposure	72	0.24	*

** = p < .01
* = p < .05

Table 6.4b
MULTIPLE REGRESSION
Dependent variable: Preverbal rate
Mainlanders

		Standardized	
R2 = .01	N	Beta	P
SES	67	0.09	
Exposure	67	0.08	

Table 6.3c
MULTIPLE REGRESSION
Dependent variable: Preverbal rate
Caribbeans

R2 = .22**	N	Standardized Beta	P
English	70	0.36	**
SES	71	0.24	*

** = p < .01
* = p < .05

Table 6.4c
MULTIPLE REGRESSION
Dependent variable: Preverbal rate
Mainlanders

R2 = .01	N	Standardized Beta	P
English	67	0.08	
SES	67	0.06	

Table 6.3d
MULTIPLE REGRESSION
Dependent variable: Preverbal rate
Caribbeans

R2 = .10a	N	Standardized Beta	P
Education	71	0.24	*
Generation	72	0.19	

* = p < .01
a = p < .10

Table 6.4d
MULTIPLE REGRESSION
Dependent variable: Preverbal rate
Mainlanders

R2 = .06	N	Standardized Beta	P
Education	67	0.18	
Generation	67	0.14	

Table 6.3e
MULTIPLE REGRESSION
Dependent variable: Preverbal rate
Caribbeans

R2 = .13*	N	Standardized Beta	P
Education	71	0.25	*
Exposure	72	0.24	*

* = p < .05

Table 6.4e
MULTIPLE REGRESSION
Dependent variable: Preverbal rate
Mainlanders

R2 = .04	N	Standardized Beta	P
Education	67	0.18	
Exposure	67	0.08	

Table 6.3f
MULTIPLE REGRESSION
Dependent variable: Preverbal rate
Caribbeans

R2 = .18**	N	Standardized Beta	P
English	70	0.37	**
Education	71	0.11	

** = p < .01

Table 6.4f
MULTIPLE REGRESSION
Dependent variable: Preverbal rate
Mainlanders

R2 = .03	N	Standardized Beta	P
Education	67	0.17	
English	67	0.03	

A few descriptive remarks about Tables 6.3 and 6.4a-f are in order before proceeding to the interpretation of the tables shown in this and the previous section.

THE REGIONS DIFFER IN SIGNIFICANCE AND AMOUNT OF VARIANCE ACCOUNTED FOR BY THE VARIABLES. Even a glimpse across the tables shows that the variables entered in the regression account for more variance in the preverbal rate among Caribbeans than among Mainlanders. The six regression models in Tables 6.3a-f, for Caribbeans, are all statistically significant but none of the models in Tables 6.4a-f are significant for Mainlanders. The R-square values for Caribbeans show that these variables account for between one tenth to one fifth of the variance in the dependent variable while they almost do not account for any variance at all in the dependent variable in the Mainland. Thus, this analysis shows that the variables Generation, Exposure, English, SES and Education are much more relevant to the study of New Yorkers with origins in the Caribbean than in the Mainland. The reason why the variance explained by these variables is low in general, and almost non-existent in the Mainland, is probably that there are a number of internal linguistic variables (not considered so far in the present study), which play a large role in the choice of subject pronoun placement and account for substantial amounts of the remaining variance. These Linguistic variables are analyzed in the following chapter.

THE REGIONS DIFFER WITH RESPECT TO VARIABLE RELEVANCE. For Mainlanders, none of the variables survive the regression to achieve conventional levels of significance. However, the Standardized Beta ranking allows for the conclusion that the variables that play the most important role in the Mainland are Education and Generation. For Caribbeans, Education does make an independent contribution, though Generation does not. In addition, the variables SES, English, and Exposure, to different extents, and involving varying degrees of confidence, make independent contributions to accounting for variance in the preverbal rate among Caribbeans.

THE REGIONS ARE SIMILAR FOR EDUCATION, DIFFERENT FOR THE REST. The variable Education produces somewhat similar results in both regions, but it achieves conventional levels of significance only in the Caribbean. English and SES are the highest ranked variables in the Caribbean, followed by Education and Exposure.

THE PLACE OF THE VARIABLE SES IN THE REGRESSIONS. The variable SES attains conventional levels of significance among Caribbeans whether the model includes Generation, Exposure, or English, and is in fact the highest ranked variable when English is not in the model. It is non-significant in all regression models among Mainlanders.

5. Interpreting multivariate regression results: Continuity and change in Spanish in New York

The whole-sample results for bivariate correlations and multivariate regressions in Table 1 and Tables 6.2a-f allow for important generalizations about Spanish in New York, which are detailed in this section. I will first interpret the results for the whole sample and then move to an interpretation of the regionally differentiated results.

5.1. Spanish in New York City as a whole

CONTINUITY AND CHANGE. The whole-sample results of Section 4.1 reveal a clear continuity between pronoun placement patterns in Latin America and the distribution of pronouns in New York City. This became evident by the fact that both in the bivariate correlations of Table 6.1 and in each of the multiple regressions of Tables 6.2a-f, the variable Region consistently occupies first place. This shows that Spanish in New York City is, at least with regard to the feature under study here, a continuation of the regionally differentiated usage of the Spanish of the Caribbean and the Latin American Mainland. The quantitative results clearly show the strength of the force of continuity when compared to the pressure of change on Spanish in New York. The bivariate correlation for Region in Table 6.1 is 13 points larger than

the following variable. In the multivariate regressions of Tables 6.2a-f, the Standardized Beta value for Region is double that of whatever variable is in second place in almost all regressions.

However, the results from the sample also clearly demonstrate that the Spanish of the City is not simply a continuation of the Spanish of Latin America, since there are strong elements of change that are giving it a different character, as evidence by the role played by the variables indicative of change: English, Generation, and Exposure. The evidence of language-contact influence from English, which is the other language of Latino bilinguals, is shown in the multivariate regressions of Tables 6.2 by the fact that the variables Generation, Exposure, and English occupy a distant, but clearly statistically significant, second place. Despite the strength of Region, the variables that indicate change in New York have endured the regression test, demonstrating that the change variables make a distinct contribution to an account of variability in the use of preverbal pronouns. The multivariate regression analysis allows me to conclude that the use of preverbal pronouns in Spanish in New York is primarily shaped by the continuing strength of the speakers' regional origin, and followed by the new and growing forces of contact. Regardless of their region of origin, the NYR use more preverbal pronouns than the LAR; and the more exposed tend to use more preverbal pronouns than the less exposed. Furthermore, these findings for the preverbal rate parallel the findings for the pronoun rate. Otheguy & Zentella (2012) also found that Region was the strongest predictor of an occurring pronoun, evidencing continuity of Latin American ways of speaking, but they also found evidence of change in the role played by the variables that predict language contact, such as Generation, Exposure, and English. The analysis and interpretation of the forces of continuity and change continues below for the whole sample of speakers as well as for the regional sub-samples.

THE TWO VARIABLES INDICATING CHANGE. The whole sample results also indicate that of the two variables that reflect change in New York (Generation and Exposure), Generation is the most predictive of differences in the preverbal rate. In the correlations in Table 6.1, Generation occupies the fourth place while Exposure occupies the last place and is only very marginally significant. In the regressions in Tables 6.2, Generation obtains a higher Standardized Beta value in its models (Tables 6.2a, d) than does Exposure in its own models (Tables 6.2b, e). These results tell us that whether or not the speaker is a newcomer and whether he or she is first or second generation does affect the use of preverbal pronouns (i.e., the variable Generation is clearly significant, independently of all the others, at .05, and the variable Exposure is somewhat significant, independently of all others, at .10). But the results indicate most clearly that, after Region, the variable that is most predictive of the changes taking place in the use of preverbal pronouns in New York is Generation. Therefore, the process of change and accommodation begins weakly in the first generation, but drastically increases with the New York-born children in the second generation, who are using preverbal subject pronouns more and sticking less to regional patterns. When comparing these results to those of Otheguy & Zentella (2012), revealing differences arise. Although both variables indicative of change were also significant in their study, Exposure came up as the most predictive of differences in the pronoun rate. Therefore, I can corroborate a pattern that had been noted in previous chapters: changes in the pronoun rate of speakers take place before changes in their preverbal rate. In other words, for speakers to begin placing more pronouns preverbally, they have to first start using more pronouns. Spanish-speakers in New York significantly augment their rate of occurring pronouns in the first generation (as evidenced by the role of Exposure in the study under comparative replication), while only slightly augmenting their rate of preverbal pronouns in the first

generation. However, the second generation of Spanish-speakers is placing occurring pronouns preverbally at a higher rate than the first generation of speakers.

THE IMPORTANCE OF THE VARIABLE ENGLISH. The position of the variable English in the regressions is especially supportive of the language contact hypothesis for Spanish in New York. English shows a higher Standardized Beta value in one of its model (Table 6.2f) than do Generation and Exposure in their models, and it occupies the second position after Region in the bivariate correlations. This finding parallels the finding for the pronoun rate in the study being compared, which led the authors to interpret that the variable that directly connects the use of occurring (and preverbal) pronouns in Spanish with knowledge of the influencing language, English, is stronger than the variables where the connection is a more indirect one related to place of birth and time spent in New York City.

SOCIO-ECONOMIC STATUS AND EDUCATION. An important discovery of this investigation is the relevance of SES to an understanding of pronoun placement variability. It ranked between the variables Generation and Exposure in the bivariate correlations, achieving a conventional level of significance (which Exposure did not). And despite ranking below Generation and English in two of its models (Tables 6.2d and 6.2f), it consistently ranked above Exposure (Tables 6.1 and 6.2e) and achieved a higher Standardized Beta than Exposure did. The regression results show that SES plays an independent and consistent role in accounting for placement variation. In one of the regression models (Table 6.2f), SES is the non-significant third-ranked predictive element of preverbal variability, but in two of the models it occupies a significant second place and a significant third place. The results make it clear that when New York City is taken as a whole, the use of preverbal pronouns cannot be understood by considering only regional origins, immigrant generation and linguistic contact. Latinos with higher levels of

education and more white-collar professions, quite independently of their generation, level of exposure, or amount of knowledge of English, tend to place pronouns before nouns (tend to have a higher preverbal rate) than Latinos who are less educated and work in more blue-collar occupations. Even more interesting is the fact that this constitutes the exact opposite of the findings for the pronoun rate. Otheguy & Zentella (2012) found that it was actually the less educated and blue-collar-occupation Latinos who were the higher pronoun users. Thus, upper SES Latinos are more resistant to increasing pronoun use, but once they do increase the pronoun rate, probably motivated by the forces of change (higher levels of education tend to translate into higher English proficiency), then they are more accepting of preverbal pronouns than lower SES Latinos (with lower levels of education and, potentially, lower English proficiency).

5.2 Spanish in New York City's regional groups

The comparison across regions afforded by Tables 6.3 and 6.4 allows for an increased understanding of continuity and change in the Spanish spoken by New York Latinos.

THE REGIONS DIFFER WITH RESPECT TO HOMOGENEITY AND STABILITY. The results of Tables 6.3 and 6.4 ratify that the New York Latino population can be divided, with respect to the use of preverbal pronouns, into a relatively homogenous and changeable group of Caribbeans and a more heterogeneous and stable group of Mainlanders[1]. For speakers who trace their origins to the Caribbean, the variables impacting the preverbal rate that are associated with language contact, namely Exposure and English, are significant, and so are the variables of SES and Education. In contrast, none of these variables is significant for Mainlanders. While the Caribbeans appear to be a group in which everyone uses preverbal pronouns in the same way, though using them more and more under the influence of English, the Mainlanders are differentiated by their country of origin, and more resistant to the influence of English for the

variable under study. Given the chain effect that the increase in the pronoun rate seems to have on pronoun placement, it appears to be the case that, since Caribbeans come to the City with higher pronoun rates, contact with English easily leads them to increase their preverbal rate. On the other hand, Mainlanders, who come to the City with lower pronoun rates, have to first increase their use of pronouns before moving on to increase their use of preverbal pronouns. Therefore, in the case of the preverbal rate, Caribbeans behave in a more homogeneous and changeable fashion whereas Mainlanders behave in a more heterogeneous and stable fashion. One could hypothesize that there is a threshold for the pronoun rate, and that once speakers reach that threshold, they begin to increase their preverbal rate. Caribbeans are closer to (or beyond) that threshold than Mainlanders, and they easily continue their path with the influence of English. Therefore, they soon begin to place more pronouns preverbally. Mainlanders, further away from the threshold, first need to get closer to it, and they do so under the influence of dialect (Caribbean Spanish) and language (English) contact. But it is only in the second generation that they reach the threshold for the pronoun rate and begin their increase of the preverbal rate.

SES. The comparison across regions proves that the differences between SES groups that are detectable in the whole sample are particularly concentrated in the Caribbeans, and absent in the Mainland. Among Mainlanders, SES differences disappear altogether in the regression, whereas among Caribbeans they remain very strong, and in fact the variable occupies first and second place in the three models where it appears. The regressions show that variability in the use of preverbal pronouns among Caribbeans can be accounted for in part by educational and occupational differences. Furthermore, I have reason to believe that these SES differences in the Caribbean population show continuity from ways of speaking in the Caribbean. In the small sub-

sample of 19 Caribbean newcomers, a regression model including English and SES shows that SES is highly significant. Furthermore, in this model, R-square is .42, almost double the size it has in the same model for the whole Caribbean group. This means that back in Cuba, the Dominican Republic, and Puerto Rico, the use of preverbal pronouns is very likely influenced by educational and occupational differences even more than in New York.

GENERATION. Another difference between the regions is accounted for by the variable Generation. As expected in view of the results analyzed above, this variable loses significance in the regressions for the Caribbean, but it is the highest ranked in the regressions for the Mainland. This indicates that change in Caribbean speakers begins strongly in the first generation, as evidence by the variable Exposure. However, change in Mainlanders only begins, albeit subtly, in the second generation of speakers. The results of Generation with respect to the regions lends further support for the threshold hypothesis illustrated above.

6. Summary and discussion

The data analyzed in this chapter has allowed me to confirm two facts previously stated in the literature on the Spanish spoken in New York: (a) there is a difference in preverbal pronouns placement between the regions, (b) the force of continuity of Latin American ways of speaking is stronger than that of change in shaping preverbal pronoun use in New York City. Furthermore, I have been able to confirm my own proposal regarding a connection between the pronoun rate and the preverbal rate: for changes in pronoun placement to occur, changes in rates of occurring pronouns have to take place first.

THE DISTRIBUTION OF PREVERBAL PRONOUNS BY REGION. The analysis of pronoun placement in the corpus has shown that Caribbeans use more preverbal pronouns than Mainlanders, which is not a new contribution of the present study. Dialectal research has noted

that in the Caribbean the rate of subject pronoun usage is larger than that of the Latin American Mainland (López-Morales 1992:137, Lipski 1994:241, Cameron 1995:8, and Otheguy & Zentella 2012, among others) and that speakers from the Caribbean tend to place pronominal and nominal subjects preverbally, in the same linguistic environments where Spanish-speakers of other areas either have a null subject or a postverbal one (Morales 1999:78, Toribio 2000:9, Ordoñez and Olarrea 2001:223-225, Goodall 2004:104). The present analysis contributes to what had been previously noted by confirming by means of the preverbal rate, in a large corpus, through both bivariate correlations and ANOVAS, and by establishing its independence of other factors through multiple regression modeling. In addition, the regional distinction has proven useful for organizing the data and it has allowed me to use it to base my conclusions regarding dialectal continuity.

THE REASON WHY CONTINUITY IS STRONGER THAN CHANGE. There are a number of speaker factors offered by Otheguy & Zentella (2012) that help explain the now firmly established primacy of regional continuity in the use of preverbal pronouns in the Spanish of New York. One is related to the pattern of immigration from the Caribbean and the Mainland into New York City. While Spanish-speakers have been in the City for more than a century, immigration from the Spanish-speaking Caribbean intensified only in the last 60 years. Dialectal diversification started later with the arrival of substantial numbers of Mainlanders from Colombia, Ecuador, and Mexico, less than two decades ago, a trend that continues (see chapter 1) and has the potential of changing the linguistic context of Spanish in the City. Furthermore, the constant influx of immigrants both from the Caribbean and the Mainland contributes to the maintenance of ways of speaking in the homeland. This is exacerbated by the fact that Latinos, the same as other immigrant groups, tend to concentrate in specific neighborhoods. Other important factors include

technological innovations that allow regular communication with the homeland, as well as the ease of travel for some Latino groups to their homeland. For these reasons, the rise of a New York Spanish linguistic norm strongly marked by the influence of English, while clearly detectable in the data, has not yet taken a definitive form to supersede the persistence of homeland usage norms and the identification of Latinos with their areas of origin in Latin America. Furthermore, it is a well-established fact in the literature that immigrants tend to abandon the heritage language by the third generation, which would be, otherwise, the one using a full-fledged version of New York Spanish.

There are also reasons for the strength of continuity that are purely linguistic in nature. For a drastic change in pronoun placement to occur, a change in the grammar of the speakers has to take place. As mentioned in earlier chapters, if one assumes that the generative analysis of pronouns is correct and relevant to natural usage, the change would have to affect the null subject parameter that has been set in the internal linguistic system of the speakers early on in the acquisition of Spanish as their first language. The creation and development of speakers' core grammar components culminates around the age of five, while the lexicon continues to develop forever. This may be part of the reason why changes in the lexicon seem to occur very fast in contact situations, with speakers using new vocabulary from the contact language early on, and borrowing new words easily. However, changes in the underlying grammar do not occur right away and may not be so easily incorporated. Changes in the rate of preverbal pronouns are directly related to a change in the syntactic component of speakers' internal grammar. Thus, if the implicit linguistic system develops at least partly from input, then it makes sense that we can see these changes in the second generation, who has been exposed to input that already contains both a higher rate of occurring pronouns and a higher rate of preverbal pronouns and who has

also been exposed to English from birth or very early on. Given the slow nature of changes to the internal grammar of speakers, the findings of this chapter lend support, in part, to the null subject parameter hypothesis. An increase in occurring pronouns, leads to a later increase in preverbal pronouns, or in other words, as the internal grammar of speakers changes to move away from that of a typical null subject language (by having less and less null subjects), it begins to lose the ability to place pronouns postverbally, which is one of the characteristics of null subject languages. Therefore, the results of this investigation suggest that an increase in the pronoun rate leading to an increase in the preverbal rate are evidence of parametric changes in the underlying grammar and not a case of incomplete acquisition of Spanish grammar (Montrul 2004, 2008).

In sum, the regression models that were presented in this chapter allow me to confirm that in order to understand the variable use of occurring subject pronouns in Spanish-speaking New Yorkers, several external conditioning variables have to be taken into consideration: the origin of the speakers, their level of education, their socio-economic status, the generation they belong to, how exposed they have been to the City, and their level of English proficiency. It is the combination of these conflicting pressures of continuity (with Latin American ways of speaking) and change (by contact with English) that helps interpret pronoun placement variation in particular, and Spanish in general, as it is spoken in New York.

[1] Mainlanders' heterogeneity is evidenced by consistently getting larger standard deviations in the ANOVAs than the Caribbeans do.

CHAPTER 7

ANALYSIS OF INTERNAL VARIABLES IN THE SUBSET CORPUS

1. Introduction

This chapter studies the role played by internal (i.e. linguistic) factors in the placement of overt pronouns in Spanish in New York. These factors are in contrast to the external or socio-demographic factors that have been considered up to this point, and which revealed that continuity with Latin American norms and language contact are independent forces and that each makes a unique contribution to the use of pronouns in New York. The social factors on which these generalizations were based included country of origin, place of birth, amount of exposure to the City environment, etc. The internal or grammatical factors that I investigate here are all related to the eligible finite and non-finite verb tokens with which overt pronouns variably co-occur. For this part of the project, a subset of the original corpus was used containing only the verbs occurring with pronouns in the sociolinguistic interviews of 22 of the original 139 informants in the corpus. The need for a subset corpus arose because two of the main variables that I wished to investigate had not been fully extracted from or systematically coded in the original corpus. These two variables, interrogative phrases and non-finite verbs, were extracted and coded especially for the work of this dissertation, along with other syntactic properties (detailed herein below) and chosen based on a review of the literature because of their potential effect on pronoun placement. Thus, the data analyzed in this chapter includes all finite and non-finite verbs produced by 22 informants with either a preverbal or a postverbal subject pronoun.

The subset of 22 interviews was chosen to represent a stratified sample of participants raised in NYC, newcomers, and established immigrants, from the Caribbean (Cuba, Puerto Rico, Dominican Republic) and the Latin American Mainland (Mexico, Colombia, Ecuadorians).

Table 7.1 shows the number of informants in the subset corpus by dialectal region of origin and immigrant generation.

Table 7.1. Classification of Informants

	Newcomer	Immigrant	NYR	**Total by Region**
Caribbean	4	4	3	**11**
Mainland	4	4	3	**11**
Total by Immigrant Generation	**8**	**8**	**6**	

Since the subset corpus was created for the purposes of analyzing the differences between finite and non-finite verbs, and declarative and interrogative phrases with regards to the placement of overt subjects, all eligible finite and non-finite verbs occurring with overt pronouns in the subset corpus were identified and entered into a separate database (the subset database). Besides, since interrogative phrases represented a highly variable environment as discussed in the literature, and most interrogative phrases were produced in the corpus by the interviewers (whose speech had been transcribed in the corpus but not coded or entered in the full database), I decided to also enter and analyze all finite and non-finite verbs occurring with pronouns in questions produced by interviewers[1] in the subset corpus. The identification of non-finite verbs with overt pronouns in the subset corpus was performed automatically by means of a computer program that I developed in the Perl programming language for this purpose (finite verbs had already been identified in the original database and corpus). Each of the verbs yielded by the program was manually checked before being entered into the database. I also developed and used a program to identify all questions by interviewers containing a verb with an overt pronoun. Then, all finite and non-finite verbs in the Subset (from interviewers and interviewees) were coded for 11 linguistic variables and 2 socio-demographic variables identified as relevant to the placement of pronouns. Figure 2 shows the total number of verbs occurring with pronouns, the

total number of questions containing verbs with pronouns, and the total number of preverbal and postverbal pronouns in the subset corpus.

Table 7.2. Classification of verbs and pronouns in the subset corpus

	Number of Verb Tokens	**Percentage of All Verbs**
Verbs with overt pronouns	1740	100%
Finite verbs with overt pronouns	1719	99%
Non-finite verbs with overt pronouns	21	1%
Preverbal pronouns	1511	87%
Postverbal pronouns	229	13%
Questions with overt pronouns	424	24%

The use of data from the interviewers, a resource that was not exploited in prior work on the Otheguy-Zentella corpus, made it possible for me to observe the effects of all linguistic variables on pronoun placement. I can also study the regional variable, since interviewers in the project were in most cases of the same national and regional origin as the informants. However, I may not have, in all cases, information on the generation to which the interviewers belong, and this may have limited some of the conclusions involving interrogatives and non-finites.

Table 7.2 shows that the number of non-finite verbs is very small and it may not be possible to draw significant conclusions with that variable. However, the number of overt pronouns occurring in questions is quite high, since it accounts for 24 percent of all the overt pronouns in the subset corpus. Also to be noted is the fact that the number of postverbal pronouns increases significantly once the interviewer data is included. Whereas in the Otheguy-Zentella database the percentage of postverbal pronouns was 4.5 percent, in the subset database the number of postverbal pronouns is constraint. Obviously this has nothing to do with the addition of non-finite verbs, given that they account for only one percent of the overall verbs occurring with an overt pronoun in the subset corpus. Thus, the addition of finite and non-finite

verbs occurring in questions uttered by the interviewer has increased the number of overt postverbal pronouns (from 4.5 percent before coding the speech of interviewers to 13 percent after coding the speech of interviewers). These numbers provided an early indication that in Spanish, there is indeed more variability in subject pronoun placement in interrogative phrases than in declaratives.

Subject pronouns in Spanish tend to distribute unevenly with respect to the linguistic environments associated with the verbs they co-occur with. It turns out that, for example, more postverbal pronouns in the corpus occur with intransitive verbs than with transitive verbs, e.g., more instances of 'I sleep' are found as ***duermo*** *yo* 'sleep I' (instead of *yo* ***duermo*** *'*I sleep*'),* than instances of 'I bought the bread' are found as ***compré*** *yo el pan* 'bought I the bread' (instead of *yo* ***compré*** *el pan* 'I bought the bread'*).* Similarly, more postverbal pronouns are found in the main clause of a simple sentence (***Estaba*** *yo sola* 'was I alone' – *007U*), than in the main clause of a complex sentence (*Me* ***pregunto yo*** *qué diría al ver el arrebato de Arelis* 'I wonder what s/he would say upon seeing Arelis's burst' – *007U*) or in subordinate (*Máquina es que le* ***decimos*** *nosotros realmente* 'Machine is what we really call it' – *007U*) and coordinate clauses (*Se ve muy buena y los demás familiares ahora pues* ***están*** *también ellos en, ¿cómo se llama? en desgracia 'It looks very good and the rest of family members are now also in, what is it called? in disgrace'*). The different conditions under which verbs appear are INDEPENDENT VARIABLES that co-vary to different extents with preverbal and postverbal pronouns. Parallel to the speaker-centered independent variables of chapter 5, such as Region, Exposure, etc., in what follows I analyze grammatical independent variables such as verb finiteness, clause type, verb transitivity, etc.

These internal independent variables differ in important ways from external ones like Region, Generation, or Exposure, which referred to characteristics that distinguished groups of speakers who showed different preverbal rates, but which were not in any direct way connected to the grammar of those speakers. As noted by Otheguy & Zentella (2012), notions like 'being from Ecuador' or 'being exposed to New York for a long time' are related to, and very likely causatively connected with, the details of grammars, but it makes no sense to think of them as actually being constitutive parts of those grammars. In contrast, internal variables like Clause and Finiteness should properly be seen as reflections of elements of performance grammars that serve to guide speakers in variably placing subject pronouns before or after verbs.

In variationist sociolinguistics, an independent variable like Clause and Finiteness may be called a FACTOR GROUP. The different levels or factors of an independent variable or factor group may be called CONSTRAINTS. So for example, the independent variable or factor group Finiteness has two factors or constraints, namely finite and non-finite; the independent variable or factor group Clause has four factors or constraints, aimed at describing the clause and phrase where the subject pronoun appears (main clause in simple sentence, main clause in complex sentence, subordinate clause, coordinate clause). In this study, I write in terms of independent variables and their constraints, to follow the terminology in the study under comparative replication.

Independent variables and their constraints are used in this chapter to further the study of continuity with Latin American ways of speaking (Reference Spanish) and change associated with language contact (Contact Spanish). Variables are compared with respect to their strength, or the extent of their influence on the choice to place pronouns before or after verbs. I want to find out, for example, whether Finiteness is a stronger or weaker independent variable than Clause. I pose the question first with respect to the whole subset corpus, and then inquire

whether the answer remains the same or is altered by region and by immigrant generation. Given the size of the subset corpus, I will not be able to make certain analyses that were conducted in the study under replication. For example, Otheguy & Zentella further investigated these questions by region in the Reference Spanish of immigrant newcomers, and the Contact Spanish of the established immigrants and the New York raised (NYR).

Regarding constraints, comparisons of strength are in terms of the probability of a pronoun being placed after the verb in the presence of a particular constraint. For example, the probability of a postverbal pronoun occurring with an intransitive verb is greater than with a transitive verb. I want to know whether the pattern of intransitive verbs encouraging postverbal pronouns is the same in the whole subset corpus, the two regions, and all three exposure groups, or whether this pattern has been disturbed by the extent of language contact, for example, for the exposure groups. I am thus attempting to determine whether internal considerations (internal variables and their constraints) can be used, as were the social considerations of previous chapters, to delineate the extent of continuity and contact-induced change in Spanish in New York.

In variationist sociolinguistics, the rankings of variables and of constraints are known respectively as VARIABLE HIERARCHIES and CONSTRAINT HIERARCHIES. When I ask whether, with respect to their impact on the placement of subject pronouns, Clause is stronger or weaker than Transitivity in a particular group of speakers, I am asking whether the groups share the same variable hierarchy. When I ask whether two groups agree in ranking intransitive verbs above transitive verbs, I am asking whether they share the same constraint hierarchy. Looking both at variable and constraint hierarchies is called comparative sociolinguistics (Tagliamonte, 2002:729). In this type of work, exemplified in such studies as Poplack (2000), Poplack &

Tagliamonte (1999) and Poplack & Sankoff (1987), the variables and constraints that condition variation are regarded as more revealing, for the purpose of identifying groups and their relationships, and for distinguishing between continuity and change, than the study of occurrence rates that we have carried out up to this point (cf. Poplack 2000:14).

A major difference between the previous chapters and the present one is the change to a new dependent variable. Previously, I was looking at a property of informants, namely their preverbal rates. Now, I am looking at a property of verb tokens, namely whether they are accompanied by a preverbal or a postverbal subject pronoun. For this reason, the co-varying independent variables of Chapters 3 – 6 were based on personal and demographic properties of individuals (Region, Exposure, etc.) whereas the co-varying independent variables here are grammatical properties associated with each verb token (Clause, Transitivity, etc.). With regards to the statistical analysis, whereas the study of social variables in the previous chapters has been conducted at the informant level, the analysis of grammatical variables of this chapter will be conducted at the item level. Another difference is that the dependent variable in previous chapters, called Preverbal Rate, was a continuous variable, that is, a percentage. The dependent variable here, abbreviated as Placement, is a categorical variable, that is, the placement of a pronoun before or after a particular verb token. This dependent variable is categorical for each verb, since each verb either appears with a preverbal or a postverbal subject pronoun. Therefore, in the multivariate analysis of this chapter, I switch from linear regression, which was used in the previous chapter, to logistic regression.

2. Grammatical variables in the study

After the dependent variable (Placement), the 10 independent variables are listed below along with their constraints, followed by their abbreviation. The variables were chosen for their

potential influence on the placement of pronouns based on existing research literature (Navarro Tomás 1948; Gili Gaya 1951; Lipski 1994; De Bruyne 1995; Butt and Benjamin 1996; Morales 1988, 1997, 1999; Toribio 2000; Ordoñez and Olarrea 2001; Zubizarreta 2001; Zagona 2002; Goodall 2004).

The list of independent variables and their constraints is mostly self-explanatory. So, for example, Finiteness, as formulated in this study, is a straightforward morphological variable based on the observation of the tense ending of the verb (e.g., whether it is tensed or not).

GRAMMATICAL VARIABLES AND THEIR CONSTRAINTS

Dependent variable:

Pronoun Placement Type (Placement)

Pronoun is preverbal

Pronoun is postverbal

Independent variables:

Verb morphology: Verb Finiteness Type (Finiteness)

Pronoun appears with a finite verb

Pronoun appears with a non-finite verb

Sentence structure: Clause Type (Clause)

Pronoun is in a main clause in a simple sentence

Pronoun is in a main clause in complex sentence

Pronoun is in a subordinate clause

Pronoun is in a coordinate clause

Sentence structure: Subordinate Clause Type (Subordinate)

Pronoun is in an Argument/Nominal Clause

Pronoun is in a Relative/Adjectival Clause

Pronoun is in a Temporal Adjunct Clause

Pronoun is in a Gerundive and Participial Adjunct Clause

Pronoun is in an Infinitival Adjunct Clause

Sentence Structure: Relative Clause Type (Relative)

Pronoun is in a Relativization of Direct Object Clause

Pronoun is in a Relativization of Indirect Object Clause

Verbal Syntax: Verb Transitivity Type (Transitivity)

Pronoun appears with a Transitive verb

Pronoun appears with a Intransitive verb

Pronoun appears with a Ditransitive

Sentence Structure: Phrase Type (Phrase)

Pronoun is in a Declarative Sentence

Pronoun is in a Non-declarative Sentence

Sentence Structure: Interrogative Phrase Type (Interrogative)

Pronoun is in a yes/no direct question

Pronoun is in a Wh question

Sentence Structure: Wh Question Type (Wh Question)

Pronoun is in a Bare Wh question

Pronoun is in a Non-bare Wh question

Pronoun is in an indirect question

Lexical Category: Pronoun Type (Pronoun)

Pronoun is *yo*

Pronoun is *tú*

Pronoun is *él, ella*

Pronoun is *nosotros, nosotras*

Pronoun is *vos*

Pronoun is *ellos, ellas*

Pronoun is *usted*

Pronoun is *ustedes*

Pronoun is *uno, una*

Pronoun is *unos, unas*

Pronoun Morphology: Pronoun Gender Type (Gender)

Pronoun is Feminine

Pronoun is Masculine

Gender is Not applicable

3. Variable and constraint hierarchies in the subset corpus

In this section I lay the groundwork for inquiring which of these independent variables is most relevant to the dependent variable Placement, that is, which linguistic considerations play the greatest role in guiding speakers in the decision to place an overt subject pronoun before or after the verb.

3.1 The use of cross-tabulations to select significant independent variables

First, I start by performing a bivariate analysis of each independent variable against the dependent variable, in order to find out which of the aforementioned independent variables are statistically significant with respect to pronoun placement variability. Below I present all cross-tabulations (significant and non-significant) because of the interesting information that they

provide. However, only those variables that yield significant results will be used in the subsequent multivariate regression analyses. Following convention, levels of significance are indicated by asterisks (** = p < .01, * = p < .05). The variables to be presented in the nine cross-tabulation tables that follow are:

Clause Type

Subordinate Clause Type

Relative Clause Type

Verb **Transitivity**

Phrase Type

Interrogative Phrase Type

Wh-question Type

Verb **Finiteness**

Pronoun Type

Pronoun **Gender**

The first significant cross-tabulation appears in Table 7.3, below, and it involves the dependent variable Placement and the independent variable Clause.

Table 7.3

Pronoun Placement by Clause

Cross-tabulation

N = 1743

	Main Clause Simple Sentence	Main Clause Complex Sentence	Subordinate Clause	Coordinate Clause
Preverbal	79%	94%	90%	86%
Postverbal	21%	6%	10%	14%

Sig. **

Table 7.3 shows that speakers use more postverbal pronouns in simple sentences (one clause, one verb) than they do when the pronoun is in the main clause of a complex sentence. It also shows that when the pronoun is in the subordinate clause, it tends to appear preverbally more often than when it appears in a coordinate clause. However, it has been claimed in the literature (Zagona 2004) that certain types of subordinate clauses (infinitival adjuncts) might encourage the use of postverbal subjects. Therefore, following Zagona's classification of subordinate clauses, I further classified all verb tokens appearing with an overt pronoun in a subordinate clause into the following five types of subordinate clauses:

Table 7.4

Pronoun Placement by Subordinate Clause

Cross-tabulation

N = 412

	Nominal Clause	Relative Clause	Temporal Adjunct Clause	Gerundive / Participial Clauses	Infinitival Adjuncts
Preverbal	87%	93%	90%	100%	80%
Postverbal	13%	7%	10%	0%	20%

Not Sig.

Although the cross-tabulation illustrated in Table 7.4 is not statistically significant, it can be seen that it is indeed in infinitival adjuncts where most postverbal subject pronouns appear. For example, speaker 096P produced a preverbal pronoun in an infinitival adjunct clause (in square brackets in the following two examples): *Yo no tengo tiempo [para **yo** <u>ponerme</u> a estudiar cuatro o cinco años]* 'I don't have time to study for four or five years', while speaker 173C produced a postverbal pronoun in an infinitival adjunct clause: …*no estaba preparado [para entrar al bachillerato y <u>empezar</u> **yo** a coger clases en español]* 'I wasn't ready to enter high school and start taking classes in Spanish'. Interestingly, the relative clause shows a much smaller proportion of postverbal subject pronouns, which is aligned with Goodall (2004:5) expectation

that preverbal subjects would be acceptable in this type of construction given that the head of the relative clause (*que*) is usually heavily d-linked (referential), such as in the question formulated by 007U: *¿Qué es lo que **tú** <u>crees</u> que…?* 'What is it that you think that…?'. Morales (1997) had also noted a low rate of postverbal subject pronouns in relative clauses, but Butt and Benjamin (1996) claimed that relative clauses very often favored postverbal subjects, which is not the case in the subset corpus. Given the variability in opinions in the literature, I decided to look more in depth into this structure to see if different types of relative clauses motivated speakers to place subject pronouns in different positions. Thus, I classified relative clauses by whether a direct or an indirect object was relativized, and the results appear below in Table 7.5.

Table 7.5

Pronoun Placement by Relative Clause Cross-tabulation

N = 72

	Relativization of Direct Object	Relativization of Indirect Object
Preverbal	94%	88%
Postverbal	6%	12%
Not Sig.		

Although it seems that relativization of indirect objects (*…en la escuela donde <u>está</u> **él**…*- 348M '…in the school where he is at...') more often promote the use of postverbal subject pronouns than relativization of direct object (*No me dejó ni pintar el apartamento al cual **yo** <u>tenía</u> -234U* 'She didn't even let me paint the apartment that I had'), the difference between the two is not statistically significant. It is possible that a larger sample would allow for clearer findings with regards to this variable.

The next significant cross-tabulation, presented in Table 7.6 below, gives an initial illustration of the role played by verbal transitivity in pronoun placement.

Table 7.6

Pronoun Placement by Transitivity

Cross-tabulation

N = 1738

	Transitive	Intransitive	Ditransitive
Preverbal	88%	83%	97%
Postverbal	12%	17%	3%

Sig. **

The foregoing table shows that speakers use more postverbal pronouns with intransitive verbs than with transitive verbs. This tendency had been noted by De Bruyne (1995) who stated that verbs that do not take a direct object (intransitive) favor the verb-subject order. Although intransitives do not really favor postverbal pronouns over preverbal pronouns, they do encourage the appearance of a postverbal pronoun more often than transitive verbs do. The analysis presented in Table 7.6 also shows that ditransitive verbs accept postverbal subject pronouns even less than transitive verbs.

The next cross-tabulation involves the analysis of non-declarative sentences (exclamative, interrogative, and imperative sentences) which are often mentioned in the literature as promoting the use of postverbal pronouns. Some authors have even claimed that postverbal subjects are required in interrogatives with a wh-word (Butt and Benjamin 1996, Zubizarreta 2001). Table 7.7 presents significant results for a cross-tabulation that classifies all applicable sentences in the corpus as declarative and not-declarative and it shows that indeed, there are more postverbal pronouns in non-declarative sentences than in declarative sentences.

Table 7.7
Pronoun Placement by Phrase
Cross-tabulation
N = 1740

	Declarative	Non-declarative
Preverbal	88%	83%
Postverbal	12%	17%

Sig. **

These results are aligned with the claims of Zagona (2002) and Goodall (2004) who noted that specific non-declarative constructions (to be analyzed in more detail below) favor the appearance of postverbal subjects. These results motivated a further classification of interrogative sentences into yes/no questions and wh-questions.

Table 7.8
Pronoun placement by Interrogative
Cross-tabulation
N = 419

	Yes/No Question	Wh-Question
Preverbal	92%	72%
Postverbal	8%	28%

Sig. **

Table 7.8 shows that wh-questions, that is interrogative sentences headed by a wh-word (e.g. *qué, cuál, cómo,* 'what, which, how',etc.) or indirect questions (e.g. *Me pregunto qué piensa…* 'I wonder what he thinks'), encourage the use of postverbal subject pronouns more than yes/no questions do, and the difference between the two types of questions is highly significant. Since there are three distinct types of wh-questions, the next step was to analyze whether wh-question type accounted for placement variability. These questions were further classified into bare wh-questions, which refer to questions headed by a wh-word followed by a verb (e.g. *¿Qué dijiste tú?*

'What did you say'), non-bare wh-questions, which refer to questions head by a wh-word with a complement noun (e.g. *¿Qué libros trajeron ellos?* 'What books did they bring'), and indirect questions, which were illustrated above.

Table 7.9
Pronoun Placement by Wh-Questions
Cross-tabulation
N = 205

	Bare Wh	Non-bare Wh	Indirect Question
Preverbal	72%	68%	79%
Postverbal	28%	32%	21%
Not Sig.			

Table 7.9 shows that the three types of interrogative phrases involving a wh-word favor postverbal subject pronouns in a similar proportion, making the cross-tabulation not significant. That is approximately 30 percent of the questions uttered by the speakers in the subset corpus that involve a wh-word and a subject pronoun, present the subject pronoun in postverbal position. These results coincide with the statements made by Goodall (2004:5) regarding matrix versus embedded contrast in wh-questions. His analysis predicted that there should be no contrast between fronted wh-questions and indirect questions. However, there are conflicting views in the literature. For example, Zubizarreta (2001:183) claims that Romance languages exhibit a constraint "that disallows the subject to intervene between a fronted wh-phrase and the verb." She claims that bare wh-questions activate this constraint while non-bare wh-questions do not. Zubizarreta (2001:184) also stated that the same constraint that applies to bare wh-questions would apply to embedded clauses (indirect questions) in Spanish. Although the data analyzed in this project clearly shows a tendency to place subject pronouns postverbally in interrogatives involving a wh-word, the fact that approximately 70 percent of wh-questions were uttered with a

preverbal subject pronoun cannot be ignored. That is, wh-questions with a preverbal subject pronoun such as *¿Cómo tú crees que es la diferencia?* (005U) 'What do you think the difference is?' are far more frequent than those with a postverbal pronoun such as *¿Cómo se llama él?* (007U) 'What is his name?' The data in the corpus suggests that a wh-phrase is a factor that encourages postverbal placement rather than a constraint on preverbal placement. Furthermore, differences proposed by Zubizarreta between the three types of wh-questions cannot be supported in this corpus. It seems that bare and non-bare wh-questions and indirect questions favor postverbal subject pronouns in a very similar fashion.

Several authors have pointed out that the type of personal pronoun involved is a determining factor in pronoun placement variability. Table 7.10 below shows that the pronoun placed more often in preverbal position is the second person singular pronoun *tú* and the pronouns more often placed in postverbal position are the formal second person singular pronoun *usted* and the second person plural pronoun *ustedes*.

Table 7.10
Pronoun Placement by Pronoun Type
Cross-tabulation
N = 1738

	yo	tú	él/ella	nosotros/as	ellos/as	usted	Ustedes	uno/a
Preverbal	87%	93%	84%	88%	86%	74%	40%	75%
Postverbal	13%	7%	16%	12%	14%	26%	60%	25%

Sig. **

The foregoing analysis suggests that degrees of formality in a conversation may affect pronoun placement given the differences in placement between formal and informal second person pronouns. It would seem that preverbal subjects are used in informal conversations, while postverbal subjects may be favored in formal conversations.

As it was mentioned earlier in this chapter, it has been suggested in the literature (Toribio 2000, Zagona 2004) that non-finite verbs may produce more variability in pronoun placement than finite verbs. Although there are only 21 non-finite verbs in the subset corpus, I still run the cross-tabulation between the dependent variable Placement and the independent variable Finiteness which contains 1743 verb tokens (1722 finite verbs and 21 non-finite verbs).

Table 7.11
Pronoun Placement by Finiteness
Cross-tabulation
N = 1743

	Finite	Non-finite
Preverbal	87%	81%
Postverbal	13%	19%

Not Sig.

The analysis presented in Table 7.11 shows slightly more postverbal pronouns with non-finite verbs. However, the difference is not statistically significant. A larger corpus is needed to confirm these results.

The last variable considered looks at whether pronoun gender plays a role in subject pronoun placement. The results appearing in Table 7.12 below show that whether the subject pronoun is masculine or feminine does not seem to affect pronoun placement in any specific manner, when the whole corpus is considered.

Table 7.12
Pronoun Placement by Pronoun Gender
Cross-tabulation
N = 553

	Feminine	Masculine	Not Available
Preverbal	86%	88%	87%
Postverbal	14%	12%	13%

Not Sig.

After evaluating the significance of each of the independent variables separately, by means of cross-tabulations with the dependent variable, I have discovered that, in the subset corpus, only the following five variables have a statistically significant bivariate relationship with the dependent variable Placement:

1. **Clause**

   Pronominal subjects appear more often in postverbal position when they are found in the main clause of a simple sentence, followed by, in this order, when they appear in a coordinate clause, a subordinate clause, and a main clause in complex sentence.

2. **Transitivity**

   Transitive verbs are more often accompanied by a postverbal pronominal subject than intransitive and ditransitive verbs, in this order.

3. **Phrase**

   Pronominal subjects are more often in postverbal position in non-declarative than in declarative sentences.

4. **Interrogative**

   There are more postverbal subjects in wh-questions than in yes/no questions, and

5. **Pronoun**

   The type of pronoun that more often appears in postverbal position is *ustedes*, followed by *usted, uno/a, él/ella, ellos/as, nosotros/nostras, ustedes, yo,* and *tú*.

In order to compare the strength of the different independent variables, I take the same multivariate approach of the previous chapter, comparing independent variables through the use of regression analysis.

3.2 The use of logistic regression to construct variable hierarchies

Because the dependent variable Placement is dichotomous (preverbal pronoun vs. postverbal pronoun) and not continuous like the dependent variable Preverbal Rate, I switch from linear regression, which was used in the previous chapter, to logistic regression, the proper statistical measure to handle dichotomous dependent variables (Menard 2002).

The regression tables that appear below present hierarchies that rank variables according to how much variance between preverbal and postverbal pronouns they cover. In order to assess the relative strength of the independent variables in a logistic regression, I look at the Wald statistic associated with each of the variables (Menard 2002:43). Variable hierarchies are rankings in terms of each independent variable's Wald value.

I start this part of the investigation by including all significant variables identified with the cross-tabulation. The goal is to investigate which of these variables better predicts the appearance of a postverbal subject pronoun in the subset corpus. The results are in Table 7.13 below.

Table 7.13

LOGISTIC REGRESSION

Dependent: Placement

N = 424

R2 = .41

Rank	Variable	Wald	Sig
1st	Pronoun	49.52	**
2nd	Interrogative	15.68	**
3rd	Clause	13.07	**
4th	Transitivity	.01	
5th	Phrase Type	0	

The figure for N on the fourth line of Table 7.13 is the number of verb tokens analyzed in this regression model. 1735 is the total number of verb tokens in the subset corpus of which 424 were included in this regression analysis (the reason why not all verb tokens are included is explained below). The table also indicates, in the coefficient for R square (R2), the amount of variance between preverbal and postverbal subject pronouns that is accounted for by this regression model. It is telling us that this model explains 41 percent of variance in the subset corpus. The Wald value assigned to each variable has its statistical significance noted with asterisks in the usual way. The regression model tells us that the variable Pronoun is the biggest predictor of postverbal subject pronouns in the whole subset corpus. Pronoun is followed by Interrogative, Clause, Transitivity, and Phrase. However, Transitivity and Phrase have lost statistical significance in the regression model.

The model also gives us information about the order of the constraints within the variables, and which of the constraints predicts the appearance of a postverbal pronoun, and which discourages the appearance of a postverbal pronoun. For brevity sake, instead of presenting five tables with the constraint hierarchy of each variable, I am going to mention below only those constraints that predict a postverbal pronoun within each variable and ignore those that do not. The model provides an Exp(B) value for each of the constraints. Given the methodology used to code the data, if the Exp(B) value is larger than 1, it means that the constraint predicts the appearance of a postverbal pronoun. Values smaller than 1 indicate that the constraint disfavors postverbal pronouns. Therefore, below and in subsequent regressions, I will present, in narrative, the predicting constraints for each variable in their hierarchical order, with their Exp(B) value in brackets.

In the case of the variable Pronoun, the pronouns that favor postverbal placement are *ustedes* (2.89), *uno* (2.55), *usted* (1.10). All other constraints for the variable Pronoun had values that were smaller than 1. In the case of the variable Clause, the constraint that favors postverbal pronouns is main clause in simple sentences (1.38), and the coordinate clause (1.22). In the case of Interrogatives, the constraint that favors postverbal pronouns is wh-questions (2.24) as opposed to yes/no questions. Regarding the variable Transitivity, the regression informs that intransitive verbs (4.97) favor postverbal pronouns more than transitive verbs (3.75), and that ditransitive verbs disfavor them. For the variable Phrase, non-declarative sentences (1.24) favor the appearance of a postverbal pronoun while declarative sentences do not.

The reason why only 424 verb tokens were included in this analysis is that the variable Interrogative only includes non-declarative interrogative phrases and ignores declarative phrases. In order to include all tokens in the analysis I run a second regression model excluding the variable Interrogative. The variable ranking changed only a little. Pronoun continues to be significant and to rank first; Clause continues to be significant and second. But in this model Phrase was significant and ranked third, and Transitivity was almost significant and ranked fourth. Obviously, there is collinearity between the variables Phrase and Interrogative, but once we remove Interrogative from the analysis, R2 diminishes and only 13 percent of variance between preverbal and postverbal pronouns is accounted for.

Thus far I have been able to identify the variables that play a statistically significant role in the placement of overt pronouns, their ranking, and the constraints within the variables that encourage the use of postverbal pronouns. Next, I will analyze whether these same variables play a role in the placement of overt pronouns in both regions, and whether they are ranked in the same way or not.

3.3 Variable hierarchies across the regional groups

In this section I analyze the same internal variables that were considered above for the whole of the subset corpus, but taking into consideration each region separately. Dialectal research based for the most part on qualitative data has noted that speakers from the Caribbean tend to place pronominal and nominal subjects before verbs in the same linguistic environments where Spanish-speakers from other geographical areas either have a null subject or a postverbal one (Morales 1999, Toribio 2000, Ordoñez and Olarrea 2001, Goodall 2004). The cross-tabulations that follow provide an initial bivariate exploration into regional differences in pronoun placement variability with the aim of providing statistical support to the claims made in the literature. One of the first findings to highlight is that each region has different significant variables, as shown by the cross-tabulations that follow.

Table 7.15a
Pronoun Placement by Pronoun Type
Cross-tabulation
CARIBBEAN
N = 1087

	yo	tú	él ella	nos.	ellos ellas	ud.	uds.	uno una
Preverbal	93%	94%	85%	82%	88%	85%	100%	76%
Postverbal	7%	6%	15%	18%	12%	15%	0%	24%

Sig. **

Table 7.15b
Pronoun Placement by Pronoun Type
Cross-tabulation
MAINLAND
N = 648

	Yo	tú	él ella	nos.	ellos ellas	ud.	uds.	uno una
Preverbal	78%	85%	84%	92%	85%	52%	38%	73%
Postverbal	22%	15%	16%	8%	15%	48%	62%	26%

Sig. **

Although the cross-tabulation between the independent variable Pronoun and the dependent variable Placement is significant in both regions, the percentage of postverbal pronouns is larger in the Mainland than in the Caribbean for every pronoun except *nosotros/as*. In the Caribbean, the two pronouns that are more often placed before the verb are *yo* and *tú*. Ordoñez and Olarrea (2001) make claims regarding differences between the first and second person singular pronouns for Caribbean Spanish that are not supported in this corpus. However, it is true, as they claimed, that the third person singular pronoun is placed before the verb less often than first and second

person. Differences between first and second person are seen in the Mainland, though, where the third person behaves like the second person and unlike the first. In the Mainland, the two pronouns more often placed after the verb are *usted* and *ustedes*.

I present next the results of the cross-tabulation for the variable Clause with regards to the dependent variable Placement.

Table 7.16a

Pronoun Placement by Clause
Cross-tabulation
CARIBBEAN
N = 1092

	Main Clause in Simple S.	Main Clause in Complex S.	Subordinate Clause	Coordinate Clause
Preverbal	85%	96%	93%	90%
Postverbal	15%	4%	7%	10%

Sig. **

Table 7.16b

Pronoun Placement by Clause
Cross-tabulation
MAINLAND
N = 648

	Main Clause in Simple S.	Main Clause in Complex S.	Subordinate Clause	Coordinate Clause
Preverbal	68%	90%	81%	81%
Postverbal	32%	10%	19%	19%

Sig. **

Once again it can be seen that although the variable Clause is significant in both regions, the percentage of postverbal pronouns is larger in the Mainland, in each clause. Both regions tend to place more pronouns after the verbs in Simple Sentences than in any other clause type.

For the variable Interrogative the results are slightly different between the Caribbean and the Mainland.

Table 7.17a

Pronoun Placement by Interrogative
Cross-tabulation
CARIBBEAN
N = 358

	Yes/No Question	Wh-Question
Preverbal	95%	79%
Postverbal	5%	21%

Sig. **

Table 7.17b

Pronoun Placement by Interrogative
Cross-tabulation
MAINLAND
N=66

	Yes/No Question	Wh-Question
Preverbal	68%	46%
Postverbal	32%	54%

NS

Although both regions behave in the same way (i.e. more postverbal subjects in wh-questions than in yes/no questions) there are more postverbal pronouns in the Mainland. Also, the cross-tabulation is only significant in the Caribbean, while in the Mainland it is only close to significant ($p < .08$). Differences on subject pronoun placement in interrogatives between the Caribbean and other dialects of Spanish have been addressed by many researchers. Lipski (1994) observed that in the Caribbean preverbal subjects are favored in wh-questions. This is indeed the case in the subset corpus given that 79 percent of occurring subjects are preverbal in wh-questions. On the other hand, Toribio (2000), who focused on the study of Dominican Spanish only and worked from speech samples gathered in New York and in Dominican Republic, claimed that in Dominican Spanish, word order is relatively rigid (SVO), irrespective of sentence type or verb class. She noted that this order is also maintained in questions, where the preverbal position is available to subject pronouns and full NPs alike. Although Toribio did not differentiate between types of questions, all her examples included wh-questions only. Given the results in Table 7.17a above, the subset corpus findings do not provide full support for Toribio's claims. Although Caribbeans place more subject pronouns before verbs, and so do Mainlanders in almost every constraint in every variable, in the case of wh-questions Caribbeans still place 21 percent of overt pronouns after verbs, which shows that word order is not that "rigid" after all, and that there is considerable variation in the region. Furthermore, the cross-tabulation results provide no support for claims of required subject-verb inversion in non-Caribbean dialects of Spanish made in the literature (Cuza, 2013).

Regarding the independent variable Phrase, there are similarities in constraint order between both regions.

Table 18a Pronoun Placement by Phrase Cross-tabulation **CARIBBEAN** N = 1092		
	Declarative	Non-declarative
Preverbal	93%	88%
Postverbal	7%	12%
Sig. **		

Table 18b Pronoun Placement by Phrase Cross-tabulation **MAINLAND** N = 647	
Declarative	Non-declarative
82%	62%
18%	38%
Sig.**	

As noted for the previous variables, Mainlanders place more pronouns postverbally, both in declarative and non-declarative phrases. But both regions use more postverbal subject pronouns in non-declarative phrases than in declaratives, and the cross-tabulation is significant in both. The findings above differentiate between declarative sentences and all non-declarative sentences which include yes/no questions, wh-questions, and exclamative sentences. The results show once again that word order is not as rigid in the Caribbean as it has been claimed. They also show that even in non-declarative phrases, Mainlanders also place more subject pronouns before verbs.

The differences between the regions begin with the variable Transitivity, the results of which are presented below.

Table 7.19a Pronoun Placement by Transitivity Cross-tabulation **CARIBBEAN** N = 1092			
	Transitive	Intransitive	Ditransitive
Preverbal	91%	93%	93%
Postverbal	9%	7%	7%
NS			

Table 7.19b Pronoun Placement by Transitivity Cross-tabulation **MAINLAND** N = 648		
Transitive	Intransitive	Ditransitive
84%	72%	79%
16%	28%	21%
Sig. **		

Whereas in the Caribbean there are no major differences in the amount of postverbal pronouns when the verb is transitive, intransitive or ditransitive, in the Mainland, speakers use more

postverbal pronouns with intransitive verbs than with ditransitive, and more with ditransitive than with transitive verbs. Furthermore, the cross-tabulation is only significant in the Mainland. Therefore, while verbal transitivity is a variable that triggers placement variability in Mainlanders, it does not seem to generate much variability in Caribbeans.

With regards to wh-questions, once again there are striking differences between the regions.

Table 7.20a
Pronoun Placement by Wh Questions
Cross-tabulation
CARIBBEAN
N = 163

	Bare Wh Question	Non-bare Wh Question	Indirect Question
Preverbal	78%	81%	70%
Postverbal	22%	19%	30%

Not Sig.

Table 7.20b
Pronoun Placement by Wh Questions
Cross-tabulation
MAINLAND
N = 41

	Bare Wh Question	Non-bare Wh Question	Indirect Question
Preverbal	45%	35%	100%
Postverbal	55%	65%	0%

Sig. *

Whereas Caribbeans use more postverbal pronouns in indirect questions than in bare and non-bare wh-questions, Mainlanders use more postverbal pronouns with non-bare and bare wh-questions. Also to be noted, Mainlanders use more postverbal than preverbal pronouns in these two types of wh-questions. In any case, I have to be cautious in the interpretation of these results given that Mainlanders only produced 41 verb tokens with overt pronouns in wh-questions.

The last independent variable presented in this section is Pronoun Gender.

Table 7.21a
Pronoun Placement by Pronoun Gender
Cross-tabulation
CARIBBEAN
N = 1089

	Feminine	Masculine	N/A
Preverbal	89%	87%	92%
Postverbal	11%	13%	8%

Not Sig.

Table 7.21b
Pronoun Placement by Pronoun Gender Cross-tabulation
MAINLAND
N = 647

	Feminine	Masculine	N/A
Preverbal	83%	89%	72%
Postverbal	17%	11%	29%

Sig. *

According to the results obtained, speakers place more feminine pronouns in postverbal position in the Mainland while they place slightly more masculine pronouns after the verb in the Caribbean. Interestingly, Caribbeans tend to place preverbally most of the pronouns that are not marked for gender, while Mainlanders place a large amount of unmarked pronouns postverbally.

The rest of the independent variables (Finiteness, Subordinate, Relative) are not presented because they were not significant in either region.

In sum, from the cross-tabulations above the following conclusions can be drawn:

- Both regions use more postverbal pronouns in simple sentences than in any other type of sentence or clause.
- Both regions use more postverbal pronouns in wh-questions than in yes/no questions.
- Both regions use more postverbal pronouns in non-declarative sentences than in declarative sentences.
- Both regions consistently place more pronouns before verbs than after verbs, except for wh-questions and the pronoun *ustedes*, where Mainlanders tend to place as many pronouns after verbs as they do before verbs.
- Mainlanders consistently place more pronouns after verbs than Caribbeans in all variables.
- Caribbeans do place a significant amount of pronouns after verbs in wh-phrases showing that word order is not rigid in the region, and that even in this region there is considerable variation in pronoun placement.
- Verbal transitivity triggers placement variability in the Mainland but not in the Caribbean.

In the same way that was done above for the whole of the subset corpus, I investigate next which of these variables better predict the appearance of a postverbal subject pronoun in the Caribbean and in the Mainland by means of separate logistic regression analysis.

Table 7.22a

LOGISTIC REGRESSION

Dependent: Placement

CARIBBEAN

N = 358

R2 = .38

Rank	Variable	Wald	Sig
1st	Pronoun	34.59	**
2nd	Interrogative	12.47	**
3rd	Clause	9.63	**
4th	Phrase	0	

Table 7.22b

LOGISTIC REGRESSION

Dependent: Placement

MAINLAND

N = 646

R2 = .21

Rank	Variable	Wald	Sig
1st	Pronoun	16.29	**
2nd	Gender	16.06	**
3rd	Clause	15.3	**
4th	Phrase	7.93	**
5th	Transitivity	4.58	A

In the regression model on table 7.22a, 358 verb tokens were analyzed. The amount of variance between preverbal and postverbal subject pronouns accounted for by this model is 38 percent in Caribbean speakers. According to the Wald value assigned to each variable, the variable Pronoun is the biggest predictor of postverbal subject pronouns in the Caribbean. Pronoun is followed by Interrogative, and Clause. However, Phrase has lost statistical significance in the regression model.

As it was mentioned earlier in this chapter, the regression analysis also gives us information about the order of the constraints within the variables. As before, I mention below only the constraints that predict a postverbal pronoun and I include their Exp(B) value in brackets next to each predicting constraint. A value larger than 1 indicates that the constraint predicts a postverbal pronouns. Constraints with values smaller than 1 (discouraging the

appearance of a postverbal pronoun) are not included. In the case of the variable Pronoun, the pronouns that favor postverbal pronouns in the Caribbean are *ustedes* (2.89), *uno* (2.55) and *usted* (1.104). In the case of Interrogatives, wh-questions (2.24) favor the appearance of postverbal pronouns, while yes/no questions do not. In the case of the variable Clause, main clause in a simple sentences (1.38) and coordinate clause in a complex sentence (1.22) both favor the appearance of a postverbal pronoun. In the case of the variable Phrase, non-declarative sentences (1.24) favor the appearance of a postverbal pronoun while declarative sentences do not.

In the case of Mainland speakers, the first regression I run contained the six variables that displayed significant results in the cross-tabulations. However, because one of the variables (wh-questions) had only 41 tokens, the regression only took into consideration this number of tokens for all variables and it was not enough to obtain a variable hierarchy. Interestingly, that regression gave an R square of 0.91, meaning that in those 41 cases, the six variables explained 91 percent of variance. The results presented in table 7.22b above represent the regression without the variable wh-question, which allowed for the inclusion of 646 verb tokens, but a much lower R square of 0.21. The analysis tells us that in the Mainland, the variable Pronoun is the most important predictor of postverbal subject pronouns. Pronoun is followed by Gender, Clause, Phrase and Transitivity.

Regarding the order of the constraints within the variables, in the case of the variable Pronoun, the pronouns that favor postverbal placement in the Mainland are *ustedes* (2.32), *usted* (1.83), *uno* (1.75), *tú* (1.22), *yo* (1.10). As for the variable Gender, it is actually the pronouns which are unmarked for Gender the ones that favor postverbal pronoun placement, whereas the feminine and masculine pronouns do not. In the case of the variable Clause, the clauses that favor postverbal pronouns are the main clause in a simple sentence (1.55), the coordinate clause

in a complex sentence (1.31), and the subordinate clause (1.07). Regarding the variable Transitivity, the regression informs that intransitive verbs (791.28) predict postverbal pronouns more than transitive verbs (494.06), and that ditransitive verbs disfavor them. For the variable Phrase, non-declarative sentences (1.64) favor the appearance of a postverbal pronoun while declarative sentences do not.

Table 7.23
Summary of variables and factors favoring postverbal placement by region

	VARIABLES	FACTORS
Caribbean		
1	Pronoun	(1) *ustedes*, (2) *uno*, (3) *usted*
2	Interrogative	(1) wh-question
3	Clause	(1) simple sent.
4	Phrase	(1) non-declarative
Mainland		
1	Pronoun	(1) *ustedes*, (2) *usted*, (3) *uno*, (4) *tú*, (5) *yo*
2	Gender	(1) unmarked
3	Clause	(1) simple sent, (2) coordinate cl., (3) subordinate cl.
4	Phrase	(1) non-declarative
5	Transitivity	(1) intransitive, (2) transitive

In summary, there are more variables and constraints affecting pronoun placement in the Mainland than in the Caribbean. However, the statistical analysis has revealed that there are still four ranked and distinct variables and six ranked and distinct constraints in the Caribbean that predict a postverbal subject pronoun. These results provide a strong contention to the claims of rigidity in word order in the Caribbean. It also shows that there are more similarities than differences between the Caribbean and the Mainland when it comes to pronoun placement: all variables and constraints that favor postverbal pronouns in the Caribbean, also favor postverbal pronouns in the Mainland, usually in the same order.

3.4 Variable hierarchies across the exposure groups

It was established in previous chapters that Latinos in New York increase their use of preverbal pronouns in the second generation. However, I wanted to find out what else differentiates the immigrant newcomers, the established immigrants, and the NYR. Therefore, once again I used cross-tabulations and variable hierarchies to explore potential differences between the three groups. I wanted to know whether, irrespective of preverbal rates, the order in which variables impact the placement of pronouns is the same or different among the three exposure groups, and whether, based on the results of the multiple regressions, I can make conclusions regarding continuity with Latin American ways of speaking, or change due to the influence of English, as it was done in the previous chapter.

The cross-tabulations showed that for the Newcomer group the variables Clause, Interrogative and Phrase were significant, and the variable Subordinate was almost significant. For the Immigrant group, the variables Clause and Interrogative were significant as well, and also Pronoun. Transitivity was almost significant for this group. However, none of these variables were significant for the NYR group.

In order to investigate variable hierarchies in a comparative way among the three groups, I decided to run regressions with the five variables that were significant for the whole subset corpus (i.e. Person, Interrogative, Transitivity, Clause, Phrase). However, there were not enough interrogative sentences produced by the NYR and therefore, the regression did not run. Thus, I present below the results for the Immigrant Newcomers and Established Immigrants only.

	Table 7.24a				Table 7.24b		
	LOGISTIC REGRESSION				LOGISTIC REGRESSION		
	Dependent: Placement				Dependent: Placement		
	N = 95				N = 105		
	R2 = .60				R2 = .56		
	IMMIGRANT NEWCOMERS				ESTABLISHED IMMIGRANTS		
Rank	Variable	Wald	Sig		Variable	Wald	Sig
1st	Pronoun	12.06	**		Pronoun	5.06	
2nd	Interrogative	3.05	*		Interrogative	2.97	a
3rd	Transitivity	1.70			Clause	.95	
4th	Clause	.03			Transitivity	.21	
5th	Phrase						

CONTINUITY AND CHANGE. The results above reveal that there is some level of continuity with Latin American ways of speaking in the first generation, given that the first two variables occupy the same position in both groups, although they are only significant for newcomers. The ranking of the rest of the variables differs. The decrease in significance of the first two variables and amount of variance explained by all variables, as well as the different variable hierarchy in the established immigrant group suggest that the change starts, albeit slightly, in the first generation. In any case, these five internal variables explain 60 percent and 56 percent of variance in the groups, respectively.

In order to be able to compare the three groups, I decided to remove the variable Interrogative. The results obtained appear in Table 7.25a-c below.

	Table 7.25a			Table 7.25b			Table 7.25c		
	LOGISTIC REGRESSION			LOGISTIC REGRESSION			LOGISTIC REGRESSION		
	Dependent: Placement			Dependent: Placement			Dependent: Placement		
	N = 492			N = 709			N = 207		
	R2 = .19			R2 = .13			R2 = .24		
	IMMIGRANT NEWCOMERS			ESTABLISHED IMMIGRANTS			NYR		
Rank	Variable	Wald	Sig	Variable	Wald	Sig	Variable	Wald	Sig
1st	Pronoun	13.09	*	Pronoun	29.72	**	Clause	1.30	
2nd	Phrase	12.59	**	Clause	14.68	**	Transitivity	1.08	
3rd	Clause	5.30		Transitivity	.19		Pronoun	.86	
4th	Transitivity	.53		Phrase	.16		Phrase	0	

The removal of the variable interrogative allowed for more tokens to be included in the analysis. However, R2 has gone down. This represents additional evidence of how much variance the variable Interrogative actually explains. The removal of Interrogative has moved up the variable Phrase in the ranking for the Immigrant Newcomers. And now there are more differences between the two first generation groups and I can also make observations with regards to the three groups. First, while two variables are significant in the Immigrant Newcomer and Established Immigrants groups, none of the variables are significant in the NYR group. Also, Pronoun remains important in the first generation but loses significance and ranking placement in the NYR. Furthermore, if we look at the position of the variables Clause and Transitivity together, we see that while they occupied 3rd and 4th place in the Immigrant Newcomer group, they move up to 2nd and 3rd place in the Established Immigrant group, and up to 1st and 2nd place in the NYR group. This evidences once again that some changes start in the first generation, with more exposure to the City, and continue in the second generation.

What we see is that when Immigrant Newcomers place pronouns before or after verbs, the most important consideration is the type of subject pronoun used, followed by whether it's a declarative or non-declarative sentence, followed by verb transitivity, followed by the type of clause the verb and pronoun appear in. When Established Immigrants place pronouns before or after verbs, they are guided first by the type of subject pronoun that they are using, followed by the type of clause the verb and pronoun appear in, followed by whether the verb is transitive, intransitive or ditransitive, followed by whether it's a declarative or non-declarative sentence. And when NYRs place pronouns before or after verbs, the most important consideration, if any given the lack of significance, is the type of clause the verb and pronoun appear in, followed by whether the verb is transitive, intransitive or ditransitive, followed by the type of subject pronoun used, followed by whether it's a declarative or non-declarative sentence. This means that while for newcomers whether the phrase is declarative or interrogative is so important in the decision to place subject pronouns before or after verbs, the type of phrase has lost total importance for the NYR group.

Regarding the issue of continuity and change, it is very clear in this small sample that changes begin in the first generation, with the established immigrants. First, they begin to use more subject pronouns than newcomers (Otheguy & Zentella, 2012). Then, they lower their preverbal rate by placing more subject pronouns after the verb than newcomers do (as established and discussed in chapter 5). When internal variables are studied, further evidence of change is revealed given that the variables' significance decreases in the established immigrant group, and the hierarchy of variables changes as well. The change continues in the second generation that has further increased the use of subject pronouns (Otheguy & Zentella, 2012), reduced the number of pronouns placed postverbally (from 14 percent in the first generation to 4

percent in the second generation), and the internal variables that guided first-generation speakers in pronoun placement do not seem to be playing much of a role for them.

4. Conclusion

One of the main goals of creating the subset corpus was to analyze the impact that non-finite verbs and interrogative phrases had on the placement of pronominal subjects. With regards to non-finite verbs, the results may be inconclusive given the small size of the sample (21 verb tokens, 1 percent of the total verb tokens analyzed). However, regarding interrogative phrases I have been able to arrive at very interesting conclusions that shed light on the differences between Caribbean and Mainland Spanish. On the one hand, I found that 21 percent of wh-questions uttered by speakers from the Caribbean had a postverbal subject pronoun. This is significant because the literature on subject placement in the Caribbean claimed that word order tends to be a rigid SVO. The facts of this corpus show that although SVO is the most frequent order in wh-questions (79 percent of wh-questions have a preverbal pronouns), Caribbeans have not lost the ability to place pronouns postverbally. On the other hand, I have also found that while Mainlanders place more subject pronouns postverbally in wh-questions, they still placed 46 percent of their pronominal subjects before the verb. This is relevant because it has been claimed in the literature that while Caribbeans placed subject pronouns preverbally in wh-questions, speakers of other varieties of Spanish either had a postverbal pronoun or a null subject. Therefore, both regions show variation in this type of phrases.

Regarding the issues of continuity and change in the exposure groups, my conclusions are limited by the size of the sample. However, I was able to find evidence for the assertion that change begins in the first generation with small adjustments in variable ranking and significant variables. Then it continues in the second generation, with changes in variable ranking and lack

of significant variables. That is, the internal variables that influenced the speakers' choice in subject pronoun placement begin to change with more time spent in NYC and more exposure to English, and lose total significance in the second generation.

In sum, these results are aligned with results obtained for the external variables in previous chapters: the real change occurs in the second generation that according to the findings in the investigations in this dissertation, are less and less motivated by either internal or external factors to place subject pronouns postverbally, and are further influenced by English given that they have been raised speaking both languages with English being the dominant language of their environment.

[1] Eight interviewers participated in the 22 interviews that make up the subset corpus. Therefore, the speech of 30 people has been coded in this portion of the investigation. The interviewers were not included in Table 7.1 because it was not possible to reach all of them to determine their exposure group. Their identity of nationality with the corresponding informant has been ascertained.

APPENDIX

PERL PROGRAMS

```
#This program opens the merged transcripts, normalizes case,
#and explodes punctuation.
#Then it counts unigrams and outputs them to a file.
open(INFILE, "<C:/scripts/transcripts.txt") or die "Cannot open file!\n";
while ($text = <INFILE>){
chomp($text);
$text .= "$_ ";
$text =~ tr/A-Z/a-z/;                                  # converts everything to lower case
$text =~ s/(\S)([^ a-z0-9áéíóúÁÉÍÓÚ])([\s])/\1 \2\3/g;  # explode punc at end of word
$text =~ s/(\S)([^ a-z0-9áéíóúÁÉÍÓÚ])$/\1 \2/g;         # explode punc at end of sentence
$text =~ s/(\S)([^ a-z0-9áéíóúÁÉÍÓÚ])([\s])/\1 \2\3/g;  # explode punc at end of word, again
$text =~ s/([\s])([^ a-z0-9áéíóúÁÉÍÓÚ])(\S)/\1\2 \3/g;  # explode punc at beginning of word
$text =~ s/^([^ a-z0-9áéíóúÁÉÍÓÚ])(\S)/\1 \2/g;         # explode punc at beginning of word
$text =~ s/([a-z0-9áéíóúÁÉÍÓÚ])(\[)/\1 \2/g;            # separate numbers/letters from [
$text =~ s/(\[)([áéíóúÁÉÍÓÚa-z0-9\-])/\1 \2/g;          # separate [ from letters/numbers
$text =~ s/(\.+)([a-záéíóúÁÉÍÓÚ])/\1 \2/g;             # separate . from letters
$text =~ s/([a-záéíóúÁÉÍÓÚ])(\.+)/\1 \2/g;             # separate . from letters
$text =~ s/(\¿)([a-záéíóúÁÉÍÓÚ])/\1 \2/g;              # separate ¿ from letters
$text =~ s/([\s]+)/ /g;                                # substitute string of whitespaces by
                                                       # a whitespace
```

```
@word_list = split(" ",$text);
# Unigram count
for ($j = 0; $j <= $#word_list; $j++){
$unifreq{$word_list[$j]}++;
$n++;
}
}
close(INFILE);
open(OUTFILE,">trans.unifreq") or die "Cannot open trans.unifreq: $!\n";
foreach $word (sort { $unifreq{$b} <=> $unifreq{$a} } keys %unifreq){
print OUTFILE "$unifreq{$word} $word\n";
}
close(OUTFILE);
# The list of unigrams extracted is filtered with a list of possible infinitival endings
# and output infinitos.txt
my ($line);
open (infile, "<C:/scripts/trans.unifreq")|| die("Could not open file!");
open (outfile, ">C:/scripts/infinitivos.txt")|| die("Could not open file!");
while ($line = <infile>){
        chomp($line);
        # applies inf filter
        if ($line =~
/.+\s.+ar$|er$|ir$|arse$|arselo$|arsela$|arsele$|arseles$|arselos$|arselas$|arme$|armelo$|armele$|ar
```

```
mela$|armelos$|armelas$|armeles$|arte$|artelo$|artela$|artele$|artelos$|artelas$|arteles$|arle$|arle
s$|arlo$|arla$|arnos$|arnoslo$|arnolo$|arnosla$|arnola$|arnosle$|arnole$|arnoslos$|arnolos$|arnos
las$|arnolas$|arnosles$|arnoles$|erse$|erselo$|ersela$|ersele$|erseles$|erselos$|erselas$|erme$|erm
elo$|ermele$|ermela$|ermelos$|ermelas$|ermeles$|erte$|ertelo$|ertela$|ertele$|ertelos$|ertelas$|ert
eles$|erle$|erles$|erlo$|erla$|ernos$|ernoslo$|ernolo$|ernosla$|ernola$|ernosle$|ernole$|ernoslos$|
ernolos$|ernoslas$|ernolas$|ernosles$|ernoles$|irse$|irselo$|irsela$|irsele$|irseles$|irselos$|irselas
$|irme$|irmelo$|irmele$|irmela$|irmelos$|irmelas$|irmeles$|irte$|irtelo$|irtela$|irtele$|irtelos$|irte
las$|irteles$|irle$|irles$|irlo$|irla$|irnos$|irnoslo$|irnolo$|irnosla$|irnola$|irnosle$|irnole$|irnoslos
$|irnolos$|irnoslas$|irnolas$|irnosles$|irnoles$/){
        print outfile "$line\n";
        }
}
close(outfile);
close (infile);

#This program opens the stoplist, stores it in a hash, and filters the list of infinitives
#to eliminate nouns with similar endings.
open (infile, "<C:/scripts/stoplist_inf.txt")|| die("Could not open file!");
while ($line = <infile>)
{
        chomp ($line);
        $stop{$line}++;
}
```

```
close (infile);
open (infile, "<C:/scripts/infinitivos.txt")|| die("Could not open file!");
open (outfile, ">C:/scripts/infinitivos_filtrados.txt")|| die("Could not open file!");
while ($line = <infile>)
{
	chomp ($line);
	if ($line =~ /(.*) (.*)/)
	{
		$freq = $1;
		$word1 = $2;
		if ($stop{$word1})
		{
		}
		else
		{
			$infinitivo = $word1;
			$stops{$infinitivo}="$freq $infinitivo";
		}
	}
}
foreach $infinitivo (sort { $stops{$b} <=> $stops{$a} } keys %stops)
{
	print outfile "$stops{$infinitivo}\n";
```

```
}
close(outfile);
close (infile);

#This program looks for the word that precedes or follows an infinitive
#from the list of infinitivos filtrados
open (infile, "<C:/scripts/infinitivos_filtrados.txt")|| die("Could not open file!");
while ($line = <infile>)
{
	chomp ($line);
	if ($line =~ /(.*) (.*)/)
	{
		$freq = $1;
		$word1 = $2;
		$inf =$word1;
		$lastinf{$inf}++;
	}
}
close (infile);
open(infile, "<C:/scripts/transcripts.txt") or die "Cannot open file!\n";
open (outfile, ">C:/scripts/inf_menos1.txt")|| die("Could not open file!");
open (outfile1, ">C:/scripts/inf_mas1.txt")|| die("Could not open file!");
while ($text = <infile>){
```

```
chomp($text);


$text .= "$_ ";
$text =~ tr/A-Z/a-z/;                                   # converts everything to lower case
$text =~ s/(\S)([^ a-z0-9áéíóúÁÉÍÓÚ])([\s])/\1 \2\3/g;  # explode punc at end of word
$text =~ s/(\S)([^ a-z0-9áéíóúÁÉÍÓÚ])$/\1 \2/g;         # explode punc at end of sentence
$text =~ s/(\S)([^ a-z0-9áéíóúÁÉÍÓÚ])([\s])/\1 \2\3/g;  # explode punc at end of word, again
$text =~ s/([\s])([^ a-z0-9áéíóúÁÉÍÓÚ])(\S)/\1\2 \3/g;  # explode punc at beginning of word
$text =~ s/^([^ a-z0-9áéíóúÁÉÍÓÚ])(\S)/\1 \2/g;         # explode punc at beginning of word
$text =~ s/([a-z0-9áéíóúÁÉÍÓÚ])(\[)/\1 \2/g;            # separate numbers/letters from [
$text =~ s/(\[)([áéíóúÁÉÍÓÚa-z0-9\-])/\1 \2/g;          # separate [ from letters/numbers
$text =~ s/(\.+)([a-záéíóúÁÉÍÓÚ])/\1 \2/g;              # separate . from letters
$text =~ s/([a-záéíóúÁÉÍÓÚ])(\.+)/\1 \2/g;              # separate . from letters
$text =~ s/(\¿)([a-záéíóúÁÉÍÓÚ])/\1 \2/g;               # separate ¿ from letters
$text =~ s/([\s]+)/ /g;                                 # substitute string of whitespaces by
                                                        # a whitespace

@word_list = split(" ",$text);
for ($j = 0; $j <= $#word_list; $j++){
if ($j > 0 && $lastinf{$word_list[$j]}){
$bigram = "$word_list[$j-1] $word_list[$j]";
$bifreq{$bigram}++;
$bigram2 = "$word_list[$j] $word_list[$j+1]";
$bifreq2{$bigram2}++;
```

```
}
}
}
foreach $bigram (sort { $bifreq{$b} <=> $bifreq{$a} } keys %bifreq)
{
	print outfile "$bifreq{$bigram} $bigram\n";
}
close(outfile);
foreach $bigram2 (sort { $bifreq2{$b} <=> $bifreq2{$a} } keys %bifreq2)
{
	print outfile1 "$bifreq2{$bigram2} $bigram2\n";
}
close(outfile1);
close (infile);

#This program checks whether the word that follows an infinitive in the corpus
#is an overt pronoun
open (infile, "<C:/scripts/pronombres.txt")|| die("Could not open file!");
while ($line = <infile>)
{
	chomp ($line);
	$pronombre{$line}++;
}
```

```
close (infile);
open (infile, "<C:/scripts/inf_mas1.txt")|| die("Could not open file!");
open (outfile, ">C:/scripts/inf_con_pron1.txt")|| die("Could not open file!");
while ($line = <infile>)
{
	chomp ($line);
	if ($line =~ /(.*) (.*) (.*)/)
	{
		$bifreq = $1;
		$word1 = $2;
		$word2 = $3;
		if ($pronombre{$word2})
		{
			$bigram = $word1." ".$word2;
			$pronombres{$bigram}="$bifreq $bigram";
		}
	}
}
foreach $bigram (sort { $pronombres{$b} <=> $pronombres{$a} } keys %pronombres)
{
	print outfile "$pronombres{$bigram}\n";
}
close(outfile);
```

```
close (infile);


#This program checks whether the word that precedes an infinitive in the corpus
#is an overt pronoun
open (infile, "<C:/scripts/pronombres.txt")|| die("Could not open file!");
while ($line = <infile>)
{
	chomp ($line);
	$pronombre{$line}++;
}
close (infile);
open (infile, "<C:/scripts/inf_menos1.txt")|| die("Could not open file!");
open (outfile, ">C:/scripts/inf_con_pron_menos1.txt")|| die("Could not open file!");
while ($line = <infile>)
{
	chomp ($line);
	if ($line =~ /(.*) (.*) (.*)/)
	{
		$bifreq = $1;
		$word1 = $2;
		$word2 = $3;
		if ($pronombre{$word1})
		{
```

```
                $bigram = $word1." ".$word2;
                $pronombres{$bigram}="$bifreq $bigram";
        }
    }
}
foreach $bigram (sort { $pronombres{$b} <=> $pronombres{$a} } keys %pronombres)
{
    print outfile "$pronombres{$bigram}\n";
}
close(outfile);
close (infile);

#This program outputs the sentence where the infinitive appears to be followed by a pronoun
#to manually check if it's really a pronoun or an article
open (infile, "<C:/scripts/inf_con_pron1.txt")|| die("Could not open file!");
while ($line = <infile>)
{
    chomp ($line);
    if ($line =~ /(.*) (.*) (.*)/)
    {
        $freq = $1;
        $word1 = $2;
        $word2 = $3;
```

```
                $inf =$word1." ".$word2;
                $bigram{$inf}++;
        }
}
close (infile);
open(infile, "<C:/scripts/transcripts.txt") or die "Cannot open file!\n";
open (outfile, ">C:/scripts/inf_mas1_oracion.txt")|| die("Could not open file!");
while ($text = <infile>){
chomp($text);
$text .= "$_ ";
$text =~ tr/A-Z/a-z/;                          # converts everything to lower case
$text =~ s/(\S)([^ a-z0-9áéíóúÁÉÍÓÚ])([\s])/\1 \2\3/g;   # explode punc at end of word
$text =~ s/(\S)([^ a-z0-9áéíóúÁÉÍÓÚ])$/\1 \2/g;          # explode punc at end of sentence
$text =~ s/(\S)([^ a-z0-9áéíóúÁÉÍÓÚ])([\s])/\1 \2\3/g;   # explode punc at end of word, again
$text =~ s/([\s])([^ a-z0-9áéíóúÁÉÍÓÚ])(\S)/\1\2 \3/g;   # explode punc at beginning of word
$text =~ s/^([^ a-z0-9áéíóúÁÉÍÓÚ])(\S)/\1 \2/g;          # explode punc at beginning of word
$text =~ s/([a-z0-9áéíóúÁÉÍÓÚ])(\[)/\1 \2/g;             # separate numbers/letters from [
$text =~ s/(\[)([áéíóúÁÉÍÓÚa-z0-9\-])/\1 \2/g;           # separate [ from letters/numbers
$text =~ s/(\.+)([a-záéíóúÁÉÍÓÚ])/\1 \2/g;              # separate . from letters
$text =~ s/([a-záéíóúÁÉÍÓÚ])(\.+)/\1 \2/g;              # separate . from letters
$text =~ s/(\¿)([a-záéíóúÁÉÍÓÚ])/\1 \2/g;               # separate ¿ from letters
$text =~ s/([\s]+)/ /g;                        # substitute string of whitespaces by
                                               # a whitespace
```

```
@word_list = split(" ",$text);
for ($j = 0; $j <= $#word_list; $j++){
if ($j > 0 && $bigram{"$word_list[$j] $word_list[$j+1]"})
{
        print outfile "$text\n";
}
}
}
close(outfile);
close (infile);
```

REFERENCES

Butt, John, Carmen Benjamin. 1996. *A new reference grammar of modern Spanish*. London: Edward Arnold / Lincolnwood, Ill.: NTC. Second Edition.

Chomsky, Noam. 1981. *Lectures on Government and Binding*. Dordrecht: Foris Publications

Cuza, Alejandro. 2013. Crosslinguistic influence at the syntax proper: Interrogative subject-verb inversion in heritage Spanish. *International Journal of Bilingualism*. 2013, No. 17, 71-96. Sage.

De Bruyne, Jacques. 1995. *A Comprehensive Spanish Grammar*. Massachusetts: Basil Blackwell.

Gili y Gaya, Samuel. 1951. *Curso superior de sintaxis española*. Barcelona: Ediciones SPES.

Goodall, Grant. 2004. On the Syntax and Processing of Wh-questions in Spanish. *Proceedings of the 23rd West Coast Conference on Formal Linguistics,* Benjamin Schmeiser, Vineeta Chand, Ann Kelleher and Angelo Rodriguez (eds.). Somerville, MA: Cascadilla Press.

Jaeggli, Osvaldo, and Kenneth J. Safir (eds.) 1989. *The null subject parameter.* Dordrecht: Kluwer.

Labov, William. 1963 [1972]. The social motivation of a sound change. *Word* 19.273-309. [Reprinted as *Sociolinguistic patterns*. University of Pennsylvania Press, 1972.]

Lipski, John M. 1994. *Latin American Spanish.* London: Longman Publishers.

López-Morales, Humberto. 1992. *El español del Caribe.* Madrid: Editorial MAPFRE.

Menard, Scott. 2002. *Applied logistic regression analysis*. 2nd edn. Sage Publications.

Montrul, Silvina. 2004. Subject and object expression in Spanish heritage speakers: a case of morphosyntactic convergence. *Bilingualism: Language and Cognition* 7.125-42.

Montrul, Silvina. 2008. *Incomplete acquisition in bilingualism: Re-examining the age factor.* Amsterdam / Philadelphia: John Benjamins Publishing Co.

Morales, Amparo. 1988. Hacia un universal sintáctico del español del Caribe: El orden SVO. *Anuario de lingüística hispánica* 5.139-152.

Morales, Amparo. 1989. Infinitivo con sujeto expreso en el español de Puerto Rico. *Studies in Caribbean Spanish Dialectology*, ed. Robert M. Hammond and Melvyn C. Resnick, Washington D.C.: Georgetown University Press, 85-96

Morales, Amparo. 1997. La hipótesis funcional y la aparición del sujeto no nominal: el español de Puerto Rico. *Hispania* 80.153-65.

Morales, Amparo. 1999. Anteposición del sujeto en el español del Caribe. *El Caribe hispánico: Perspectivas lingüísticas actuales. Homenaje a Manuel Alvarez Nazario*. Madrid: Iberoamericana.

Navarro-Tomás, Tomás. 1948 [1974]. *El español de Puerto Rico*. San Juan: Editorial Universitaria.

Newton, Rae R., and Kjell Eric Rudestam. 1999. *Your statistical consultant: Answers to your data analysis questions.* Sage Publications.

Ordóñez, Francisco and Antxon Olarrea. 2001. Weak Subject Pronouns in Caribbean Spanish and XP Pied-Piping. In Julia Herschensohn, Enrique Mallen,& Karen Zagona (eds.), *Features and Interfaces in Romance: Essays in Honor of Heles Contreras*., Amsterdam: John Benjamins, pp 223-238

Ortiz López, Luis A. 2009. Pronombres del sujeto en el español (L2 vs. L1) del Caribe. *Español en Estados Unidos y otros contextos de contacto*, ed. Manel Lacorte and Jennifer Leeman. Iberoamericana: Verbuert. 85-110.

Otheguy, Ricardo, and Ana Celia Zentella. 2007. Apuntes preliminares sobre el contacto lingüístico y dialectal en el uso pronominal del español en Nueva York. *Spanish in contact: educational, social and linguistic inquiries*, ed. by Richard Cameron and Kim Potowski, 275-96. Amsterdam: John Benjamins.

Otheguy, Ricardo; Ana Celia Zentella; and David Livert. 2007. Language and dialect contact in Spanish in New York: Toward the formation of a speech community. *Language* 83.770-802.

Otheguy, Ricardo; Ana Celia Zentella; and David Livert. 2008. Factores gramaticales y sociodemográficos en la evolución de los pronombres sujetos entre los hispanohablantes de Nueva York. *Actas del XV Congreso de la Asociación de Lingüística y Filología de América Latina.* Universidad de la República, Monteviedo, Uruguay.

Otheguy, Ricardo; and Ana Celia Zentella. 2012. *Spanish in New York.* Oxford University Press.

Poplack, Shana (ed.) 2000. *The English history of African American English*. Oxford: Blackwell Publishers.

Poplack, Shana, and David Sankoff. 1987. The Philadelphia story in the Spanish Caribbean. *American Speech* 62.291-314.

Poplack, Shana, and Sali Tagliamonte. 1999. The grammaticalization of *going to* in (African American) English. *Language Variation and Change* 11.315-42

Raña Risso, Rocío. Subject Pronoun Placement as Evidence of Contact and Leveling in Spanish in New York. *International Journal of the Sociology of Language.* May-June 2010, No. 203. 101-114. De Gruyter.

Rizzi, Luigi. 1982. *Issues in Italian Syntax*. Dordrecht: Foris Publications.

Tagliamonte, Sali. 2002. Comparative sociolinguistics. *The handbook of language variation and change,* ed. by J. K. Chambers, Peter Trudgill, and Natalie Schilling-Estes, 729-63. Oxford: Blackwell.

Tagliamonte, Sali. 2006. *Analysing sociolinguistic variation.* Cambridge University Press.

Toribio, Almeida Jacqueline. 2000. Setting parametric limits on dialectal variation in Spanish. *Lingua* 10.315-41.

Zagona, Karen. 2002. *The syntax of Spanish*. Cambridge University Press.

Zubizarreta, María Luisa. 1998. *Prosody, Focus, and Word Order*. Cambridge, Massachusetts: The MIT Press.

Zubizarreta, María Luisa. 2001. “The constraint on preverbal subjects in Romance interrogatives: A minimality effect.” In Aafke Hulk and Jean-Yves Pollock (eds.), *Subject Inversion in Romance and the Theory of Universal Grammar*. Oxford: Oxford University Press.