Mood Choice and Context Availability: A Variationist Approach to the Subjunctive in New York City Spanish

Joanna Birnbaum

The Graduate Center, City University of New York

Recommended Citation
https://academicworks.cuny.edu/gc_etds/3489
Mood choice and context availability:
A variationist approach to the Subjunctive in New York City Spanish

by

Joanna Birnbaum

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

2019
Mood choice and context availability:
A variationist approach to the Subjunctive
in New York City Spanish

by

Joanna Birnbaum

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.

Date                Ricardo Otheguy
                    Chair of Examining Committee

Date                Gita Martohardjono
                    Executive Officer

Supervisory Committee

Ricardo Otheguy
Gita Martohardjono
Beatriz Lado

THE CITY UNIVERSITY OF NEW YORK
Abstract

Mood choice and context availability:
A variationist approach to the Subjunctive in New York City Spanish

by

Joanna Birnbaum

Adviser: Professor Ricardo Otheguy

This dissertation investigates the variable treatment of the Subjunctive in Spanish in New York City. Both Mood choice (Subjunctive versus Indicative) and Linguistic context availability (the presence and absence of Subjunctive-inducing contexts in speech) are studied. Data are from sociolinguistic interviews with 142 informants, stratified with respect to immigrant generation, gender, age, socio-economic status, national origin, etc. Subjunctive rates are analyzed, at the macro-level, in nine linguistic contexts and, at the micro-level, in the four most popular contexts (Modal, Protasis Si, Temporal, and Apodosis Si). Results of bivariate Pearson correlations and Chi-square tests reveal consistent usage patterns of the Subjunctive in all informants; speakers who show a relative preference for the Subjunctive in one context are also Subjunctive-inclined in another context, as well as overall. Different speakers have different preferences for particular linguistic contexts. Results of independent samples t-tests demonstrate that the presence or absence of particular contexts shapes speakers’ Subjunctive rates in two of the most popular contexts, suggesting that Subjunctive rates and Linguistic context availability should be analyzed together. Regression analyses show that women, the young, and the New Yorkers (long-term residents or New York raised) have lower Subjunctive rates than men, the Newcomers
and older consultants. These three groups’ greater usage of the Indicative is linked to the composition of their social networks, which largely comprise non-Spanish speakers. These denser contacts with the out-group are held to intensify their social aspirations. Similarly, the messages that some of these same groups convey might also hinge on their standings in the New York City social order. It is posited that group differences may be indicative of a language change in progress.

This study’s methodology and findings contribute to the field of variationist sociolinguistics and have implications for language pedagogy. An exploration of Linguistic context availability represents a new, critical component in an analysis of Subjunctive rates. It was shown that the availability of particular contexts functions as a dependent variable, offering additional insight into speakers’ Subjunctive usage. It was noted that all informants, as well as specific subgroups, have consistent patterns of Subjunctive usage, showing that variation is anything but arbitrary. It was discovered that generation is not the only predictor of Subjunctive usage; other variables (e.g., gender and age) are just as useful. Certain linguistic contexts tend to be more available to New York Raised participants; both of these findings taken together suggest that the grammar of New York Raised speakers is not defective.
To my grandparents,

whose chutzpah, as refugees,
always made everything seem possible,
and whose ordinary multilingualism

I hope to pass on

to my extraordinary son Gabriel
Acknowledgments

My work on New York City Spanish would not have been possible without the generosity of Ricardo Otheguy and Ana Celia Zentella, who have shared their substantial corpus with their graduate students. I was fortunate to be among them. I would also like to thank the informants for their very personal, and often eye-opening, stories of immigration, past and present. Many of them remind me of my College of Staten Island students, whose experiences as Spanish/English bilinguals, and whose fieldwork observations and instruction in Spanish language classrooms, have helped inform my understanding of Spanish in New York City.

A sincere thank you to my committee members, Ricardo Otheguy, Beatriz Lado, and Gita Martohardjono, who have each taught a course or courses that have shaped my views on language while in graduate school. From Gita, I learned about, and learned to question, theories of second language acquisition (SLA). In Gita’s course, I was given the chance to conduct my own SLA experiment, which was a very worthwhile experience. Gita also taught me that even established linguists can be ambivalent, and that there is value in changing one’s mind. I deeply admire her courage and humility. From Beatriz, I learned the importance of bridging the theory of SLA with the practice of teaching, which is something that I attempt to emulate in my own courses. Beatriz’s passion for teaching is contagious, and, through her, I have observed the importance of treating graduate and undergraduate students equally well. Beatriz’s recent scholarly work in a different subfield is a great example of how exciting new inquiries can be. I can only hope to be as audacious as these two female professors and scholars someday.
From Ricardo, I have learned so many things. I have become a better thinker, thanks to his challenging and judicious questions. His rigorous editing has turned me into a better writer. I have become a better linguist, through his Sociolinguistics courses, and through our long meetings together. In fact, it was Ricardo’s profound curiosity about sociolinguistic questions that attracted me to the field in the first place, and to him as an advisor. I have benefited from his commitment, kindness, encouragement, and insights since I started graduate school, and I know that they will stay with me, even as my time in graduate school comes to an end.

I extend my heartfelt appreciation to my colleagues and friends. I am so glad that I was persuaded to join in on writing group with Emily Long Olsen, Michelle McSweeney (its creator), Ignacio Montoya, Ian Phillips, and Nadav Sabar. Over the years, my writing companions’ encouragement, good judgment, and practical advice became as indispensable as being held accountable for my weekly goals. I am especially appreciative of Ignacio, my unofficial advisor for all things linguistic, and my caring friend, whose intelligence and integrity have never ceased to impress me since our very first day together in the Linguistics program.

I feel very fortunate for the support of my non-linguist friends, near and far. Many people have inspired me, from the talented colleagues whom I taught alongside in France, to my brilliant and creative friends in Madrid, to the dedicated multilingual activist friends that I made in Tel Aviv, to my more recent New York City friends, and the list goes on. I am greatly indebted to them, and to their work, which continues to energize me. Thank you to Alexa, Courtney, Jessica, and Lucy for the life-long friendships, and for being such great models of what women and mothers are capable of accomplishing. Many thanks to
Murielle and Laetitia, for years of insightful and entertaining conversations about language, teaching, relationships, and politics, among many other topics. Thank you to some of my oldest and closest friends, Maud and Alessandra, for keeping me grounded from afar, and well fed when close by. It is difficult to find the right words to thank Rachel, my most spirited cheerleader, who always found ways to motivate me and crack me up in moments of doubt. I cannot envision life in New York City (and in general) without our friendship.

Finally, I owe my family (Patrick et Laure, Judy and Ed, Jules and Ellen, Franck et Marina, Sophie et Greg, among many others) and my pseudo-family (Lurdes, Nadine, Wendy, among many others) a debt of gratitude as well. I am very lucky to call Susan Sullivan my mentor and model. Susan has offered invaluable assistance and guidance in my development as a college instructor. She is the most dedicated educator I have ever met. From her, I have learned how to support our CUNY students in and outside of the classroom. My son Gabby (Gabriel) is fortunate to have found a wonderful adopted grandmother in Susan, as his maternal grandmother passed away just a couple of years before he was born. To my late mother, I am deeply appreciative of the high expectations that she held, and the lack of pressure that she exerted. She, Beatrice, always encouraged my pursuits, even when they led to my living on another continent while she was ill. I am fortunate to have grown up with a self-confident, avant-garde mother who was passionate about her very full-time job, and led by example. My sister, Becky, bears a great resemblance to our mother. I thank Becky for her unwavering espousal of my efforts, even when the specifics went undisclosed. Thank you to her husband Gaetano as well. My father, Daniel, has also offered constant encouragement throughout my life. My love of language learning, and of the communities that speak the languages that I study, comes
from him. Both my parents, and grandparents, to whom I dedicate my dissertation, have shown me the value of adopting several languages and cultures. I cannot thank them enough for this gift.

Last but not least, my husband, Yann, has been a vital source of love, support, and patience, since the beginning of my PhD. I am so grateful for the Java algorithms, and for the fact that he entertained our son on the weekends when I worked on my dissertation. Gabriel does not realize yet how lucky he is to have such a loving and caring father. Merci.

And to those that I have not explicitly named, thank you and ¡gracias!
# Table of Contents

Abstract ........................................................................................................................................iv

Acknowledgments .................................................................................................................... vii

Chapter 1: Introduction ............................................................................................................... 1
  1. Introduction and research questions .................................................................................. 1
  2. The Spanish Subjunctive in variation ................................................................................. 11
  3. The Spanish Subjunctive in bilingual New York City: A preview .................................. 13

Chapter 2: Literature Review .................................................................................................... 16
  1. Immigrant generations and the Spanish Subjunctive ......................................................... 16
  2. Syntactic/semantic/pragmatic contexts ............................................................................. 19
  3. Obligatory and optional contexts ..................................................................................... 26
  4. The bilingual context ........................................................................................................ 29
     4.1 Language contact .......................................................................................................... 30
     4.2 Degree and type of Spanish use .................................................................................. 33
     4.3 Simplification ............................................................................................................... 33
     4.4 Incomplete acquisition ............................................................................................... 34
     4.5 Attrition ...................................................................................................................... 38
     4.6 Minimal pairs ................................................................................................................ 39
     4.7 Cognitive complexity .................................................................................................. 41
  5. Subgroups by socio-demographic feature ...................................................................... 41
     5.1 Age .............................................................................................................................. 42
     5.2 Gender and Generation ............................................................................................... 43
     5.3 Socio-economic status ............................................................................................... 49

Chapter 3: Methodology ............................................................................................................ 53
  1. Introduction ......................................................................................................................... 53
  2. Informants and interviews ................................................................................................. 53
  3. Linguistic contexts ............................................................................................................ 56
     3.1 Concept of linguistic context ...................................................................................... 56
     3.2 Reconsidering obligatory and optional linguistic contexts ..................................... 58
     3.3 Linguistic context occurrence .................................................................................... 66
  4. Envelope of variation ........................................................................................................ 67
     4.1 Mood choice ................................................................................................................ 68
     4.2 Linguistic contexts included in the analysis of Mood choice ................................... 71
  5. Dependent variables .......................................................................................................... 72
     5.1 Subjunctive rate .......................................................................................................... 72
     5.2 Linguistic context availability ..................................................................................... 74
  6. Independent variables ........................................................................................................ 77
     6.1 Simple independent variables ..................................................................................... 78
     6.2 Interaction variables ................................................................................................... 84
  7. Data collection .................................................................................................................... 87
     7.1 Java .............................................................................................................................. 87
     7.2 Manual checking .......................................................................................................... 93
     7.2.1 Word(s) between number and verb ................................................................. 94
Chapter 4: Results ................................................................. 102
1. Introduction ........................................................................ 102
2. Bivariate results ................................................................ 103
   2.1 Bivariate analysis of Subjunctive rates.......................... 104
      2.1.1 Correlations between overall Subjunctive rate and Subjunctive rates in the four most popular linguistic contexts ...................................................... 106
      2.1.2 Correlations between Subjunctive rates in the four most popular linguistic contexts .......................................................... 108
      2.1.3 Subjunctive rate by Generation, Gender and SES ...... 109
      2.1.4 Subjunctive rate by Age ............................................ 112
   2.2 Bivariate analysis of Linguistic context availability .......... 113
      2.2.1 Relationships between the extents of availability of different linguistic contexts .......................................................... 114
      2.2.2 Linguistic context availability by Generation, Gender and SES .......................................................... 116
      2.2.3 Linguistic context availability by Age ................................ 118
   2.3 Bivariate inquiry into Subjunctive rate by Linguistic context availability ......................................................... 120
   2.4 Bivariate analyses: Summary ......................................... 123
3. Multivariate results ............................................................. 125
   3.1 Multivariate analysis of Subjunctive rates ....................... 129
      3.1.1 Gender affects Subjunctive rate, overall and in the Modal context .......... 130
      3.1.2 Generation and Age shape Subjunctive rate in the Protasis Si context ..... 134
      3.1.3 No effect of socio-demographic features on Subjunctive rate in two linguistic contexts .......................................................... 136
   3.2 Multivariate analysis of Linguistic context availability .......... 139
      3.2.1 Gender affects the Availability of the Concessive context .......... 141
      3.2.2 Generation shapes the Availability of the Possibility context .......... 143
      3.2.3 Gender and Generation together influence the Availability of the Volition context 145
      3.2.4 No effect of socio-demographic features on the Availability of the Hypothetical Como si context ......................................................... 147
3.3 Multivariate analyses: Summary ......................................... 148
4. Conclusion ........................................................................ 149

Chapter 5: Discussion .............................................................. 152
1. Introduction ........................................................................ 152
2. Linguistic relationships ...................................................... 153
   2.1 Interpreting relationships between Subjunctive rates overall and in the four most popular linguistic contexts ......................................................... 153
   2.2 Interpreting relationships between Subjunctive rates in the four most popular linguistic contexts .......................................................... 154
   2.3 Interpreting relationships between the extents of availability of different linguistic contexts .............................. 157
Chapter 6: Conclusion ........................................................................................................ 184

1. Contributions and applications ............................................................................... 184
   1.1 Contributions to the study of Spanish in the U.S.: Patterns of Subjunctive usage in New York City .................................................. 184
   1.2 Applications of this study to variationist sociolinguistic research ..................... 188
   1.3 Contributions to Second Language Acquisition research ................................. 192
   1.4 Applications to language pedagogy and teacher training: Spanish Heritage Speakers as prospective Spanish teachers ................................ 194

2. Limitations of the current study .............................................................................. 199
   2.1 No SES effect: Socio-economic status could be approached differently .......... 199
   2.2 Analysis did not consider interviewer speech .................................................. 201

3. Directions for future work ...................................................................................... 201
   3.1 Additional variables: Degree of Spanish use, English skills and Spanish skills 202
   3.2 Exploring Subjunctive verb diversity ............................................................... 203
   3.3 Adding verb tense to the exploration of Subjunctive verb diversity ................. 204

Appendix A: Linguistic contexts .................................................................................. 206

1. Nine linguistic contexts used for overall Subjunctive rate, with glosses .............. 206
2. Four most popular linguistic contexts used for Subjunctive rate ......................... 207
3. Four linguistic contexts used for Linguistic context availability .......................... 207

Appendix B: Coding manual ...................................................................................... 208

1. Variables .................................................................................................................. 208
   1.1 Verb file ........................................................................................................... 208
   1.2 Informant file .................................................................................................. 208
2. Factors (informant file) ........................................................................................... 210

References .................................................................................................................. 215
List of Tables

Table 2-1 Monolingual acquisition of the Subjunctive By linguistic context and age of acquisition.................................................................36
Table 3-1 Linguistic contexts: Number and Percentage of informants where each appears, from most to least.........................................................67
Table 3-2 Mood distribution in all (9) linguistic contexts All informants .....................73
Table 3-3 Subjunctive rate in the most popular linguistic contexts ..................................................74
Table 3-4 Linguistics contexts: Number and Percentage of informants where each context appears and does not appear from most to least..............................77
Table 3-5 Consultants, by immigrant generation ..........................................................79
Table 3-6 Consultants, by gender .............................................................................82
Table 3-7 Consultants, by socio-economic status (SES) .................................................83
Table 3-8 Consultants, by generation and gender .........................................................86
Table 4-1 Correlations between overall Subjunctive rate and Subjunctive rates in four most popular contexts...............................................................107
Table 4-2 Correlations between Subjunctive rates in four most popular contexts............108
Table 4-3 Subjunctive rate (overall and in four most popular contexts) By Generation, Gender and SES..................................................................................110
Table 4-4 Subjunctive rate (overall and in four most popular contexts) By Age .............112
Table 4-5 Associations between extents of Linguistic context availability (Part 1).........115
Table 4-6 Associations between extents of Linguistic context availability (Part 2).........115
Table 4-7 Linguistic context availability By Generation, Gender and SES (Part 1)........117
Table 4-8 Linguistic context availability By Generation, Gender and SES (Part 2).........117
Table 4-9 Linguistic context availability by Age ...........................................................119
Table 4-10 Subjunctive rate by Linguistic context availability ........................................121
Table 4-11 Overall Subjunctive rate (Model 2) All speakers ........................................131
Table 4-12 Overall Subjunctive rate (Model 1) All speakers ........................................131
Table 4-13 Subjunctive rate in the Modal context (Model 2) .........................................132
Table 4-14 Subjunctive rate in the Modal context (Model 1) .........................................133
Table 4-15 Subjunctive rate in the Protasis Si context (Model 2) ....................................134
Table 4-16 Subjunctive rate in the Protasis Si context (Model 1) ....................................135
Table 4-17 Subjunctive rate in the Temporal context (Model 2) ....................................137
Table 4-18 Subjunctive rate in the Temporal context (Model 1) ....................................137
Table 4-19 Subjunctive rate in the Apodosis Si context (Model 2) .................................138
Table 4-20 Subjunctive rate in the Apodosis Si context (Model 1) .................................139
Table 4-21 Availability of the Concessive context (Model 2) ..........................................142
Table 4-22 Availability of the Concessive context (Model 1) ..........................................143
Table 4-23 Availability of the Possibility context (Model 2) ..........................................144
Table 4-24 Availability of the Possibility context (Model 1) ..........................................144
Table 4-25 Availability of the Volition context (Model 2) .............................................146
Table 4-26 Availability of the Hypothetical Como si context (Model 2) .......................147
Table 4-27 Availability of the Hypothetical Como si context (Model 1) .......................148
List of Figures

Figure 2-1 Most frequent syntactic and semantic contexts for occurrence of the Spanish Subjunctive, as identified by Silva-Corvalán (1994a), cf. Lynch (1999).................20
Chapter 1: Introduction

1. Introduction and research questions

Linguists have been studying the Spanish Subjunctive for over a century now. As early as the nineteenth century, the grammarian Andrés Bello wrote that the manifestation of the Subjunctive mood in a dependent clause was a product of the class of verb found in the matrix clause (Terrell & Hooper 1974). In his view, Subjunctive verbs did not possess a semantic function; the presence of the mood was solely due to the syntactic relationship between the two clauses. Research on the Spanish Subjunctive has undergone three major phases since Bello's time and over the course of the twentieth century (Lynch 1999). Until the 1970s, the Spanish Subjunctive was the subject of purely syntactic analyses (e.g., Cressey 1971; Farley 1970; Gili Gaya 1969; Shawl 1975). This approach privileged the obligatory/categorical versus optional/variable perspective, whereby the occurrence of the Subjunctive was considered, as in many current analyses, obligatory in certain syntactic contexts and optional in others. When optional, the Subjunctive is thought to be in variation with the Indicative mood. Terrell and Hooper’s 1974 article on the semantic basis of mood variability paved the way for other semantic analyses to follow (e.g., Blake 1981; Lantolf 1978). By the 1980s, pragmatic interpretations had inserted themselves into the syntactic-semantic debate. Guitart (1982), Lavandera (1983) and Lunn (1989), among others, expanded the discussion on Subjunctive/Indicative mood variability.

To date, much of the research on the Spanish Subjunctive (including investigations that take a semantic and/or pragmatic approach) continues to espouse the
division between obligatory/categorical contexts and optional/variable contexts that originated in the nineteenth century. Moreover, the fact that the classification of contexts into one or the other type differs depending on the investigator is not considered problematic.

In the twentieth century, immigration from Latin America to the United States (U.S.) shifted the focus of research on the Subjunctive onto U.S.-born second generation speakers. Much of the literature on the Spanish Subjunctive in the United States views the obligatory and optional categories in terms of cross-generational differences. The dominant conclusion is that the Subjunctive rates of the first and second immigrant generations tend not to diverge overall, that is, when all linguistic contexts are examined together, but do differ in particular syntactic/semantic/pragmatic contexts (e.g., Bookhamer 2013; Gutiérrez 1990; Lynch 1999; Torres 1989). The U.S.-born second generation is generally believed to have a reduced usage of the Subjunctive in a greater number of these contexts (henceforth labeled linguistic contexts) than the Latin American Raised (LAR) first immigrant generation, whose grammar tends to be regarded as the norm.

The literature embraces several theories to account for this intergenerational change, which is often regarded as an instance of simplification (Ocampo 1990; Silva-Corvalán 1994). These include internal language changes (Silva-Corvalán 1994b; Torreblanca 1997; Torres 1989; Zentella 1997); language contact (Romaine 1995; Silva-Corvalán 1994b; Thomason & Kaufman 1988; Weinreich 1953); incomplete acquisition (Montrul 2009); attrition (Merino 1976, 1983; Montrul 2014); minimal pairs (Zentella 1997); and cognitive complexity (Ocampo 1990; Silva-Corvalán 1986). Renowned
scholars (e.g., Gutiérrez 1990; Montrul 2009; Ocampo 1990; Silva-Corvalán 1994) even go so far as suggesting that the Subjunctive/Indicative mood distinction is altogether disappearing from the Spanish grammar of second generation speakers in the United States.

The central role assigned by researchers to speakers’ immigrant generation (henceforth *generation*) has somewhat obscured the effects of other socio-demographic variables in predicting speakers’ Subjunctive rates. Studies of the Spanish Subjunctive and other grammatical features of Spanish in contact settings (such as New York City) that include Gender, Socio-economic status (SES) and Age have shown that these three variables can also significantly affect the presence and absence of particular variants in variable linguistic contexts. For instance, Bookhamer (2013) found that women\(^1\) use the Subjunctive less than men in New York City (NYC). Shin & Otheguy (2013) discovered that NYC Established Immigrant women have higher rates of subject pronouns than the men. NYC Colombians of high socio-economic status employ the periphrastic future more than their peers of low socio-economic status (Orozco 2007). Finally, Lantolf (1978) established that NYC Puerto Rican youth exploit the Subjunctive less than older NYC Puerto Ricans. Although these results are telling, socio-demographic variables other than Generation still constitute a rare occurrence in studies of the Spanish Subjunctive situated in bilingual settings.

Despite these advances, there's still much we do not understand about the Spanish Subjunctive. First, we are missing a comprehensive understanding of how individual speakers treat the Subjunctive. Most studies consider only overall rates in a sample,

---

\(^1\) Bookhamer (2013) examined the gender effect on Subjunctive rates in obligatory and optional contexts, in particular ethnonational groups and within different immigrant generations, but not on the whole.
distinguishing between different linguistic contexts, but not examining the number (or proportion) of informants that actually use these verbs in each context. Yet establishing the number of speakers in whose speech each context is or is not found, and concentrating on the most popular while setting aside the least popular ones, may lead to results that are more representative of the speech of the Latino population as a whole.

Second, very few studies examine relationships between rates of occurrence of different linguistic contexts in speech. Yet doing so would offer greater insight into speakers’ patterns of Subjunctive usage. Third, more studies on the Spanish Subjunctive in the United States that include external variables other than generation are needed in order to identify which socio-demographic characteristics best predict Subjunctive usage.

The present investigation tackles these issues, and aims to contribute to the understanding of how the socio-demographic characteristics of speakers shape their Subjunctive usage. I study here: (a) overall and context-specific speaker rates of Subjunctive usage, as well as (b) what I call *Linguistic context availability*, which deals with the occurrence and non-occurrence of each Subjunctive-inducing linguistic context in the different speakers, and (c) the socio-demographic characteristics that influence both rates and availability. The investigation derives its data from 142 bilingual Spanish speakers in New York City, whose transcripts constitute the Otheguy-Zentella Corpus of Spanish in New York (OZC-NY), details of which are given in Chapter 3. These speakers’ Subjunctive verbs are analyzed in nine linguistic contexts. The *overall Subjunctive rate* refers to individual speakers’ Subjunctive rate when all nine linguistic contexts are grouped together. A detailed description of the different contexts is offered in Chapter 3. The following questions are explored in the study:
(RQ1) Are different linguistic contexts interrelated in speech? More specifically:

a. Is there a statistical association between individual speakers’ overall Subjunctive rate and their Subjunctive rate in the most popular linguistic contexts?

b. Are Subjunctive rates in the most popular linguistic contexts statistically associated with one another?

c. Is there a statistical association between the extents of availability of different linguistic contexts?

d. Is there a statistical association between speakers’ Subjunctive rates (overall and in the most popular linguistic contexts) and the availability of different linguistic contexts?

(RQ2) Is it the case that particular socio-demographic characteristics shape Subjunctive rates, overall and in the most popular linguistic contexts? That is, who uses the Subjunctive more and who uses it less, overall and in the most popular linguistic contexts?

(RQ3) Is it the case that particular socio-demographic characteristics affect the availability of certain linguistic contexts? In other words, who tends to use each linguistic context more and who tends to use each one less?
In order to answer these three main research questions, we must start by considering the availability of each of the nine linguistic contexts in the speech of all 142 informants (taken together). The concept of availability should be understood in the context of the OZC. The availability and non-availability of a particular linguistic context is not necessarily representative of the speaker’s full linguistic repertoire. Indeed, an informant’s breadth of linguistic knowledge is probably much greater than that reflected in an hour-long interview. Still, our exploration of Linguistic context availability is suggestive of speaker tendencies, and is meaningful in that respect. Thus, we ask the following question: *In how many informants does each linguistic context occur and fail to occur?* This inquiry is resolved in Chapter 3.

The results of this investigation show that there exists a significant association between rates of Subjunctive in the most popular linguistic contexts and overall (Chapter 4, section 2.1); between rates of Subjunctive in different linguistic contexts (Chapter 4, section 2.1); between the availability of certain linguistic contexts (Chapter 4, section 2.2); and between Subjunctive rate and Linguistic context availability (Chapter 4, section 2.3). These relationships establish that the treatment of the Subjunctive is consistent for all speakers in the sample. There are speakers who show a steady preference for the Subjunctive mood and for Subjunctive-inducing linguistic contexts, and there are speakers who have a reliable aversion to the Subjunctive mood and to Subjunctive-inducing linguistic contexts (Chapter 5, section 4). Overall Subjunctive rates and Subjunctive rates in the Modal context (e.g. *Como quieran, lo que quieran, como que quieran*) are predicted by informants’ gender, with women drawing on the Subjunctive less than men in both cases (Chapter 4, section 3.1).
In addition to describing these statistical associations, I make an attempt in each case to propose an explanation for my findings. To start, I call the association between female gender and reduced Subjunctive use the *Women effect*. This effect is attributed to women’s desire for social change (Chapter 5, section 4.1). Age and Generation shape Subjunctive rates in the Protasis *Si* context (e.g. *Si vinieran, si hubieran venido*). More specifically, the younger the speaker, the less s/he uses the Subjunctive in that linguistic context (Chapter 4, section 3.1). This *Youth effect* is explained by young people’s desire for a greater social standing (Chapter 5, section 4.3). Similarly, the New Yorkers (i.e. the New York Raised and the Established Immigrants) have lower rates of Subjunctive than the Newcomers in the Protasis *Si* context (Chapter 4, section 3.1). The *New Yorker effect* refers to the New Yorkers’ pursuit of a higher social status, which they seek to obtain by displaying their North American (U.S.) identity (Chapter 5, section 4.2).

Notably, Subjunctive rates do not vary based on speakers’ socio-economic status, in any linguistic context or overall (Chapter 4, section 4). Furthermore, speakers’ Subjunctive rates in the Temporal context (e.g. *Hasta que vengan, antes de que vengan, cuando vengan*) and in the Apodosis *Si* context (e.g. *Si..., quisiera ir, si..., hubiera querido ir*) are not affected by any socio-demographic variable. In other words, there are no group differences with respect to these two linguistic contexts (Chapter 4, section 3.1).

Socio-demographic characteristics also affect Linguistic context availability. The availability of the Concessive context (e.g. *Aunque quieran*) hinges on speakers’ gender, as it is more available to men than to women (Chapter 4, section 3.2). It is more probable

---

2 Our terminology regarding the different generation groups is spelled out in Chapter 3. To summarize, what we call the *Newcomers* are first generation Latin American Raised informants who recently arrived to New York City at an older age, the *Established Immigrants* are first generation Latin American consultants who have been living in New York City for longer and from a younger age, and the *New York Raised* are second generation speakers who were born and/or raised in New York City.
that the Possibility context (e.g. *Es posible que quieran, tal vez/quizá(s) quieran, a lo mejor quieran*) will appear in the speech of New York Raised informants than in that of the Newcomers (Chapter 4, section 3.2). Finally, New York Raised women have a greater tendency to use clauses introduced by *querer que* and *esperar que* (i.e. the Volition context) than do Newcomer women (Chapter 4, section 3.2). None of the socio-demographic variables affect the availability of the Hypothetical *Como si* context (e.g. *Como si vinieran*). I propose that the availability and non-availability of all of the linguistic contexts cited in this paragraph hinge on the messages conveyed by certain types of speakers, or certain groups of speakers (Chapter 5, section 5).

The present investigation’s research questions, methodology and results weigh in on the body of work on the Spanish Subjunctive, and more generally on the fields of (variationist) Sociolinguistics and Second Language Acquisition, particularly in the context of bilingualism.

First, this study challenges the notion that mood distinctions are lost to, or highly diminished in, second generation speakers. In fact, my results show that the association between Generation and Subjunctive rates is not nearly as strong as has been argued, as Generation is found to predict Subjunctive rates in only one linguistic context. Accordingly, the notions that second generation speakers have an incompletely acquired or attrited grammar of the Subjunctive are called into question. This singular result, along with my other results, which tend to depart from the literature, is related to this investigation’s distinct methodology. The present investigation includes other socio-demographic variables besides Generation (e.g., Gender, Age, Socio-economic status); a comparatively large number of tokens (i.e. verbs in the Indicative and the Subjunctive)
collected with the help of a computer programming language designed for this purpose; and linguistic contexts that were selected for analysis based on the number of informants in whose speech each context occurs (rather than on the number of tokens found in the sample as a whole). These three elements point to ways in which variationist sociolinguistic studies that are situated in bilingual contexts can more closely reflect the speech of the communities that they intend to represent.

Secondly, speakers’ treatment of the Subjunctive tends to be viewed in the literature through the lens of particular linguistic contexts. Yet to the best of my knowledge the relationships that may exist between these linguistic contexts, which can help explain patterns of Subjunctive usage, have not been investigated. The present study shows that mood variability is better understood when we explore the relationships that exist between different linguistic contexts, in terms of their availability and in terms of the occurrence of the Subjunctive mood in them. Furthermore, when the obligatory/categorical versus optional/variable distinction is removed, linguistic contexts can be viewed more comprehensively.

The idea at the heart of this atypical perspective is that all of the contexts in these two categories are variable, meaning that the speaker is always faced with a choice—the choice to use the Subjunctive or the Indicative mood, depending on the type of message that s/he wishes to express. The result is that speakers are consistent with respect to both their mood and linguistic context preferences. More generally, this finding supports the theory that sociolinguistic variation is anything but haphazard. Although the present investigation does not attempt to explore the notional content of messages conveyed by
the Subjunctive, it makes a modest attempt at paving the way for such an analysis in the future.

Third, our results point to the fact that sociolinguistic research in bilingual settings (such as New York City) should include external variables other than Generation, such as Gender, Age, Socio-economic status (SES) and the interaction of Gender and Generation. While it has been shown that these socio-demographic variables help predict the variability of grammatical features in monolingual settings (Gutiérrez 1994; Labov 2001; Lastra & Butragueño 2012; Milroy 1987; Milroy & Milroy 1992; Serrano 1995), their presence is still too rare in research on the Subjunctive in bilingual settings. By establishing the significance of some of these socio-demographic variables, the present investigation contributes to the relatively uncharted field of variationist research in bilingual settings. Furthermore, if these effects are revealing of a language change in progress, then we contend that this language change is not merely generational. Instead, the Subjunctive grammars of women (particularly New York Raised women) and youth are affected as well. The broader sociopolitical explanations that we rely on embrace perspectives from other fields (such as sociology, anthropology, etc.), which can expand our understanding of this linguistic phenomenon.

The dissertation is organized in the following way. In the remainder of the chapter, I illustrate what it means for the Spanish Subjunctive to be in variation (section 2), and I offer a preview of how the Spanish Subjunctive is used in bilingual New York City (section 3). Chapter 2 reviews the literature on the Spanish Subjunctive in the United States. Chapter 3 describes the study’s methodology. Chapter 4 presents the bivariate and multivariate results. In Chapter 5, these results are discussed in light of the
present investigation and literature’s hypotheses. Chapter 6 concludes with the study’s contributions, pedagogical applications, limitations, and directions for future work.

2. The Spanish Subjunctive in variation

Over the centuries, French grammarians have ascribed 76 meanings to the Subjunctive and Indicative moods in that language (Poplack 2018). In his letter commenting on Poplack's work, Otheguy (2018) attributes this “record of human failure” (p.36), exposed by Poplack, to the misguided view of “a form-message pairing that exists beyond the scope of a particular instance of use” (p.37). Instead, Otheguy argues that “the forms of languages do not bear a symmetrical relationship with the notional categories of messages or parts of messages. They bear a symmetrical relationship only with their meaning” (p.37). This theory can be applied to our study of Subjunctive/Indicative mood variation. As we will see in Chapter 2, there is no consensus in the field regarding the meaning of the Subjunctive mood, which has been described as denoting non-assertion or presupposition (Lunn 1989; Terrell & Hooper 1974); previously known information (De Mello 1974; Guitart 1982); irrelevance in argumentative style (Lavandera 1983; Lunn 1989); insignificance (Lunn 1989); uncertainty (Torres 1989); and probability and irrealis (Ocampo 1990), among other notions. Similarly, the common division between categorical/obligatory and variable/optional linguistic contexts discussed above, and developed further in Chapters 2 and 3, suggests that the Subjunctive and Indicative moods always have the same meaning in obligatory/categorical contexts but not in optional/variable ones.

Yet, following Otheguy’s line of thought, even when the two forms (Subjunctive and Indicative) have the same reference, and “may be saying the same thing” (p.40), they
are not necessarily equivalent. And even when the forms are not saying the same thing, when the forms communicate different messages, they can still have roughly similar references, and be in variation. Let us take a look at examples (1) and (2) below, from the interviews. To protect their identity, all informants are anonymous. Here, we refer to them through their immigrant group (i.e. LARN for Newcomers, LARI for Established Immigrants or NYR for New York Raised), transcript number (e.g., 427) and ethnonational origin abbreviation (C for Colombian, U for Cuban, D for Dominican, E for Ecuadorian, M for Mexican and P for Puerto Rican). In example (1), the informant is a Puerto Rican Newcomer. In example (2), the informant is a Mexican Established Immigrant.

(1)  Lo que me da es como un un dolor en la pierna, es como si la circulación **fuera** bien rápida, que se me pone como entumecida…  
– LARN 427P  
‘What happens is like a a a pain in my leg, it’s as if my blood were flowing really fast, and my leg goes numb…’

(2)  No que no puedes mover las manos, sí las puedo mover vean, y aguantaron todo el dolor adentro, apenas las pude mover así, pero bien despacito, les dije ven como si las **muevo**? – LARI 352M  
‘It’s not that you can’t move your hands, I can move them, see, and they bore all of the pain inside, I almost couldn’t move them like that, but very slowly, I told them, look how/as if I can move them?’

In examples (1) and (2), both informants are describing their physical pain or discomfort.

The Hypothetical *Como si* context appears in both examples. In example (1), *como si* introduces the verb *fuera* ‘were’, which is in the Subjunctive. In example (2), on the other hand, the verb that follows *como si, mover* ‘I move’, is in the Indicative mood. I
believe that the speaker is faced with a viable choice in each instance. In (1), the informant could have employed the Indicative va ‘goes’, while in (2) the informant could have selected the Subjunctive verb moviera (or moviese) ‘I moved’. Even though the two possible forms in each example have roughly similar references (the informant’s physical pain), the two may convey different messages. Or, the messages might be similar. An in-depth analysis (in a different type of study) could determine this. Irrespective of the messages conveyed, fuera and va represent two possible variants in variation, as do movero and moviera/moviese. They are considered variants in a variable context because they represent two viable choices for the speaker.

This variationist study thus addresses the distribution of forms. In my dissertation, I seek to understand not just how these forms are generally distributed in speech, but in whom (in which types of speakers) they emerge, are likely to emerge, and fail to emerge. A modest foray into the why sets the groundwork for a future probe into a more detailed analysis centered on meanings and messages.

3. The Spanish Subjunctive in bilingual New York City: A preview

As discussed in section 2 above, and further along in Chapters 4 and 5, Spanish speakers in New York City draw on the Subjunctive mood differently depending on who they are (their socio-demographic characteristics) and on the linguistic context at hand (on the message type they want to convey). It is important to note that these differences emerge even though the 142 interviews that this study is based on comprise similar questions and themes. While the sociolinguistic interviews were not formally structured per se (i.e.
there was no list of questions\textsuperscript{3}, the interviewers were trained through a series of workshops prior to conducting the interviews (Otheguy 2019, personal communication). As reported in Chapter 3, this preparation ensured comparability between interviews, as “all informants were prompted to talk on similar topics including personal stories and opinions” (Varra 2013: 41). The following examples, gathered directly from the interviews, are meant to elaborate on the point made above that speakers can draw on different moods even when the reference is similar. Here, we add information about these speakers’ socio-demographic features. In other words, we consider one of the several significant group differences that will become apparent in our statistical analyses in Chapter 4, and which will be discussed in Chapter 5. These examples are a preview of how the Spanish Subjunctive is used by bilinguals in New York City.

Examples (3) and (4) below were selected because the two informants, one male and one female, broach a similar topic. They are both discussing their personal experiences with non-Latinos in New York City.

(3) ¿Entre ellos y nosotros? […] Unos se pegan mucho a las cosas, a las cosas que logra obtener aquí o logra obtener en los países de uno, ellos no, ellos como que estuvieran acostumbrados siempre a eso, y no les llama mayormente la atención… – MALE 338E

‘Between them and us? […] Some of us get attached to things, to the things that we manage to obtain here or manage to obtain in our own countries, they don’t, for them it’s like they were used to this all along, and they don’t think twice about it…’

\textsuperscript{3} There was, however, a two-part oral questionnaire, which was administered to all informants at the end of the interview. ‘The first part probed participants for the words they used for common things which were known to have different names in different Spanish-speaking Latin American countries (such as bizcocho in Puerto Rico and pastel in Colombia for ‘birthday cake’). The second part gathered demographic information on each informant” (Varra 2013: 41, 42). Only the conversational part of the interview (which was transcribed) was used to collect our Subjunctive and Indicative verbs.
It is evident from the examples above that the two informants have had negative encounters with non-Latinos in New York City. When voicing what it feels like to be ‘us’ (Latinos) versus ‘them’ (non-Latinos), both speakers resort to the Modal context, introduced by como que (in these examples). Yet, the verbs that follow como que are not conjugated in the same mood. In example (3), the male informant draws on the Subjunctive verb estuvieran ‘were’, whereas the female informant in example (4) employs the verb existe ‘exist’ in the Indicative mood.

While men and women draw at times on the Subjunctive mood, and at times on the Indicative mood in the Modal context (e.g., Como quieran, lo que quieran, como que quieran), examples (3) and (4) illustrate a trend that was found to be significant. Indeed, example (3) is representative of what men tend to do, and example (4) of what women are likely to do. It turns out that men have a greater tendency than women to use the Subjunctive mood in this linguistic context, and overall. These groups’ distinct mood preferences, driven by gender, represent one of the several patterns that we have encountered in the Spanish of New York City bilinguals.
Chapter 2: Literature Review

1. Immigrant generations and the Spanish Subjunctive

In the fields of corpus linguistics and second language acquisition, it has been argued that speakers of different immigrant generations differ in their treatment of the Subjunctive. It has also been argued that usage of the Subjunctive is, more generally, different in bilinguals than in monolinguals. Studies on the Spanish Subjunctive in the United States often compare two or three immigrant generations to one another. While the first generation normally encompasses speakers who were born and raised in Latin America, the second and third generations usually comprise speakers who were born and/or raised in the United States. For example, Otheguy & Zentella (2012) divide their informants into two immigrant generations, and subdivide the first generation into two distinct groups. The first generation of recent immigrants is labeled Latin American Raised Newcomers, as they arrived to New York City (NYC) after age seventeen and have been here for five years or less. The other first generation group comprises the Latin American Raised Established Immigrants. They arrived to New York by age seventeen or have been living in the city for over five years. The second generation is designated as the New York Raised, and consists of informants who were either born in NYC or arrived to the city before age three. Immigrant generations are generally split into groups according to their place of birth, age of arrival and/or number of years in the United States. It appears that growing up in a monolingual or in a bilingual environment matters. Taken together, under the label Generation, these characteristics are often regarded as engendering cross-generational differences in Subjunctive usage.

While some studies (e.g., Bookhamer 2013; Gutiérrez 1990; Lynch 1999; Torres
1989) contend that the overall frequency of the Spanish Subjunctive is constant across two or three generations of Spanish/English bilinguals in the United States, other studies (e.g., Fernández Pedraza 2014; Montrul 2009; Silva-Corvalán 1994) maintain that the mood’s general occurrence decreases from one generation to the next. Fernández Pedraza (2014) asserts that there is a clear connection between Generation and the Subjunctive’s rate of occurrence. Montrul (2009: 258) would concur, as the results of one of her tasks show that use of the Subjunctive among heritage speakers (i.e. the second generation) increases by proficiency level. In her study, a written Spanish proficiency test, “consisting of a cloze part and a vocabulary part” (p.248), is used to ascertain the heritage speakers’ proficiency level, which determines their distribution into three groups (advanced, intermediate, and low). Additional details on this written proficiency test are not provided.

In her view, the advanced heritage speakers most resemble the native speakers\(^4\), the equivalent of the first generation in her research. While a little over a quarter of the verbs produced by the intermediate and advanced groups are in the Indicative, the low proficiency heritage speakers draw on the Indicative half of the time. Montrul (2009) concludes that the Indicative may have supplanted the Subjunctive among low proficiency heritage speakers.

In this same investigation, the researcher also finds that intermediate and low proficiency heritage speakers produce a higher rate of infinitives than the advanced

\(^4\) Although heritage speakers are considered native speakers in the present study, Montrul distinguishes between heritage speakers and native speakers (also called the control group). In her 2009 study, heritage speakers are second generation bilingual speakers of varying proficiency levels, while her native speakers correspond to her first generation monolingual control group. The author’s terminology is used when referring to her research.
heritage speakers and the native speakers. The use of infinitives as an alternative for
Subjunctives in Montrul (2009) may very well constitute an avoidance strategy
comparable to Gutiérrez (1990)’s interesting observation of the increasing avoidance
(with each generation) of subordinate clauses. Direct speech tends to replace indirect
speech with direct objects in subordinate clauses. Clearly, all of the aforementioned
studies show that usage of the Subjunctive mood diminishes from one generation to the
next.

Yet, other studies establish that there is little, if any, cross-generational change in
overall frequency of the Subjunctive. For example, in Torres (1989), the first
generation’s Subjunctive output represents 4.7 percent of all verbs, while the second
generation uses the Subjunctive 4 percent of the time, that is, at a similar rate as the
previous generation. Lynch (1999) confirms this finding, as the three generations in his
study use the Subjunctive to the same extent in all but one of his categorical linguistic
environments. Similarly, Gutiérrez (1990) finds that the first and second generations in
his investigation yield the same amount of adverbial phrases. Bookhamer (2013) also
establishes that his two generations draw on the Subjunctive to similar degrees. Indeed,
his results show a negligible difference of 1.6 percentage points (first generation: 6.8
percent of all finite verbs, second generation: 5.2 percent of all finite verbs).

Interestingly, Bookhamer adds that the two groups are nearly identical in terms of
Subjunctive tense distribution as well. Both generations conjugate 82 percent of their
Subjunctive verbs in the present tense (e.g., vaya), while 15 to 16 percent of them are in

---

5 The Volition context (e.g., Mis padres siempre esperaban que yo…) is the only exception in his study.
6 Bookhamer (2013) analyzed the distribution of four tenses in the Subjunctive: present, imperfect,
pluperfect and present perfect. The last two were hardly used by either generation.
the imperfect Subjunctive (e.g., *viniera*). Thus, it is clear that both mood and tense rates are nearly identical cross-generationally.

Though the research on cross-generational rates of occurrence of the Subjunctive may vary, most studies agree that the mood is not altogether disappearing from Spanish in the U.S.-born generation. Gutiérrez (1990) observes a decrease in Subjunctive usage in temporal, causal and final clauses in third generation Spanish in Los Angeles, but shows that these are still appearing with Subjunctives. In these constructions, the Subjunctive simply occurs less frequently in third generation Spanish (19%) than in the first (42%) and second (30%) generations. Thus, it appears that overall frequency alone does not contribute to potential cross-generational differences in Subjunctive usage. Other criteria, such as contexts of use, should be considered when making cross-generational comparisons.

2. **Syntactic/semantic/pragmatic contexts**

In addition to analyzing frequency, most studies have also examined the occurrence of the Subjunctive mood in specific linguistic contexts. The number and type of contexts of use differ depending on the study. The average number of contexts in the studies examined here is 12, ranging from just one in Mikulski (2010) to 24 in Silva-Corvalán (1994). Rather than listing all of the different contexts analyzed in the many investigations on the Spanish Subjunctive in Spanish in the U.S., Figure 2.1 reflects the most common occurring ones, according to Silva-Corvalán (1994a: 264-268), and reproduced in Lynch (1999: 30):
Figure 2-1
Most frequent syntactic and semantic contexts for occurrence of the Spanish Subjunctive, as identified by Silva-Corvalán (1994a), cf. Lynch (1999)

<table>
<thead>
<tr>
<th>Matrix</th>
<th>Example with Subjunctive form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volitional</td>
<td>Quiero que hable</td>
</tr>
<tr>
<td>Purpose clause</td>
<td>Para que hable</td>
</tr>
<tr>
<td>Concessive clause</td>
<td>Aunque hable</td>
</tr>
<tr>
<td>Comment</td>
<td>Lamento que hable</td>
</tr>
<tr>
<td>Modal (main clause)</td>
<td>Debiera hablar</td>
</tr>
<tr>
<td>Mental act</td>
<td>No advierte que hable</td>
</tr>
<tr>
<td>Temporal clause</td>
<td>Cuando hable</td>
</tr>
<tr>
<td>Apodosis</td>
<td>…le hablara</td>
</tr>
<tr>
<td>Protasis</td>
<td>Si hablara</td>
</tr>
<tr>
<td>Uncertainty</td>
<td>No sé si hable</td>
</tr>
<tr>
<td>Modal clause</td>
<td>Así como hable</td>
</tr>
<tr>
<td>Locative clause</td>
<td>Donde hable</td>
</tr>
<tr>
<td>Adjectival clause</td>
<td>El que hable</td>
</tr>
<tr>
<td>Assertive</td>
<td>Sé que hable</td>
</tr>
</tbody>
</table>

Figure 2.1 represents the 14 most commonly occurring contexts that favor the Subjunctive, according to Silva-Corvalán (1994a). As Lynch (1999) points out, Silva-Corvalán actually identified 18 such contexts. Four of them, however, appeared fewer than 30 times in her sample. These were hypothetical manner (e.g., como si), impossibility, causative and adverbials of manner/time/place.

Although the contexts tend to overlap in the different studies, there is no consensus in the field regarding motivation for the choice between the Subjunctive and the Indicative. Silva-Corvalán (1994: 255) effectively encapsulates the differing schools of thought on mood (Subjunctive/Indicative) variation as follows:

- Syntactically and lexically motivated (e.g., Bello 1847/1982, Lakoff 1968)
- Semantically based (e.g., Guitart 1990, Terrell & Hooper 1974)
- Linguistically meaningless distinction in French (e.g., Poplack 1990), but
meaningful in Spanish (e.g., Klein-Andreu 1980)

One of the initial analyses of mood selection in Spanish, conducted by Terrell & Hooper (1974), provides much insight on the difference between a syntactic or semantic context for the Subjunctive. They refer to the Colombian grammarian Andrés Bello, along with other linguists and textbook writers in the United States, who hold that a certain type of verb in the matrix clause causes the Subjunctive to occur in the sentential complement (Terrell & Hooper 1974). To clarify the distinction, the authors provide the two following examples (p.484):

(5) Sé que Ud. **tiene** que trabajar mucho (*Indicative*)

‘I know that you have to work a lot’

(6) Me alegro que Ud. no **tenga** que trabajar tanto (*Subjunctive*)

‘I’m happy that you don’t have to work so much’

In example (6), the matrix verb *alegrarse* ‘to be happy’ belongs to the syntactic and/or semantic class of verbs that requires the Subjunctive, which shows up in the conjugation of the embedded verb *tenga* ‘you have’ (Subjunctive). This is not the case with *saber* ‘to know’ in example (5). In Bello’s view, the Subjunctive mood only emerges because of the co-occurrence relationship, and not because it possesses any sort of semantic function (Terrell & Hooper 1974).

Terrell & Hooper (1974: 485) highlight the differences between the syntactic and the semantic approaches: “The first analysis claims that the subjunctive or indicative
forms do not function meaningfully, because the choice of mood is determined
automatically by the type of phrase found in the matrix. This analysis is syntactically
based. Even though the matrices are classified in terms of their meaning, the result is
considered to be only syntactic: the mood of the embedded verb is merely a
morphological reflex of the class of the matrix phrase. The second analysis is
semantically based in that it claims that the mood of the embedded verb can be freely
chosen and thus carries meaning. The hypothesis is that there are several basic attitudes
that a speaker can adopt toward a proposition. These attitudes govern the choice of verb
form and the choice of matrix.” Thus, unlike syntactic approaches that chiefly consider
the matrix verb, semantic approaches such as Terrell & Hooper’s examine the sentence as
a whole.

Similarly, Hadlich (1971: 188, cf. Terrell & Hooper 1974) asserts that the
semantic content of the whole utterance should be taken into consideration, rather than
that of the matrix verb alone. Terrell & Hooper (1974: 488) coincide with Hadlich in that
semantic properties are crucial, and that the syntactic characteristics of matrices and
complements “are an automatic consequence of their semantic properties”. Torres (1989)
corroborates this theory by showing that the function of the Subjunctive has changed for
her second generation Puerto Rican speakers in New York City, who do not draw on the
Subjunctive after Si (in an if clause, or conditional clause category), but do, instead, in the
result clause, as in example (7), from Torres (1989: 72):

(7)  

    [...] Si me **había ido** a trabajar **estuviera** bien [...]  

    ‘[...] If I had gone to work, it would’ve been good [...]’
Since the mood (the imperfect Subjunctive) is still in use, Torres claims that the second generation is making a decision that is semantically based. To explain the semantic distinction between the two moods, Terrell & Hooper (1974, cf. Lynch 1999: 13) propose a system that consists of assertion (assertion or report), presupposition (mental act or comment) or neither (doubt or imperative). Still, semantic explanations remain at the sentence-level domain, and fail, according to some (e.g., Lynch 1999) to explain all of the distinctions in mood.

Pragmatic-level explanations such as Lavandera (1983) expand the perspective to the discourse context. As Lynch (1999: 16) affirms, “By doing so, she was able to account for further variation which previous analyses at the clause and sentence level had been unable to explain.” Through her research on Buenos Aires Spanish, Lavandera added the idea of relevance in argumentative style to Terrell & Hooper (1974)’s concepts of assertion (requiring the Indicative mood) and non-assertion (requiring the Subjunctive mood), and to De Mello (1974) and Guitart (1982)’s principle of known (prompting the Subjunctive) versus unknown (triggering the Indicative). A few years later, Lunn (1989) further developed Lavandera’s pragmatic theories of assertion and relevance (cf. Lynch 1999: 18).

In the same vein, Silva-Corvalán (1994: 259) believes that “the different meanings of Indicative and Subjunctive determine their distribution.” However, advocates of the semantic approach disagree on what these meanings are. Ocampo (1990: 43) argues that the Subjunctive/Indicative mood variation allows the speaker to express semantic nuances by means of a scale that extends from the domain of reality (Indicative) to those of probability and irrealis (Subjunctive). Fernández Pedraza (2014:
17), however, contends that this division is too rigid, and that mood choice cannot be limited to the distinctions between objectivity/subjectivity, certainty/uncertainty, concrete/non-concrete or real/unreal.

It is interesting to note that Silva-Corvalán herself, along with the researchers who adopt her theory (e.g., Ocampo), allow for exceptions to a purely semantic theory. In her study (1994), she only labels three contexts as dependent on the speaker’s degree of “freedom of choice, regardless of pragmatic factors” (p.264). In her view (prior to analyzing her results), there are instances that unconditionally require the use of either the Subjunctive or the Indicative mood. Likewise, Ocampo (1990), who bases his study on Silva-Corvalán’s oral corpus of Spanish in Los Angeles, claims that there are linguistic environments that only license the Subjunctive form. When it is grammatically conditioned, the Subjunctive loses its semantic content, and the speaker has no choice but to employ that particular mood (Ocampo 1990).

Similarly, even though Lynch (1999: 105) examines eight variable contexts, “conditioned by the semantic and pragmatic values associated with the discourse context in which the verb is found [...]”, he identifies four categorical contexts that only allow for the Subjunctive: volition (e.g. Quieren que sus hijos hablen bien ‘They want their children to speak well’), purpose (e.g. Para que aprendan español ‘so that they learn Spanish’), temporal with futurity (Mañana cuando llegue ‘when I arrive tomorrow’) and

---

7 In the literature as in the present study, linguistic contexts can comprise one or several environments. For instance, the Volition context includes two environments, introduced by querer que and esperar que. Environments that convey similar messages are grouped together into linguistic contexts. See Chapter 3 for a complete discussion.

8 A fifth context, which Bookhamer (2013) touches on as well, is the idiomatic usage of the Subjunctive (e.g., O sea, vaya, lo que sea, lo que fuera, comoquiera). This context is not discussed much in either Lynch (1999) or Bookhamer (2013), since it differs from the other contexts in that it only includes fixed expressions.
hypothetical manner (*Como si estuviera en Cuba* ‘as if I/you (formal)/he/she were in Cuba’). Likewise, Montrul (2009)’s experiment on relative clauses assumes that the Subjunctive is always obligatory in her examples.

Yet Torres (1989)’s data challenge the assumption that the Subjunctive is the only acceptable mood in certain contexts, such as Lynch’s temporal with futurity alluded to above. Indeed, one of Torres’ first generation informants uses at times the Subjunctive, and at times the Indicative in conjunctive expressions of time that refer to the future (with *cuando*). While prescriptive grammars maintain that the Subjunctive is required in that environment, Torres contends that the context should be factored into the equation, as degrees of certainty in this case, and implicit and explicit meanings more generally, may influence the speaker’s decisions with respect to mood.

Furthermore, the Subjunctive can occur and fail to occur in unexpected contexts, that is, in the ones that prescriptive grammars fail to identify. Torres (1989) found that her first generation informants use the imperfect Indicative in both the *if* clause and the result clause, as in my example (8) below, rather than the prescribed norm which requires the Subjunctive in the *if* clause and the Conditional in the result clause.

(8)  

   Si *era* rica, *me compraba* una casa  

   ‘If I was/were rich, I bought/would buy myself a house’

In example (8), the verbs *era* ‘I was’ and *me compraba* ‘I bought myself’ are in the imperfect Indicative, instead of in the prescribed Subjunctive and Conditional (*fuera* ‘were’ and *me compraría* ‘would buy myself’). As a result, Torres argues that pragmatic
factors should be taken into consideration when contemplating variation.

Finally, certain studies label their contexts ‘syntactic-semantic’ (Bookhamer 2013; Fernández Pedraza 2014; Silva-Corvalán 1994; Torres 1989), thereby creating some common ground between the two distinct approaches. Irrespective of the preferred theory, much of the literature divides syntactic/semantic/pragmatic contexts into two distinct categories: obligatory and optional (contexts).

3. Obligatory and optional contexts

Besides the syntactic and semantic/pragmatic classifications, nearly all of the studies on the Spanish Subjunctive in the United States divide their contexts into two categories: obligatory (also called categorical) and optional (also called variable). The terms obligatory and optional are quite confusing, as they are used in two different ways. As noted in the previous section, some researchers (e.g., Fernández Pedraza 2014; Lynch 1999; Montrul 2009; Ocampo 1990; Silva-Corvalán 1994) seemingly describe their informants’ Subjunctive usage with prescriptive grammar rules in mind, or based on their own intuitions. They deem that the Subjunctive is obligatory in certain linguistic environments, prior to analyzing their data. The terms obligatory and optional refer to an a priori notion, in this case. On the other hand, many linguists (e.g., Bookhamer 2013; Lynch 1999; Ocampo 1990; Silva-Corvalán 1994) —some of whom simultaneously use the a priori notion for a portion of their contexts—examine their data first, and then decide whether the Subjunctive has a tendency to occur in the linguistic context at hand. Here, the obligatory category is a tendency rather than a rule, as variability is taken into consideration. Bookhamer (2013), for example, judges that an environment is obligatory when the Subjunctive mood arises over 90 percent of the time. These contexts are
designated as obligatory and optional a posteriori, following the data analysis.

It is difficult to compare the different studies’ classifications of obligatory and optional contexts, due to the informants’ diverse characteristics (e.g. varied ethnic origins, assorted locations, etc.). Yet it may still be of interest to note that the two types of analyses (a priori and a posteriori) yield divergent outcomes. For example, Fernández Pedraza (2014)’s a priori classification characterizes comment clauses (e.g., *es bueno que tengas* ‘it’s good that you have’), possibility clauses (e.g., *es posible que tengas* ‘it’s possible that you have’) and negative subordinate clauses (e.g., *no es que no tengas* ‘it’s not that you don’t have’) as obligatory, whereas Bookhamer (2013)’s a posteriori categorization places them in the optional category for first generation Latin American Raised Newcomers.

Although this may seem like a mere question of methodology, the procedure is in fact theoretical in nature. Researchers that mark a certain context as obligatory prior to examining the data subscribe to the traditional view, frequently found in Spanish grammar books, in which the occurrence of the Subjunctive is subject to a compulsory rule. The mood’s illicit presence or absence is therefore often regarded as ungrammatical. Montrul (2009: 263) conveys this idea when describing “very high error rates in obligatory contexts” among Spanish heritage speakers.

Interestingly, studies grounded in similar theories also place linguistic environments into distinct obligatory and optional categories. As noted, within the same generation, linguistic contexts that are deemed obligatory in some studies are categorized as optional in others. For example, Silva-Corvalán (1994) found that her informants really only drew on the Subjunctive categorically in just one of eighteen contexts (in
comment clauses), whereas most studies on the Subjunctive indicate that the mood is obligatory in many more, particularly with respect to the first generation (e.g., Bookhamer 2013; Lynch 1999; Ocampo 1990; Torres 1989). For instance, Torres (1989: 71)’s first generation uses the Subjunctive over 90 percent of the time in nine out of ten contexts. The degree of mood variability in each linguistic context, observed within a same generation, appears to greatly vary depending on the investigation.

Furthermore, the obligatory and optional categories greatly diverge depending on whether the speaker belongs to the first or to the second generation. Fernández Pedraza (2014)’s findings are the exception to the rule, as her different generations of Milwaukee Puerto Ricans only differ with respect to three of her eighteen obligatory and optional contexts. Cross-generational inconsistencies are much more common. In Torres (1989), first generation Puerto Ricans in New York City use the Subjunctive over 90 percent of the time in nine contexts, whereas second generation speakers only employ the mood to that extent in four (or five, if the context with 89 percent Subjunctive is included). Bookhamer (2013) confirms this trend over a decade later, by showing that second generation New York Raised Latinos of diverse ethnic origins (Mainlander and Caribbean) have a reduced number of obligatory contexts (5 instead of 9), and a higher number of optional contexts (14 instead of 10), relative to the first generation. In other words, his second generation informants show greater variability than his first generation informants in both types of contexts. In four of the nine contexts deemed obligatory in Bookhamer’s study, his New York Raised second generation speakers draw on the Subjunctive less than 90 percent of the time (meaning, in the author’s view, that they have become optional contexts for these consultants). Finally, third generation East L.A.
Mexicans’ Subjunctive usage is reduced not only in categorical contexts, but also more generally (Ocampo 1990). While still active, employing the Subjunctive in particular contexts seems to become less compulsory, or necessary, for subsequent generations of Latinos in the United States.

The Indicative mood appears to be progressively replacing the Subjunctive mood in certain contexts. Fernández Pedraza (2014) notes an expansion of the Indicative mood in the Protasis Si clause (e.g. Si...., hablo con ella ‘If..., I would talk to her’) among her second generation informants. Lynch (1999) emphasizes the extension of the Indicative in place of the Subjunctive in variable contexts. Bookhamer (2013) notes that his New York Raised informants draw on the Indicative in a greater number of linguistic environments than his Latin American Raised informants. For Montrul (2009), low proficiency heritage speakers produce very few instances of the Subjunctive, and use the Indicative in Subjunctive contexts, as do some intermediate and advanced speakers. The literature makes clear that mood variability (Subjunctive/Indicative) expands to a greater number of contexts with each subsequent generation. This cross-generational change is overwhelmingly associated with the second generation’s unique experience of growing up in a bilingual environment.

4. The bilingual context

In the literature, various theories attempt to explain the observed differences in Subjunctive usage between the different generations of Spanish speakers in the United States. These differences have mostly been attributed to specific circumstances inherent to bilingual contexts, such as language contact (section 4.1), and degree and type of Spanish use (section 4.2). The cross-generational changes are espoused by theories of
simplification (section 4.3), incomplete acquisition (section 4.4), attrition (section 4.5), minimal pairs (section 4.6), and cognitive complexity (section 4.7). Each of these concepts is examined one by one.

4.1 Language contact

Language contact has been invoked to describe the changes in the grammar of the Subjunctive of U.S. raised bilingual speakers. Language contact can either have a direct or an indirect effect on a given linguistic feature. In bilingual contexts, it has been argued that language contact has an indirect effect on bilingual speakers’ treatment of the Subjunctive, meaning that “spaces of variability” in the grammar of a minority language can be affected by contact with the grammar of the majority language (Lynch 1999: 180, cf. Thomason & Kaufman 1988, Romaine 1995, Silva-Corvalán 1994b). However, language contact does not necessarily imply that a particular linguistic feature of the majority, or dominant, grammatical system encroaches upon, interferes with, or is transferred onto the same linguistic feature in the minority language of the bilingual speaker. Interference and transfer refer to the application of linguistic features from one language system to another (Weinreich 1953). With respect to the Subjunctive, Lynch (1999) argues that there are no instances of interference or transfer between bilingual speakers’ two grammatical systems (English and Spanish). Lynch (1999: 179) raises the question of a direct effect in the following way:
Does the evidence presented here support the notion that the lack of subjunctive in the English verb system causes the reduction of subjunctive in the Spanish verb system? In the case of Miami Cuban Spanish, I believe that, for the most part, it does not. The variability of the subjunctive/indicative opposition among successive generations of Cuban Spanish speakers in Miami appears to be linguistically unrelated to direct contact with the English system [...].

Thus, language contact has not been deemed directly responsible for the observed changes in the second generation grammar of the Subjunctive. Lynch (1999) reaches this conclusion by arguing that second generation bilinguals’ reduced Subjunctive usage in speech and in writing is not pervasive. The diminished use that he notices only concerns particular linguistic contexts (or, as the author calls them, discourse semantic contexts)--specifically those that “already demonstrate variability in non-contact varieties of Spanish” (Lynch 1999: 179). In his view, direct language contact with the English system would affect Subjunctive/Indicative variability indiscriminately, in all linguistic contexts alike. Yet, Silva-Corvalán shows that change is not sudden, but rather “gradual and context-selective”, as it does not affect all linguistic environments at the same time (1994: 270).

Furthermore, Otheguy & Zentella (2012), which we refer to henceforth as O&Z, argue that the grammars of the two languages do not have to be identical in order for a contact-induced change to happen. In O&Z’s view, “the [...] paradigms of the two languages” have to be “significant enough for bilinguals to equate them and, consequently, for the possibility to arise that the forms from the paradigm of one language will influence those from that of the other” (p.17). Thus, interference or transfers are not the only outcomes of language contact. Contact-induced change due to pre-existing similarities between two languages (Silva-Corvalán 1994), which Bullock &
Toribio (2004) call *convergence*, can occur as well.

Rather than contact with English, Torres (1989) contends that the reduction in Subjunctive use in certain linguistic environments could be the result of internal language changes. Silva-Corvalán (1994b: 214, cf. Lynch 1999: 180) would agree, as “changes in the minority language [are] already present in the non-contact ancestor variety, i.e. they have intralinguistic roots.” Zentella (1997, cf. Lynch 1999: 29) observes that the imperfect Indicative has a tendency to supplant the Subjunctive in prescriptive contexts on the island of Puerto Rico. Torreblanca (1997: 137, cf. Lynch 1999: 180) summarizes this notion:

> Al igual que lo ocurrido en otras lenguas románicas, en español ha habido una batalla contínua entre el modo indicativo y el subjuntivo en ciertas estructuras sintácticas; ha existido una tendencia popular a la simplificación de la gramática mediante la eliminación del modo subjuntivo… El español hablado en los Estados Unidos es un fiel reflejo de esta tendencia popular⁹…

Silva-Corvalán (1994) agrees that language change is part of an evolutionary change that is not specific to Spanish. It is noticeable in other Romance languages as well.

Furthermore, the researcher (1994: 269) argues that the effect of English is indirect, since “changes occur rather as a result of reduction of both exposure to and use of a complete variety of a subordinate language in contact with a superordinate one.” Language contact, then, has the effect of accelerating this internal change, especially when associated with “reduced access or lack of access to formal varieties of the language or those institutions that maintain conservative and prescriptive language

⁹ ‘Just as in other Romance languages, there has been a constant battle between the Indicative and the Subjunctive moods in certain syntactic structures in Spanish. There has been a popular tendency toward grammatical simplification through the elimination of the Subjunctive mood… The Spanish spoken in the United States is a clear reflection of this popular tendency…’ (My translation)
norms” (Silva-Corvalán 1994: 270). Bilinguals’ degree and type of Spanish use are thus examined in the following section.

### 4.2 Degree and type of Spanish use

Another theory is that change happens because many second generation speakers have not received any sort of formal education in Spanish (Otheguy 2013, based on Menken & Kleyn 2010). Although, as Bookhamer (2013: 9) asserts, “formal education is not required for command of mood use”, the development of this grammatical feature may very well hinge on the second generation’s degree and type of Spanish use. If Spanish is mostly absent from U.S. raised bilinguals’ schooling, it is plausible that these speakers would be more likely to draw on the more prevalent Indicative mood where the Subjunctive is expected. Lynch (1999) attributes the changes in second generation grammar of the Subjunctive to the weakening or absence of monolingual societal norms and the already present mood variability inherent to Spanish. U.S. born bilinguals’ reduced Spanish use and minimized access to formal varieties of Spanish, relative to prior generations, may lead to a form of language change called *simplification*.

### 4.3 Simplification

Silva-Corvalán (1994: 257) defines *simplification* as “a complex process involving the expansion of a form to a larger number of contexts (i.e., generalization) at the expense of a form undergoing simplification, which is used with increasingly lower frequency.” In her view, the Indicative is slowly replacing the Subjunctive in contexts traditionally reserved to the latter, thereby leading to the simplification of her informants’ speech. An earlier study of hers showed that the Subjunctive was undergoing simplification in the second generation, and that the third generation did not use the imperfect Subjunctive at
all. Ocampo (1990)’s conclusion coincides with Silva-Corvalán (1994)’s, in that he observes a decrease in the Subjunctive form in categorical contexts, and more generally.

The reasons for simplification are unclear. Both Torres (1989) and Silva-Corvalán (1994) believe that contact with English could be part of the answer. Indeed, the singularity of the bilingual context has led some researchers (e.g., Montrul 2009) to regard the supposed reduction in bilingual speakers’ Subjunctive usage as an instance of *incomplete acquisition.*

### 4.4 Incomplete acquisition

In Montrul (2009)’s view, U.S. raised bilinguals’ acquisition of the Subjunctive has been disrupted because of their lack of schooling in Spanish. The researcher considers that the second generation’s failure to command the mood represents a case of *incomplete acquisition.* In essence, Montrul (2009: 265) subscribes to the idea that heritage speakers “may have missed the opportunity to develop productive use and written comprehension of more complex structures typically developed during the school-age period. These include subordination and coordination, relative clauses and subordinating conjunctions expressing time, result, purpose, cause and so on, as well as counterfactuals (Nippold 1998). All of these structures call for uses of Subjunctive in Spanish, depending on pragmatic and contextual considerations.” In the researcher’s view, heritage speakers’ reduced use of Spanish at home in addition to their schooling in English (rather than in Spanish) are responsible for their incomplete acquisition of the Subjunctive mood. She refers to her heritage speakers’ high error rates in several tasks to support her view.

The more complex structures that Montrul (2009) mentions correspond to the different linguistic contexts that prompt the occurrence of the Subjunctive mood. In the
literature on the monolingual acquisition of the Spanish Subjunctive, children who were raised in Latin America, and schooled in Spanish, have been found to acquire the mood in stages, rather than in one fell swoop (e.g., Blake 1983; Gili Gaya 1972). It appears that the Subjunctive develops at around age two in the environments (e.g., *querer que* and *esperar que*) that comprise the Volition context (Gili Gaya 1972, Hernández-Pina 1984, López Ornat, Fernández, Gallo, & Mariscal 1994). The Subjunctive is acquired around age seven/eight in the Hypothetical *Como si* context (e.g., *Como si vinieran*) (Gili Gaya 1972), and, much later, at age 10, in the environments found in the Possibility context (e.g., *Es posible que, tal vez/quizá(s), a lo mejor*) (Blake 1983). The approximate age of acquisition for each linguistic context is summarized in Table 2.1. The first column corresponds to the linguistic context; the second to the age of acquisition (which sometimes differs depending on the investigation); and the third column includes the investigations’ references.

---

10 Table 2.1 only presents those linguistic contexts of interest to the present investigation. The labels correspond to the ones used in the current study. A discussion of the Concessive context (e.g. *Aunque quieran*) was not found in the literature, even though this linguistic context appears in the present study. This context could therefore not be included in the table.
As is clear from Table 2.1, complex grammatical features such as the Subjunctive are acquired in stages. The contexts that are acquired later may not be fully mastered by U.S. raised bilingual speakers, according to Montrul (2009), because these speakers are
educated in English and use Spanish less frequently (than Latin American raised
generations) during this crucial developmental stage. Thus, according to this view, U.S.
raised bilinguals would be more likely to draw on the Subjunctive in the Volition context,
which is acquired early on, than in the Possibility context, which is acquired much later
in childhood by monolingual children, and may therefore not have been acquired at all by
bilingual children.

Similarly, Bookhamer (2013) refers to his New York Raised second generation
speakers’ partial development of the mood in particular linguistic contexts as fragmented
mood grammar. Bookhamer rejects the notion of incomplete acquisition for two reasons.
First of all, both the Subjunctive and the Indicative moods are available to these speakers.
Secondly, if there were such a thing as an incomplete grammar, then its counterpart
would have to be a complete grammar. Yet, the researcher discards such a notion by
stating that “Any grammar from which communication is achieved is, in its own respect,
complete” (Bookhamer 2013: 108).

Instead, Bookhamer (2013: 107) argues that a fragmented mood grammar is one
with a higher degree of variability, “wherein some syntactic and semantic contexts would
be under strong command by the speakers, while other contexts would show more
inconsistent alteration between indicative and subjunctive verb forms.” The author
concedes, however, that his fragmented mood grammar hypothesis is only of use when
the grammars of the different generations are compared to each other.

In reality, the theories of incomplete acquisition and fragmented mood grammar
appear almost analogous. If the idea of a complete grammar is questionable, so is the
concept of a fragmented mood grammar. Attrition, on the other hand, is a theory that
posits that grammatical features that were once acquired can recede in the bilingual brain.

4.5 Attrition

As Montrul (2009) points out, the Subjunctive is not used as frequently as the Indicative in general in Spanish. Thus, in her view, it is not surprising that the more infrequent mood would recede. So-called errors could be due to attrition, a phenomenon that Montrul (2014) and others (e.g., Polinsky 2006) describe as the loss of certain grammatical features after they have been fully acquired by adolescence or early adulthood. In order to justify her heritage and native speaker disparities, Montrul (2009) invokes Jakobson (1941)’s Regression Hypothesis, which states that “language acquisition takes place in stages or in a specific order, and linguistic features learned later in said acquisition are the first to undergo loss” (cf. Bookhamer 2013: 15).

Merino (1976, 1983)’s research “was the first which longitudinally documented the attrition of subjunctive verb forms among Spanish-English bilinguals in the U.S.” (Lynch 1999: 27). Merino (1976)’s first study focused on 41 San Francisco Chicano children in grades K-4 in a bilingual program, whose syntax of the Subjunctive underwent attrition over the years, across grades. Through a delayed imitation task, in which a child had to repeat what the experimenter said about a particular picture, Merino found that the fourth graders’ Subjunctive usage was close to that of the kindergartners, even though all of the children had been identified as balanced bilinguals when they entered kindergarten. In a follow-up study two years later, Merino (1983) discovered that the children ($n = 32$)’s “accuracy in the subjunctive forms fell from 70% in the first investigation to 55% in the follow-up study two years later” (Lynch 1999: 28). Merino (1983: 291) explains that:
In the Spanish Subjunctive, changes in performance from one administration to the next were largely due to the items testing control of the dubitative, *tal vez* construction... In the purposive construction, children who had previously produced *El señor saca un libro para que lea* (The man takes out a book for him to read), at the second administration said, *El señor saca un libro para que leer*\(^{11}\).

Merino (1983) claims that her results support the theory which stipulates that features that are acquired last are lost first in the speech of bilinguals born and/or raised in the United States (cf. Lynch 1999: 29).

Merino’s research focuses on attrition in children. Research on attrition in adults suggests that erosion does not affect all aspects of language competence in the same way (Montrul 2009). Syntax and morphology have been found to remain mostly intact (Köpke 2002), while phonetics values (Major 1992), lexical retrieval (Hulsen 2000), and “some referential and discourse-related aspects of language (Gürel 2004; Tsimpli, Sorace, Heycock, & Filiaci 2004)” can undergo loss (Montrul 2009: 241). Montrul (2009: 241) asserts that the difference “between attrition in adults and incomplete acquisition is the range and extent of the loss.” It is her belief that incomplete acquisition, which affects people’s grammatical knowledge in addition to all other areas, has deeper consequences than attrition in adults, which is more confined.

Other theories besides those pertaining to acquisition and loss are called upon to explain generational differences in treatment of the Subjunctive. The high incidence of minimal pairs and cognitive complexity offer a distinct perspective.

4.6 Minimal pairs

Given the greater role played by oral (versus written) Spanish in U.S. raised bilinguals’

---

\(^{11}\) In the first administration, the verb *leer* ‘to read’ is in the Subjunctive mood (*lea*), whereas it is in the infinitive (*leer*) in the second administration.
lives (see section 4.2 of this chapter), Indicative and Subjunctive forms that closely resemble each other may be more difficult to distinguish for these speakers. Zentella (1997: 190-194) argues that *minimal pairs* complicate the distinction in speech between present Indicative and present Subjunctive, as in my example (9), and imperfect Indicative and imperfect Subjunctive, as in my example (10), for bilingual speakers in the United States. In her study, Zentella argues that three of her five bilingual Puerto Ricans in New York City have an *irregular* usage of the present, past, and pluperfect forms of the Subjunctive, meaning, it seems, that they fail to regularly employ them (cf. Lynch 1999: 29).

(9)  Come – Coma

(10) Comían – Comieran

In example (9), the only difference between the present Indicative (*come*) and the present Subjunctive (*coma*) is the ending, where just one vowel (*e-a*) changes. This divergence is not very salient in speech. Similarly, the imperfect Indicative *comian* and the imperfect Subjunctive *comieran* in example (10) are phonologically similar.

The synonymy or lack of synonymy of minimal pairs is not the issue here. Instead, the saliency of distinct forms is significant because, as mentioned in section 4.2, most U.S. raised bilinguals have not received any formal education in Spanish. They therefore have had little chance to notice these differences in writing. And, as noted, depending on speech to grasp the phonological nuances may not always be helpful (Zentella 1997). Yet, phonological barriers may not be the only difficulties that bilingual
speakers face in terms of the Subjunctive. The Indicative/Subjunctive distinction is also cognitively complex.

4.7 Cognitive complexity

As an alternative to incomplete acquisition and attrition, Ocampo (1990) blames cognitive complexity for the cross-generational differences that he observes. Echoing Silva-Corvalán (1986)’s theory, the author suggests that selecting between two forms (Indicative vs. Subjunctive) that are very similar in meaning is not as simple as employing grammaticalized or fixed forms that do not allow for variation. In fact, the latter favor the preservation of the Subjunctive, as mechanic associations are easier to internalize and produce than semantic ones (Ocampo 1990: 45). Ocampo asserts that having to make a tricky choice is the real issue.

The cognitive complexity of mood choice could intimate that the Subjunctive is progressively disappearing from second generation grammar, as proposed by some of the aforementioned studies (e.g., Gutiérrez 1990; Montrul 2009; Ocampo 1990). Yet Silva-Corvalán (1994), who observes the progression of the Indicative on Subjunctive terrain, remains optimistic. Even if a process of simplification is taking place, the Spanish Subjunctive does not seem to be endangered. While not at risk of extinction, the literature on the Subjunctive evidences the fact that different subgroups, shaped by their particular socio-demographic characteristics, tend to employ the mood differently.

5. Subgroups by socio-demographic feature

Section 5 explores the effect of certain socio-demographic characteristics on Subjunctive usage. Additional linguistic features (other than the Subjunctive) are included in the review, as the independent variables Age and Socio-economic status are mostly absent
from investigations on the Spanish Subjunctive, where the generation effect is emphasized instead. Section 5.1 examines how age shapes the occurrence (or lack thereof) of specific linguistic features. Section 5.2 explores the roles played by gender and generation, separately and together. In section 5.3, speakers’ socio-economic status is considered. Each of these socio-demographic characteristics is explored in both monolingual and bilingual settings.

5.1 Age
Age has been found to shape Subjunctive usage in monolingual settings. Lastra & Butragueño (2012) discovered, in their Mexico City study, that the younger the adult informant (the youngest consultants were 20 years old), the greater the proportion of Subjunctive verbs in their speech. Conversely, Gutiérrez (1994) found that younger Mexican informants, this time in Michoacán (Mexico), employed the Indicative mood more than older informants. Among his 25 adult monolingual speakers, those aged 51 and over failed to use Subjunctive verbs in 11 percent of variable contexts, compared to 24 percent for those aged 30 to 50, and 23 percent for speakers who were age 29 or younger (cf. Lynch 1999). Interestingly, these two opposing phenomena--the first illustrating an increase in Subjunctive usage among younger speakers, and the second a decrease and “possible loss of the subjunctive mood in younger speakers” (Gutiérrez 1994: 117, cf. Lynch 1999: 22)--occur in the same country, Mexico, in a non-contact variety of Spanish.

Research conducted on other linguistic features in Latin America supports Gutiérrez (1994)’s findings. Younger informants tend to spearhead sound changes in progress in monolingual settings. For instance, in Buenos Aires, the devoicing of /z/ is
much more prevalent among younger informants (e.g., high school students) than among older adults (Wolf & Jimenez 1979, cf. Labov 1990). It appears that in some monolingual settings, youth tend to be at the forefront of language changes in progress.

In investigations in bilingual contexts, the independent variable Age has hardly been brought forth, despite the fact that Lantolf (1978) discovered an age effect among his New York City Puerto Rican informants. In his investigation, younger speakers were more likely than older speakers to draw on the Indicative in all categories (cf. Lynch 1999: 15). However, it appears that since the 1970s most investigations have replaced the Age variable with the Generation variable. This may be due to the fact that an adult’s age, which is unrelated to an informant’s age of arrival or generation, is not perceived as contributing to theories about bilingual speakers’ distinct treatment of the mood (see section 4 for a discussion of the different theories). On the other hand, age of arrival, which is one of the main features of the independent variable Generation, is overwhelmingly regarded as shaping linguistic variability. The ways Generation and Gender (separately and together) influence Subjunctive usage are explored in the following section. Gender is described in greater detail than Generation, since the latter was already discussed in section 1 of this chapter.

5.2 Gender and Generation

Several studies on the Spanish Subjunctive in Spanish-speaking countries have shown that gender affects mood choice. Lastra & Butragueño (2012)’s analysis of *el corpus sociolingüístico de la ciudad de México* (the sociolinguistic corpus of Mexico City) led the authors to conclude that Mexican women’s Subjunctive rate is roughly 10 percentage points greater than that of men. This result coincides with their belief that women
possess richer verbal skills than men. According to Chambers (1995: 136-137), “it is plausible to speculate that... the neuropsychological verbal advantage of females results in sociolinguistic discrepancies such that women use a larger repertoire of variants and command a wider range of styles than men of the same social groups even though gender roles are similar or identical” (cf. Labov 2001: 276).

Despite the different context, Serrano (1995)’s study of Spanish women’s Subjunctive use corroborates the gender gap. Indeed, women from the Canary Island of la Laguna lead the trend by employing the standard Subjunctive-Conditional combination in the Protasis Si context and in the Apodosis Si context (e.g., Si hiciera calor, saldría ‘If it were hot, I would go out’), instead of the more common traditional or vernacular Indicative-Indicative (e.g., Si hace calor, salgo ‘If it’s hot, I’m going out’) 12. Yet, contrary to Lastra & Butragueño (2012), Serrano (1995) concludes that women tend to employ the standard form more than men (Subjunctive in the Protasis Si and Conditional in the Apodosis Si), because of its prestige. While Serrano (1995) perceives a change in progress, with Canary women’s increased production of the Subjunctive, Lastra & Butragueño (2012) are more cautious. In their view, their study simply confirms that the Subjunctive mood is alive and well in the Spanish of Mexico City.

The relationship between gender and prestige is paramount in Labov (2001)’s renowned principles on language change, which are based on Philadelphia English. In Labov’s view, women of the lowest social classes use stigmatized variables as much as men. However, in what he calls linguistic change from below, “women use higher frequencies of innovative forms than men do” (Labov 2001: 292). Linguistic change from below refers to changes that occur below the level of consciousness. When

12 My examples.
speakers are *not* aware of a language change that is occurring, they tend to employ innovative forms of the *unstable* sociolinguistic variables. For *stable* sociolinguistic variables, when language change is *not* taking place, women show “a lower rate of stigmatized variants and a higher rate of prestige variants than men” (Labov 2001: 266). According to Labov, this can be explained through women’s social role in the speech community.

Eckert (1989), on the other hand, rejects the idea that gender-based linguistic variation is related to prestige. Instead, she concentrates on power dynamics not only between the two genders, but also within a same gender. Eckert (1989: 254) maintains that members of either sex category are unlikely to “evaluate their status in relation to members of the other” sex category. Rather, they are more likely to view themselves in relation to members *within* the same sex category—“by and large, men perceive their social status in relation to other men, whereas women largely perceive their social status in relation to other women” (Eckert 1989: 254). In the author’s view, “men and women compete to establish their social status in different ways, as dictated by the constraints placed on their sex for achieving status” (Eckert 1989: 254). She proposes that differences on the basis of gender should be pursued “within, rather than between, sex groups” (Eckert 1989: 254). Thus, it is Eckert (1989: 256)’s belief that women’s innovative and conservative patterns rest mainly “in their need to assert their membership in all of the communities in which they participate.” In essence, it is women’s authority in their communities that ensures their membership, rather than their power or their linguistic prestige.

Both social and biological theories have been drawn on to account for women’s

---

13 My emphasis.
adherence to prestige norms for stable sociolinguistic variables. Wolfram & Schilling-Estes (1998) argue that women must rely on symbolic capital to make up for their lower economic power (cf. Labov 2001: 275). In Gordon (1997)’s view, women steer clear of stigmatized speech patterns because social conventions connect them to sexual looseness (cf. Labov 2001: 275). Labov discounts Chambers (1995)’s biological theory, introduced above, that attributes female conformity to greater verbal abilities, as this notion predicts that women are more attentive to language than men. This turns out not to be the case, as “women over-report their use of the prestige norm much more than men” (Labov 2001: 277). Furthermore, Labov (2001: 263) argues that “gender is a social factor—language is not differentiated by the biological aspects of sex differences.” Instead, the sociolinguist proposes a social theory akin to that of Wolfram & Schilling-Estes (1998). To him, women’s linguistic conformity reveals their desire for upward mobility. As the primary caregivers, women feel responsible for their children’s future.

As language learning tends to happen more with mothers (or other female caregivers) than with fathers in the early stages of development, it is not surprising that “the vernacular that we speak, the first language that we have mastered perfectly and use without doubt or hesitation—is our mother’s vernacular” (Labov 2001: 306-307).

Potowski (2008)’s study of Chicago children of mixed Puerto Rican and Mexican parentage (which she labels MXPR) is a case in point. Although her MexiRican informants were raised by both their parents, and were therefore in contact with both the Mexican and Puerto Rican dialects, 20 out of 27 of them use their mother’s variety of Spanish. Instead of a ‘mother tongue’, she speaks of a ‘mother dialect’ (Potowski 2008: 217). Because women tend to be more engaged in child rearing than men, they are at
some stages a full generation ahead of men, according to Gauchat and Wolf & Jiménez (cf. Labov 2001). Labov asserts that “the logical inference is difficult to avoid: men are at the level of linguistic change characteristic of their mothers because they acquired their first use of these variables from their mothers” (p. 306-307).

While women in the monolingual Spanish contexts seem to employ the Subjunctive more than men (as discussed above), they appear to use it less than men in bilingual environments. The only two bilingual studies (to the best of my knowledge) that associate gender and Subjunctive usage obtain differing results. Although Fernández Pedraza (2014)’s study of bilingual speakers in Milwaukee does not establish any correlation between gender and Subjunctive usage, Bookhamer (2013)’s investigation lends support to the significant role that gender plays in shaping Subjunctive rates. Bookhamer (2013: 68) finds that in the Latin American Raised Newcomers Mainlander group, men use the Subjunctive more frequently than women in obligatory contexts, and that in the New York Raised Mainlander group, men employ the Subjunctive more than women in optional contexts. Thus, in both the first and second generations, women draw on the Subjunctive mood less than men.

Although still scarce, there are also signs of a gender effect in research on other linguistic features in bilingual contexts. Orozco (2007) established that women’s distinct expression of futurity and possession in Spanish, through the increased use of a periphrastic construction, indicate that they are leading an ongoing language change (cf. Shin & Otheguy 2013). Alfaraz (2010)’s study in a bilingual environment in Lansing, Michigan also points to a female-led language change, with women increasingly using the “copula estar in contexts where the ser copula is expected” (Shin & Otheguy 2013:}
Finally, Shin & Otheguy (2013) find that immigrant first generation women employ more Spanish subject pronouns than first generation men. In bilingual settings where other languages are spoken, women have been found to spearhead changes as well. Van Ness (1995) showed that young Amish women in Ohio increasingly use the neuter pronominal form *es* instead of the feminine form *sie* in Pennsylvania German (cf. Shin & Otheguy 2013). In all of these studies located in bilingual settings, women are the ones to lead language change.

Shin & Otheguy (2013) posit that the women effect they find can either be explained by first generation women’s extensive contact with US-born bilinguals or by women’s greater susceptibility to direct external influences. In the first case, the authors maintain, “this mother-to-child dialect transmission is a strong indicator that U.S.-born Latinos have more conversations in Spanish with their mothers than with their fathers” (Shin & Otheguy 2013: 447). In the second case, they feel that “there is reason to believe that women tend to adapt and change their use of language more readily” (Shin & Otheguy 2013: 447). They support this idea by citing research findings that show that women establish interpersonal sensitivity through language, more than men (Hall & Mast 2008; Leaper & Friedman 2007; Leaper & Robnett 2011, cf. Shin & Otheguy 2013), “a tendency that might well translate into an increased susceptibility to change” (Shin & Otheguy 2013: 447). Moreover, Shin & Otheguy (2013: 447) affirm that “there is evidence that women more than men use language to demarcate social differences among themselves (Eckert & McConnell-Ginet 2003: 302), indicating that women have a heightened sensitivity to linguistic form and a tendency to change their own speech.” The authors advise that both factors (women’s increased contact with U.S.-born
bilinguals and their greater openness to change) be taken into consideration when accounting for the women effect in bilingual contexts.

Finally, Shin & Otheguy (2013: 447) conclude their investigation by offering the following recommendations:

Both Labov (2001) and Eckert & McConnell-Ginet (2003) explain women’s innovative use of language in monolingual settings as a way of indexing nonconformity. But projecting a nonconformist social identity is not a likely explanation in bilingual settings, where language contact itself is a primary source of change. To gain a better understanding of why women are so important in advancing language change in BOTH monolingual and bilingual settings, future research should examine women’s roles in various social networks, as well as the possibility that women are more susceptible than men to external influences on speech patterns.

Social networks may influence more than men and women’s linguistic differences. The association between social networks and speakers’ socio-economic status has also been found to shape linguistic variability.

5.3 Socio-economic status

Besides age and gender, Lastra & Butrageño (2012)’s study points to the fact that speakers’ level of education plays a role in their greater Subjunctive rates. They find that the higher the level of education, the greater their Mexico City informants’ rates of Subjunctive.

Bookhamer (2013) found an education effect in just one specific instance in his study on the Spanish Subjunctive in New York City. His New York Raised consultants with a secondary school level of education produced a greater number of Indicative verbs in the Apodosis Si clause (e.g., Si..., volveré a casa) than his New York Raised informants who were college educated. The lower the level of education, the higher the rates of Indicative in this specific linguistic context in second generation speech (only).
Bookhamer (2013) did not find any effect for socio-economic status or social class\textsuperscript{14}. This is interesting because although socio-economic status has rarely been explored in the context of the Spanish Subjunctive, the socio-demographic variable has been found to affect the variability of other linguistic features.

According to Labov (1990), prestige markers are correlated with speakers’ high socio-economic status; “the higher a speaker’s socioeconomic status, the higher the frequency of use. For stigmatized markers, the reverse is true” (Labov 1990: 220). Labov has found that the upper working and lower middle classes (the intermediate groups) tend to employ a new form more than upper middle and lower working classes (the extreme groups).

Labov (1990)’s conclusions, based on studies set in monolingual contexts, contrast with the findings of the few investigations (that include socio-economic status) located in bilingual settings. Orozco (2007) found that New York City Colombians with high socio-economic status were the ones to lead the change with respect to the expression of futurity. Shin & Otheguy (2013)’s study of personal subject pronouns supports these trends. In their investigation, the most affluent Latinos are the innovators, while the conservative speakers are the poorest. According to the authors (2013: 431), “persistence of ancestral Latin American usage patterns in Spanish in New York is strongly associated with lower socioeconomic status, while the more Anglicized usage is connected to the more affluent.” They attribute the difference, spurred by speakers’ socio-economic status, to the denser social networks that have been found to “correlate with more conservative patterns of language use” (Milroy 1987; Milroy & Milroy 1992, 14)

\textsuperscript{14} In Bookhamer (2013), the variable Social class has three factors (high, middle and low). Bookhamer’s Socio-economic status (SES) variable is the same as the one used in the present study (see Chapter 3 for details). SES merges the informants’ level of education with their occupation (social class).
cf. Shin & Otheguy 2013: 431). Indeed, the more affluent Latino national groups “are more likely to have looser social networks and be therefore more susceptible to external influence and change” (Shin & Otheguy 2013: 431).

The denser social networks that people of lower socio-economic status tend to have also engender a greater retention of minority languages (Bills, Hudson, & Hernández Chávez 2000; Hudson, Hernández Chávez, & Bills 1995, cf. Shin & Otheguy 2013: 431). The increased retention and the more conservative language patterns found among the lower social classes could also be ascribed to the fact that recent immigrants from Latin America (i.e. the Newcomers) tend to be poor (Shin & Otheguy 2013). Shin & Otheguy (2013)’s most innovating informants are Colombians and Cubans, two groups that are wealthier on average than other Hispanic communities. “Affluence”, the researchers contend, “tends to be correlated with more interaction with the dominant Anglophone communities, thereby decreasing connection to the Spanish-speaking communities and very likely increasing bilingualism, resulting in greater susceptibility to change in Spanish usage patterns” (Shin & Otheguy 2013: 443).

In their view, the extent to which speech changes hinges more on “the overall socioeconomic make-up of a community and the strength of its social network” than “each individual’s place within the community” (Shin & Otheguy 2013: 443). They ascertain that “with respect to social class, patterns of change in bilingual communities diverge from those in monolingual ones” (Shin & Otheguy 2013: 448). As is clear from the research that considers social class, the effect of speakers’ socio-economic status varies in monolingual and bilingual settings (Shin & Otheguy 2013). In monolingual contexts, the working class usually initiates and spreads the changes found in speech. In
bilingual contexts, on the other hand, wealthier communities tend to spearhead language changes.
Chapter 3: Methodology

1. Introduction

This chapter describes the methods used in the present study. Section 2 describes the informants and the interviews. Section 3 introduces the concept of the linguistic context and discusses questions related to frequency. Sections 4, 5 and 6 describe the study’s envelope of variation, dependent variables, and independent variables, respectively. Finally, the data collection process is presented in detail in section 7. This includes details of the automated procedure using the Java program that was utilized to gather and tag the data, as well as an overview of the statistical methods employed.

2. Informants and interviews

The current study is based on the Otheguy-Zentella Corpus of Spoken Spanish in New York (henceforth OZC). Three hundred sociolinguistic interviews were conducted between the years 2000 and 2005. A stratified subset of these interviews, totaling 142, was selected for the corpus. O&Z (2012) describe a stratified sample of a population (in this case, Latino New Yorkers) as containing balanced proportions of participants from “several key sectors, or strata of the relevant population” (p. 26). The stratification criteria used to create the OZC:

- Gender
- National origin
- Areal origin
- Age of arrival

---

15 The Otheguy-Zentella Corpus of Spoken Spanish in New York was developed with funding from the National Science Foundation, grant number 0004133.
• Years in NYC
• Social class
• Years of education
• English skills
• Amount of Spanish use

The OZC contains over 300 hours of recorded conversations lasting approximately one hour each\textsuperscript{16}, almost always conducted by an interviewer whose Latin American national origin was the same as that of the participant (O&Z 2012: 22). The interviews were conducted entirely in Spanish, but they contain English loanwords and code-switches, which is to be expected in Spanish in New York.

A snowballing technique (Oppenheim 1992: 43) was used to attract informants of the OZC, starting with the friends, families and acquaintances of the research team (Varra 2013). All of the informants belong to six ethnonational groups: Colombian, Cuban, Dominican, Ecuadorian, Mexican, and Puerto Rican. Each ethnonational group was assigned a letter (C, U, D, E, M and P, respectively) to enable informant identification. Each letter follows the informant’s identification number (e.g., 005U corresponds to a Cuban participant).

Labovian sociolinguistic techniques (Labov 1963, 1966) were used to compile the OZC. “To encourage a relaxed and informal speech style and to minimize the amount of mutual accommodation of the speech of the interlocutors, several strategies were used (O&Z 2012: 25, 40). First, the majority of interviews were conducted in the home of the informant. Second, interviewers were native Spanish speakers of the same national

\textsuperscript{16} According to Varra (2013: 42), “While interviews generally took at least an hour, the flow of conversations varied. Some interviews consisted almost entirely of informant talk; some demonstrated a pattern of regular turn-taking between informant and interviewer.”
origin as the informant. Occasionally, two or three informants participated in a single interview” (Varra 2013: 41). Finally, even though informants were given the liberty to express themselves, similar topics were introduced so as to ensure comparability between interviews. The interview questions encouraged informants to share their opinions and personal stories, such as “their arrival in the U.S., living in New York City, memories of life or experiences in the home country, school, bilingual education in New York City, work, family, then-popular musicians (such as Michael Jackson, Ricky Martin and Menudo), personal relationships, U.S. politics, religion and the scariest or most memorable experiences of their life” (Varra 2013: 41). After each interview, a questionnaire (in Spanish) was administered to each participant in order to obtain language and personal background data. The questionnaire is not considered a part of the OZC.

While the corpus was designed for the purpose of research on Spanish subject pronouns by O&Z, many other researchers, the authors’ former graduate students, have exploited it for their own investigations on different linguistic features (e.g., variable coda /s/ production, subject pronoun placement, word order and intonation, lexical borrowing, etc.). The current investigation is the second study on the Spanish Subjunctive in New York City based on the corpus; but whereas the first one made use of a portion of the corpus (52 informants in Bookhamer 2013), the present study makes use of the entire OZC. This is just one of a number of differences between the current investigation and all prior studies on the Spanish Subjunctive in the United States. Indeed, one of the present study’s unique contributions is its very large database of 6576

---

17 As noted, the OZC includes 142 interviews in the present investigation (two more than O&Z’s 140 interviews exploited for the purpose of their 2012 study on Spanish subject pronouns).
Subjunctive and Indicative verbs present in nine different linguistic contexts of occurrence for Indicatives and Subjunctives.

3. Linguistic contexts

3.1 Concept of linguistic context

As discussed in Chapter 2, nearly all studies on the Spanish Subjunctive in the United States comprise a variety of linguistic contexts, but their label, number, and type vary depending on the investigation. What does the term linguistic context refer to here? A linguistic context is the “discourse that surrounds a language unit and helps to determine its interpretation” (Dictionary, Encyclopedia and Thesaurus - The Free Dictionary). In the case of the present study, the language unit is the finite verb that is conjugated in the Subjunctive mood. The Subjunctive verb is often found in a subordinate or dependent clause, as vengas in quiero que vengas 'I want you to come'. The discourse that surrounds the Subjunctive verb vengas, which is quiero que, helps to determine its interpretation. In the present study, querer que 'to want to' is an example of what the present study calls an environment. Sometimes environments are grouped together to form a linguistic context. For example, the linguistic context dubbed the Volition context comprises two environments: querer que 'to want to' and esperar que 'to hope to, to wish to'. These two environments have been grouped together under the same context because analysts see them as conveying similar messages. With the help of the linguistic elements in these two environments, speakers can express hope and longing or aspiration.

A separate exploration of each individual environment was attempted in the present study’s initial stages. For example, rather than grouping the two environments that comprise the Volition context together, each environment was examined separately.
(i.e. in the initial stage of the study, querer que and esperar que were treated as different linguistic contexts\textsuperscript{18}). The widespread view in the literature that environments that are grouped together express similar messages has not been tested. The current study thus preferred to take a more rigorous approach and start by studying each environment separately. However, the effort to explore individual environments (e.g., querer que) rather than linguistic contexts (e.g., the Volition context) failed for several reasons. First, it became quickly clear that the small numbers involved in the analysis of individual environments rendered the inquiry statistically unrealizable. Second, even if the task had been statistically viable, a bigger picture emerges when grouping several environments into one linguistic context. For these reasons, the present study follows prior studies’ unverified position that comparable environments can be usefully grouped under the same linguistic contexts.

Some linguistic contexts comprise only one environment. For example, the Protasis Si context contains a single environment, namely the clause introduced by Si ‘if’. These linguistic contexts consisting of a single environment are given the label linguistic context rather than the label environment for ease of presentation, even though in these cases the two labels are interchangeable. A highly frequent environment like the Si clause can be counted in the statistical analysis as its own context precisely because it appears in the speech of a great number of informants, as will be discussed in greater detail in the following section. This situation is rather unique, however. The great majority of environments fail to provide viable statistical material when taken on their

\textsuperscript{18} The literature (e.g., Bookhamer 2013) includes the matrix clause desear que in the Volition context, but this environment was not found in the OZC. Thus, it is not included in the description of the Volition context in the present study.
The nine linguistic contexts included in the current investigation are listed below (see Appendix A for glosses):

- Volition context (e.g. Quiere que vengan, espera que vengan)
- Temporal context (e.g. Hasta que vengan, antes de que vengan, cuando vengan)
- Protasis Si context (e.g. Si vinieran, si hubieran venido)
- Hypothetical Como si context (e.g. Como si vinieran)
- Apodosis Si context (e.g. Si..., quisiera ir, si..., hubiera querido ir)
- Possibility context (e.g. Es posible que quieran, tal vez/quizá(s) quieran, a lo mejor quieran)
- Concessive context (e.g. Aunque quieran)
- Uncertainty context (e.g. No creo que quieran, no sé si quieran)
- Modal context (e.g. Como quieran, lo que quieran, como que quieran)

As is evident from the list, five linguistic contexts comprise more than one environment, while four linguistic contexts (the Protasis Si context, the Hypothetical Como si context, the Apodosis Si context, and the Concessive context) only include one environment. It will become clear further along (in sections 4 and 5) that some of these single environment linguistic contexts do not represent viable contenders for certain statistical analyses in the present investigation.

### 3.2 Reconsidering obligatory and optional linguistic contexts

The literature on the Spanish Subjunctive in the United States argues that Subjunctive verbs occur to different degrees in distinct syntactic/semantic/pragmatic contexts. The variability in Subjunctive rates is largely attributed to the linguistic context in which the

---

own, thus the decision of the present study to explore mood variability in broad contexts rather than in narrower individual environments.

The nine linguistic contexts included in the current investigation are listed below (see Appendix A for glosses):

- Volition context (e.g. Quiere que vengan, espera que vengan)
- Temporal context (e.g. Hasta que vengan, antes de que vengan, cuando vengan)
- Protasis Si context (e.g. Si vinieran, si hubieran venido)
- Hypothetical Como si context (e.g. Como si vinieran)
- Apodosis Si context (e.g. Si..., quisiera ir, si..., hubiera querido ir)
- Possibility context (e.g. Es posible que quieran, tal vez/quizá(s) quieran, a lo mejor quieran)
- Concessive context (e.g. Aunque quieran)
- Uncertainty context (e.g. No creo que quieran, no sé si quieran)
- Modal context (e.g. Como quieran, lo que quieran, como que quieran)

As is evident from the list, five linguistic contexts comprise more than one environment, while four linguistic contexts (the Protasis Si context, the Hypothetical Como si context, the Apodosis Si context, and the Concessive context) only include one environment. It will become clear further along (in sections 4 and 5) that some of these single environment linguistic contexts do not represent viable contenders for certain statistical analyses in the present investigation.

### 3.2 Reconsidering obligatory and optional linguistic contexts

The literature on the Spanish Subjunctive in the United States argues that Subjunctive verbs occur to different degrees in distinct syntactic/semantic/pragmatic contexts. The variability in Subjunctive rates is largely attributed to the linguistic context in which the
verb appears. These contexts are often split into two categories: obligatory contexts (also called categorical) and optional contexts (also labeled variable). While the allocation of contexts to one or the other category varies depending on the investigation, the distinction between obligatory and optional remains remarkably similar throughout the literature. Under this view, certain linguistic contexts require the use of the Subjunctive (almost) invariably. This notion, based on long-standing rules found in traditional grammars, holds that the use of a mood other than the Subjunctive in these obligatory linguistic contexts renders the sentence ungrammatical (Montrul 2009). For example, were an informant to draw on the Indicative mood in the obligatory context of Volition, saying *quiero que vengas* instead of *quiero que vienes*, this informant’s grammar of the Subjunctive would be regarded either as incomplete (Montrul 2009; Polinsky 2006), attrited (Montrul 2009; Tsimpi, Sorace, Heycock, & Filiaci 2004), or simplified (Silva-Corvalán 1995). These notions have been discussed in Chapter 2.

Conversely, moods other than the Subjunctive are accepted in the literature in linguistic contexts that are deemed optional (e.g., Bookhamer 2013; Fernandez Pedraza 2014; Gutiérrez 1994; Lynch 1999; Ocampo 1990; Silva-Corvalán 1994). Saying *aunque vinieras* 'even if you came' with a Subjunctive is considered as correct as saying *aunque vienes* using an Indicative. In this common line of analysis, contexts such as the Concessive, which are considered optional, allow for variability because of the differing messages conveyed by the two moods. This theory thus stipulates that *aunque vinieras* and *aunque vienes* do not have the same meaning, but lead to the same message. Mood variability in these optional contexts is associated with notions of futurity, relevance, argument and (ir)realspeness (Lynch 1999), as well as with degrees of assertiveness (Silva-
Corvalán 1994b). In contrast to the obligatory contexts, then, the Indicative in these optional contexts is not considered ungrammatical. Chapter 2 offers a full account of these theories.

Another layer of complexity is added in the literature when informants’ immigrant generation is included in the analysis (e.g., Bookhamer 2013; Fernandez Pedraza 2014; Gutiérrez 1994; Lynch 1999; Ocampo 1990; Silva-Corvalán 1994). These studies generally find that first-generation, Latin American-raised immigrants have a significantly greater tendency to employ the Subjunctive in both types of contexts than second-generation U.S.-raised speakers. The alleged errors or ungrammaticality involving the use of other moods in obligatory Subjunctive context, which are usually only studied in the speech of second generation consultants, are always ascribed to an informant’s immigrant generation, usually defined in terms of place of birth and, in some studies, age of arrival and number of years in the United States.

The classification of contexts into obligatory and optional persists in the literature even in the rare instances where the distinction is not based on traditional grammar or native speaker intuition (usually the researcher’s). Bookhamer (2013) offers an unusual interpretation of the linguistic context dichotomy that relies on the use of the Subjunctive as found in first generation Newcomers. In his study, a context is classified as obligatory if 90 percent of verbs appear in the Subjunctive in the speech of Newcomers (e.g., the Volition context introduced by querer que and esperar que is obligatory for Bookhamer because over 90 percent of verbs in this context appear with a Subjunctive). Conversely, the contexts where less than 90 percent of the Newcomers’ verbs are in the Subjunctive
are classified as optional (e.g., the Concessive context, introduced by aunque). Bookhamer (2013) does not consider usage of the Indicative, even in his obligatory contexts, and even by first generation informants, as erroneous or ungrammatical; instead, the presence of the two moods is simply viewed as variation.

It is clear that the binary categorizations based on traditional rules or intuitions and the ones based on usage (that is, on what speakers actually say) have distinct theoretical foundations. Yet, both approaches are challenged in the present study. The idea that there are obligatory or categorical contexts on the one hand, and optional or variable contexts on the other, is shown here to be problematic. In essence, the binary division hinges on the idea that Subjunctive and Indicative verbs always have the same meaning (and express similar messages) in the obligatory contexts while they never do in the variable ones. But this characterization should be questioned. First of all, it is not clear that the two variants in a so-called obligatory context such as quiero que, with the supposedly required Subjunctive vengas and the presumably erroneous Indicative vienes, have the same meaning. This claim would have to be tested. Indeed, “as with any other unobservable, one must be reminded, once again, that a meaning is a hypothesis that has to be treated as testable and expected to yield testable predictions” (Otheguy 2018: 38). Similarly, it is unclear whether two variants (e.g., vinieras and vienes) in a so-called optional context, such as aunque vinieras (Subjunctive) and aunque vienes (Indicative) in the Concessive context, have distinct meanings, as is implied in the literature. The focus

---

19 Although Bookhamer (2013)’s division of contexts into obligatory and optional can be considered usage-based, knowing the number of first generation Newcomers in whose speech each of these so-called obligatory contexts occurs would offer a more reliable picture of these speakers’ treatment of the Subjunctive. Indeed, it could very well be that the obligatory contexts are available (i.e. present) in the speech of only a portion of his Newcomer consultants. If this were the case (i.e. only a small number of these informants employ the Subjunctive over 90 percent of the time), there would be fewer grounds to designate obligatory contexts as such. Essentially, the division would be futile.
on errors or grammaticality rather than on variability in most of the literature seems misguided, as does the allocation of linguistic contexts to one category or the other, even when, as in Bookhamer, this allocation is based on usage and the two moods are seen as variants free of prescriptive judgment.

The present study attests to the fact that contexts that have been classified in the literature as obligatory as well as contexts that have been classified as optional can all present instances of Subjunctive/Indicative variability. For example, see in Group I below the Hypothetical *Como si* context (examples 11a, 11b, 12a, 12b). See also in Group II the Possibility context (examples 13a, 13b) and the Modal context (examples 14a, 14b). Moreover, as shown in these examples, mood variability is observed in the speech of both second and first generation informants. The following examples from the OZC, produced by first-generation, Latin American Raised Newcomer consultants (henceforth labeled Newcomers or LARN) and second-generation New York Raised (NYR) consultants, illustrate these two claims:
I. Variability in a so-called obligatory linguistic context, Newcomer and NYR speech

(11a) Yo sentí como si [mi] propia madre se hubiera muerto - LARN 422P
    ‘I felt as if [my] own mother had died’

(11b) Es como si mi trabajo es entrevistar al deportista - LARN 323E
    ‘It’s as if my work is to interview the athlete’

(12a) Dices tú, “como si estuvieras hablando con tu amiga” - NYR 340M
    ‘You say, “as if you were talking to your friend”’

(12b) Es como si uno estaba en un sitio hispano - NYR 401P
    ‘It’s as if one was in a Hispanic place’

II. Variability in two so-called optional linguistic contexts, Newcomer and NYR speech

(13a) A lo mejor, pueda ser que algún pollo salte por ahí ¿no? - LARN 316E
    ‘Maybe, it could be that a chicken jumps right around there, don’t you think?’

(13b) Nos quedamos así como ah!!! a lo mejor no limpia… - LARN 351M
    ‘We were like ah!!! maybe s/he/it/you doesn’t/don’t clean…’

(14a) Entonces me dejaban ir pero como quiera - NYR 331.2D
    ‘So they were letting me go but like I want to’

(14b) Y mi mamá como estaba bajo el mando de mi papá - NYR 315M
    ‘And my mom since she followed my dad’s orders’

In Group I, the Newcomer consultants and the New York Raised consultants sometimes draw on the Subjunctive, and sometimes employ the Indicative in a linguistic context that

---

20 As mentioned in Chapter 1 and in section 2 of this chapter, all examples followed by a number and letter are from the OZC.
has been classified as obligatory in the literature (Bookhamer 2013; Lynch 1999, among others). This is evident in the Hypothetical Como si context in examples (11a), (11b), (12a) and (12b). Unsurprisingly, the same is true of two optional linguistic contexts, as in the Possibility context introduced by a lo mejor in examples (13a) and (13b), and the Modal context, introduced by como in examples (14a) and (14b). The fact that a Newcomer uses an Indicative verb in a context that allegedly requires the Subjunctive, as in (11b), is not discussed in much of the literature, which tends to scrutinize only the grammar of second generation informants. Yet, example (11b) is not unique. Subjunctive/Indicative variability is regularly found in so-called obligatory contexts even in the grammar of first generation immigrants.

The idea that linguistic contexts should not be split into these two categories is related to a larger debate in the field of sociolinguistics. At issue is the question of whether (two) morphological/syntactic variants have to be synonymous (in a particular context of variation) in order to be considered variants (of a same variable). It is the supposed synonymy of Subjunctive and Indicative variants (e.g. quiero que vengas/vienes) in obligatory contexts that leads some linguists (e.g., Montrul) to view speakers’ usage of the Indicative (e.g., quiero que vienes) as ungrammatical. In effect, since the Indicative and Subjunctive are not regarded as bearing different meanings, usage of the Indicative is not portrayed as intentional, or as a viable choice. In this case, it is only possible for the speaker to express one message, and it must be expressed through the usage of a Subjunctive verb. This approach appears inconsistent, since these analysts appear happy to regard the two different meanings of the forms and the two different resulting messages as acceptable in optional contexts. In those contexts,
speakers are viewed as having a choice, and as deliberately opting for a particular mood, a privilege then denied in the so-called obligatory contexts.

To address the larger sociolinguistic question of what constitutes a variable context, the present study abides by Otheguy’s view that “we should not have variation on one side and differences of meaning on the other. When meanings are properly understood, we have both variation and different meanings, and different messages too, but for the rough superficial referential equivalence” (Otheguy 2018: 41). Otheguy asserts that variable contexts are contexts “where two forms constitute two viable expressive choices, contexts where the user of the language finds, systematically, that two forms can be usefully deployed (to produce subtle differences in the message or to attain other important communicative goals)” (Otheguy 2018: 41). The last part of the quote is key, in my view, since even when the two verb forms are found in two roughly synonymous expressions (as, possibly, quiero que vengas/vienes, and/or aunque vengas/vienes), they should be considered variable contexts, just as they would if they were phonological features (e.g., coda /s/ in variation with aspiration or absence). Perhaps the speaker who uses the Indicative vienes in quiero que vienes is signaling a particular aspect of their identity (like the fact that they feel more integrated into U.S. society, for example), even if they're not necessarily expressing a different message from the one expressed if they had said quiero que vengas (Subjunctive).

For all of these reasons, this investigation’s nine linguistic contexts are not split into the two different categories of optional and obligatory. They are examined as a single group of contexts. The hypothesis is that variants (Subjunctive and Indicative verbs) can be found in synonymous expressions, in any or all of these linguistic contexts.
What matters in the present investigation, and which is something that has never been examined to date, is how popular each linguistic context is, in the speech of all (142) informants irrespective of their generation. In other words, the analysis includes an exploration of the number of informants in whose speech each linguistic context occurs. This question will be examined in the following section.

3.3 Linguistic context occurrence

One of this investigation’s most interesting questions is one that has not been addressed in prior studies on the Spanish Subjunctive, to my knowledge. It concerns the number of informants in whose speech each linguistic context occurs. It is very plausible that speakers do not make use of all linguistic contexts in the same way or to the same extent. In other words, some linguistic contexts are surely more popular, and others less popular. Furthermore, it seems likely that a linguistic context’s popularity would be significantly associated with speakers’ Subjunctive rates, since linguistic contexts were selected based on their propensity to give rise to the mood.

Table 3.1 below lists the nine linguistic contexts, and shows, in descending order, the number (N) and percentage (Pct.) of informants (out of a total of 142) where each is found.
Table 3.1 shows great disparities in the occurrence of particular linguistic contexts in the speech of the 142 informants. It is evident that the nine linguistic contexts occur to different degrees in the informants’ speech. For instance, the Modal context is present in the speech of 141 informants (or 99.3 percent), while the Uncertainty context is found in only 12 informants (or 8.5 percent). The table shows that there is a 90.8 percentage-point difference between the linguistic context with the greatest number of informants and the one present in the speech of the fewest number of consultants. It turns out that the hypothesis that different linguistic contexts are not found with equal frequency in the different informants is quite accurate. Some linguistic contexts are indeed more popular than others, a fact that is likely to affect informants’ rates of Subjunctive.

4. **Envelope of variation**

The current study, like all variationist research, includes an *envelope of variation*. The envelope of variation (Labov 1972; Tagliamonte 2006) comprises “...the items that are legitimate candidates for coding and statistical treatment in the study of a linguistic
variable” (O&Z 2012: 23). The envelope of variation delimits the cases where the analyst finds instances of a study's dependent variable, that is, instances of the variants under study. The present study includes two dependent variables. The first variable is Mood choice, consisting of the variants Subjunctive and Other than Subjunctive. As we proceed with the statistical treatment the name of this variable will be changed to Subjunctive rate. To be noted now is that not all moods and not all linguistic contexts are included in the study of Mood choice; details are given below. The second variable of the study is Linguistic context availability. It will be discussed in section 5.

4.1 Mood choice

In the study of our first dependent variable, Mood choice (later called Subjunctive rate), the envelope of variation would ideally comprise all of the verbs in our nine linguistic contexts. In the early stages of this investigation, all of the verbs that occurred in the nine contexts were collected for the study. Three moods emerged in these contexts: the Subjunctive, the Indicative, and the Conditional. Initially, Conditional verbs were included in the envelope of variation along with Subjunctives and Indicatives. This is because, as Bookhamer (2013: 40) confirms, “most grammarians agree that the conditional is an indicative verb form”. However, as Bookhamer (2013: 40) skillfully demonstrates, Conditionals and Indicatives can serve different functions. Compare examples (15) and (16).
The two verbs, *sería* in example (16) and *tuvieron* in example (15) are in hypothetical sentences (in Apodosis *Si* clauses). Yet, Bookhamer (2013: 40) points out that there “is a clear difference between the use of the conditional *sería* […] and the indicative […] *tuvieron*” as “the use of the conditional in the apodosis has a distinct function when compared to the use of the indicative in the same context. That is, if we were to count the two as indicatives, we would be unable to differentiate between the use of indicative *tuvieron* and the more conventional conditional form *sería*”. Consequently, if Conditionals were to be included in the envelope of variation, the study would actually include three moods, namely the Subjunctive, the Indicative (without Conditional verbs), and the Conditional. The inclusion of the Conditional is of definite interest for a future study that would build on the results of the present one. But for now, it is more judicious to proceed step by step, and to begin with a simpler binary mood choice variable as we will do in the present study.

The following examples illustrate the types of verbs that are included in and excluded from the envelope of variation. In examples (17) to (21), the bold and underlined verbs are included in the envelope of variation, while the verbs in italics are not. An explanation follows the examples.
(17) Si él _estuviese_ aquí y las cosas _fueran_… - LARN 381D

‘If he were here and things were…’

(18) Es lo que él _dice_ - LARN 381D

‘It is what he says’

(19) Que Dios _haga_ que él _pueda_ venir antes de yo ir - LARN 381D

‘God willing he will be able to come before I go’

(20) Bueno _hay_ dos cosas que _son_…, una _es_ el idioma, y otra _es_ - LARN 381D

‘Well there are two things that are…, one is language, and the other is’

(21) _Diría_ yo, algo así, pero nada - LARN 381D

‘I would say, something like that, but nothing/never mind’

In example (17), the two verbs _estuviese_ and _fueran_ are inside the envelope of variation because they are in the Subjunctive and occur in one of this study’s nine linguistic contexts (the Protasis _Si_ context). The Indicative _dice_, introduced by _lo que_ in example (18), appears in a Modal context, a linguistic context that is also included in the present study. On the other hand, the verbs in italics in examples (19), (20) and (21) are outside of the envelope of variation, because these examples are found in excluded linguistic contexts. In addition, the verb _diría_ is in the Conditional, which, as mentioned above, is excluded from the present investigation.

Following Bookhamer (2013: 40), “regarding the process involved in determining
the inclusion and exclusion of a token and the contexts in which to situate it”, the current investigation complies with the principle of accountability (Bayley 2002; Labov 1982; Tagliamonte 2006, cited in O&Z 2012: 47), “which states that the collection must be consistent, legitimate and as neutral as possible. In other words, the researcher is required to consider every instance of the token under study, irrespective of its abnormality or unconventional usage.”

4.2 Linguistic contexts included in the analysis of Mood choice

The envelope of variation for Mood choice²¹ comprises 6576 Subjunctive and Indicative verbs in nine linguistic contexts. (The nine contexts are listed in section 3.1 of this chapter, and glosses are provided in Appendix A). Subjunctive/Indicative variability is examined twice, at two different levels. At a macro level, we study all nine linguistic contexts taken together. We then zero in at the micro level on the four most popular contexts and study each of them separately. The four most popular contexts are those found in at least 50 percent of informants, namely Modal contexts, found in 141 participants (99.3 percent of the sample), Protasis Si contexts, found in 140 participants (98.6 percent of the sample), Temporal contexts, found in 131 informants (92.3 percent of the sample), and Apodosis Si contexts, found in 129 informants (90.8 percent of the sample). The five remaining linguistic contexts each appear in less than 50 percent of the sample (see Table 3.1 in section 3.3 of this chapter for details).

Since Cohen’s seminal 1992 paper on the necessity of statistical power analysis, it is impossible to ignore the vital role played by sample size in determining the rejection of the null hypothesis (here, the hypothesis that there is no significant difference between

²¹ Our original Mood choice variable comprised several hundred additional verbs, since Conditional verbs were included in the present study’s initial phase, as explained in this chapter.
our informants with respect to the use of the Subjunctive). Cohen (1992) argues that researchers should not just be concerned with “the statistical test result and its accompanying p value” (p.155), but also, and more importantly, with “the phenomenon under study” (p.155). As the author states, it is “most useful to determine the N [sample size] necessary to have a specified power […]” (p.156). For high statistical power, a sample size of N > 80 is considered good for regression modeling.

We follow Cohen, then, in basing our study in only four linguistic contexts. A statistical analysis of Subjunctive rates in the five other linguistic contexts would inevitably fail to represent the speech patterns of the whole community and would undermine the value of the research. As a result, the analysis of mood choice is conducted in all nine linguistic contexts taken together, and in each of the four most popular linguistic contexts (i.e. the Modal context, the Protasis Si context, the Temporal context, and the Apodosis Si context).

5. Dependent variables

5.1 Subjunctive rate

The dependent variable used to determine the frequency with which an informant draws on a Subjunctive verb is called Subjunctive rate. Subjunctive rate is a token-level measure. It counts how many times a speaker uses a Subjunctive verb in the linguistic contexts included in the envelope of variation. The number of Subjunctive verbs is then divided by the total number of verbs in the envelope of variation, and turned into a percentage. Thus, the dependent variable Subjunctive rate is a continuous variable. The overall Subjunctive rate (number of verbs, or N, and percentage of verbs, or Pct.) for all nine linguistic contexts combined is shown in Table 3.2.
Table 3-2
Mood distribution in all (9) linguistic contexts
All informants

<table>
<thead>
<tr>
<th>Mood</th>
<th>N</th>
<th>Pct.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subjunctive</td>
<td>684</td>
<td>10.4</td>
</tr>
<tr>
<td>Indicative</td>
<td>5892</td>
<td>89.6</td>
</tr>
<tr>
<td>Total</td>
<td>6576</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 3.2 establishes that, for all informants and linguistic contexts taken together, the overall Subjunctive rate hovers around 10 percent (10.4 percent to be exact), while the Indicative mood is drawn on about 90 percent of the time (89.6 percent, precisely). That is, in the corpus overall, 90 percent of verbs in the envelope of variation are in the Indicative and 10 percent are in the Subjunctive.

The micro-level dependent variable Subjunctive rate measures separately the frequency with which the informants drew on a Subjunctive verb in each of the four most popular linguistic contexts (Modal, Protasis Si, Temporal, and Apodosis Si). The Subjunctive rate for each context only considers the informants in whose speech the context is available (i.e. those informants who used the linguistic context in question). Table 3.3 below displays the informants’ Subjunctive rate (percentage of Subjunctive verbs, or Pct.) in each of these four contexts.
Table 3-3
Subjunctive rate in the most popular linguistic contexts

<table>
<thead>
<tr>
<th>Linguistic contexts</th>
<th>Pct.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Modal context</strong></td>
<td></td>
</tr>
<tr>
<td>(e.g. Como quieran, lo que quieran, como que quieren)</td>
<td>9.6</td>
</tr>
<tr>
<td><strong>Protasis Si context</strong></td>
<td>8.7</td>
</tr>
<tr>
<td>(e.g. Si vinieran, si hubieran venido)</td>
<td></td>
</tr>
<tr>
<td><strong>Temporal context</strong></td>
<td>8.8</td>
</tr>
<tr>
<td>(e.g. Hasta que vengan, antes de que vengan, cuando vengan)</td>
<td></td>
</tr>
<tr>
<td><strong>Apodosis Si context</strong></td>
<td>6.3</td>
</tr>
<tr>
<td>(e.g. Si..., quisiera ir, Si..., hubiera querido ir)</td>
<td></td>
</tr>
</tbody>
</table>

It is evident from Table 3.3 that speakers’ Subjunctive rate in each of these four linguistic contexts is lower than it is overall (the overall Subjunctive rate totaling 10.4 percent, see Table 3.2). This indicates that even in the most popular linguistic contexts, speakers still have very high rates of Indicative verbs (90.4 percent of Indicatives in the Modal context, 91.3 percent in the Protasis Si, 91.2 in the Temporal, and 93.7 in the Apodosis Si).

5.2 Linguistic context availability

The other measure developed in the present study to explore Subjunctive usage is Linguistic context availability. This dependent variable gauges the number of informants in whose speech each linguistic context occurs and fails to occur. Recall that the term availability does not refer to an informant’s full linguistic repertoire. Instead, the availability and non-availability of a context exemplifies speaker tendencies. As stated earlier, the exploration of Linguistic context availability relates to our study of the Subjunctive, as speakers’ Subjunctive rates are surely related to their propensity to use the contexts in which these verbs occur. Any given linguistic context, such as for example the Volition context, either does or does not occur in the speech of an informant.
in the OZC. Thus, the dependent variable Linguistic context availability is categorical and binary.

Many of the present study’s linguistic contexts comprise more than one environment. The Volition context, for example, encompasses two environments, namely clauses introduced by querer que and esperar que. If at least one of the environments appeared at least once in an informant’s speech, it was determined that the corresponding linguistic context was available to that informant. For example, informant 334E never used esperar que but did use querer que, as in (22), at least once:

(22) Yo no quiero que me lo regale – LARI 334E
    ‘I don’t want him/her to give it to me’

The conclusion is that the Volition context is available to this speaker (in the context of the OZC). In contrast, if none of the environments that pertain to a particular linguistic context appear in an informant’s speech, the linguistic context in question is deemed not available (in the context of the OZC). For example, informant 002U never used aunque, so the Concessive context is considered not available for this speaker. It is important to note that at least one linguistic context occurred in the speech of each individual informant. That is, every one of our 142 informants made use of at least one environment (e.g., Si ‘if’). On average, each informant has five linguistic contexts available in his/her repertoire. (In other words, around four contexts are absent in individual speakers.)
In order to compare the availability (i.e. presence) and non-availability (i.e. absence) of a given linguistic context in the informants’ speech, we started by examining the number of informants in each group of users and non-users. In other words, the comparison between users (those informants for whom a context is available) and non-users (those informants for whom the same context is not available) is only interesting (and statistically viable) if there are a significant number of informants in each of the two groups. For instance, the Modal context is available to 141 out of our 142 informants, and thus unavailable to just one informant. Comparing the 141 informants in whose speech the Modal context occurs to the one informant in whose speech the Modal context does not occur seems futile. Thus, we decide on 15 informants as the minimum number of informants (that is, a little over 10 percent of our sample) required in each of the two groups. That is, our exploration of Linguistic context availability includes only those linguistic contexts that are available to 15 or more informants, and not available to 15 or more informants. As is evident in Table 3.4, four linguistic contexts match this description. The four corresponding contexts (Concessive, Possibility, Volition, and Hypothetical Como si) are highlighted and bolded for emphasis.
Table 3-4
Linguistics contexts: Number and Percentage of informants where each context appears and does not appear from most to least

<table>
<thead>
<tr>
<th>Rank</th>
<th>Linguistic context</th>
<th>Informants</th>
<th>Where context appears</th>
<th>Where context does not appear</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Modal (e.g. Como, lo que, como que)</td>
<td>141</td>
<td>99.3</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Protasis Si (e.g. Si)</td>
<td>140</td>
<td>98.6</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>Temporal (e.g. Hasta que, antes de que, cuando)</td>
<td>131</td>
<td>92.3</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>Apodosis Si (e.g. Si..., quisiera ir)</td>
<td>129</td>
<td>90.8</td>
<td>13</td>
</tr>
<tr>
<td>5</td>
<td>Concessive (e.g. Aunque)</td>
<td>61</td>
<td>43.0</td>
<td>81</td>
</tr>
<tr>
<td>6</td>
<td>Possibility (e.g. Es posible que, tal vez, quizás, a lo mejor)</td>
<td>45</td>
<td>31.7</td>
<td>97</td>
</tr>
<tr>
<td>7</td>
<td>Volition (e.g. Quiere que, espera que)</td>
<td>18</td>
<td>12.7</td>
<td>124</td>
</tr>
<tr>
<td>8</td>
<td>Hypothetical Como sí (e.g. Como sí)</td>
<td>17</td>
<td>12.0</td>
<td>125</td>
</tr>
<tr>
<td>9</td>
<td>Uncertainty (e.g. No creo que, no sé sí)</td>
<td>12</td>
<td>8.5</td>
<td>130</td>
</tr>
</tbody>
</table>

As a result, the dependent variable Linguistic context availability is considered only with respect to these four linguistic contexts. The five other contexts are disregarded.

6. Independent variables

An independent variable bears an associative relation to a dependent variable. In statistical terms, an independent variable is said to influence or predict the observed values of a dependent variable. Following convention, all of the independent variables in this study are labeled external independent variables, since they make reference to the informants’ socio-demographic features. As discussed in section 2 of this chapter, the OZC is made up of 142 sociolinguistic interviews conforming a stratified sample of the New York City Spanish-speaking population. For the purpose of their 2012 study, O&Z turned the speakers' background data into 34 independent variables. The present

---

22 For more on the interviews and the stratification process, see O&Z (2012: 22-23, 25-27).
study includes a select number of these independent variables. The rationale is explained in the following section.

6.1 Simple independent variables

The current study explores four simple independent variables corresponding to four of the informants’ socio-demographic characteristics. They are (1) Generation, (2) Gender, (3) Age, and (4) Socio-economic status (SES). These variables are discussed one by one.

(1) Generation. Generation refers to an informant’s immigrant generation.

Typically, the literature divides Spanish/English bilinguals in the United States into two or three categories, referred to as the first, second and third generation. Although the divisions between generations vary from one study to another, the first generation was always raised (and mostly educated) in Latin America, the second generation was born and/or raised in the United States, and the third generation was born in the United States to U.S. born and/or raised parents.

The present investigation employs O&Z’s definition of generation because it happens to align perfectly with patterns of Subjunctive acquisition (discussed below). Participants are divided into two immigrant generations. Generation one is subdivided into two, giving us a total of three generational groups. Generation one is called the Newcomers (LARN\(^{25}\)); Generation two is called the New York Raised (NYR).\(^{26}\) The first generation, which represents over three quarters of the informants, is divided into two groups: the Newcomers and the Established Immigrants (LARI). The Newcomers

\(^{23}\) Information on the way they were coded can be found in Appendix B.

\(^{24}\) The term simple variable is assigned to non-interaction variables in the present investigation.

\(^{25}\) We refer to these three groups as the Newcomers, Established Immigrants, and the New York Raised. It is helpful to keep the acronyms in mind as they appear in tables and examples. LAR stands for Latin American Raised.

\(^{26}\) For a complete picture, see O&Z (2012: 30-34).
arrived to New York at age seventeen or older and have lived in the city for five years or less. The Established Immigrants arrived before age 17 and/or have lived in New York City for more than five years. The New York Raised were either born in the U.S. or arrived by age three. The distribution of informants into the three groups is shown in Table 3.5 below.

<table>
<thead>
<tr>
<th>Consultants, by immigrant generation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Newcomers</td>
</tr>
<tr>
<td>Established Immigrants</td>
</tr>
<tr>
<td>New York Raised</td>
</tr>
<tr>
<td><strong>Total</strong></td>
</tr>
</tbody>
</table>

Table 3.5 shows that the majority of informants are first generation, as the two first generation groups together represent 82.4 percent of informants. The Established Immigrants constitute the largest group overall, as they represent 54.9 percent of all informants. A little under one fifth of the participants are second generation, New York Raised consultants.

The average Newcomer is 30 years old and has spent less than three years in the city. The average Established Immigrant is 38.5 years old and has spent at least 17 years in the city. The average New York Raised consultant is 29 years old. The number of years spent in the city for the New York Raised is essentially the same as their age, since they were born in the city or arrived by age three. The separation into these three groups is consistent with the main language in which the consultants were schooled (mostly in Spanish for the two first generation groups, only in English for the New York Raised); with the degree of language contact they can be assumed to have experienced based on
the number of years they have spent in New York (the New York Raised have been in the City longer than the Established Immigrants, who in turn have been in the City longer than the Newcomers); and with the gradual acquisition of the Spanish Subjunctive (by stages corresponding to different linguistic contexts). O&Z justify the division into three categories based on differences in exposure to both English and Spanish, by asserting that “the established immigrants represent a middle group between the NYR, who started being exposed to English and, in many cases, to out-group forms of Spanish when they were very little children and their own Spanish was still developing, and the immigrant newcomers, who have not been in NYC for long and were linguistically quite mature when they first arrived” (p.34).

There is an undeniable connection between language contact and language acquisition in bilingual environments like New York City. In both monolingual and bilingual settings, primary linguistic development is deemed to take place between birth and age four, while later language development ensues between ages four and 13, corresponding to when children are in school (Blake 1980; Mikulski 2010; Montrul 2009, cf. Bookhamer 2013). With respect to the monolingual acquisition of the Spanish Subjunctive in Spanish-speaking contexts such as Latin America, the consensus in the literature is that the mood is acquired in phases, rather than all at once (see Chapter 2 for a full account). These different phases of Subjunctive acquisition correspond to the distinct messages that speakers express, classified as linguistic contexts in the present study. According to Montrul (2009: 247) “the full spectrum of uses and subtleties of subjunctive knowledge […] in both production and comprehension” are acquired “well after the age of 8 years”. For Gili Gaya (1972) and Blake (1983) full Subjunctive
acquisition occurs by age 10. As the average age of arrival to New York City for all of our first generation informants is 24, it is clear that they have lived in monolingual environments\textsuperscript{27} well past the age of 10, and therefore should have a full command of the different contexts of use of the Spanish Subjunctive. Montrul and Gili Gaya’s assessments may well apply to first generation informants who were raised and schooled in monolingual environments. However, the situation is entirely different with respect to the New York Raised, who were born and/or raised in a bilingual setting. These second generation speakers acquired a different system from that of their Latin American Raised parents. Thus, it is our view that these two groups are not comparable in terms of their Subjunctive acquisition (see Chapter 6 for a more complete appraisal of second language acquisition research on the Subjunctive).

At any rate, the present investigation does not focus on questions of acquisition. Instead, our exploration centers on our three groups’ distribution of Subjunctive verbs, which may vary based on these speakers’ distinct experiences with Spanish and English. It is this expectation of different Subjunctive usage based on generation that justifies the inclusion of this variable in the present study.

(2) Gender. Informants were asked to define themselves in terms of gender. The term gender is used rather than “sex”, as gender refers to a social construction, that is, not necessarily to the sex the informant was born with. All informants responded as either male or female. Thus, the categorical variable Gender is binary, and has two factors:

\textsuperscript{27} Puerto Rico is included in the monolingual environments, even though Spanish and English are both official languages on the island. According to the 1990 U.S. Census, fewer than 20 percent of Puerto Ricans in Puerto Rico speak English fluently. In the 2000 U.S. Census, 85.4 percent of Puerto Ricans (in Puerto Rico) say that they speak Spanish at home, and 71.9 percent assert that they speak English “less than very well”. Furthermore, a 1997 New York Times article (written just a few years before the start of the OZC data collection) confirms that all school subjects (except for English) are taught in Spanish in Puerto Rico.
male and female. Table 3.6 displays the distribution of male and female informants in the present investigation.

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Pct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>79</td>
<td>55.6</td>
</tr>
<tr>
<td>Female</td>
<td>63</td>
<td>44.4</td>
</tr>
<tr>
<td>Total</td>
<td>142</td>
<td>100</td>
</tr>
</tbody>
</table>

As discussed in section 2 of this chapter, the informants represent a stratified sample of the population, meaning that an attempt was made to balance the informants in terms of their socio-demographic characteristics. Table 3.6 exhibits a slight unevenness in terms of the proportion of male and female informants.

Although the variable Gender is rarely included in investigations on the Spanish Subjunctive (see Chapter 2), it is expected to have a significant bearing on Subjunctive usage as captured by the two dependent variables, Subjunctive rate and Linguistic context availability, of the present study. This hypothesis rests on the results of the few studies on the topic that do include the socio-demographic feature (Bookhamer 2013; Lastra & Butragueño 2012; Serrano 1995) but also on the research conducted on other linguistic features, in which women and men’s linguistic behavior tends to differ, both in monolingual and in bilingual settings (Alfaraz 2010; Eckert 1989; Eckert & McConnell-Ginet 2003; Labov 2001; Orozco 2007; Shin & Otheguy 2013; Van Ness 1995).

(3) Age. Age refers to an informant’s age at the time of the interview. The variable Age is the only continuous independent variable in the study. In this investigation, informants range from age 12 to age 80 (with the mean age being 34.6). A
continuous variable was chosen in the place of a categorical one because there is no rational justification for splitting Age into separate categories, as has been pointed out by Eckert (1989: 246) and Eckert, Edwards, & Robins (1985). The variable Age was selected based on the effect that it has been found to have on the occurrence of the Spanish Subjunctive (Gutiérrez 1994; Lantolf 1978; Lastra & Butragueño 2012).

(4) Socio-economic status (SES). The variable Socio-economic status is a combination of an informant’s educational level and category of occupation. The present investigation adopts O&Z (2012)’s SES variable. Our informants fall into two categories: Low and Middle SES. The missing values in Table 3.7 correspond to missing SES data on three of the 142 informants.

<table>
<thead>
<tr>
<th>Table 3-7</th>
<th>Consultants, by socio-economic status (SES)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
</tr>
<tr>
<td>Low</td>
<td>57</td>
</tr>
<tr>
<td>Middle</td>
<td>82</td>
</tr>
<tr>
<td>Missing</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>142</td>
</tr>
</tbody>
</table>

O&Z (2012) acknowledge that it was difficult to stratify, i.e. balance, the sample with respect to speakers’ socio-economic status, which is evident in Table 3.7 above, as Middle SES informants are more numerous (N = 82) than Low SES informants (N = 57). The authors explain that the snowballing technique (described in section 2) which was used to facilitate the recruitment of participants for their 2012 study, by calling on their own and their graduate student assistants’ friends and acquaintances, resulted in greater
numbers of educated\textsuperscript{28} informants. Yet, the sample can still be viewed as representative of the U.S. immigrant situation, since, as Feliciano (2017, in press) asserts, “the reality is that only a select few from any one country migrate to the United States and those that do tend to be more educated and of a higher class than their non-migrating counterparts.” Rather than exploring the Social class and Years of education variables\textsuperscript{29} separately, the more comprehensive variable SES was selected for this study. This decision was motivated by the fact that immigrants “often have to take jobs below their previous status levels because their education may not transfer or they have language difficulties” (Feliciano, 2017 in press). Feliciano and Lanuza (2017) argue that immigrants often come from countries with less educated populations, so a high school degree, for instance, often translates into a higher class status in other countries than it does in the United States. Thus, SES, which involves a combination of social class and educational level, seems like a more reliable variable for a study centered on immigrants.

To conclude, four simple independent variables (i.e. Generation, Gender, Age and Socio-economic status) were drawn on in the present study. Speakers’ other socio-demographic characteristics are not exploited, as they are not expected to have an effect on the two main dependent variables associated with Subjunctive usage in this study. As a result, they are not included in the present study.

### 6.2 Interaction variables

Two interaction variables were included in the present investigation. The interaction of the variable Generation and the variable Gender, split into the two interaction variables

\textsuperscript{28} “The percentage of the NYC Latino population age twenty-five and older who obtained a college degree or higher level of education was, in 1999 […], 13.7 percent” (O&Z 2012: 34).

\textsuperscript{29} O&Z (2012) turned the OZC informants’ socio-demographic and linguistic characteristics into independent variables for their study. Two of these variables were Social class and Years of education.
(the reasons for this split are explained in Appendix B), is examined based on the hypothesis that women and men of diverse generations behave differently with respect to Subjunctive usage\textsuperscript{30}. This inkling is largely based on Shin & Otheguy (2013)’s New York City study of Spanish subject pronouns, which uncovered significant differences between first generation women and men. In addition, Bookhamer (2013)’s exploration of the Spanish Subjunctive in New York City showed preliminary signs of a gender and generation effect, when analyzed together, though the interaction between the two was not examined. His study points to the fact that both Newcomer\textsuperscript{31} men from the Mainland and New York Raised men from the Mainland draw on the Subjunctive to a greater extent than their female counterparts in obligatory contexts (in the case of the Newcomers) and in optional contexts (in the case of the New York Raised).

Consequently, an interaction between Generation and Gender, in the form of two interaction variables, was created for the current investigation. An interaction variable considers the relationship between at least two variables. The effect of one variable (e.g., Generation) on the outcome (e.g., Subjunctive usage) depends on the state of a second variable (e.g., Gender). An asterisk is used to denote the interaction between two variables, as shown below.

- LARI*Female (i.e. Established Immigrants*Female)
- NYR*Female (i.e. New York Raised*Female)

The presence of the interaction variables in the statistics should help elucidate the Subjunctive usage of the female Newcomer, female Established Immigrant and female New York Raised informants. As laid out in Appendix B, the statistical results pertaining

\textsuperscript{30} Recall that the general term \textit{Subjunctive usage} encompasses both rates of Subjunctive and Linguistic context availability, both discussed in section 5 of the present chapter.

\textsuperscript{31} Bookhamer (2013)’s OZC Newcomer informants constitute a subset of our larger Newcomer group.
to the interaction LARN*Female, or female Newcomers, are inferred from those of the other two interaction variables. Table 3.8 below shows the division of men and women into each of the three groups. (Total rates are absent as they are provided in the discrete generation and gender tables above).

<table>
<thead>
<tr>
<th></th>
<th>LARN</th>
<th></th>
<th>LARI</th>
<th></th>
<th>NYR</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>20</td>
<td>14.1</td>
<td>48</td>
<td>33.8</td>
<td>11</td>
<td>7.7</td>
</tr>
<tr>
<td>Female</td>
<td>19</td>
<td>13.4</td>
<td>30</td>
<td>21.1</td>
<td>14</td>
<td>9.9</td>
</tr>
</tbody>
</table>

Table 3.8 evidences the fact that, once again, the sample was stratified (i.e. balanced), as discussed in section 2 of this chapter. Male and female consultants are fairly evenly distributed in each immigrant group.

To the best of my knowledge, the literature on the Spanish Subjunctive in the United States does not incorporate interaction variables. More generally in linguistics research, it appears that interaction variables are less frequent than simple independent variables. This may be because interaction variables only offer a reduced assessment of the data, which is condensed when the relationship between two variables is taken into consideration. Note, for instance, that there are only 14 New York Raised women. Thus, when investigations such as this one decide to include an interaction variable, they must proceed with caution, bearing in mind that including more than one or two interaction variables in a study may not be statistically sound. More interaction variables imply a narrowing of the picture. For these reasons, LARI*Female and NYR*Female were the only interaction variables included in the present study.
7. Data collection

7.1 Java

Because the OZC is a very large corpus, many of the investigations that are based on the OZC have only exploited a portion of the corpus. For example, Bookhamer (2013)’s dissertation on the Spanish Subjunctive in New York City includes 52 of the 142 consultants that comprise the whole corpus. Similarly, Erker & Guy (2012)’s study of variable Subject Personal Pronoun Expression focuses on 12 participants (6 Mexican and 6 Dominican). The reasons for this seem clear. Searching for tokens manually in a sizeable corpus is very time-consuming. Furthermore, manual data collections are liable to human error. For these reasons, the present investigation, which studies all of the 142 informants of the OZC, makes use of Java, one of the most popular computer programming languages in the world. The use of Java has proved indispensible for the processing of this study’s very large dataset.

Although Java has facilitated the process of collecting data, the algorithms for the computer programming language had to be specifically created for this study. The first step consisted in designing algorithms capable of detecting finite verbs. Fortunately, this first step was made possible thanks to the number in parentheses that precedes each finite verb in the OZC transcripts. As described in section 2 of this chapter, Otheguy and Zentella, together with their graduate students and fellow researchers, transcribed the oral interviews that comprise the OZC. Every transcript includes a number in parentheses that precedes each finite verb and its subject pronoun when applicable, as in example (23):
 Como (178) yo espero que (179) venga – LARN 002U

‘Like I hope s/he comes’

In example (23), the first finite verb espero ‘I hope’ is introduced by the personal subject pronoun yo ‘I’. The second finite verb venga ‘s/he/you/it come(s)’ has a null subject. Both verbs are introduced by a number in parentheses indicating that a finite verb will follow. Whether the number in parentheses is placed before the subject pronoun (as with 178) or right before a verb without a subject pronoun (as with 179) is of little import in Java, which can be taught about the subject pronoun between the number and the verb. What matters is that the algorithm can recognize the presence of a finite verb thanks to the consistent pattern of numbers in parentheses.

Next, the program had to learn which verbs to extract from the corpus. Obviously, only the verbs in the envelope of variation are of interest. Yet, as described in section 4.1, the envelope of variation was larger in the initial stages of the study, and was later reduced. The program was first taught to recognize all of the verbs that the study started out with. This was done by assigning a number to each linguistic context, resulting in the assignment of the same number to the different environments within a context. For example, the Volition context, introduced by the clauses querer que and esperar que was assigned the number 3. An algorithm was then fashioned to recognize the different environments in this linguistic context.

This was not as simple as it may seem, however. Several important questions had to be considered to ensure a successful outcome. First of all, the verbs in the matrix clauses, e.g., querer, esperar, ser (posible que), (no) creer (que), (no) saber (si), are

---

32 See section 4 of this chapter for details on the study’s envelope of variation.
almost always conjugated in the corpus and are seldom found in the infinitive. For instance, in example (23) above, the verb esperar ‘to hope’ is conjugated in the first person singular, espero ‘I hope’. Thus, an algorithm had to be designed in such a way that it would learn to detect all moods, tenses and declensions of each of the matrix verbs, in each environment, within each linguistic context. For example, the verb esperar ‘to hope’ had to be recognizable in the infinitive, as well as in all of its finite forms, that is, in the first person singular present Indicative (espero), second person singular present Indicative (esperas), third person singular present Indicative (espera), and so on for all person-number forms, then in the first person singular preterite (esperé), and so on for all person-number forms, then in the first person singular plucumperfect (había esperado), and so on for all person-number forms, as well as for all other tenses and moods. This was achieved by giving the program all of the matrix verbs’ possible forms. Once Java was familiarized with them, the program was instructed that all of these forms, in addition to the conjunction que ‘that’ and the following finite verb, were to be assigned the number 3 (corresponding to the Volition context). Below is an example of the algorithm’s procedure (with example 23 from above):

(23) Como (178) yo espero que (179) venga – LARN 002U
>>> (179) >>> espero que + venga >>> 3

To recapitulate, the algorithm recognized the finite verb espero ‘I hope’ thanks to the number 178 that precedes it in parentheses. Then, it noticed that the verb espero is part of the thread espero + que, and that another verb, venga, follows it (recognized
through the number 179 that precedes it in parentheses). Finally, it classified the verb (venga) that follows the matrix clause (espero que) as belonging to linguistic context number 3, that is, the Volition context. More complex scenarios, such as sentences that include additional words between the matrix and embedded verbs, are discussed further along in section 7.2.1.

Once the linguistic context was detected, the next step consisted in identifying the nature of the verb inside the dependent (or subordinate) clause. In example (23), the embedded verb venga ‘s/he/you/it come(s)’ is in the present Subjunctive. One would assume that Java would already know how to detect the embedded verb’s tense, mood, and aspect (TMA), since the program had already learned to do this for verbs in the matrix clause (e.g., espero, esperaste, habíamos esperado, quiere, quería, quisiera, sé, supo, creyeron, etc.). However, this was not the case, because the matrix verbs and the embedded verbs are not necessarily the same. The matrix verbs were already known to the researcher, and to Java, since they are limited in number and are present in the linguistic contexts selected for the study. In contrast, the embedded verbs are much more varied and abundant. At first, it seemed impossible to identify all embedded verbs without going through the whole corpus manually, which would have defeated the purpose of the automated process. Thus, the program had to be taught to recognize the

---

33 Recall that the embedded verbs in the nine linguistic contexts were not only in the Indicative and Subjunctive, but also in the Conditional (see discussion in section 4.1). The 6576 tokens (i.e. Indicative and Subjunctive verbs) collected by Java only constitute a portion of all of the verbs found in the study’s nine linguistic contexts. Thus, although the additional verbs are not part of the envelope of variation, they were still present in the corpus (OZC) that Java had to contend with.
TMA\textsuperscript{34} of each embedded verb found in the linguistic contexts that comprise the envelope of variation.

The process of tagging verbs in dependent clauses was achieved in two steps. The first step involved feeding the program a list of eligible verbs from the OZC whose TMA had already been tagged. The eligible tagged verbs (i.e. the verbs that were already tagged and that fit into our envelope of variation) and their tag (i.e. the number assigned to each, see list below) were fed to Java, so that the program could learn to recognize the mood, tense and aspect of the same finite verbs each time they reoccurred. Below are the tags, which combine those already available from the O&Z study with those created for the present one.\textsuperscript{35}

**List: Tense/Mood/Aspect (TMA) form of the verb**

11 = Present Indicative, *canto*  
12 = Preterite Indicative, *canté*  
13 = Imperfect Indicative, *cantaba, estaba cantando*  
14 = Periphrastic future Indicative, *voy a cantar*  
15 = Future Indicative, *cantaré*  
16 = Conditional, *cantaría*  
17 = Present Subjunctive, *cante*  
18 = Past Subjunctive, *cantara*  
19 = Imperative use of any form, *canta, cante, sal, ten*  
20 = Perfect Indicative, *he cantado, había cantado*

\textsuperscript{34} Even though verb tense is not explored (as an independent internal variable) in the present investigation, it was necessary for Java to know a verb’s tense in order to fit with O&Z’s tags. The inclusion of tense was considered a time saver. Furthermore, a verb’s aspect was brought in because verbs in the perfect and progressive comprise two words, in both moods. For example, the verb *hablar* can take the forms *he hablado, haya hablado, estaba hablando, estuvieras hablando*, etc. Java had to be instructed to pay attention to two words (instead of one) when it encountered the auxiliary *haber* (in all of its forms) for the perfect, and the verb *estar* (in all of its forms) for the progressive. In both cases, Java had to be taught to determine whether the word that followed *haber* or *estar* was a verb (since both *haber* and *estar* can stand on their own), in order to either collect the following word (if a verb) or discard it (if not a verb).

\textsuperscript{35} The Conditional appears in the list, as Conditional verbs were collected by O&Z as well as in the initial stages of the present investigation. Verbs in the Conditional were later excluded from the envelope of variation and discarded from the current study (as explained in section 4.1). The four TMA add-ons (22-25) were incorporated a posteriori, when it was discovered-- through Java’s misclassifications-- that O&Z’s TMA tags were too limited for the present investigation. The list is not an exhaustive rendering of all Indicative and Subjunctive tenses. Instead, it represents the tenses that informants tended to draw on in both moods in our work.
21 = Perfect Subjunctive, *haya cantado, hubiera cantado*
22 = Present progressive Indicative, *estoy diciendo, sigue tocando, van enseñando*
23 = Past progressive Indicative, *estaba diciendo, iba diciendo, fuimos corriendo*
24 = Past progressive Subjunctive, *estuviera pasando, anduviéramos sufriendo*
25 = Present progressive Subjunctive, *esté hirviendo, vaya enseñando*

For example, O&Z had tagged the verb *chequeas* ‘you check’ as 11. The verb and its tag were shared with the program. Every time the algorithm came across *chequeas* in the envelope of variation, it categorized its TMA as belonging to 11, as in example (24):

(24)  Si (535) tú *chequeas*... - LARI 318D
      chequeas >> 11

The verbs that were not tagged by O&Z also had to be identifiable by Java. Thus, the second step consisted in searching online for the most common spoken Spanish verbs, and finding pre-made lists of their infinitives and their roots. Once this was accomplished, I made a list of all of the possible morphological verb endings in the TMA’s found in the list above. The plan was to have Java recognize the TMA of the most frequent verbs through a combination of their infinitives, roots and endings. However, this time-consuming task went wrong, when I realized that the same morphological endings corresponded to too many different tenses/moods. For example, the first person plural ending –*amos* in a verb like *hablar* corresponds at once to the present (*hablamos*), preterite (*hablamos*) and imperfect (*hablábamos*) Indicative, and to the imperfect Subjunctive (*habláramos*). Supplying Java with the letters “ár”

---

36 The fact that Java does not recognize accent marks is discussed in section 7.2. This means that rather than telling the program that ár + amos constitutes a morphological Subjunctive ending, ar (without an accent) would have been entered.
indicate that ár + amos corresponds to a Subjunctive mood ending would have been a
solution. However, coming up with all of the different possible combinations in order to
help the program distinguish between moods seemed more onerous than tagging the
remaining verbs manually.

7.2 Manual checking

Java was able to collect all of the data, and tag most of it. That is, Java was able to tag
all of linguistic contexts, matrix verbs, and verb TMA's that matched the verbs already
collected by O&Z. But some of the data (e.g., the remaining verbs’ TMA discussed in
the previous section) had to be manually tagged. Fortunately, the manual tagging could
be performed at the same time as the manual checking.

While it is virtually impossible for a computer programming language to be error
free, it is possible to reduce the number of errors by resorting to manual checking. It may
seem contradictory to have used a programming language if manual work is required
anyway. But this is not the case here. Collecting data manually for this study’s very
large dataset would have proved extremely tedious and more time-consuming than the
combination of Java and manual checking. Thus, Java’s output was checked manually
following the completion of the whole automated process. As the manual verification
took place, so did the fine-tuning of the instructions conveyed to the program. Moreover,
it was decided that half of the output would be surveyed manually, which would provide
ample opportunities to correct any oversights. Several different types of errors emerged.
The errors and their solutions are described below.
7.2.1 Word(s) between number and verb

In the OZC, the number in parentheses does not always immediately precede the verb in the dependent clause. Sometimes, there is another word (often an adverb) or several words between the number in parentheses and the verb in the dependent clause, e.g. no, nunca, siempre, ni, etc. Sometimes, there are three words (e.g., en ese tiempo ‘at the time’) in addition to the subject pronoun, as in example (25):

(25) Y como (655) yo en ese tiempo contestaba los teléfonos… – LARI 320E

‘And since I would answer the phones back then…’

Thus, when Java was told to collect the word that immediately follows the number in parentheses, except if that word was a subject pronoun, it picked up the first word that came after the number in parentheses, or after the subject pronoun, if there was one. In the case of example (25), Java “jumped over” the subject pronoun yo ‘I’ which, as with all other subject pronouns, the program had been instructed to disregard. It then collected the word en, instead of the verb contestaba ‘would answer’. To remedy this, the program had to be given a list of words (e.g., no, le, en, ese, etc.) that could possibly occur between the number in parentheses and the verb. The algorithm was then instructed to skip those words (with the maximum number of words being three, excluding the subject pronoun), and to pay attention to the following one, that is, the verb. After the new algorithm was created, manual checking corroborated that this type of error hardly ever occurred anymore. The few errors that remained were corrected manually.
7.2.2 Two contiguous linguistic contexts

Linguistic context misclassification was the most common type of error. This happened mainly because of the contiguous presence of two environments belonging to two different linguistic contexts. Just as Java had to be taught to skip certain specific words after the number in parentheses to find the embedded verb, the program had to learn to look for certain words before the number in parentheses to locate the correct environment belonging to the appropriate linguistic context. Were the program to look no further than immediately before the number in parentheses as it was taught to do at first, it would categorize a linguistic context such as the Uncertainty context, which contains the environment no saber si, as a Protasis Si context that had been introduced simply by si.

This type of misclassification occurred, for instance, with example (26):

(26) … una buena amiga (no sé si (1043) usted la conocerá…) – LARN 183U

‘… a good friend (I don’t know if you know her…)’

The program had to be taught to look for no + the verb saber ‘to know’ (in all of its forms) before si ‘if’. The two words’ presence would indicate that the verb in the dependent clause, in this case conocerá, should belong to an Uncertainty context introduced by no sé si ‘I don’t know if’ rather than to a Protasis Si context, introduced simply by si ‘if’. The new instructions were successful with respect to this environment.

In order to avoid this type of error more generally, Java was told to always look for the environment in all the words that preceded the number in parentheses rather than just paying attention to the word that immediately preceded the number, and to stop when it came across the preceding number in parentheses (so as to avoid any overlapping with
a preceding verb in a different environment). The new instructions led Java to start with the farthest word preceding the number in parentheses of interest, but within the limits of the preceding number in parentheses. This means that Java picked up the first environment that it encountered within the words between the two numbers in parentheses. In example (27), seven words appear between the number in parentheses of interest, which is 600, and the preceding number in parentheses, which is 599: *uno* (first), *no* (second), *sabe* (third), *quizás* (fourth), *a* (fifth), *lo* (sixth), and *que* (seventh).

(27) A veces (599) uno no sabe... quizás a lo que (600) uno le **llama** no tener suerte – LARI 332.1D

‘Sometimes one doesn’t know… maybe what one calls not being lucky’

Java started with the first word in the string after 599, which is *uno* in example (27). This word is not in any of this study’s environments, so it was ignored. The algorithm also ignored the two following words (*no* and *sabe*), which are not of interest either. Finally, the program encountered the word *quizás*, correctly identified *quizás* ‘maybe’ as an environment that belongs to the Possibility context, and matched the verb *llama* with that linguistic context. However, the verb *llama* is actually in the Modal context introduced by *lo que* ‘what’. Although this constituted an error, the program had properly followed the instructions. The unforeseen presence of two contiguous linguistic contexts was challenging. Asking the program to select the second environment (*lo que* instead of *quizás* in example 27) between the two numbers in parentheses proved possible. As a result, this type of error was remedied both through manual checking and the steady fine-tuning of the algorithm.
7.2.3 Typos and grammatical errors in the OZC

Different types of miscategorizations can be attributed to typos and errors in the OZC transcripts. Examples are offered below.

(1) **Missing accent marks.** When the adverb *sí* ‘yes’ was misspelled as the conjunction *si* ‘if’ in the transcript, the algorithm correctly categorized the *si* + verb string as belonging to the Protasis *Si* context. However, this constituted an error, since the speaker was not using the conjunction *si* ‘if’, but the affirmative adverb *sí* ‘yes/absolutely/for sure’ instead, as in example (28). The glosses in examples (28), (29) and (30) correspond to what the speaker said, rather than to the inaccurate transcription.

(28) Mi amigo dijo que *sí* también que **iba** – NYR 315M

‘My friend said that he was also going for sure’

In example (28), *si* should be spelled *sí*, and the verb *iba* ‘was going’ should not be analyzed.

Similarly, the missing accent mark on *cómo* ‘how’ meant that certain verbs (e.g., *se dice* ‘you say’ in example 29) were categorized as belonging to a Modal context, which can be introduced by *como* ‘like’, when they shouldn’t have been collected at all, since *cómo* + verb is not part of this study.

(29) Estaba bien contenta y bien, este…*como* **se dice**…no sorprendida – NYR 403P

‘She was very happy and very, um…how do you say…I mean surprised’
The misclassifications due to missing accent marks could only be corrected manually.

(2) **Two words instead of one.** Transcribers sometimes wrote *si no* ‘if not/don’t’ when they meant *sino* ‘but rather/instead’, for instance. This type of error in the corpus produced additional cases of Protasis *Si* context misclassifications, as with example (30):

(30) Hicieron una apuesta, no fue una apuesta, irracional, a favor de tal o cual candidato, *si no*, **fue** un proceso muy complejo – LARN 339M

‘They made a bet, it wasn’t an irrational bet, backing this or that candidate, instead, it was a very complex process’

In example (30), the verb *fue* was categorized as belonging to a Protasis *Si* context because *si no* is misspelled. It should be *sino* (i.e. *sino fue un proceso muy complejo* ‘instead it was a very complex process’), without a comma, which does not correspond to any of the linguistic contexts in this study.

### 7.2.4 Accent marks not recognized

It was discovered that Java does not recognize accent marks, which often change the meaning of a word in Spanish. Thus, had the transcriptions been devoid of accent mark errors, certain issues would still have come up. For example, Java mistook *cómo* ‘how’ for *como* ‘like’, *sí* ‘yes’ for *si* ‘if’, *sé* ‘I know’ for *se* ‘each other/one another’ or ‘yourself/himself/herself/themselves/itself’, etc. A missed accent on a verb such as *sé* ‘I know’ means that it was not categorized as a finite verb (as *se* without the accent mark is not a verb), when it should have been. And, as discussed above (in section 7.2.3), the

---

37 **As an aside,** Java was later instructed to disregard instances where a comma was found between the conjunction *si* ‘if’ and the corresponding finite verb. When the comma was absent, *si* + finite verb led to the correct categorization of a Protasis *Si* context. A comma in addition to a neighboring finite verb (signaled by a number in parentheses, found before or after the *si* clause) denoted the presence of an Apodosis clause.
difference between \textit{si} and \textit{sí} resulted in linguistic context misclassifications. Fortunately, this issue was not as prevalent as some of the others, and was quickly remedied manually.

Prior to the manual checking, it was discovered that Java had an error rate of 2.5 percent, which is minimal. Following the manual treatment of half of the data, the error rate was lowered to zero percent for the half that was checked, and remained at 2.5 percent for the other half. As a result, the whole database (6576 tokens) has an error rate of 1.25 percent, meaning that it is nearly 99 percent accurate.

### 7.3 Statistical methods

Following most variationist studies in the field of sociolinguistics, the present investigation uses both quantitative and qualitative measures to analyze the data. While the quantitative analysis considers all 142 informants that comprise the OZC, the qualitative exploration is pursued through examples from the speech of individual informants. The specific qualitative examples are exploited in order to illustrate the more general quantitative data.

The quantitative analysis is conducted in SPSS (Statistical Package for the Social Sciences)\textsuperscript{38}. Since the current study relies on external independent variables, an informant-level file consisting of 142 lines (one per informant) was entered in SPSS. This informant-level file has one column per dependent variable and one column per independent variable. The dependent variables are the overall Subjunctive rate (in the nine linguistic contexts) and the Subjunctive rate in each of the four most popular linguistic contexts, as well as the Linguistic context availability in each of four linguistic contexts. The independent variables are Generation, Gender, Age, Socio-economic

\textsuperscript{38} For more information on how SPSS works, see Antonius (2003) and Shannon and Davenport (2001).
status, LARI*Female and NYR*Female (the interaction variables, used only in the multivariate analyses, as explained in Chapter 4 and in Appendix B). The data collected with Java (and manually checked) were inserted into SPSS, as was the tagged socio-demographic information (i.e. independent variables). After that, the consultants’ Subjunctive usage was finally ready to be statistically handled and analyzed. The present investigation’s statistical exploration began with a bivariate analysis of the data, and was followed by a multivariate analysis (Chapter 4).

Measures of Subjunctive usage (i.e. overall Subjunctive rate, Subjunctive rate in the four most popular linguistic contexts, and Linguistic context availability in each of four linguistic contexts—as well as the relationships between all of these) were calculated. Relationships between all of the informants’ rates of Subjunctive overall and in the four most popular linguistic contexts, and among these same contexts, were brought to light by correlations. Cross-tabulations (Chi-square tests) revealed the associations that exist (and fail to exist) between the availability of different linguistic contexts in all of the informants’ speech. Independent samples t-tests uncovered the effects of Linguistic context availability and non-availability on the dependent variables overall Subjunctive rate and Subjunctive rate in each of the four most popular linguistic contexts. One-way ANOVAs were called upon in the bivariate exploration in order to gauge the import of relations between Subjunctive rates and socio-demographic variables in advance of the multivariate examination. Multiple linear regressions were used to establish the association between the dependent measures of overall Subjunctive rate and independent variables like Gender and Socio-economic status. Linear regressions were also used to determine the relationship between the dependent measure Subjunctive rate
in each of the four most popular linguistic contexts and independent variables (such as Generation and Age, for instance). Binary logistic regression analysis was used to understand the association between the dependent variables Linguistic context availability in each of four linguistic contexts and the independent variables.

For all statistical tests, the probability value is set to $p < .05$, but results with $p$ values above .05 are not rejected. This is done in order to avoid Type II errors “whereby the researcher erroneously rejects a relevant finding” (Shin & Otheguy 2013: 436, cf. Newton & Rudestam 1999: 65). Results with $p$ values between .05 and 1 are presented, as they may “reflect a true relationship between variables holding in the New York population” (Shin & Otheguy 2013: 436). Finally, as Bookhamer (2013) aptly points out, the fact that the OZC was carefully stratified implies that the present study’s findings should be seen as legitimately reflecting New York City’s Spanish-speaking Latino population as a whole.
Chapter 4: Results

1. Introduction

The analysis of the data collected for the present study comprises two steps: a bivariate exploration, followed by a multivariate analysis. Bivariate statistics are essentially descriptive measures, meant to gain a better understanding of the relationships that exist between the dependent and independent variables selected for the study. Once preliminary associations are established through bivariate analyses, multivariate statistical measures can be used to enhance the initial findings. The data analysis in section 2 of this chapter is purely descriptive, but is an important and informative first step, leading to the multivariate analysis in section 3, which may corroborate or invalidate the initial results. Both of these sections are divided into two parts, the first introducing the results associated with Subjunctive rate (sections 2.1 and 3.1), and the second presenting the results linked to Linguistic context availability (sections 2.2 and 3.2). Furthermore, section 2 includes a subsection (section 2.3) on Subjunctive rate by Linguistic context availability.

The results of the statistical analyses introduced in this chapter elucidate all of the present study’s research questions. As hypothesized in Chapter 1, it turns out that particular socio-demographic characteristics do indeed influence speakers’ Subjunctive rates, as well as the availability of specific linguistic contexts in their speech. Recall from the previous chapter that what I call Linguistic context availability refers to the presence of particular linguistic contexts (i.e. a clause or cluster of clauses) in speech. The reader should bear in mind that Linguistic context availability is explored in the context of the OZC, which means that it is not necessarily representative of an
informant’s entire linguistic repertoire. Moreover, it is the case that different linguistic contexts are interrelated, both in terms of Subjunctive rate, and in terms of Linguistic context availability. The study’s findings are presented in the current chapter (Chapter 4), and discussed in the following chapter (Chapter 5).

2. Bivariate results

Section 2, which presents the results of the bivariate statistics, is divided into two sections. Section 2.1 introduces the results of bivariate analyses with respect to Subjunctive rate. The objective is to understand the effects that socio-demographic variables (e.g., Gender) may have on overall Subjunctive rate, and on Subjunctive rate in the four most popular linguistic contexts. The associations between rates of Subjunctive overall and in the four most popular linguistic contexts are also explored. Section 2.2 displays the results of bivariate analyses with respect to Linguistic context availability. Again, the idea is to explore how particular socio-demographic variables shape the availability of each of four linguistic contexts in speech. Section 2.3 investigates the effect of Linguistic context availability (in each of four linguistic contexts) on Subjunctive rate (overall and in the four most popular linguistic contexts).

All of the bivariate analyses that explore the association between speakers’ socio-demographic characteristics and Subjunctive rate or Linguistic context availability exploit the same four variables\(^{39}\), discussed in detail in Chapter 3. Interaction variables, such as the two introduced in Chapter 3, cannot be included in bivariate analyses. Thus, the four variables discussed in the present section (section 2) are the following:

\(^{39}\) Bivariate analyses generally measure relationships between variables. The terms dependent and independent are therefore only really useful with respect to variables included in multivariate analyses (which measure the effects of independent variables on a dependent variable).
• Generation
• Gender
• Socio-economic status (SES)
• Age

As a reminder, three of the variables (Generation, Gender and SES) are categorical, and one is continuous (Age). For a full description of these variables, see Chapter 3. Below is a list of the variables’ constraints:

• Generation:
  - LARN (Newcomers)
  - LARI (Established immigrants)
  - NYR (New York Raised)

• Gender: Male
  Female

• SES: Low
  Middle

• Age: Continuous

Generation comprises three constraints (LARN, LARI and NYR). The Gender and SES variables are binary, as they each encompass two constraints (Male and Female, and Low and Middle, respectively). The Age variable does not include any constraints, as it is continuous.

2.1 Bivariate analysis of Subjunctive rates

The bivariate results of Subjunctive rate are summarized in four tables, one in each subsection. The first table answers the research question (RQ1a) about associations between rates of Subjunctive overall and in the four most popular linguistic contexts. The second table addresses the research question (RQ1b) about relationships between Subjunctive rates in the four most popular linguistic contexts. Both of these tables present a series of bivariate Pearson correlations. The next two tables answer the
research question (RQ2) about group differences in terms of Subjunctive rate (overall and in the four most popular linguistic contexts), based on four socio-demographic variables. These results are divided into two subsections, as the categorical and continuous variables require distinct statistical analyses. The categorical variables Generation, Gender, and SES are studied through one-way ANOVAs. The continuous variable Age is studied through bivariate Pearson correlations.

Recall from Chapter 3 that only four out of nine linguistic contexts are examined with respect to speakers’ Subjunctive rate. These four linguistic contexts were selected based on their popularity, or availability, in speech. Following Cohen (1992), only those linguistic contexts that appear in the speech of at least 80 informants are worth exploring in a regression analysis. See Chapter 3 for a complete explanation. As a reminder, the four most popular linguistic contexts are:

- The Modal context (e.g. Como quieran, lo que quieran, como que quieran)
- The Protasis Si context (e.g. Si vinieran, si hubieran venido)
- The Temporal context (e.g. Hasta que vengan, antes de que vengan, cuando vengan)
- The Apodosis Si context (e.g. Si..., quisiera ir, si..., hubiera querido ir)

Recall that overall Subjunctive rate refers to speakers’ Subjunctive rate in all nine linguistic contexts taken together (the nine linguistic contexts are listed in Chapter 3, as well as in Appendix A). In each table, statistically significant (and approaching) values are bolded. The level of significance (or p value) is represented in the following way:

*** $p < .001$

** $p < .01$

* $p < .05$

a $p < .1$
The following section explores the relationships between speakers’ overall Subjunctive rate and their Subjunctive rate in each of the four most popular linguistic contexts.

2.1.1 Correlations between overall Subjunctive rate and Subjunctive rates in the four most popular linguistic contexts

Table 4.1 below displays the series of bivariate Pearson correlations between overall Subjunctive rate and Subjunctive rate in the four most popular linguistic contexts. The table addresses the question of whether the speakers who tend to use more Subjunctives overall also tend to use them more in particular contexts. Each of the table’s horizontal tiers comprises two rows. Row 1 gives the Pearson correlation coefficient, and row 2 gives the number of complete pairwise observations (i.e. number of informants, or N). The Pearson correlation coefficient, also known as “Pearson’s r”, is a number between -1 and +1. A negative value indicates a negative correlation (meaning that the variables go in opposite directions), while a positive value represents a positive correlation (meaning that the variables tend to go in the same direction). The closer the coefficient is to -1 or to +1, the greater the strength of the association.
As is clear from Table 4.1, speakers’ overall Subjunctive rate is significantly correlated with their Subjunctive rate in three of the four most popular linguistic contexts. That is, there is a statistically significant linear relationship between the overall Subjunctive rate and the Subjunctive rate in the Modal context (e.g. Como quieran, lo que quieran, como que quieran), in the Protasis Si context (e.g. Si vinieran, si hubieran venido), and in the Apodosis Si context (e.g. Si..., quisiera ir, si..., hubiera querido ir). Moreover, there is a linear relationship that approaches statistical significance between the overall Subjunctive rate and the Subjunctive rate in the Temporal context (e.g. Hasta que vengan, antes de que vengan, cuando vengan). The direction of the relationship between overall Subjunctive rate and Subjunctive rate in each of the four most popular linguistic contexts is positive, meaning that these variables go in the same direction. The results tell us that speakers with a higher overall Subjunctive rate tend to be the same as those with higher rates in three of the contexts (Modal, Protasis Si, and Apodosis Si), and marginally so in the fourth (Temporal). Thus, the answer to (RQ1a) is affirmative; there
is indeed a significant association between speakers’ overall Subjunctive rate and their rates of Subjunctive in each of the four contexts.

2.1.2 Correlations between Subjunctive rates in the four most popular linguistic contexts

Table 4.2 below displays the series of bivariate Pearson correlations between Subjunctive rates in the four most popular linguistic contexts. This table addresses the research question (RQ1b) about whether there are significant associations between Subjunctive rates in the different linguistic contexts, that is, whether having a higher rate in one context tends to indicate that there will be a higher rate in another. The empty cells (N/A) correspond to unfeasible correlations (between a same context).

<table>
<thead>
<tr>
<th></th>
<th>Modal</th>
<th>Protasis Si</th>
<th>Temporal</th>
<th>Apodosis Si</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Modal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>N/A</td>
<td>-0.06</td>
<td>-0.09</td>
<td>-0.09</td>
</tr>
<tr>
<td><strong>Protasis Si</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>140</td>
<td>N/A</td>
<td>-0.04</td>
<td>0.37***</td>
</tr>
<tr>
<td><strong>Temporal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>130</td>
<td>129</td>
<td>N/A</td>
<td>0.29**</td>
</tr>
<tr>
<td><strong>Apodosis Si</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>129</td>
<td>129</td>
<td>120</td>
<td>N/A</td>
</tr>
</tbody>
</table>

*** p < .001  ** p < .01  * p < .05  a p < .1

There is a positive, statistically significant linear relationship between Subjunctive rate in the Temporal context (e.g. Hasta que vengan, antes de que vengan, cuando vengan) and Subjunctive rate in the Apodosis Si context (e.g. Si..., quisiera ir, si..., hubiera querido ir). The direction of the effect is positive. In other words, speakers with greater
Subjunctive rates in the Temporal context tend to have higher Subjunctive rates in the Apodosis *Si* context, and vice versa.

There is also a statistically significant linear relationship between Subjunctive rate in the Protasis *Si* context (e.g. *Si vinieran, si hubieran venido*) and Subjunctive rate in the Apodosis *Si* context. Subjunctive rate in these two linguistic contexts is positively correlated, meaning that speakers with higher Subjunctive rates in the Protasis *Si* context tend to also have higher Subjunctive rates in the Apodosis *Si* context, and vice versa.

None of the other relationships reach statistical significance. This means that, apart from the statistically significant ones described above, speakers’ Subjunctive rate in one linguistic context is not necessarily associated with their Subjunctive rate in another linguistic context.

Next, group differences in consultants’ Subjunctive rate, overall and in different linguistic contexts (the four most popular), are explored through socio-demographic features. This is done by means of a series of one-way ANOVAs (analyses of variance) with the categorical variables (Generation, Gender and SES), and a series of bivariate Pearson correlations with the continuous variable (Age).

### 2.1.3 Subjunctive rate by Generation, Gender and SES

Table 4.3 shows five one-way ANOVAs for Subjunctive rate (overall and in the four most popular linguistic contexts) by Generation, Gender and Socio-economic status (SES). The variable Age requires a distinct statistical analysis because it is continuous. That analysis is presented in the following subsection. Table 4.3 is addressing the question (RQ2) of whether a person's Generation, Gender, or SES influences the extent to which they use Subjunctives in general and in the four linguistic contexts under study. In
Table 4.3, each horizontal tier is divided into four rows. Row 1 presents the number of informants in each group (N), row 2 shows the Subjunctive rate (Mean), row 3 displays the Standard Deviation (SD), which is the standard measure of variation, and row 4 is the F-statistic, which is bolded when it yields or approaches significance.

Table 4-3
Subjunctive rate (overall and in four most popular contexts)
By Generation, Gender and SES

<table>
<thead>
<tr>
<th></th>
<th>Generation</th>
<th>Gender</th>
<th>SES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LARN</td>
<td>LARI</td>
<td>NYR</td>
</tr>
<tr>
<td>Overall N</td>
<td>39</td>
<td>78</td>
<td>25</td>
</tr>
<tr>
<td>Mean</td>
<td>11.7</td>
<td>9.9</td>
<td>10.1</td>
</tr>
<tr>
<td>SD</td>
<td>9.2</td>
<td>8.5</td>
<td>8.2</td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modal N</td>
<td>39</td>
<td>77</td>
<td>25</td>
</tr>
<tr>
<td>Mean</td>
<td>10.8</td>
<td>8.8</td>
<td>10.3</td>
</tr>
<tr>
<td>SD</td>
<td>14.2</td>
<td>10.9</td>
<td>12.5</td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protasis Si N</td>
<td>38</td>
<td>77</td>
<td>25</td>
</tr>
<tr>
<td>Mean</td>
<td>12.1</td>
<td>8</td>
<td>5.8</td>
</tr>
<tr>
<td>SD</td>
<td>16.7</td>
<td>9.9</td>
<td>9.8</td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Temporal Si N</td>
<td>37</td>
<td>72</td>
<td>22</td>
</tr>
<tr>
<td>Mean</td>
<td>5.9</td>
<td>10.7</td>
<td>7.4</td>
</tr>
<tr>
<td>SD</td>
<td>10.4</td>
<td>16.2</td>
<td>21.8</td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apodosis Si N</td>
<td>36</td>
<td>70</td>
<td>23</td>
</tr>
<tr>
<td>Mean</td>
<td>3.5</td>
<td>6.8</td>
<td>9.2</td>
</tr>
<tr>
<td>SD</td>
<td>7.5</td>
<td>15.8</td>
<td>22.4</td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*** p < .001   ** p < .01   * p < .05   a p < .1

With respect to Generation, the table shows that there are no statistically significant differences between group means in overall Subjunctive rate, Subjunctive rate in the Modal context, Subjunctive rate in the Temporal context, and Subjunctive rate in the
Apodosis Si context as determined by a series of one-way ANOVAs. With respect to Subjunctive rate in the Protasis Si context by Generation, the group mean differences approach statistical significance as determined by one-way ANOVA \((F (2, 137) = 2.4, p = .098)\). A Tukey post hoc test reveals that the difference that approaches statistical significance \((p = .1)\) is between the LARN (Newcomers) and the NYR (New York Raised). The Newcomers’ Subjunctive rate is over six percentage points higher than that of the New York Raised in the Protasis Si context (see row 2 for the Mean).

Regarding Gender, Table 4.3 shows that there are no statistically significant differences between any of the group means in Subjunctive rate in the Protasis Si context, or in the Temporal context, or in the Apodosis Si context. On the other hand, there are statistically significant differences between group means with respect to men and women’s overall Subjunctive rate \((F (1, 140) = 4.7, p < .05)\), and their Subjunctive rate in the Modal context \((F (1, 139) = 5, p < .05)\). In both cases, men have a higher Subjunctive rate than women. Men’s Subjunctive rate is nearly three percentage points higher than that of women’s overall, and it is four and half percentage points higher than that of women’s in the Modal context (e.g. Como quieran, lo que quieran, como que quieran).

Concerning Socio-economic status, Table 4.3 shows that there are no statistically significant differences between group means in overall Subjunctive rate, Subjunctive rate in the Modal context, Subjunctive rate in the Temporal context, and Subjunctive rate in the Apodosis Si context as determined by a series of one-way ANOVAs. With respect to the Protasis Si context, the differences between group means approaches statistical significance \((F (1, 135) = 3.3, p = .072)\). Low SES informants’ Subjunctive rate is nearly
four percentage points higher than that of Middle SES informants in the Protasis *Si* context (e.g. *Si vinieran, si hubieran venido*).

To summarize, statistically significant results point to the fact that men use the Subjunctive more than women overall, and in the Modal context (e.g. *Como quieran, lo que quieran, como que quieran*). Results that are marginally significant indicate that, in the Protasis *Si* context (e.g. *Si vinieran, si hubieran venido*), the Newcomers use the Subjunctive more than the New York Raised, and Low SES participants employ the Subjunctive more than Middle SES participants.

### 2.1.4 Subjunctive rate by Age

Table 4.4 below presents a bivariate Pearson correlation between Subjunctive rate (overall and in the four most popular linguistic contexts) and Age. Table 4.4 attends to the question (RQ2) of whether speakers’ age shapes the extent to which they use Subjunctives in general and in the four linguistic contexts under study. The Pearson correlation coefficient is in row 1, and the number of complete pairwise observations (i.e. number of informants, or N) is in row 2.

<table>
<thead>
<tr>
<th>Age</th>
<th>Correlation</th>
<th>Overall</th>
<th>Modal</th>
<th>Protasis <em>Si</em></th>
<th>Temporal</th>
<th>Apodosis <em>Si</em></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>-0.07</td>
<td>-0.07</td>
<td>0.13</td>
<td>0.04</td>
<td>0.09</td>
</tr>
<tr>
<td>N</td>
<td>142</td>
<td>141</td>
<td>140</td>
<td>131</td>
<td>129</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.4 establishes that the linear relationship between Age and Subjunctive rate (overall and in each of the four most popular linguistic contexts) is not statistically

---

112
significant. This means that there is no association between an informant’s age and his/her rates of Subjunctive, overall or in any of the four linguistic contexts.

The following section (section 2.2) examines the results of the bivariate tests pertaining to Linguistic context availability.

2.2 Bivariate analysis of Linguistic context availability

This section reports on the possible associations between the extents of availability of different linguistic contexts (section 2.2.1), and between the availability of different linguistic contexts and speakers’ socio-demographic variables (categorical variables in section 2.2.2, and continuous in section 2.2.3). Recall that only the linguistic contexts that appeared \textit{and} the ones that did not appear in over 10 percent of the informants (rounded up to N =15) were selected for the analysis of Linguistic context availability. This was so because only those linguistic contexts with a sufficient number of informants in each group (users and non-users) were deemed viable contenders for the statistical analyses. The four most popular linguistic contexts, explored in section 2.1, are not the same as the four linguistic contexts examined with respect to Linguistic context availability. This is because the most popular contexts have fewer than 15 informants in the non-user group. See Chapter 3 for the complete discussion. The following four linguistic contexts were chosen for the analysis of Linguistic context availability:

- Volition context (e.g. \textit{Quiere que vengan, espera que vengan})
- Hypothetical \textit{Como si} context (e.g. \textit{Como si vinieran})
- Possibility context (e.g. \textit{Es posible que quieran, tal vez/quizá(s) quieran, a lo mejor quieran})
- Concessive context (e.g. \textit{Aunque quieran})
Furthermore, the same four socio-demographic variables (Generation, Gender, SES and Age) that were exploited with respect to Subjunctive rate are employed with respect to Linguistic context availability. Just as in the previous sections, levels of significance (p values) are denoted in bold and by asterisk(s) or a superscript a.

2.2.1 Relationships between the extents of availability of different linguistic contexts

Table 4.5 and Table 4.6 below display the results of twelve cross-tabulations (Chi-square tests of independence) between the extents of availability of four linguistic contexts. The summary of these cross-tabulations is divided into two tables for ease of presentation. The results of the cross-tabulations correspond to the research question (RQ1c) on the relationship between the extents of availability of different linguistic contexts in speech. The question is whether the availability of a particular linguistic context is significantly associated with the availability of a different linguistic context. For each cross-tabulation between two linguistic contexts, the Yes column gives the number (N) and percentage (in parentheses) of consultants in whose speech the linguistic context is available, and the No column represents the number (N) and percentage in parentheses of consultants in whose speech the linguistic context is not available. The third row (“Total”) corresponds to the total number (N) and percentage (in parentheses) of informants for each condition. The third column shows the Chi Square (X²) statistic, bolded when it reaches or approaches statistical significance. Note that there is only one such measure that is bolded in the table. The empty cells (N/A) correspond to unfeasible cross-tabulations (between a same linguistic context).
### Table 4-5
Associations between extents of Linguistic context availability (Part 1)

<table>
<thead>
<tr>
<th></th>
<th>Volition</th>
<th></th>
<th>Hypothetical <em>Como si</em></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>N (%)</td>
<td>N (%)</td>
<td>X²</td>
<td>N (%)</td>
</tr>
<tr>
<td>Volition</td>
<td>Yes</td>
<td>2 (1.4)</td>
<td>17 (12.0)</td>
<td>2 (1.4)</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>15 (10.6)</td>
<td>109 (76.7)</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>17 (12.0)</td>
<td>125 (88.0)</td>
<td></td>
</tr>
<tr>
<td>Hypothetical Como si</td>
<td>Yes</td>
<td>2 (1.4)</td>
<td>15 (10.6)</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>16 (11.3)</td>
<td>109 (76.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>18 (12.7)</td>
<td>124 (87.3)</td>
<td></td>
</tr>
<tr>
<td>Possibility</td>
<td>Yes</td>
<td>6 (4.2)</td>
<td>39 (27.5)</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>12 (8.5)</td>
<td>85 (55.9)</td>
<td>9 (6.3)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>18 (12.7)</td>
<td>124 (87.4)</td>
<td>17 (12.0)</td>
</tr>
<tr>
<td>Concessive</td>
<td>Yes</td>
<td>9 (6.3)</td>
<td>52 (36.6)</td>
<td>0.42</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>9 (6.3)</td>
<td>72 (50.7)</td>
<td>10 (7.0)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>18 (12.7)</td>
<td>124 (87.3)</td>
<td>17 (12.0)</td>
</tr>
</tbody>
</table>

*** p < .001  ** p < .01  * p < .05  a p < .1

### Table 4-6
Associations between extents of Linguistic context availability (Part 2)

<table>
<thead>
<tr>
<th></th>
<th>Possibility</th>
<th>Concessive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>N (%)</td>
<td>N (%)</td>
</tr>
<tr>
<td>Volition</td>
<td>Yes</td>
<td>6 (4.2)</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>39 (27.5)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>45 (31.7)</td>
</tr>
<tr>
<td>Hypothetical Como si</td>
<td>Yes</td>
<td>8 (5.6)</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>37 (26.1)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>45 (31.7)</td>
</tr>
<tr>
<td>Possibility</td>
<td>Yes</td>
<td>N/A</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td>Concessive</td>
<td>Yes</td>
<td>25 (17.6)</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>20 (14.1)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>45 (31.7)</td>
</tr>
</tbody>
</table>

*** p < .001  ** p < .01  * p < .05  a p < .1
The association between the availability of the Possibility context (e.g. *Es posible que, tal vez, quizá(s), a lo mejor vengan*) and the availability of the Concessive context (e.g. *Aunque vengan*) is statistically significant, \(X^2\ (1, N = 142) = 4.27, p < .05\). This means that the speakers in whom the Possibility context is available tend to be the same as those in which the Concessive context is available. None of the other Chi-square tests in Tables 4.5 and 4.6 yield statistical significance. This means that there is no association between the extents of availability of those different linguistic contexts.

The following section delves into the possible relations between Linguistic context availability and speakers’ socio-demographic features.

**2.2.2 Linguistic context availability by Generation, Gender and SES**

Table 4.7 and Table 4.8, which are organized like Tables 4.5 and 4.6 above, display the cross-tabulations (Chi-square tests of independence) between Linguistic context availability and Generation, Gender and Socio-economic status (SES). These cross-tabulations are divided into two tables for ease of presentation. The results of the cross-tabulations (Chi-square tests) answer the question (RQ3) about whether speakers’ generation, gender and SES affect the availability of each of the four contexts. In other words, we want to know who tends to use each context more and who tends to use each one less.
### Table 4-7
Linguistic context availability
By Generation, Gender and SES (Part 1)

<table>
<thead>
<tr>
<th></th>
<th>Volition</th>
<th>Hypothetical Como si</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes N (%)</td>
<td>No N (%)</td>
</tr>
<tr>
<td><strong>Generation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LARN</td>
<td>7 (4.9)</td>
<td>32 (22.5)</td>
</tr>
<tr>
<td>LARI</td>
<td>7 (4.9)</td>
<td>71 (50.0)</td>
</tr>
<tr>
<td>NYR</td>
<td>4 (2.8)</td>
<td>21 (14.8)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>18 (12.7)</td>
<td>124 (87.3)</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>7 (4.9)</td>
<td>72 (50.7)</td>
</tr>
<tr>
<td>Female</td>
<td>11 (7.7)</td>
<td>52 (36.6)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>18 (12.7)</td>
<td>124 (87.3)</td>
</tr>
<tr>
<td><strong>SES</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>8 (5.8)</td>
<td>49 (35.3)</td>
</tr>
<tr>
<td>Middle</td>
<td>10 (7.2)</td>
<td>72 (51.8)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>18 (12.9)</td>
<td>121 (87.1)</td>
</tr>
</tbody>
</table>

*** p < .001  ** p < .01  * p < .05  a p < .1

### Table 4-8
Linguistic context availability
By Generation, Gender and SES (Part 2)

<table>
<thead>
<tr>
<th></th>
<th>Possibility</th>
<th>Concessive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes N (%)</td>
<td>No N (%)</td>
</tr>
<tr>
<td><strong>Generation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LARN</td>
<td>16 (11.3)</td>
<td>23 (16.2)</td>
</tr>
<tr>
<td>LARI</td>
<td>25 (17.6)</td>
<td>53 (37.3)</td>
</tr>
<tr>
<td>NYR</td>
<td>4 (2.8)</td>
<td>21 (14.8)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>45 (31.7)</td>
<td>97 (68.3)</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>26 (18.3)</td>
<td>53 (37.3)</td>
</tr>
<tr>
<td>Female</td>
<td>19 (13.4)</td>
<td>44 (31.0)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>45 (31.7)</td>
<td>97 (68.3)</td>
</tr>
<tr>
<td><strong>SES</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>15 (10.8)</td>
<td>42 (30.2)</td>
</tr>
<tr>
<td>Middle</td>
<td>28 (20.1)</td>
<td>54 (38.8)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>43 (30.9)</td>
<td>96 (69.1)</td>
</tr>
</tbody>
</table>

*** p < .001  ** p < .01  * p < .05  a p < .1

Tables 4.7 and 4.8 show that the Chi-square tests performed to examine the relationship between Generation and the availability of each of four linguistic contexts do not yield
statistical significance. There is no association between speakers’ generation and the availability of the Volition context, or of the Hypothetical Como si context, or of the Possibility context, or of the Concessive context.

While no statistically significant association is found between Gender and the availability of three linguistic contexts (the Volition context, the Hypothetical Como si context, and the Possibility context), the cross-tabulation between Gender and the availability of the Concessive context (e.g. Aunque quieran) yields statistical significance. This tells us that there is an association between speakers’ gender and the availability of this linguistic context.

None of the Chi-square tests of association between Socio-economic status (SES) and any of the Available contexts are statistically significant. There is no relation between speakers’ socio-economic status and the availability of any of the four linguistic contexts.

### 2.2.3 Linguistic context availability by Age

Table 4.9 displays the summary of four independent samples t-tests of Linguistic context availability by Age (one t-test per linguistic context). Each horizontal tier is divided into two rows: the Yes row concerns the informants for whom the linguistic context is available, and the No row shows the informants for whom it is not. Column 1 represents the number (N) of informants in each group; column 2 gives the Mean age of the informants; column 3 presents the Standard Deviation (SD), and column 4 gives the t statistic, bolded when significant. Table 4.9 corresponds to the same research question (RQ3) as Tables 4.7 and 4.8 above, with a focus on speakers’ age here. That is to say, we explore whether speakers’ age affects the availability of each of four contexts.
Three of the four independent samples t-tests fail to yield statistical significance, meaning that an informant’s age does not shape the availability for them of the Hypothetical \textit{Como si} context (e.g. \textit{Como si vinieran}) or of the Possibility context (e.g. \textit{Es posible que quieran, tal vez/quizá(s) quieran, a lo mejor quieran}) or of the Concessive context (e.g. \textit{Aunque quieran}).

A statistically significant difference \textit{is} found in Age for the Volition context ($M=29.8$, $SD=8.4$) and the no Volition context conditions\textsuperscript{40} ($M=35.3$, $SD=14.4$); $t(33.7)=2.33$, $p<.05$. In other words, the informants’ age shapes the availability to him or her of the Volition context (e.g. \textit{ Quiere que vengan, espera que vengan}). Younger informants (whose mean age is a little under 30) have a significantly greater tendency to draw on these clauses than older informants (whose mean age is slightly above 35).

\textsuperscript{40} Both conditions can be found in all of the tables that display the results of independent samples t-tests. For instance, in Table 4.9, the Volition context condition is presented in the column marked “Yes” (where the Mean is 29.8 and the SD is 8.4) and the no Volition context condition appears in the column marked “No” (where the Mean is 35.3 and the SD is 14.4).
2.3 Bivariate inquiry into Subjunctive rate by Linguistic context availability

A series of independent samples t-tests were performed in order to answer the research question (RQ1d) on how Linguistic context availability (i.e. the availability of four linguistic contexts) affects speakers’ Subjunctive rate (overall and in the four most popular linguistic contexts). That is, we want to look into the question of whether speakers who use, or have available to them, more linguistic contexts tend to be the same who use more Subjunctives. The summary of the five independent samples t-tests is introduced below, in Table 4.10, whose format is similar to the ones we've just seen. The availability of each linguistic context (Volition, Hypothetical Como si, Possibility and Concessive) is represented in the columns. Column 1, marked Yes, gives the number of consultants for whom the linguistic context is available. Column 2, marked No, gives the number for whom it is not available. The rows correspond to speakers’ overall Subjunctive rate and Subjunctive rate in the four most popular linguistic contexts (Modal, Protasis Si, Temporal, and Apodosis Si). Row 1 gives the number of informants (N); row 2 the Mean Subjunctive rate; row 3 the Standard Deviation (SD), and row 4 the t statistic.
Table 4-10
Subjunctive rate by Linguistic context availability

<table>
<thead>
<tr>
<th></th>
<th>Volition</th>
<th>Hypothetical Como si</th>
<th>Possibility</th>
<th>Concessive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td><strong>Overall</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>18</td>
<td>124</td>
<td>17</td>
<td>125</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>11.5</td>
<td>10.3</td>
<td>11.4</td>
<td>10.3</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>5.5</td>
<td>9</td>
<td>10.9</td>
<td>8.3</td>
</tr>
<tr>
<td><strong>t</strong></td>
<td>-0.56</td>
<td>-0.49</td>
<td>0.79</td>
<td>-0.84</td>
</tr>
<tr>
<td><strong>Modal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>18</td>
<td>123</td>
<td>16</td>
<td>125</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>13.2</td>
<td>9.1</td>
<td>6.5</td>
<td>10</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>11.4</td>
<td>12.2</td>
<td>5.5</td>
<td>12.7</td>
</tr>
<tr>
<td><strong>t</strong></td>
<td>-1.33</td>
<td><strong>1.98</strong></td>
<td>1.03</td>
<td>-0.32</td>
</tr>
<tr>
<td><strong>Protasis Si</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>18</td>
<td>122</td>
<td>16</td>
<td>124</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>7.4</td>
<td>8.9</td>
<td>7.6</td>
<td>8.9</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>7.6</td>
<td>12.8</td>
<td>11.1</td>
<td>12.4</td>
</tr>
<tr>
<td><strong>t</strong></td>
<td>0.50</td>
<td>0.41</td>
<td>0.14</td>
<td>-1.08</td>
</tr>
<tr>
<td><strong>Temporal</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>17</td>
<td>114</td>
<td>15</td>
<td>116</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>6.2</td>
<td>9.2</td>
<td>6.9</td>
<td>9.1</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>10.9</td>
<td>16.6</td>
<td>10.5</td>
<td>15.9</td>
</tr>
<tr>
<td><strong>t</strong></td>
<td>0.72</td>
<td>0.51</td>
<td>0.12</td>
<td><strong>-1.76</strong></td>
</tr>
<tr>
<td><strong>Apodosis Si</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>18</td>
<td>111</td>
<td>15</td>
<td>114</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td>6.6</td>
<td>6.3</td>
<td>6.9</td>
<td>6.2</td>
</tr>
<tr>
<td><strong>SD</strong></td>
<td>8.5</td>
<td>16.4</td>
<td>11.8</td>
<td>15.9</td>
</tr>
<tr>
<td><strong>t</strong></td>
<td>-0.07</td>
<td>-0.14</td>
<td><strong>-0.22</strong></td>
<td>0.69</td>
</tr>
</tbody>
</table>

*** p < .001  ** p < .01  * p < .05  ^ p < .1

Table 4.10 shows that there is no statistically significant difference in overall Subjunctive rate, or Subjunctive rate in the Protasis Si context, or in the Temporal context, or in the Apodosis Si context, for the Volition context and the no Volition context conditions; for the Hypothetical Como si context and the no Hypothetical Como si context conditions; for the Possibility context and the no Possibility context conditions; and for the Concessive context and the no Concessive context conditions. These results tell us that the availability of these linguistic contexts does not shape speakers’ Subjunctive rates overall and in any of the linguistic contexts cited above. In other words, the frequency
with which speakers use the Subjunctive is unrelated to whether or not these linguistic contexts appear in their speech.

Furthermore, there is no statistically significant difference in Subjunctive rate in the Modal context (e.g. Como quieran, lo que quieran, como que quieran) for the Volition context and the no Volition context conditions; or for the Possibility context and the no Possibility context conditions; or for the Concessive context and the no Concessive context conditions. Here too, the amount of Subjunctive verbs that speakers employ in the Modal context does not depend on the occurrence or lack of occurrence of these linguistic contexts.

A statistically significant difference is found in Subjunctive rate in the Modal context (e.g. Como quieran, lo que quieran, como que quieran) for the Hypothetical Como si context (M= 6.5, SD= 5.5) and the no Hypothetical Como si context conditions (M=10, SD=12.7); t (40) = 1.98, p = .05. There is a difference in percentage of Subjunctives in the Modal context between when the Hypothetical Como si context is available and when it is not available. Indeed, speakers’ Subjunctive rate in clauses introduced by como, lo que and como que is lower (by 3.5 percentage points) when como si appears in their speech. This means that speakers who tend to say como si use fewer Subjunctives after como, lo que and como que than speakers who don’t tend to say como si.

Moreover, a difference that approaches statistical significance is found in Subjunctive rate in the Temporal context for the Concessive context condition (M=11.5, SD= 16.2) and the no Concessive context condition (M=6.6, SD=15.5); t (129) = -1.76, p = .08. Thus, there is a difference in percentage of Subjunctives in the Temporal context
(e.g. *Hasta que vengan, antes de que vengan, cuando vengan*) between when the Concessive context (e.g. *Aunque quieran*) is available and when it *is not* available. Speakers’ Subjunctive rate in Temporal contexts is nearly five percentage points higher when *aunque* clauses appear in their speech than when they do not. In other words, speakers who tend to say *aunque* use more Subjunctives after *hasta que, antes de que* and *cuando* than speakers who don’t tend to say *aunque*.

### 2.4 Bivariate analyses: Summary

The results of all of the bivariate analyses presented in the previous sections offer preliminary answers to the present study’s three main research questions (research question 1 was divided into four parts (a, b, c and d)). The results of bivariate Pearson correlations resolved parts (a) and (b) of the first research question, about the relationships between different rates of Subjunctive overall and in the four most popular linguistic contexts (sections 2.1.1 and 2.1.2). It was discovered, unsurprisingly, that overall Subjunctive rate is associated with speakers’ Subjunctive rates in all four linguistic contexts. In other words, speakers who tend to employ Subjunctive verbs in those four contexts also tend to use the mood overall (in all nine contexts). On the other hand, it was striking to discover that speakers who tend to use more Subjunctives in the Apodosis *Si* context (e.g. *Si..., quisiera ir, si..., hubiera querido ir*) also tend to use more Subjunctives both in the Protasis *Si* context (e.g. *Si vinieran, si hubieran venido*) and in the Temporal context (e.g. *Hasta que vengan, antes de que vengan, cuando vengan*), and vice versa.

The second research question (RQ2), about which socio-demographic variables affect consultants’ Subjunctive rate (overall and in the four most popular linguistic
contexts), was examined through a series of one-way ANOVAs and bivariate Pearson correlations (sections 2.1.3 and 2.1.4). While speakers’ age is not determinant of their rates of Subjunctive in any context or overall, their gender, generation and socio-economic status are. Indeed, speakers’ gender is strongly associated with their overall Subjunctive rate and their Subjunctive rate in the Modal context (e.g. *Como quieran, lo que quieran, como que quieran*), while their generation and socio-economic status are marginally related to their Subjunctive rate in the Protasis *Si* context (e.g. *Si vinieran, si hubieran venido*). Women draw on the Subjunctive less than men overall and in the Modal context, while Newcomer and Low SES participants employ the Subjunctive more than New York Raised and Middle SES participants, respectively, in the Protasis *Si* context.

Part (c) of the first question concerned the connections between the extents of availability of four different linguistic contexts. The results of twelve cross-tabulations (Chi-square tests) established (in section 2.2.1) that there is a statistically significant association between the extents of availability of two linguistic contexts: the Possibility context (e.g. *Es posible que quieran, tal vez/quizás quieran, a lo mejor quieran*) and the Concessive context (e.g. *Aunque quieran*). Speakers who tend to say *es posible que, tal vez, quizás*, and *a lo mejor* also tend to say *aunque*.

Cross-tabulations (Chi-square tests) and independent samples t-tests were also drawn on to determine the possible relationships between speakers’ socio-demographic features and the availability of four linguistic contexts (sections 2.2.2 and 2.2.3). This constituted our third research question (RQ3). It turns out that speakers’ gender is significantly related to the availability of the Concessive context (e.g. *Aunque quieran*),
and that age shapes the availability of the Volition context (e.g. *Quiere que vengan, espera que vengan*).

Finally, a series of independent samples t-tests were conducted to explore how the availability of four linguistic contexts might affect speakers’ Subjunctive rate, overall and in the four most popular linguistic contexts (section 2.3). In answer to part (d) of our first research question, we discovered that the availability of the Hypothetical *Como si* context significantly shapes Subjunctive rate in the Modal context (e.g. *Como quieran, lo que quieran, como que quieran*), and that the availability of the Concessive context (e.g. *Aunque quieran*) slightly influences speakers’ Subjunctive rate in the Temporal context (e.g. *Hasta que vengan, antes de que vengan, cuando vengan*).

The results of bivariate tests represent a reliable first step; they serve as a good indication of how the research questions will ultimately be resolved. Yet, bivariate statistics only offer a partial understanding of the data, as they only inform about the association between two variables. To obtain a more complete picture of how the independent variables shape Subjunctive rate and Linguistic context availability (the two dependent variables), we must conduct multivariate regression analyses, which consider the contribution of the different predictor variables together (rather than one at a time). This is what the following section aims to do.

3. **Multivariate results**

The current investigation’s two main research questions (RQ2 and RQ3) concern the effect of speakers’ diverse socio-demographic characteristics on their rates of Subjunctive (discussed in section 3.1), and on the availability of specific linguistic contexts in speech (covered in section 3.2). To answer the first main research question (RQ2), we run five
linear regressions, given that the dependent variable, Subjunctive rate, is continuous. Rates of Subjunctive vary depending on whether they refer to speakers’ overall Subjunctive rate (i.e. the percentage of Subjunctive verbs in nine linguistic contexts, in the 142 participants), or to their Subjunctive rate in each of the four most popular linguistic contexts. Recall that the four most popular linguistic contexts are the following (see Chapter 3 for the full discussion on how they were selected):

- The Modal context (e.g. *Como quieran, lo que quieran, como que quieran*)
- The Protasis *Si* context (e.g. *Si vinieran, si hubieran venido*)
- The Temporal context (e.g. *Hasta que vengan, antes de que vengan, cuando vengan*)
- The Apodosis *Si* context (e.g. *Si..., quisiera ir, si... hubiera querido ir*)

Four logistic regressions were performed to answer the second core research question (RQ3), on how socio-demographic variables might affect the availability of four distinct linguistic contexts in speech. Logistic regressions are warranted since the dependent variable, Linguistic context availability, is categorical. Chapter 3 specifies how these four linguistic contexts were selected. Recall that the four linguistic contexts in question are the following:

- Volition context (e.g. *Quiere que vengan, espera que vengan*)
- Hypothetical *Como si* context (e.g. *Como si vinieran*)
- Possibility context (e.g. *Es posible que quieran, tal vez/quizá(s) quieran, a lo mejor quieran*)
- Concessive context (e.g. *Aunque quieran*)

The same independent variables are exploited in the linear and in the logistic regressions. The bivariate analyses presented in section 2 included four socio-
demographic variables: Generation, Gender, Socio-economic status (SES) and Age. Yet, we selected five independent external or socio-demographic variables for the regressions. As explained in Chapter 3, having three constraints (e.g., Newcomers, Established Immigrants, and New York Raised) is unviable for regression analysis. We therefore dummy coded the two variables Established Immigrants (LARI) and New York Raised (NYR). The Newcomers (LARN) became the reference category. See Appendix B for details. As a result, the linear and logistic regressions comprise five independent variables:

- LARI
- NYR
- Gender
- SES
- Age

Moreover, in the multivariate analyses, we add two independent variables in the form of interactions\(^{41}\) between LARI and Female, and between NYR and Female. Our seven independent external variables are the five listed above (LARI, NYR, Gender, SES and Age) in addition to the following two (asterisk indicate interacting variables):

- LARI*Female
- NYR*Female

Reference categories, such as the Newcomers (LARN), do not appear in regressions. Thus, the linear and logistic regressions only show the results of two Generation variables: LARI and NYR. However, the results for the LARN can be

\(^{41}\) See Chapter 3 for an explanation regarding the choice of interaction variables.
inferred from those of the LARI and the NYR. Similarly, the results of the interaction
between LARN and Female (i.e. Newcomer women) can be gathered from the results of
the two interaction variables (LARI*Female and NYR*Female) included in the
regressions.

As discussed in Chapter 3, all of the independent variables are categorical except
for Age, which is a continuous variable. Each categorical variable has two factors (or
constraints), while the continuous variable presents each informant’s exact age. The
categorical variables are binary, meaning that each one has two factors. They are the
following:

- **LARI:**
  - Not LARI\(^{42}\)
  - LARI

- **NYR:**
  - Not NYR
  - NYR

- **Gender:**
  - Male
  - Female

- **SES:**
  - Low
  - Middle

- **Age:**
  - Continuous

- **LARI*Female:**
  - Not LARI Female
  - LARI Female

- **NYR*Female:**
  - Not NYR Female
  - NYR Female

Each linear and logistic regression is divided into two Models\(^{43}\), with five
variables entered in Model 1 (LARI, NYR, Gender, SES, Age), and the same five

\(^{42}\) As explained in Appendix B, the factors “Not LARI” and “Not NYR” stand for the LARN (i.e.
Newcomers), which is the reference category. Similarly, the factors “Not LARI Female” and “Not NYR
Female” denote LARN Female informants (i.e. Newcomer women).
variables plus the two interaction variables (NYR*Female and LARI*Female) entered in Model 2. The five variables in Model 1 probe for main effects; the interaction variables probe for associations that may hold between Gender (Female) and the two generational variables, LARI and NYR. Model 2 is analyzed first because the two interaction variables offer more information than the three individual variables (LARI, NYR, Gender) analyzed separately. If at least one of the interaction variables yields statistical significance, we adopt regression Model 2. However, if neither of the interaction variables reaches statistical significance, we set aside Model 2 and adopt regression Model 1. The decision to analyze the results of Model 1 versus Model 2 is therefore entirely based on whether the interaction variables are significant, rather than on the level of significance of any of the other variables. Values that yield or approach statistical significance are bolded in all tables.

The following section (section 3.1) presents the multivariate analysis pertaining to Subjunctive rate.

3.1 Multivariate analysis of Subjunctive rates

Five linear regressions were run in order to answer the second research question (RQ2) on whether and how much the independent external variables affect the informants’ overall Subjunctive rate, as well as their Subjunctive rate in the four most popular linguistic contexts (the Modal context, the Protasis Si context, the Temporal context, and the Apodosis Si context). The R-square ($R^2$) below each linear regression shows the amount of variance of Y (the dependent variable) explained by X (the independent

---

43 The current investigation takes a predictive approach to regression analysis. This means that each Model’s level of significance is not of import here. Following convention, the $p$ value for each Model is presented under the linear and logistic regression tables. However, in a predictive approach, only the independent variables’ levels of significance matter.
variables). Each linear regression has three columns: column 1 shows the
Unstandardized B (which predicts the dependent variable from the independent variable),
column 2 is the Standard Error (SE) coefficient, and column 3 is the level of significance
(or p value), which is bolded when $p < .05$ (is statistically significant) or when $p < .1$
(approaches statistical significance). The first row in every linear regression is the
Constant, also known as the y-intercept. The Constant is the expected mean value of Y
when all $X = 0$. For example, we coded the variable Gender as $0 =$ Male and $1 =$ Female,
and the variable SES as $0 =$ Low SES and $1 =$ Middle SES. The code of the reference
category, or LARN (Newcomer), is $X = 0$. The Constant is a mean, so the Constant in
row 1 column 1 is the rate of Subjunctive verbs when all $X = 0$, meaning that the value
denotes the Subjunctive rate for an imaginary participant who is a Newcomer male of low
socio-economic status. As the Unstandardized B values represent the mean differences
between the Constant ($X = 0$) and $X=1$ for the categorical variables, it is possible to infer
the remaining participants’ Subjunctive rates, i.e. the Subjunctive rates of the LARI,
NYR, women, and the Middle SES informants. Finally, if $X$ never equals 0, as is the
case with our Age variable, then the intercept (or Constant) has no intrinsic meaning (as
no informant is 0 years old). This will all become clearer to the reader with specific
examples in the linear regressions presented below.

### 3.1.1 Gender affects Subjunctive rate, overall and in the Modal context

Tables 4.11 and 4.12 below display the results of a linear regression where the dependent
variable is overall Subjunctive rate, with Table 4.11 corresponding to Model 2 of the
regression and Table 4.12 corresponding to Model 1. The independent variables in Table
4.11 (Model 2) are LARI, NYR, Gender, SES, Age, NYR*Female and LARI*Female.
The independent variables in Table 4.12 (Model 1) are the same minus the two interaction variables. That is, in Table 4.12 (Model 1) the independent variables are LARI, NYR, Gender, SES and Age.

### Table 4-11
Overall Subjunctive rate (Model 2)
All speakers

<table>
<thead>
<tr>
<th></th>
<th>Unstandardized B</th>
<th>Coefficient SE</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>12.99</td>
<td>2.99</td>
<td>0.00</td>
</tr>
<tr>
<td>LARI</td>
<td>-1.27</td>
<td>2.41</td>
<td>0.60</td>
</tr>
<tr>
<td>NYR</td>
<td>0.59</td>
<td>3.33</td>
<td>0.86</td>
</tr>
<tr>
<td>Gender</td>
<td>-1.96</td>
<td>2.85</td>
<td>0.49</td>
</tr>
<tr>
<td>SES</td>
<td>-0.20</td>
<td>1.59</td>
<td>0.90</td>
</tr>
<tr>
<td>Age</td>
<td>-0.01</td>
<td>0.06</td>
<td>0.87</td>
</tr>
<tr>
<td>NYR*Female</td>
<td>-3.53</td>
<td>4.55</td>
<td>0.44</td>
</tr>
<tr>
<td>LARI*Female</td>
<td>-1.28</td>
<td>3.52</td>
<td>0.72</td>
</tr>
</tbody>
</table>

$R^2 = 0.047 \quad F(7, 131) = 0.915, p = .497$

Table 4.11 above (Model 2) shows that neither of the interaction variables yields statistical significance. We must therefore consider Model 1 (Table 4.12 below).

### Table 4-12
Overall Subjunctive rate (Model 1)
All speakers

<table>
<thead>
<tr>
<th></th>
<th>Unstandardized B</th>
<th>Coefficient SE</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>13.68</td>
<td>2.70</td>
<td>0.00</td>
</tr>
<tr>
<td>LARI</td>
<td>-1.88</td>
<td>1.81</td>
<td>0.30</td>
</tr>
<tr>
<td>NYR</td>
<td>-1.31</td>
<td>2.24</td>
<td>0.56</td>
</tr>
<tr>
<td>Gender</td>
<td>-3.28</td>
<td>1.52</td>
<td>0.03</td>
</tr>
<tr>
<td>SES</td>
<td>-0.12</td>
<td>1.57</td>
<td>0.94</td>
</tr>
<tr>
<td>Age</td>
<td>-0.01</td>
<td>0.06</td>
<td>0.83</td>
</tr>
</tbody>
</table>

$R^2 = 0.042 \quad F(5, 133) = 1.173, p = .326$
Table 4.12 (Model 1) shows that one of the independent variables, Gender, yields statistical significance ($p < .05$). In fact, we notice (via the Unstandardized B) that there is a 3.28 percentage point difference in men and women’s rates of Subjunctive, with men drawing on the mood more than three times as much as women. While men employ the Subjunctive 13.68 percent of the time (see the Constant), women do so just 10.4 percent of the time. Put differently, women draw on the Indicative mood more than men, overall. Furthermore, the fact that none of the other variables (LARI, NYR, SES, Age) reach statistical significance signals that only speakers’ gender shapes their overall Subjunctive rate.

The same research question is examined with respect to speakers’ Subjunctive rate in the Modal context (introduced by *Como, lo que, como que*), the most popular linguistic context. In order to find out which socio-demographic features shape speakers’ Subjunctive rate in the Modal context, we run a linear regression with two Models. Table 4.13 (below) shows the results of the linear regression (Model 2) of Subjunctive rate in the Modal clause by LARI, NYR, Gender, SES, Age, NYR*Female and LARI*Female.

<table>
<thead>
<tr>
<th></th>
<th>Unstandardized B</th>
<th>Coefficient SE</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Constant</strong></td>
<td>11.98</td>
<td>4.18</td>
<td>0.01</td>
</tr>
<tr>
<td><strong>LARI</strong></td>
<td>-2.16</td>
<td>3.39</td>
<td>0.53</td>
</tr>
<tr>
<td><strong>NYR</strong></td>
<td>1.93</td>
<td>4.65</td>
<td>0.68</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td>-4.47</td>
<td>3.98</td>
<td>0.26</td>
</tr>
<tr>
<td><strong>SES</strong></td>
<td>1.70</td>
<td>2.22</td>
<td>0.44</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td>-0.01</td>
<td>0.08</td>
<td>0.91</td>
</tr>
<tr>
<td><strong>NYR*Female</strong></td>
<td>-3.51</td>
<td>6.36</td>
<td>0.58</td>
</tr>
<tr>
<td><strong>LARI*Female</strong></td>
<td>0.75</td>
<td>4.92</td>
<td>0.88</td>
</tr>
</tbody>
</table>

$R^2 = 0.049 \quad F(7, 130) = 0.965, p = .459$
Table 4.13 (Model 2) shows that neither of the interaction variables yield statistical significance. We must therefore consider Model 1 (Table 4.14 below).

<table>
<thead>
<tr>
<th></th>
<th>Unstandardized B</th>
<th>Coefficient SE</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Constant</strong></td>
<td>12.12</td>
<td>3.78</td>
<td>0.00</td>
</tr>
<tr>
<td><strong>LARI</strong></td>
<td>-1.87</td>
<td>2.54</td>
<td>0.46</td>
</tr>
<tr>
<td><strong>NYR</strong></td>
<td>-0.03</td>
<td>3.14</td>
<td>0.99</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td>-4.72</td>
<td>2.13</td>
<td>0.03</td>
</tr>
<tr>
<td><strong>SES</strong></td>
<td>1.78</td>
<td>2.20</td>
<td>0.42</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td>-0.01</td>
<td>0.08</td>
<td>0.89</td>
</tr>
</tbody>
</table>

\[ R^2 = 0.045 \quad F \left(5, 132\right) = 1.253, \quad p = .288 \]

Table 4.14 indicates that the variable Gender yields statistical significance \((p < .05)\). This means that speakers’ gender shapes their Subjunctive rate in the Modal context (e.g. *Como quieran, lo que quieran, como que quieran*). Men draw on the Subjunctive mood to a much greater extent than women in Modal clauses. There is a nearly five-percentage point (4.7 percent, see Unstandardized B) difference between the sexes. While men employ the Subjunctive a little over 12 percent of the time (see Constant) in this linguistic context, women do so 7.4 percent of the time. In other words, women exploit the Indicative mood to a much greater degree than men in Modal clauses. We also note that since Gender is the only statistically significant variable, none of the other variables (i.e. LARI, NYR, SES, Age) influence Subjunctive rate in this linguistic context. In other words, speakers’ generation, socio-economic status and age are not connected to their usage of the Subjunctive mood in the Modal context. Finally, we learned from Model 2 that the interaction variables (NYR*Female and LARI*Female) fail to yield statistical
significance. This indicates that speakers’ rates of Subjunctive in the Modal context are influenced by their gender alone.

3.1.2 Generation and Age shape Subjunctive rate in the Protasis Sí context

This section explores the effects of the different independent variables on Subjunctive rate in the second most popular linguistic context, the Protasis Sí context (e.g. Sí vinieran, si hubieran venido). As with the linear regressions described in the previous section, we begin by analyzing Model 2. If neither of the interaction variables (NYR*Female and LARI*Female) yields statistical significance, we discard Model 2 and interpret the results from Model 1, which only includes the main effects. Table 4.15 shows the results of the linear regression (Model 2) of Subjunctive rate in the Protasis Sí context by LARI, NYR, Gender, SES, Age, NYR*Female and LARI*Female.

<table>
<thead>
<tr>
<th></th>
<th>Unstandardized B</th>
<th>Coefficient SE</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Constant</strong></td>
<td>10.51</td>
<td>4.18</td>
<td>0.01</td>
</tr>
<tr>
<td><strong>LARI</strong></td>
<td>-6.41</td>
<td>3.35</td>
<td>0.06</td>
</tr>
<tr>
<td><strong>NYR</strong></td>
<td>-5.96</td>
<td>4.60</td>
<td>0.20</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td>-0.87</td>
<td>4.00</td>
<td>0.83</td>
</tr>
<tr>
<td><strong>SES</strong></td>
<td>-3.28</td>
<td>2.20</td>
<td>0.14</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td>0.16</td>
<td>0.09</td>
<td>0.07</td>
</tr>
<tr>
<td><strong>NYR*Female</strong></td>
<td>-1.25</td>
<td>6.34</td>
<td>0.84</td>
</tr>
<tr>
<td><strong>LARI*Female</strong></td>
<td>-0.13</td>
<td>4.93</td>
<td>0.98</td>
</tr>
</tbody>
</table>

$R^2 = 0.089 \quad F (7, 129) = 1.797, p = .093$
Neither of the interaction variables yields statistical significance. Model 1, displayed in Table 4.16 below, must be considered.

Table 4-16
Subjunctive rate in the Protasis Si context
(Model 1)

<table>
<thead>
<tr>
<th></th>
<th>Unstandardized B</th>
<th>Coefficient SE</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>10.68</td>
<td>3.74</td>
<td>0.01</td>
</tr>
<tr>
<td>LARI</td>
<td>-6.47</td>
<td>2.56</td>
<td>0.01</td>
</tr>
<tr>
<td>NYR</td>
<td>-6.63</td>
<td>3.11</td>
<td>0.04</td>
</tr>
<tr>
<td>Gender</td>
<td>-1.17</td>
<td>2.11</td>
<td>0.58</td>
</tr>
<tr>
<td>SES</td>
<td>-3.25</td>
<td>2.18</td>
<td>0.14</td>
</tr>
<tr>
<td>Age</td>
<td>0.16</td>
<td>0.09</td>
<td>0.07</td>
</tr>
</tbody>
</table>

$R^2 = 0.089 \quad F (5, 131) = 2.544, p = .031$

Table 4.16 (Model 1) establishes that two independent variables reach statistical significance (LARI: $p < .05$, NYR: $p < .05$), and that one independent variable approaches statistical significance (Age: $p < .1$). In other words, the variables LARI, NYR and Age shape Subjunctive rate in the Protasis Si context (e.g. Si vinieran, si hubieran venido). Both the Established Immigrants (LARI) and the New York Raised (NYR) have significantly lower rates of Subjunctive than the Newcomers in this linguistic context. The Newcomers’ Subjunctive rate is 10.7 percent (see Constant in Table Y), whereas the LARI’s Subjunctive rate is about 6.5 percentage points lower (see Unstandardized B). Similarly, the NYR’s Subjunctive rate is 6.6 percentage points lower than that of the Newcomers. Put differently, the Subjunctive rates of the Established Immigrants and the New York Raised are equal to 4.2 and 4.1 percent, respectively. This implies that the Established Immigrants and the New York Raised use the Indicative more than the Newcomers in this context.
With respect to the variable Age, we know from the positive Unstandardized B value ($B = 0.16$) that the older the informant, the greater his/her Subjunctive rate in the Protasis $Si$ context. Since the Age variable is continuous, and the Unstandardized B value is a percentage, we infer that informants’ Subjunctive rate increases by 0.16 percent for every year they age. In other words, an informant’s Subjunctive rate increases by 1.6 percent every 10 years. The older the speaker, the more Subjunctives s/he uses. For instance, there is a 10.9 percentage-point difference in Subjunctive rate between the oldest informant, who is 80 years old, and the youngest informant, who is 12. Evidently, the 80-year old consultant draws on the Subjunctive considerably more than the 12 year old, in this linguistic context. Finally, the fact that none of the other variables (Gender, SES, and the interaction variables) yield statistical significance indicates that only speakers’ generation and age influence their Subjunctive rate in Protasis $Si$ clauses.

### 3.1.3 No effect of socio-demographic features on Subjunctive rate in two linguistic contexts

In this section, Subjunctive rate in the third and fourth most popular linguistic contexts is explored through the lens of speakers’ socio-demographic features. We start with the Temporal context (e.g. *Hasta que vengan, antes de que vengan, cuando vengan*), and then we examine the Apodosis $Si$ context (e.g. *Si..., quisiera ir, si... hubiera querido ir*). As with the linear regressions described in the previous section, we begin by analyzing Model 2. If neither of the interaction variables (LARI*Female and NYR*Female) yields statistical significance, we discard Model 2 and interpret the results from Model 1, which only includes the main effects. Table 4.17 shows the results of the linear regression (Model 2) of Subjunctive rate in the Temporal context by LARI, NYR, Gender, SES, Age, NYR*Female and LARI*Female.
Table 4-17
Subjunctive rate in the Temporal context
(Model 2)

<table>
<thead>
<tr>
<th></th>
<th>Unstandardized B</th>
<th>Coefficient SE</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>5.44</td>
<td>5.79</td>
<td>0.35</td>
</tr>
<tr>
<td>LARI</td>
<td>4.23</td>
<td>4.58</td>
<td>0.36</td>
</tr>
<tr>
<td>NYR</td>
<td>4.79</td>
<td>6.42</td>
<td>0.46</td>
</tr>
<tr>
<td>Gender</td>
<td>-2.72</td>
<td>5.42</td>
<td>0.62</td>
</tr>
<tr>
<td>SES</td>
<td>0.10</td>
<td>3.06</td>
<td>0.97</td>
</tr>
<tr>
<td>Age</td>
<td>0.06</td>
<td>0.12</td>
<td>0.58</td>
</tr>
<tr>
<td>NYR*Female</td>
<td>-5.97</td>
<td>8.84</td>
<td>0.50</td>
</tr>
<tr>
<td>LARI*Female</td>
<td>-1.04</td>
<td>6.75</td>
<td>0.88</td>
</tr>
</tbody>
</table>

$R^2 = 0.039 \quad F(7, 120) = 0.690, p = .680$

Neither of the interaction variables yields statistical significance. Model 1, shown in Table 4.18 below, must be considered.

Table 4-18
Subjunctive rate in the Temporal context
(Model 1)

<table>
<thead>
<tr>
<th></th>
<th>Unstandardized B</th>
<th>Coefficient SE</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>6.39</td>
<td>5.19</td>
<td>0.22</td>
</tr>
<tr>
<td>LARI</td>
<td>3.69</td>
<td>3.44</td>
<td>0.29</td>
</tr>
<tr>
<td>NYR</td>
<td>1.60</td>
<td>4.37</td>
<td>0.71</td>
</tr>
<tr>
<td>Gender</td>
<td>-4.30</td>
<td>2.96</td>
<td>0.15</td>
</tr>
<tr>
<td>SES</td>
<td>0.14</td>
<td>3.04</td>
<td>0.96</td>
</tr>
<tr>
<td>Age</td>
<td>0.06</td>
<td>0.11</td>
<td>0.61</td>
</tr>
</tbody>
</table>

$R^2 = 0.035 \quad F(5, 122) = 0.877, p = .499$

None of the variables in Model 1 yield statistical significance. This indicates that speakers’ socio-demographic characteristics fail to predict their Subjunctive rate in the Temporal context (e.g. Hasta que vengan, antes de que vengan, cuando vengan). This
also signals that speakers’ rates of Subjunctive in this linguistic context do not significantly differ based on their generation, gender, socio-economic status or age.

Table 4.19 shows the results of the linear regression (Model 2) of Subjunctive rate in the Apodosis *Si context by LARI, NYR, Gender, SES, Age, NYR*Female and LARI*Female.

### Table 4-19
**Subjunctive rate in the Apodosis *Si context**
(Model 2)

<table>
<thead>
<tr>
<th></th>
<th>Unstandardized B</th>
<th>Coefficient SE</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-5.50</td>
<td>6.01</td>
<td>0.36</td>
</tr>
<tr>
<td>LARI</td>
<td>5.81</td>
<td>4.62</td>
<td>0.21</td>
</tr>
<tr>
<td>NYR</td>
<td>10.54</td>
<td>6.12</td>
<td>0.09</td>
</tr>
<tr>
<td>Gender</td>
<td>4.85</td>
<td>5.34</td>
<td>0.37</td>
</tr>
<tr>
<td>SES</td>
<td>1.66</td>
<td>3.02</td>
<td>0.58</td>
</tr>
<tr>
<td>Age</td>
<td>0.18</td>
<td>0.12</td>
<td>0.14</td>
</tr>
<tr>
<td>NYR*Female</td>
<td>-9.12</td>
<td>8.51</td>
<td>0.29</td>
</tr>
<tr>
<td>LARI*Female</td>
<td>-8.50</td>
<td>6.64</td>
<td>0.20</td>
</tr>
</tbody>
</table>

R² = 0.046    F (7, 118) = 0.820, p = .573

Neither of the interaction variables reaches statistical significance. Model 1, shown in Table 4.20 below, is analyzed.
Table 4-20
Subjunctive rate in the Apodosis *Si* context
(Model 1)

<table>
<thead>
<tr>
<th></th>
<th>Unstandardized B</th>
<th>Coefficient SE</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-1.87</td>
<td>5.39</td>
<td>0.73</td>
</tr>
<tr>
<td>LARI</td>
<td>1.94</td>
<td>3.52</td>
<td>0.58</td>
</tr>
<tr>
<td>NYR</td>
<td>5.81</td>
<td>4.22</td>
<td>0.17</td>
</tr>
<tr>
<td>Gender</td>
<td>-1.30</td>
<td>2.86</td>
<td>0.65</td>
</tr>
<tr>
<td>SES</td>
<td>1.81</td>
<td>3.01</td>
<td>0.55</td>
</tr>
<tr>
<td>Age</td>
<td>0.16</td>
<td>0.12</td>
<td>0.19</td>
</tr>
</tbody>
</table>

R² = 0.031  $F (5, 120) = 0.774, p = .570$

Table 4.20 (Model 1) establishes that none of the variables yield statistical significance.

In other words, speakers’ Subjunctive rate in the Apodosis *Si* context (e.g. *Si..., quisiera ir*, *si..., hubiera querido ir*) is not determined by any of their socio-demographic characteristics.

3.2 Multivariate analysis of Linguistic context availability

In this section, we investigate the second leading research question (RQ3), regarding the socio-demographic characteristics that exert an influence on whether or not a particular linguistic context is found in a particular informant. That is, what are the socio-demographic variables that predict whether a particular linguistic context will appear in an informant’s speech? To answer this query, we run four logistic regressions (logistic regressions are warranted here because the dependent variable is binary and categorical). The dependent variable in each of the four logistic regressions is *Linguistic context availability*. The dependent variable is binary because the linguistic context either appeared or did not appear. As such, the dependent variable in each logistic

---

44 Recall that the first research question, RQ1, was answered with the help of bivariate analyses. The two leading research questions, RQ2 and RQ3, which were also explored via bivariate statistics, are further investigated through the multivariate analyses.
regression is coded as 0 = Not Available, and 1 = Available (see Appendix B). As Linguistic context availability is explored in four different linguistic contexts, the dependent variable varies in each logistic regression. The selection process regarding the four linguistic contexts is explained in detail in Chapter 3. The dependent variables in the logistic regressions are the following:

- Availability of the Volition context (e.g. *Quiere que vengan, espera que vengan*)
- Availability of the Hypothetical *Como si* context (e.g. *Como si vinieran*)
- Availability of the Possibility context (e.g. *Es posible que quieran, tal vez/quizá(s) quieran, a lo mejor quieran*)
- Availability of the Concessive context (e.g. *Aunque quieran*)

The independent external variables are the same in these logistic regressions as in the linear regressions described in the previous sections. They are LARI, NYR, Gender, SES, Age, NYR*Female, and LARI*Female. Again, the reference category for Generation is the group comprised of Newcomers (LARN). The reference category does not show up in the logistic regressions. Just as with the linear regression above, the logistic regressions are divided into two Models. As above, the first Model includes the two interaction variables, while the second Model comprises the five independent variables, excluding the interactions. The results of the logistic regressions are analyzed in the same way as they were in the linear regressions (section 3.1 above), considering Model 1 only when neither of the two interaction variables in Model 2 yields statistical significance. Again, Model 2 is analyzed first because the two interaction variables offer more information than the three individual variables (LARI, NYR, Gender) analyzed separately.
In each table (logistic regression), column 1 shows the B (the log-odds units, which predict the dependent variable from the independent variables), column 2 is the Standard Error coefficient, column 3 is the Wald Chi-square value, column 4 presents the Exp(B), which is the odds-ratio for the independent variables, and column 5 is the level of significance, which is bolded when $p < .1$ (when approaching or yielding statistical significance). When significant, the odds ratio, or Exp(B), is of utmost interest because it gives us the direction of the effect. When an Exp(B) value is above 1, there is a greater likelihood for the linguistic context to be available. When an Exp(B) value is below 1, the linguistic context has a greater chance of being unavailable. See Appendix B for a list of the variables’ coding. For example, the variable Gender is coded 0 = Male, 1 = Female. An Exp(B) value that is greater than 1 would indicate that the linguistic context in question tends to be more available to women. Finally, it is important to keep in mind that logistic regressions are about tendencies, meaning, in our case, that they are about the likelihood that a linguistic context is or is not available in the speech of particular informants. Thus, the Exp(B) value also tells us how much more or less likely the availability will be. To continue with the example above, if the significant Exp(B) value were equal to 1.5, the linguistic context would tend to be one and a half times more available to women than it would be to men.

3.2.1 Gender affects the Availability of the Concessive context
Table 4.21 and Table 4.22 below display the results of a logistic regression where the dependent variable is the Availability of the Concessive context, with Table 4.21 corresponding to Model 2 of the regression and Table 4.22 corresponding to Model 1. The independent variables in Table 4.21 (Model 2) are LARI, NYR, Gender, SES, Age,
NYR*Female and LARI*Female. The independent variables in Table 4.22 (Model 1) are the same minus the two interaction variables. That is, in Table 4.22 (Model 1) the independent variables are LARI, NYR, Gender, SES and Age. Table 4.21 below presents the logistic regression (Model 2) of the dependent variable Availability of the Concessive context (e.g. *Aunque quieran*) by the independent variables LARI, NYR, Gender, SES, Age, NYR*Female and LARI*Female.

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>Exp(B)</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Constant</strong></td>
<td>-0.93</td>
<td>1.01</td>
<td>0.86</td>
<td>0.39</td>
<td>0.35</td>
</tr>
<tr>
<td><strong>LARI</strong></td>
<td>0.67</td>
<td>0.58</td>
<td>1.33</td>
<td>1.95</td>
<td>0.25</td>
</tr>
<tr>
<td><strong>NYR</strong></td>
<td>0.22</td>
<td>0.79</td>
<td>0.08</td>
<td>1.25</td>
<td>0.78</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td>-0.56</td>
<td>0.67</td>
<td>0.70</td>
<td>0.57</td>
<td>0.41</td>
</tr>
<tr>
<td><strong>SES</strong></td>
<td>-0.48</td>
<td>0.38</td>
<td>1.53</td>
<td>0.62</td>
<td>0.22</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td>0.02</td>
<td>0.02</td>
<td>2.28</td>
<td>1.02</td>
<td>0.13</td>
</tr>
<tr>
<td><strong>NYR*Female</strong></td>
<td>0.09</td>
<td>1.07</td>
<td>0.01</td>
<td>1.10</td>
<td>0.93</td>
</tr>
<tr>
<td><strong>LARI*Female</strong></td>
<td>-0.27</td>
<td>0.83</td>
<td>0.11</td>
<td>0.76</td>
<td>0.74</td>
</tr>
</tbody>
</table>

\[ \text{Chi}^2 (7, N=139) = 9.647, p = 0.209 \]

Neither of the interaction variables reaches statistical significance in Model 2. Model 1, shown in Table 4.22 below, is considered.
Table 4.22
Availability of the Concessive context
(Model 1)

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>Exp(B)</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-0.99</td>
<td>0.83</td>
<td>1.44</td>
<td>0.37</td>
<td>0.23</td>
</tr>
<tr>
<td>LARI</td>
<td>0.54</td>
<td>0.43</td>
<td>1.57</td>
<td>1.72</td>
<td>0.21</td>
</tr>
<tr>
<td>NYR</td>
<td>0.29</td>
<td>0.54</td>
<td>0.28</td>
<td>1.33</td>
<td>0.59</td>
</tr>
<tr>
<td>Gender</td>
<td>-0.68</td>
<td>0.36</td>
<td>3.55</td>
<td>0.51</td>
<td>0.06</td>
</tr>
<tr>
<td>SES</td>
<td>-0.48</td>
<td>0.38</td>
<td>1.55</td>
<td>0.62</td>
<td>0.21</td>
</tr>
<tr>
<td>Age</td>
<td>0.02</td>
<td>0.02</td>
<td>2.30</td>
<td>1.02</td>
<td>0.13</td>
</tr>
</tbody>
</table>

$\chi^2(5, N=139) = 9.456, p = .092$

Table 4.22 establishes that the variable Gender approaches statistical significance ($p < .1$). This means that Gender shapes the availability of the Concessive context (e.g. *Aunque quieran*). Since the Exp(B) is below 1 (= 0.51), it is clear that this linguistic context is less likely to appear in the speech of women than in that of men. More specifically, it is half as likely to occur in the speech of women. In other words, the Concessive context tends to be twice as available to men as it is to women. So, for every *aunque* a woman utters, a man will most likely utter two (at any given time).

### 3.2.2 Generation shapes the Availability of the Possibility context

Tables 4.23 and 4.24 below exhibit the results of a logistic regression where the dependent variable is the Availability of the Possibility context, with Table 4.23 corresponding to Model 2 of the regression and Table 4.24 corresponding to Model 1. The independent variables in Table 4.23 (Model 2) are LARI, NYR, Gender, SES, Age, NYR*Female and LARI*Female. The independent variables in Table 4.24 (Model 1) are the same minus the two interaction variables. That is, in Table 4.24 (Model 1) the independent variables are LARI, NYR, Gender, SES and Age. Table 4.23 below represents the logistic regression (Model 2) of the dependent variable Availability of the
Possibility context (e.g. *Es posible que quieran, tal vez/quizás quieran, a lo mejor quieren*) by the independent variables LARI, NYR, Gender, SES, Age, NYR*Female and LARI*Female.

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>Exp(B)</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-2.26</td>
<td>1.17</td>
<td>3.71</td>
<td>0.10</td>
<td>0.05</td>
</tr>
<tr>
<td>LARI</td>
<td>0.10</td>
<td>0.58</td>
<td>0.03</td>
<td>1.10</td>
<td>0.86</td>
</tr>
<tr>
<td>NYR</td>
<td>1.03</td>
<td>0.92</td>
<td>1.25</td>
<td>2.80</td>
<td>0.26</td>
</tr>
<tr>
<td>Gender</td>
<td>-0.16</td>
<td>0.67</td>
<td>0.06</td>
<td>0.86</td>
<td>0.82</td>
</tr>
<tr>
<td>SES</td>
<td>-0.42</td>
<td>0.41</td>
<td>1.03</td>
<td>0.66</td>
<td>0.31</td>
</tr>
<tr>
<td>Age</td>
<td>0.00</td>
<td>0.02</td>
<td>0.00</td>
<td>1.00</td>
<td>0.97</td>
</tr>
<tr>
<td>NYR*Female</td>
<td>0.42</td>
<td>1.29</td>
<td>0.11</td>
<td>1.52</td>
<td>0.75</td>
</tr>
<tr>
<td>LARI*Female</td>
<td>0.48</td>
<td>0.85</td>
<td>0.33</td>
<td>1.62</td>
<td>0.57</td>
</tr>
</tbody>
</table>

*Chi*²(7, N=139) = 5.847, *p* = .558

Neither of the two interaction variables yields statistical significance in Model 2. We therefore examine Model 1 in Table 4.24 below.

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>Exp(B)</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-1.93</td>
<td>0.93</td>
<td>4.32</td>
<td>0.15</td>
<td>0.04</td>
</tr>
<tr>
<td>LARI</td>
<td>0.31</td>
<td>0.44</td>
<td>0.52</td>
<td>1.37</td>
<td>0.47</td>
</tr>
<tr>
<td>NYR</td>
<td>1.24</td>
<td>0.64</td>
<td>3.75</td>
<td>3.46</td>
<td>0.05</td>
</tr>
<tr>
<td>Gender</td>
<td>0.16</td>
<td>0.39</td>
<td>0.17</td>
<td>1.17</td>
<td>0.68</td>
</tr>
<tr>
<td>SES</td>
<td>-0.42</td>
<td>0.41</td>
<td>1.08</td>
<td>0.66</td>
<td>0.30</td>
</tr>
<tr>
<td>Age</td>
<td>0.00</td>
<td>0.02</td>
<td>0.00</td>
<td>1.00</td>
<td>0.99</td>
</tr>
</tbody>
</table>

*Chi*²(5, N=139) = 5.511, *p* = .357

The variable NYR is statistically significant in Table 4.24 (*p* = .05). Belonging to the New York Raised group influences the availability of the Possibility context (e.g. *Es
 posible que quieran, tal vez/quizás quieran, a lo mejor quieran). As the Exp(B) = 3.46 (which is above 1), the Possibility context is nearly three and a half times more available to New York Raised participants than to the Newcomers. Put differently, there is a much greater likelihood that this linguistic context will occur in the speech of the New York Raised than in that of the Newcomers. So, every time a Newcomer informant says a lo mejor (or any of the other Possibility clauses), a New York Raised informant will most likely have done the same three or four times.

### 3.2.3 Gender and Generation together influence the Availability of the Volition context

Table 4.25 below exhibits the results of a logistic regression where the dependent variable is the Availability of the Volition context, with Table 4.25 corresponding to Model 2 of the regression. The independent variables in Table 4.25 (Model 2) are LARI, NYR, Gender, SES, Age, NYR*Female and LARI*Female. Table 4.25 presents the logistic regression (Model 2) of the dependent variable Availability of the Volition context (e.g. Quiere que vengan, espera que vengan) by the independent variables LARI, NYR, Gender, SES, Age, NYR*Female and LARI*Female.
Table 4-25
Availability of the Volition context
(Model 2)

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>Exp(B)</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-2.41</td>
<td>1.54</td>
<td>2.43</td>
<td>0.09</td>
<td>0.12</td>
</tr>
<tr>
<td>LARI</td>
<td>-0.71</td>
<td>1.16</td>
<td>0.37</td>
<td>0.49</td>
<td>0.54</td>
</tr>
<tr>
<td>NYR</td>
<td>-1.83</td>
<td>1.25</td>
<td>2.17</td>
<td>0.16</td>
<td>0.14</td>
</tr>
<tr>
<td>Gender</td>
<td>-2.24</td>
<td>1.16</td>
<td>3.77</td>
<td>0.11</td>
<td><strong>0.05</strong></td>
</tr>
<tr>
<td>SES</td>
<td>0.47</td>
<td>0.54</td>
<td>0.75</td>
<td>1.60</td>
<td>0.39</td>
</tr>
<tr>
<td>Age</td>
<td>-0.04</td>
<td>0.02</td>
<td>2.63</td>
<td>0.96</td>
<td>0.11</td>
</tr>
<tr>
<td>NYR*Female</td>
<td>3.71</td>
<td>1.71</td>
<td>4.70</td>
<td>40.76</td>
<td><strong>0.03</strong></td>
</tr>
<tr>
<td>LARI*Female</td>
<td>1.52</td>
<td>1.38</td>
<td>1.21</td>
<td>4.56</td>
<td>0.27</td>
</tr>
</tbody>
</table>

$\text{Chi}^2 (7, N=139) = 11.976, p = .101$

In Table 4.25, two variables yield statistical significance: Gender ($p < .05$) and the interaction variable NYR*Female ($p < .05$). First of all, as noted, the interaction variable offers more precise information about women than the single variable. As a result, only the interaction variable can be considered. Secondly, as one of the two interaction variables yields statistical significance, Model 1 is not taken into consideration and thus not displayed. Being a New York Raised woman affects the availability of the Volition context (e.g. *Quiere que vengan, espera que vengan*). In fact, this linguistic context is over 40 times more likely to occur (see the Exp(B)) in the speech of New York Raised women than in that of Newcomer women (the reference category). In other words, the Volition context tends to be *much* more available to second generation NYR women than to first generation Newcomer women. This suggests that New York Raised women are much more likely than Newcomer women to express volition (with *querer que* and *esperar que*).
3.2.4 No effect of socio-demographic features on the Availability of the Hypothetical Como si context

Tables 4.26 and 4.27 below exhibit the results of a logistic regression where the dependent variable is the Availability of the Hypothetical Como si context, with Table 4.26 corresponding to Model 2 of the regression and Table 4.27 corresponding to Model 1. The independent variables in Table 4.26 (Model 2) are LARI, NYR, Gender, SES, Age, NYR*Female and LARI*Female. The independent variables in Table 4.27 (Model 1) are the same minus the two interaction variables. That is, in Table 4.27 (Model 1) the independent variables are LARI, NYR, Gender, SES and Age. Table 4.26 below shows the results of the logistic regression (Model 2) of the dependent variable Availability of the Hypothetical Como si context (Como si) by the independent variables LARI, NYR, Gender, SES, Age, NYR*Female and LARI*Female.

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>Exp(B)</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-1.13</td>
<td>1.42</td>
<td>0.63</td>
<td>0.32</td>
<td>0.43</td>
</tr>
<tr>
<td>LARI</td>
<td>0.15</td>
<td>0.92</td>
<td>0.03</td>
<td>1.16</td>
<td>0.87</td>
</tr>
<tr>
<td>NYR</td>
<td>-0.54</td>
<td>1.09</td>
<td>0.24</td>
<td>0.58</td>
<td>0.62</td>
</tr>
<tr>
<td>Gender</td>
<td>-0.41</td>
<td>0.98</td>
<td>0.18</td>
<td>0.66</td>
<td>0.68</td>
</tr>
<tr>
<td>SES</td>
<td>-0.43</td>
<td>0.58</td>
<td>0.55</td>
<td>0.65</td>
<td>0.46</td>
</tr>
<tr>
<td>Age</td>
<td>-0.01</td>
<td>0.02</td>
<td>0.22</td>
<td>0.99</td>
<td>0.64</td>
</tr>
<tr>
<td>NYR*Female</td>
<td>0.10</td>
<td>1.42</td>
<td>0.01</td>
<td>1.11</td>
<td>0.94</td>
</tr>
<tr>
<td>LARI*Female</td>
<td>0.16</td>
<td>1.27</td>
<td>0.02</td>
<td>1.18</td>
<td>0.90</td>
</tr>
</tbody>
</table>

*Chi² (7, N=139) = 3.238, p = .862*

Neither of the interaction variables yields statistical significance in Model 2. Model 1 in Table 4.27 below must be taken into consideration.
Table 4.27 establishes that none of the variables reach statistical significance. This means that speakers’ socio-demographic characteristics do not predict the availability of the Hypothetical Como si context (e.g. Como si vinieran).

### 3.3 Multivariate analyses: Summary

The current study’s two main research questions (RQ2 and RQ3) were answered through two types of multivariate tests presented in the previous sections. The first core research question (RQ2) concerned the effect of speakers’ socio-demographic features on their rates of Subjunctive, overall and in the four most popular linguistic contexts. This question was explored through the analysis of five multiple linear regressions (section 3.1). It was discovered that Gender shapes overall Subjunctive rate as well as Subjunctive rate in the Modal context (e.g. Como quieran, lo que quieran, como que quieran). In both cases, men have higher rates of Subjunctive than women. In other words, women draw on the Indicative more than men in both instances.

Furthermore, the linear regressions demonstrated that New York Raised and Established Immigrants have lower rates of Subjunctive than Newcomers in the Protasis
Si context (e.g. *Si vinieran, si hubieran venido*). Age was also found to have an effect. The older the informant, the greater his/her Subjunctive rate in this linguistic context. Finally, speakers’ socio-demographic characteristics fail to shape their Subjunctive rate in two linguistic contexts: the Temporal context (e.g. *Hasta que vengan, antes de que vengan, cuando vengan*), and the Apodosis Si context (e.g. *Si..., quisiera ir, si..., hubiera querido ir*).

The second key research question (RQ3), relating to the socio-demographic characteristics that predict the availability of four linguistic contexts in the informants’ speech, was answered by means of four binary logistic regressions (section 3.2). Gender was found to influence the availability of the Concessive context (e.g. *Aunque quieran*), which tends to be more available to men than to women. The Possibility context (e.g. *Es posible que quieran, tal vez/quizá(s) quieran, a lo mejor quieran*) tends to be more available to New York Raised informants than to the Newcomers. The Volition context (e.g. *Quiere que vengan, espera que vengan*) has greater chances of occurring in the speech of New York Raised women than in that of Newcomer women. Finally, it was discovered that speakers’ socio-demographic features do not affect the availability of the Hypothetical Como si context.

4. Conclusion

This chapter presented the results of bivariate and multivariate analyses, conducted to inform our three research questions. The first research question, split into four parts (a, b, c, and d), was only answered by means of bivariate analyses, as these questions involved associations between two variables and differences among group means. The relationship between rates of Subjunctive overall and in the most popular linguistic
contexts (RQ1a), and between these most popular contexts (RQ1b), the association between the extents of availability of distinct linguistic contexts (RQ1c), and the connection between Subjunctive rate and Linguistic context availability (RQ1d) were only explored in section 2. We had hypothesized that some of these relationships would be significant, and this turns out to be the case. Indeed, speakers’ Subjunctive rate in certain linguistic contexts is connected to their Subjunctive rate in others, the availability of a particular linguistic context can relate to that of another, and Subjunctive rates in a particular linguistic context can indeed vary based on the availability of a different linguistic context.

Furthermore, the study’s two main research questions (RQ2 and RQ3) were examined in sections 2 and 3, that is, both through bivariate and multivariate tests. Indeed, certain binary relationships can be further explored by means of more complex statistical analyses (e.g. linear and logistic regressions) that offer a more complete picture. While several of the bivariate test results are corroborated as a result of the multivariate tests, others are discarded. Gender and Generation, for instance, were found to play separate significant roles both in the bivariate and multivariate analyses of Subjunctive rate. Socio-economic status approached statistical significance in the one-way ANOVA (that is, in the bivariate analysis), but was not found to influence Subjunctive rate in the multivariate approach. It is therefore discarded. On the other hand, a linear regression corroborated that speakers’ age influences their Subjunctive rate in one context, while this was not the case in the original bivariate Pearson correlation.

The discrepancy between the bivariate and multivariate results is more salient with respect to Linguistic context availability. The only result that yields significance in
both statistical approaches relates to the availability of the Concessive context (e.g. *Aunque quieran*), which is predicted by speakers’ gender. Speakers’ age is found to shape the availability of the Volition context (e.g. *Quiere que vengan, espera que vengan*) in the independent samples t-tests, but not in the logistic regressions. As logistic regressions are more revealing than t-tests because they consider all socio-demographic variables together, this t-test result is not taken into consideration in our discussion (Chapter 5). On the other hand, only the logistic regressions ascertained the roles played by Generation and the interaction between Generation and Gender (Female) with respect to Linguistic context availability. Evidently, this can be attributed to the fact that multivariate analyses tend to be more informative than bivariate analyses. In addition, our interaction variables could only be introduced in the multivariate analyses (because of the nature of these statistical tests).

These comparisons between the bivariate and multivariate results confirm that bivariate analyses represent an important first step in any analysis, but that they should be supplemented by multivariate analyses, which offer the bigger picture. Finally, the results of both the bivariate and multivariate analyses largely confirm the present study’s hypotheses, articulated via our three research questions. The descriptive results outlined in this chapter will now be discussed in detail in Chapter 5.
Chapter 5: Discussion

1. Introduction

This chapter discusses the bivariate and multivariate statistical results presented in Chapter 4. The first part of this chapter (section 2) discusses the associations holding between Subjunctive rates in different linguistic contexts; the relationships between the extents of availability of the different linguistic contexts; and the connection between Subjunctive rate and Linguistic context availability. These associations reveal that all speakers have consistent patterns of Subjunctive usage, both in terms of rates and context availability. In section 2, it is posited that a comprehensive analysis of speakers’ treatment of the Subjunctive should encompass both the frequency of Subjunctive verbs in particular speakers and the availability of different linguistic contexts in those speakers. In addition, it is argued, as already suggested in Chapter 3, that studies that pay heed only to the number and percentages of Subjunctive verbs, without taking into account the number of informants in whose speech these verbs occur, cannot accurately reflect the linguistic behavior of the New York City Latino population as a whole. The second part of the present chapter (sections 3, 4 and 5) explores why certain socio-demographic characteristics are associated with the availability of particular linguistic contexts, and with the presence and absence of the Subjunctive mood. More specifically, section 3 questions the critical role allotted to speakers’ generation in prior research, and challenges the deficit hypothesis proposed for the second generation’s grammar of the Subjunctive. Section 4 proposes a theory of aspiring social change that simultaneously accounts for what I have called the Women effect, the New Yorker effect and the Youth

45 The reader should keep in mind that a linguistic context either comprises a group of clauses (e.g., Quiero que vengan and espero que vengan) or a single environment, that is, just one clause (e.g., Aunque quieran).
effect. The main idea is that social change and social status represent the impetus for the comparatively low Subjunctive rates encountered in these three groups. The broad social change and social status hypotheses presented in section 4 are qualified by a much narrower consideration, in section 5, of the types of messages conveyed in particular linguistic contexts. The goal is to gauge how and why particular socio-demographic features influence the availability of certain linguistic contexts, that is, why certain contexts are more available in the speech of some speakers than in that of others. Finally, this chapter’s third part (sections 6 and 7) assesses whether the present investigation’s findings support the theories of simplification (section 6) and language change (section 7) that are frequently espoused in the literature. The chapter closes, in section 8, with the idea that speakers’ treatment of the Subjunctive is illuminated when its study embraces a wide-ranging linguistic and sociopolitical approach.

2. Linguistic relationships

In this section I discuss the linguistic relationships that were discovered in Chapter 4. Many of the hypotheses presented here are new, as very little research has been conducted on the possible associations that may exist between different linguistic contexts.

2.1 Interpreting relationships between Subjunctive rates overall and in the four most popular linguistic contexts

Bivariate Pearson correlations showed that the overall Subjunctive rate (in the nine linguistic contexts taken together) is significantly associated with the Subjunctive rate in
the four most popular linguistic contexts\textsuperscript{46}, that is, with Subjunctive rate in the Modal context (e.g. \textit{Como quieran, lo que quieran, como que quieran}), in the Protasis \textit{Si} context (e.g. \textit{Si vinieran, si hubieran venido}), in the Temporal context (e.g. \textit{Hasta que vengan, antes de que vengan, cuando vengan}) and in the Apodosis \textit{Si} context (e.g. \textit{Si..., quisiera ir, si..., hubiera querido ir}). This finding suggests that speakers who have a tendency to draw on the Subjunctive in one of these four linguistic contexts also draw on the mood by and large, in all of the linguistic contexts that are available to them. This finding will be further interpreted following the discussion, in the next subsection, of relationships between Subjunctive rates in the four most popular linguistic contexts.

\textbf{2.2 Interpreting relationships between Subjunctive rates in the four most popular linguistic contexts}

A more striking discovery is the fact that the speakers’ Subjunctive rates in the different linguistic contexts are interrelated. Subjunctive rates in the Temporal context (e.g. \textit{Hasta que vengan, antes de que vengan, cuando vengan}) are statistically connected to Subjunctive rates in the Apodosis \textit{Si} context (e.g. \textit{Si..., quisiera ir, si..., hubiera querido ir}), as in examples (31) and (32) below, from the same informant. A significant association was also found between Subjunctive rates in the Protasis \textit{Si} context (e.g. \textit{Si vinieran, si hubieran venido}) and in the Apodosis \textit{Si} context. Furthermore, both pairs are positively correlated, that is, the speakers for whom the rates are comparatively higher in one context are the same as the speakers for whom the rates are comparatively higher in the other.

\textsuperscript{46}Recall that only the four most popular linguistic contexts were explored with respect to speakers’ Subjunctive rate (in addition to their overall Subjunctive rate), as each of these contexts is available to over 50 percent of our sample. More than 80 informants use each of these four linguistic contexts. See Chapter 3 for details on the decision to examine the four most popular linguistic contexts (and to discard the other five).
Ahora mismo, aquí hay millones de latinos, que cuando el inglés sea lo único… - LARI 229D

‘Right now, there are millions of Latinos here, but when English is the only one…’

Si hubieran conseguido trabajo, yo la hubiera dejado… - LARI 229D

‘If they had found work, I would have let her’

The significant correlation between Subjunctive rates in the first pair (Temporal context, as in example (31), and Apodosis Si context, as in example (32)) has not been discussed in the literature, and is difficult to grasp without an in-depth analysis of the message types conveyed in each of the two contexts. Yet, we expected to find associations for three reasons: first, because we examined usage of verb mood in a very large communicative context (i.e. each informant’s entire interview); secondly, because this study does not adhere to prescriptive grammar rules (which means that the Subjunctive does not go unnoticed in contexts in which its presence is unexpected); and thirdly, because the Subjunctive is anticipated whenever it helps to convey a particular message.

In contrast to the pair of linguistic contexts shown above, the relationship between the Protasis Si context and the Apodosis Si context has been considered in prior studies (Serrano 1995; Torres 1989). These two linguistic contexts have often been examined in conjunction with each other, as they have been thought to co-occur in the same sentence, or in the same communicative context (although this hypothesis is usually implicit). Prescriptive grammar rules maintain that the Subjunctive mood is only warranted in the Protasis Si context, while the Conditional is expected in the Apodosis Si context (e.g., Si lloviera, tomaría mi paraguas ‘If it rained, I would take my umbrella’).
Contrary to these rules, our results show that the Subjunctive mood is also employed in the Apodosis Si context (as in example (33) below) and that its presence increases in proportion to its occurrence in the Protasis Si context (and vice versa). Furthermore, this happens in the speech of first generation speakers, as shown in example (33), and not merely in the speech of second generation informants, which challenges the prescriptive view that considers this usage to be erroneous.

(33) Si hubieran conseguido trabajo, yo la hubiera dejado… – LARI 229D

‘If they had found work, I would have let her’

In example (33) above, the verbs hubieran conseguido ‘had found’ in the Protasis Si context, and hubiera dejado ‘would have let’ in the Apodosis Si context, are both in the Subjunctive.

The associations between different linguistic contexts are revealing of speakers’ consistent patterns of Subjunctive usage. We have found that if a speaker prefers to use the Subjunctive in one context, s/he will also prefer to use it in another (that is, another context with which it is significantly associated). In fact, it is never the case that a speaker will prefer the Subjunctive in one linguistic context, and disfavor the mood in another. We know this to be the case because speakers’ Subjunctive rates in the two remaining pairs of linguistic contexts (out of the four most popular pairs) are not significantly correlated, which indicates that speakers neither prefer nor disfavor the Subjunctive in those contexts. We did not find any negative correlations, with mood preferences going in opposite directions. Put differently, if a speaker shows a relative
preference for the Subjunctive in one linguistic context, s/he will never show a relative preference for the Indicative in another context. Thus, we have discovered that certain speakers *always* show a relative preference for the Subjunctive mood. I call these speakers *Subjunctive-inclined*. This finding establishes that individuals are consistent with respect to their mood choice. The reasons for this consistency may be due to the different meanings that speakers assign to the Indicative and to the Subjunctive, but the answer to this question warrants a different type of investigation.

Finally, our discovery, described in the previous section, that speakers’ overall Subjunctive rate is significantly correlated with their Subjunctive rates in the different linguistic contexts, lends additional support to our *consistent patterns hypothesis*, as the same speakers who are Subjunctive-inclined in the most popular linguistic contexts are Subjunctive-inclined overall.

### 2.3 Interpreting relationships between the extents of availability of different linguistic contexts

Of the twelve cross-tabulations that were conducted between the extents of availability of different linguistic contexts, one relationship was found to be statistically significant. The Chi-square tests point to the fact that there is a significant association between the availability of the Possibility context (*e.g.* *Es posible que quieran, tal vez/quizá(s) quieran, a lo mejor quieran*) and the availability of the Concessive context (*e.g.* *Aunque quieran*). This suggests that speakers who have recourse to the contexts involved with Possibility tend to be the same that utilize contexts involved with Concession, because Possibility and Concession may be closely related communications, or closely related message types. Individuals who speak in terms of Possibilities are also the ones who
speak in terms of Concession. This unexpected finding merits a detailed analysis that would fully explain how the types of messages conveyed in the different contexts produce the correlation that we observe. This analysis is, however, beyond the scope of the current investigation.

Nevertheless, we propose that our theory of consistent patterns (formulated in the previous sections of this chapter) would be relevant if Possibility and Concession are indeed two closely related message types (which is plausible given their significant correlation). If this is the case, then it is evident that people tend to be consistent in what they have to say, and not just in how they say it (that is, which mood they select). In other words, speakers are consistent with respect to the contexts that they choose to make use of, and the kinds of messages that they (the speakers) convey.

One of the reasons for examining Linguistic context availability is to explore how it relates to Subjunctive rate. This relationship is explored in the following section.

2.4 Interpreting relationships between Subjunctive rate and Linguistic context availability

The independent samples t-tests of Subjunctive rate (overall and in the four most popular linguistic contexts) by Linguistic context availability showed that Subjunctive rate in the Modal context (e.g. Como quieran, lo que quieran, como que quieran) is statistically connected to the availability of the Hypothetical Como si context. It turns out that speakers’ Subjunctive rate in the Modal context is significantly lower when they draw on clauses introduced by como si in their speech. Conversely, speakers’ Subjunctive rate in the Modal context is higher when clauses introduced by como si are absent. In other
words, speakers who do not have *como si* in their repertoire tend to use the Subjunctive more after *como, lo que* and *como que*.

Similarly, the independent samples t-tests established that the availability of the Concessive context (e.g. *Aunque quieran*) is marginally related to speakers’ Subjunctive rate in the Temporal context (e.g. *Hasta que vengan, antes de que vengan, cuando vengan*). Their Subjunctive rate in this linguistic context is significantly greater when *aunque* clauses appear in their speech than when they do not. In other words, the speakers who express Temporal message types also tend to prefer the Subjunctive in Concessive communications.

The connection between these two linguistic contexts’ availability and rates of Subjunctive is difficult to grasp, as the two linguistic contexts in each pair are seemingly unrelated to one another. However, an in-depth analysis of the messages that speakers convey through these linguistic contexts might help shed light on their relationship. While this is beyond the scope of the present investigation, the fact that these relationships exist is, in and of itself, a noteworthy finding. Indeed, the notion that speakers’ rates of Subjunctive can be conditioned by the presence or absence of a distinct linguistic context is unprecedented. Not only are discrete linguistic contexts connected in terms of their presence or absence, as shown in the previous section, their presence or absence also affects speakers’ Subjunctive rates.

This additional discovery evidences the fact speakers’ Subjunctive rates should not be examined in a vacuum, as is often the case. Analyses that only consider percentages of Subjunctive verbs (relative to verbs in other moods) are missing a vital component. It seems evident now that an analysis of Subjunctive usage should comprise
both an examination of linguistic context availability and Subjunctive rates, as the two are highly connected. In other words, the researcher should pay attention to Subjunctive rates and the types of communication, or the message types, that informants are interested in. When both of these types of analyses are considered, it becomes clear that the theoretical framework that must be abided by is one that goes beyond the mere clause or sentence. Instead, the entire communicative context, which in our case corresponds to the whole interview, must be explored. Indeed, taking a bird’s eye view leads to a much more comprehensive understanding of the ways in which speakers treat the Subjunctive mood. It is only by studying what people want to say that one can make sense of the greater or lesser use they make of a certain linguistic form, in this case the Subjunctive.

3. Demystifying second generation Spanish deficit hypotheses

Because a speaker's immigrant generation (whether they're first or second) plays such a significant role in the literature on the Spanish Subjunctive in the United States, this section is entirely devoted to this socio-demographic feature. The present investigation makes clear, however, that an informant’s generation is merely one of several characteristics that contribute to shaping their Subjunctive rates and Linguistic context availability. In fact, and contrary to what is implied by the literature’s focus on its effects, the generation that a speaker belongs to does not carry any more weight than the other significant predictor variables. In addition to discussing the limited role played by the Generation variable in predicting Subjunctive rates (section 3.1), in this section we also discuss the theories of Subjunctive avoidance, incomplete acquisition, attrition, and second-rate Spanish (sections 3.2 and 3.3). In general, this section challenges the view,
commonly upheld in the literature, that second generation informants’ grammar of the Spanish Subjunctive is defective.

3.1 On the role of Generation in predicting Subjunctive rates

In the literature, speaker generation emerges as the most influential, and sometimes the only, socio-demographic variable impacting Subjunctive rates. To be noted is that the overall frequency of Subjunctive verbs has been deemed comparable in first and second generation speech (Bookhamer 2013; Gutiérrez 1990; Lynch 1999; Torres 1989). The cross-generational discrepancies that much of the literature discusses emerge in particular linguistic contexts. Overwhelmingly, the second generation has been charged with drawing on the Subjunctive mood to a lesser degree than the first generation in particular linguistic contexts.

The present study supports this finding, but to a much lesser extent than has been portrayed in prior research. In the current investigation, we have found that the differences in cross-generational Subjunctive rates only apply to one linguistic context. The results of the linear regressions established that the second generation New York Raised participants and the first generation Established Immigrants employ the mood significantly less than the first generation Newcomers in the Protasis Si context (e.g. Si vinieran, si hubieran venido). But, importantly, Generation fails to predict the overall Subjunctive rate and it fails to predict too the Subjunctive rate in any of the three remaining linguistic contexts. Thus, the first major difference between the current investigation and the majority of the research lies in the weight ascribed to a speaker's generation.
There are several possible reasons for the divergence. First, the present study includes other socio-demographic variables, which are considered jointly in the regressions, whereas much of the literature only considers the Generation variable. Secondly, this study’s dataset is much larger than most, with several thousand tokens (i.e. several thousand tokens of verbs in the Subjunctive). This enhances the reliability of our results (see Chapter 3 for a full discussion). Thirdly, Subjunctive verbs are collected via sociolinguistic interviews that depict authentic speech, rather than through grammaticality judgment tests which rely on less dependable speaker intuitions. Finally, only those linguistic contexts that appear in the speech of a minimum of 80 informants (more than half of the sample of 142 informants) are considered here, as described in detail in Chapter 3. Whereas the present investigation examines Subjunctive rate in the most available (or popular) linguistic contexts, the concept of Linguistic context availability is non-existent in the rest of the literature. This suggests that the cross-generational discrepancies in terms of Subjunctive rates discovered in previous studies may be based in the speech of only a handful of informants. Put differently, when the number of users in whose speech Subjunctive verbs occur is not taken into account, the Subjunctive verbs that are counted as occurring in a particular linguistic context might belong to very few (first or second generation) consultants. The number of tokens, no matter how great, is not equivalent to number of speakers, even when cross-generational differences yield statistical significance. Thus, claiming that cross-generational differences are illustrative of the behavior of the U.S. Latino population as a whole appears very problematic when the number of informants is not taken into consideration.
As evident here, speakers’ generation does predict rates of Subjunctive, but does so in only one linguistic context, that is, not nearly to the extent maintained in previous studies.

3.2 On the theory of Subjunctive avoidance

While our second generation consultants follow the trend described in previous studies with respect to their reduced Subjunctive rate in one (and only in one) specific linguistic context, their behavior in terms of Linguistic context availability is a discovery of the present study and not previously found in the literature. Indeed, the Possibility context (e.g. *Es posible que quieran, tal vez/quizá(s) quieran, a lo mejor quieran*) tends to be more available to New York Raised speakers than to the Newcomers. Similarly, the Volition context (e.g. *Quiere que vengan, espera que vengan*) is more available to New York Raised women than to Newcomer women.

Although the analyses of Linguistic context availability and Subjunctive rates concern distinct linguistic contexts (see Chapter 3 for a detailed explanation), second generation speakers do not appear to be avoiding the Subjunctive mood through alternative syntactic constructions, by means of infinitives or direct speech (Gutiérrez 1990; Montrul 2009, see Chapter 2). Indeed, the fact that these two Subjunctive-inducing contexts tend to abound in New York Raised speech (relative to Newcomer speech) challenges the notion that second generation speakers avoid using the Subjunctive mood because of their alleged lack of grammatical knowledge or proficiency (Montrul 2009). Furthermore, they do not appear to circumvent subordinate clauses, as Gutiérrez (1990) maintains. If that were the case, these linguistic contexts—selected by the speaker based on their propensity to give rise to the Subjunctive—would not be present, let alone likelier, in their speech than in that of the Newcomers.
### 3.3 On theories of incomplete acquisition, attrition and second-rate Spanish

As mentioned in the previous section, the New York Raised draw on the Subjunctive mood to the same degree as the other generational groups, overall and in all but one linguistic context (Protasis *Si*). This fact alone should render theories of incomplete acquisition, attrition and imperfect Spanish questionable. As laid out in Chapter 2, second generation informants are often implicitly charged with speaking second-rate Spanish. This perception is based on the belief that these speakers have little experience with formal varieties of Spanish (mainly acquired in school), that their grammar of Spanish has been supplanted by English syntax during their childhood (usually coinciding with when they started school), and that the bilingual environment in which they grew up has lessened the *monolingual pressures* (Lynch 1999) that could have made them Spanish dominant. Although these theories may explain certain linguistic phenomena, they are incompatible with our findings.

Second generation informants are not alone in drawing on the Subjunctive to a lesser extent than the Newcomers. Established Immigrants have lower rates of Subjunctive in the Protasis *Si* context (e.g. *Si vinieran, si hubieran venido*) relative to Newcomers as well. In fact, it has been shown as early as the 1980s that first generation New York City (NYC) Puerto Ricans preferred the Indicative mood in this linguistic context (cf. Torres 1989, see Chapter 2). Yet first generation informants are never faulted with having incompletely acquired, or for mishandling, the Spanish Subjunctive (or any other linguistic feature, for that matter).

Because most studies on the Subjunctive compare the U.S. born and/or raised second generation to recently arrived first generation immigrants, such as our
Newcomers, the intermediate immigration stage embodied by first generation Established Immigrants is often omitted (e.g., by Bookhamer 2013, etc.). As a result, the linguistic similarities that may arise between Established Immigrants and the second generation (e.g., our New York Raised) are lost in the process.

Yet, just as their Newcomer peers, Established Immigrants grew up in monolingual Spanish-speaking environments and attended Latin American schools. Both groups are believed to have acquired the Subjunctive mood in the Protasis Si context during their childhood, by age 3 or 7/8 (depending on the study), as shown in Table 2.1 in Chapter 2. Thus, the Established Immigrants are familiar with the mood’s presence in this linguistic context (as in all of the others) by the time they arrive to New York City, at age twenty, on average. If second generation speakers are thought to have incompletely acquired the Spanish Subjunctive (Montrul 2009), then the Established Immigrants should be faulted as well.

It could be argued that the reduced Subjunctive rate of these first generation informants is due to attrition, meaning that they would have lost this grammatical feature despite having fully acquired it by adolescence or early adulthood (e.g., Merino 1976, 1983; Montrul 2014, see Chapter 2). Jakobson (1941)’s Regression Hypothesis, which Montrul (2014) and Merino (1976, 1983) seem to adhere to, stipulates that features that are acquired late (in childhood) are the first to undergo loss. However, as stated above, the Subjunctive in the Protasis Si context is acquired fairly early on. Thus, erosion does not appear to be responsible for the Established Immigrants’ distinct grammar of the Subjunctive (relative to Newcomers) in this linguistic context.
Finally, theories of incomplete acquisition, attrition, and second-rate Spanish are problematic because speakers’ generation is not alone in predicting the Subjunctive rates of different speakers in the Protasis *Si* context. Indeed, speakers’ age also shapes their Subjunctive rate in this context. As nearly all of our informants are adults, the age effect that we find is unrelated to questions of acquisition. The answer must therefore lie elsewhere.

4. **Language variation and social change**

In section 2, we determined that all speakers are consistent with respect to their usage of the Subjunctive. After having treated our entire sample of 142 informants, we are now interested in examining whether our theory of patterns of consistency also applies to subsets of people. This section explores the role of the three statistically significant socio-demographic variables in predicting Subjunctive rates (the role of significant socio-demographic variables in predicting Linguistic context availability is explored in the following section, section 5). The three variables are constituted by a speaker's membership in one or more of three groups (women, New Yorkers, and the young), and will lead us to speak presently of the Women effect, the New Yorker effect, and the Youth effect. Similar hypotheses, albeit with some distinctions, are offered to explain the comparatively low Subjunctive rates overall and in particular contexts that are found in all three groups. The theory advanced here is based on the idea that the distinct Subjunctive rates found in women, the New Yorkers and the young is representative of their desire for social change and greater social status. This *social change and social status hypothesis* is introduced by means of three similar (yet slightly distinct) stances:
the Women effect\textsuperscript{47} (in section 4.1), the New Yorker effect (in section 4.2), and the Youth effect (in section 4.3).

\textbf{4.1 The Women effect}

The results of the linear regression conducted with overall Subjunctive rate as the dependent variable established that Gender was the only variable to yield statistical significance when all linguistic contexts are analyzed together. In other words, only a speaker’s gender shapes his/her overall rate of Subjunctive verbs. Men have an overall higher rate of Subjunctive verbs than women (three percentage points higher), which also means that women have an overall higher rate of Indicative verbs than men.

As shown in the previous chapter, the same gender effect was encountered with respect to the Modal context (e.g. \textit{Como quieran, lo que quieran, como que quieran}), with an even greater gender disparity in this linguistic context than overall. Men exploit the Subjunctive nearly five percentage points more than women in Modal contexts. In other words, the Indicative mood is significantly more prevalent in Modal contexts that occur in the speech of women than in that of men.

First of all, it could be argued that a large amount of variability is to be expected in the Modal context due to its classification in many studies as an optional or variable Subjunctive context (e.g., Bookhamer 2013; Lynch 1999, see Chapters 2 and 3 for discussions of optional/variable contexts). But recall that the artificial distinction between obligatory versus optional Subjunctive contexts is rejected here, based on the plain fact that all linguistic contexts are variable, and that statistically significant disparities in Subjunctive rates should be explored irrespective of how each linguistic context is classified.

\textsuperscript{47} The \textit{Women effect} is a label and concept developed by Shin & Otheguy (2013) in reference to their first generation male and female consultants’ distinct rates of Spanish subject pronouns.
context has been classified in the literature. The examples below illustrate the use of the Indicative mood by a female informant (example 34), and the Subjunctive mood by a male informant (example 35), in the Modal context (introduced by lo que here):

(34) Ya nosotros sabíamos lo que significaban en español… - FEMALE 112D  
‘We already knew what they meant in Spanish’

(35) Y me decía […] lo que falte aquí o lo que se acabe… - MALE 198P  
‘And he would say whatever is missing here or whatever runs out’

The mood choices displayed in the two examples above are representative of men and women’s disparate tendencies in the Modal context, but also overall, when all nine linguistic contexts are examined as a whole. As women are never Subjunctive-inclined (in any linguistic context), we infer that women are consistent in their preference for the Indicative, relative to men. Thus, the patterns of consistency that were uncovered with respect to all of the informants’ Subjunctive usage (section 2 of this chapter) also apply to the subgroup of women. We find that women feature in the Subjunctive-disinclined group.

The few previous studies that include the speaker's gender point to the fact that gender can affect the occurrence of the Spanish Subjunctive (as reviewed in Chapter 2). Lastra & Butragueño (2012) observe that women’s overall Subjunctive rate is significantly higher than that of men in Mexico City. Serrano (1995) finds a similar gender effect on the Canary Island of La Laguna with respect to the Protasis Sí and Apodosis Sí contexts. In her study, women have a greater tendency than men to employ
what Serrano calls the “standard” Subjunctive-Conditional combination in Protasis *Si* and Apodosis *Si* clauses (e.g. *Si hiciera calor, saldría* ‘If it were hot, I would go out’) rather than the more common traditional or vernacular Indicative-Indicative (e.g. *Si hace calor, salgo* ‘If it were hot, I would go out’).

Furthermore, research on other variable linguistic features confirms that gender can play an important role in terms of linguistic variability in bilingual environments. For example, Shin & Otheguy (2013) found a higher rate of Spanish subject pronouns in the speech of first generation women than in that of men. The authors argue that first generation Established Immigrant women have more conversations than men with the New York Raised (such as their children), and, as a result, picked up the innovative feature from these second generation speakers. Why are women at the vanguard of this language change? In their view, the reasons could lie in women’s looser social networks within their own Spanish-speaking communities, and “the possibility that women are more susceptible than men to external influences on speech patterns” (p.447).

The gender disparities that we find with respect to Subjunctive usage may be caused by women’s inferior social position. According to Eckert (1989), women are more conscious of their gender because of their less powerful position in society. She adds that “whenever one sees sex differences in language, there is nothing to suggest that it is not power that is at issue rather than gender per se” (p. 256). As such, women may unintentionally use language, and in this case the Subjunctive/Indicative contrast, to signal to men that they enjoy (or should enjoy) a more powerful position in the social hierarchy. While women’s sense of power does not directly stem from employing more Indicative verbs where the Subjunctive is expected, their greater usage of the Indicative
may serve to signal their greater contact with the more powerful out-groups (i.e. non-Latino whites), and looser connections to their own Spanish-speaking communities (Shin & Otheguy 2013).

There is evidence that Latina women network less with the in-group, that is, with their Spanish-speaking peers than Latino men. In fact, Falcón (1995)’s study on social networks and employment in Boston shows that there is a difference of 10 percentage points between genders, with Latino men having relied much more heavily on their in-group networks to find their current job. In his study, this gender effect is found in all Latino national or regional categories (i.e. Dominicans, Puerto Ricans, Central and South Americans). Interestingly, Falcón (1995) shows that in-group networking is not necessarily advantageous for Latinos. As a matter of fact, relying on other Latinos to secure jobs may actually lead to lower rewards in terms of earnings and prestige. The sociologist asserts that “education and English-speaking ability are both inversely correlated with network48 usage—the higher the level of Education or English ability, the lower the use of a network. In general, the data are consistent with the notion that those with lower-human capital characteristics are more likely to have relied on a network to obtain a current job” (p.24). This “significant negative effect of [in-group] networking on earnings” identified by Falcón (1995: 27) supports the idea that if the goal is to gain power and social status in the New York City pecking order, Latinas are actually better off if they steer clear of the connections that their own communities can afford.

In this line of thinking, it may be that Latina women’s increased connections to the out-group, signaled via their comparatively low usage of the Subjunctive, could help

---

48 In Falcón (1995), the term “network” appears to refer to in-group networks only. That is, networks within Latino communities.
raise their social status in their own communities. It is the “construction of gender in
[the] community” (Eckert 1989: 246, 247)—or the broadly defined distinct roles that
each gender plays in the economy of the Latino community-- that is responsible for
women’s apparent linguistic innovation. Thus, theories of local prestige norms (Labov
2001) do not seem to hold in the present study.

Because women spend more time with non-Spanish speakers, it could be that the
larger presence of the Indicative in women’s grammar (relative to men) is brought about
by language contact. It is plausible that convergence, defined in Chapter 2 as the pre-
eexisting similarities between two languages (Bullock & Toribio 2004), is responsible for
the greater usage of the Indicative mood in the subset of women who are more
susceptible (than men and other women) to external influences (Shin & Otheguy 2013).
As previously stated, the Indicative mood is much more prevalent than the Subjunctive in
both English and Spanish. Yet, the mood’s greater presence may not affect the speech of
all women alike. Some women may establish interpersonal sensitivity through language
more than others (Hall & Mast 2008; Leaper & Friedman 2007; Leaper & Robnett 2011,
cf. Shin & Otheguy 2013), and thus use the Indicative to a greater extent for that reason.

Even though language contact may be a factor within a diverse group of women
(of different socio-economic statuses, ages, etc.), the significant gender differences that
we find seem to point in a different direction. Indeed, we did not find any significant
differences in the Subjunctive rates of our three generational categories of women, which
suggests that women’s varying exposure to English is not the cause. Instead, it appears
much more likely that women are challenging the status quo by unconsciously flip-
flopping the linguistic expectation in these variable contexts. Even though men and
women draw on both moods overall (when all linguistic contexts are combined) and in the Modal context, the fact that women’s mood choice significantly differs from that of men’s represents a way for women to upset the unequal power relations that exist between genders. Rather than merely a result of language contact, it is women’s desire to equalize the playing field that leads to the comparatively greater presence of the Indicative mood in their speech.

4.2 The New Yorker effect

The results of the linear regressions in Chapter 4 showed that the New York Raised and the Established Immigrants (henceforth labeled the New Yorkers when referring to both groups together, as in the title of this section) have lower rates of Subjunctive than the Newcomers in the Protasis Si context (e.g. Si vinieran, si hubieran venido).

Just as with our male and female consultants in the previous section, the nature of the immigrant groups’ social network, either primarily in-group or primarily out-group, may significantly contribute to their differing treatment of the Subjunctive. As Falcón (1995) asserts, “Social networks are critical in the settlement process of immigrants, providing not only information but also connection to employers and jobs. Immigrants in general tend to rely on [in-group] networks to obtain jobs in urban settings” (p.19). The sociologist finds that the more prestigious the job, the less Latinos used their networks to obtain it. For example, Latino employees in semiprofessional, technical and professional jobs relied less on their in-group networks than workers in labor, service and operative jobs. Falcón (1995: 23) attributes this difference to the importance of credentials as a screening device for the more high-status professions. As recent immigrants from Latin America are disproportionately represented in poorer Hispanic communities (Shin &
Otheguy 2013: 442, cf. data from the 2010 Census), our Newcomers are most likely in low-status professions, and may have counted on their connections with their Spanish-speaking peers to secure their job.

There is evidence that immigrant Latinos are less socially integrated than U.S.-born Latinos, and that “they have smaller and less diverse social networks than their U.S.-born counterparts” (Viruell-Fuentes, Morenoff, Williams, & House 2013: 11). According to a 2017 report from the Pew Hispanic Center, “The composition of networks of friends varies widely across immigrant generations. Most (77%) immigrant Latinos say all or most of their friends are Latinos. But this share drops to 55% among second-generation self-identified Latinos and only 37% among third or higher generation self-identified Latinos” (Lopez, Barrera, & López 2017). Although controlling for education and income lessens the “magnitude of the social ties differential between the foreign- and U.S.-born Latinos”, the generations’ levels of informal integration, network diversity and network size still vary significantly (Viruell-Fuentes, Morenoff, Williams, & House 2013: 9). Indeed, time spent in the United States is significantly associated with social ties, according to these researchers. The longer Latinos live in the United States, the greater the diversity and size of their social networks. This includes U.S.-born Latinos, who have more diverse and larger networks than the Established Immigrants.

The denser social ties that the New Yorkers (NYR and LARI) maintain with the out-group may intensify their aspirations of social mobility, that is, their desire to enjoy a higher social status in the New York City social hierarchy. Their reduced Subjunctive rates may therefore represent a subtle way of signaling these differences. Contact with non-Latino English speakers is essential for social advancement, as, in New York City,
English represents “the language of education, social institutions and power. It is the de facto official language. Spanish, as well as other non-English immigrant languages, are often perceived as obstacles to social advancement and economic success” (Varra 2013: 118). In fact, Latinos who speak English at home earn higher wages49 than those who do not. Yet, although Latinos with higher English proficiency may have greater connections to the out-group (or, vice versa, greater connections to the out-group could enhance English proficiency), language contact is not deemed responsible for the New Yorkers’ reduced Subjunctive rates relative to the Newcomers (as argued with respect to gender differences in the previous section).

Instead, the linguistic differences that we observe may be a way for the New Yorkers to inadvertently signal their U.S. identity to the Newcomers, since being perceived as North American (from the U.S.) increases one’s chances of economic success. It has been shown that the category “American” (i.e. U.S.) is associated with (non-Latino) whiteness50, and that most Latinos do not consider themselves white51. Thus, it is distinguishing themselves from the Newcomers primarily through human contact with non-Latino whites that must play the most significant role in the New Yorkers’ apparent linguistic innovation.

49 “Latinos who speak only English earn higher wages ($41,000) than Latinos who speak a language other than English at home, but their wages are still lower than Whites ($50,000)”, according to Carnevale & Fasules (2017: 12).
50 Devos & Banaji (2005)’s very interesting investigation confirms the long-standing ideology in the United States that “to be American is implicitly synonymous with being White” (p.2).
51 According to a 2018 Washington Post article by Ed Morales, “On the census forms most Latinos prefer not to identify as either black or white, and 97 percent of those who check the “some other race” category are Hispanic.”
4.3 The Youth effect

The linear regressions in Chapter 4 established that the older the informant, the greater his/her Subjunctive rate in the Protasis Si context (e.g. Si vinieran, si hubieran venido). As discussed in Chapter 2, Gutiérrez (1994) also found that younger informants exploit the Subjunctive less than older informants in Michoacán, Mexico and Lantolf (1978) discovered that younger New York City Puerto Ricans tend to use the Indicative more than their older peers. Similarly, research on other linguistic features, also discussed in Chapter 2, has shown that youth can spearhead language changes in progress.

The social status hypothesis developed in the previous sections could also apply to differences with respect to age and Subjunctive rates. It is probable that older Latinos have stronger ties to their own Spanish speaking communities than younger Latinos, who have increased contact with the non-Spanish speakers that they go to school with. Indeed, there is evidence that older age (for all types of people) is associated with “smaller, less frequently seen, and less proximal networks that have a higher proportion of kin” (Ajrouch et al. 2001: 116).

Similarly to women and New Yorkers as discussed above, younger speakers may draw on the Subjunctive less than older speakers as a way to demonstrate their connections to the out-group. Younger Latinos’ greater social ties with non-Spanish speakers may serve to signal their real or imagined superior social standing to their older peers.

Finally, we do not view groups of speakers as uniform. Language contact could be responsible for the comparatively greater presence of the Indicative in a subset of younger informants, whose Spanish grammar may be affected by contact with English.
grammar, as it was argued with respect to individual women (section 4.1). Furthermore, contact between bilinguals’ two grammars may be particularly productive in the Protasis Si context, as mood variability is prevalent in both languages in this context. Consider the following examples. The first, example (36), is from a relatively younger informant (20 years old), and the second, example (37), is from a relatively older informant (64 years old). Recall that the median informant age is 34.6 years in our sample (see Chapter 3).

(36) Si la mujer *estaba* aquí y *se casaban*, se demoraban muchos años para traer al esposo… - YOUNGER 38C

‘If the wife was here and they got married, it took many years to bring the husband over…’

(37) Pero si *estuviera* … ella *estuviera* viva, yo le explicara todas las cosas, que este es un país […] para la juventud. Aquí dan mucha prebenda para la juventud que … para que estudie, para que aprenda, que sean algo… en la vida… pero si […] quieren ir a fumar marihuana, y quieren ir a hacer todas esas mala de cosas que hacen… – OLDER 11U

‘If she were… were alive, I would explain all of the things to her, that this is a country […] for youth. Many cushy jobs are given to youth here that… so that they study, so that they learn, so that they become someth… in life… but if […] they want is to go smoke marijuana, and they want to go do all those bad kinds of things that they do…’

The younger speaker, in example (36) is representative of all younger informants’ tendency to employ the Indicative (e.g., *estaba* and *se casaban*) in the Protasis Si context, while the older informant’s usage of the Subjunctive (e.g., *estuviera*), in example (37), is characteristic of older speakers in the same linguistic context. The glosses that I have provided illustrate the possibility of drawing on either mood in English in this context.
For instance, English speakers can opt for the Indicative ‘If she was’ or for the Subjunctive ‘If she were’. In English, the Indicative is much more widespread than the Subjunctive in this context, especially in speech. Thus, those younger informants with out-group connections who also show greater susceptibility to external influences may be transferring the Indicative mood from English *if* clauses to the corresponding Spanish *si* clauses. On the other hand, since the Indicative mood is also prevalent in this context (as it is in general) in Spanish, this could be viewed as a case of convergence between the two languages, rather than as an instance of grammatical transfer.

Yet, when taking a closer look at example (37) above, we cannot help but notice this relatively older informant’s cynical attitude toward youth who grow up in the U.S. Her conviction that youth squander the professional and educational opportunities that are afforded to them may be revealing of a broader rift between younger and older Latinos in New York City. In that sense, our initial hypothesis offers a better explanation for this group, as a whole. Indeed, if faced with such negative attitudes on a larger scale, it is easy to see how younger speakers might exploit the Indicative (consciously or unconsciously) to distinguish themselves from their older peers. Greater reliance on the Indicative mood would therefore allow younger speakers to signal to older speakers that they are deserving of a more prominent social position.

5. **Linguistic context availability and message types**

Although an in-depth analysis that goes beyond the scope of the present investigation would be needed in order to explore the messages associated with each linguistic context, we undertake a preliminary inquiry into the relationship between speakers’ socio-demographic characteristics and Linguistic context availability. It is possible to predict
the availability of three different linguistic contexts (the Concessive context, the Possibility context, and the Volition context) based on speakers’ gender, generation, and gender and generation together, respectively. The specifics (e.g., degrees of availability) are discussed in Chapter 4.

The Concessive context (e.g. *Aunque quieran*) tends to be more available to men than to women. This may be due to the fact that the message type associated with this linguistic context is unrelated to the ideas of possibility and aspiration, which are linked, in turn, to our social change hypothesis introduced in the previous sections. It is evident, therefore, that women who are aspiring to change the social order do not feel the need to draw on clauses introduced by *aunque* as much as men do.

Conversely, the greater availability of the Possibility context (e.g. *Es posible que quieran, tal vez/quizá(s) quieran, a lo mejor quieran*) in New York Raised speech than in Newcomer speech, and the greater availability of the Volition context (e.g. *Quiere que vengan, espera que vengan*) in the speech of New York Raised women than in that of Newcomer women, may bear connections to the social status and social change hypotheses. Both of these linguistic contexts seem to emerge when consultants express notions of possibility, desire, appeal or hope. Perhaps the tendency that second generation informants, and particularly second generation female informants, have to avail themselves more of these particular contexts is an indication that they view their life prospects as less static and more dynamic than the Newcomers (and particularly Newcomer women). On the other hand, Newcomers, and Newcomer women especially, have no other choice than to be pragmatic in order to withstand their unfamiliar situation as new immigrants to New York City.
It is possible that the messages of desire and aspiration conveyed in this particular linguistic context speak more to second generation women, who have more hopeful prospects of improving their social condition than Newcomer women who tend to be poorer (Shin & Otheguy 2013).

What is most fascinating is that we find this marked difference within a same gender. Rather than signaling their aspirations solely to men (which women do with through their usage of the Subjunctive mood), women are using their distinct immigrant position to subtly distance themselves from one another. In fact, New York Raised women are instinctively tapping into their familiarity with New York City to remind recent Newcomer women that they have been here longer, and that they are therefore worthy of greater social or socio-economic rewards.

6. Limited simplification

Can the comparatively lower Subjunctive rates of women, New Yorkers and the young, and the comparatively higher presence of the Indicative in its place be considered an instance of simplification? As outlined in Chapter 2, simplification represents “the expansion of a form to a larger number of contexts (i.e., generalization) at the expense of a form undergoing simplification, which is used with increasingly lower frequency” (Silva-Corvalán 1994: 257). In the current investigation, it was discovered that the presence of the Indicative mood is expanding, but only in a limited number of linguistic contexts. Besides men and women’s distinct overall Subjunctive rates, the only other divergences in Subjunctive frequency (by gender, generation and age) were found in the

---

52 It is important to note, however, that New York Raised women and Newcomer women represent a small portion of the sample (of 142 informants), as shown in Chapter 3. Thus, these statistically significant differences in Volition availability should be further investigated with larger numbers of first and second generation women.
Modal context (e.g. *Como quieran, lo que quieran, como que quieran*) and in the Protasis *Si* context (e.g. *Si vinieran, si hubieran venido*). Thus, if the Subjunctive mood is indeed undergoing simplification, the process appears to be very restricted.

Simplification is typically viewed as being brought about by speakers’ incomplete acquisition, attrition or lack of familiarity with standardized features of Spanish, or by language contact (see Chapter 2 for a discussion of each concept). All of these hypotheses have one common factor: simplification is habitually viewed in a neutral or negative light, as if it were an adverse outcome of living in a bilingual setting. Yet, this possible case of limited simplification should neither be frowned upon nor remedied, if drawing on the Indicative where the Subjunctive is expected represents an unconscious attempt by less powerful (i.e. women and youth) and more confident (i.e. the New Yorkers) groups to effect change in the social order. If simplification is a conscious or subconscious attempt by women, the New Yorkers and youth to signal tangible or coveted social advancement, then this linguistic development should be celebrated.

7. Language change

The question of language change is also relevant to the present investigation. As shown in Chapter 2, there is a debate between advocates of *internal* language change and those that consider speakers’ increased usage of the Indicative in Subjunctive-inducing contexts as the sign of an *external* language change. Supporters of the internal language change hypothesis argue that the increased presence of the Indicative is not unique to Spanish, as the same phenomenon has been shown to occur in other Romance languages (*Silva-Corvalán* 1994), or that these changes are already occurring in non-contact varieties of Spanish (*Torreblanca* 1997; *Zentella* 1997). These scholars claim that language contact
(which is an external factor) may accelerate the process, but that the change begins internally. Yet, the fact that we have a well-articulated external explanation, coupled with the very uneven distribution of the change in the population, speaks much more in favor of external causation than of internal change.

Consistent with our findings is the idea that this possible language change can only be external, as the mood variations described in the previous sections are shaped by the bilingual immigrant context in which they occur. Even though gender inequality and the tensions surrounding age are not unique to New York City or to the United States, women, youth and the New Yorkers’ (sense of) prominence is attained through the immigrant versus native opposition that exists in bilingual New York City.

Furthermore, even if our findings point to an external language change, there is no indication that the change is in progress. Indeed, as evident in Chapter 2, few investigations on the Spanish Subjunctive in the United States include socio-demographic characteristics other than generation. Previous studies incorporating the variables Gender, Age, etc. would have been needed in order to determine whether the differences we observe between our informants are indeed an instance of on-going change.

8. Conclusion

A clear picture of bilingual speakers’ behavior emerges when Linguistic context availability and Subjunctive rates are considered jointly. It was discovered that all speakers have consistent patterns of Subjunctive usage, and that these patterns also exist in certain subgroups of speakers. Speakers were also shown to own a particular repertoire of linguistic contexts, meaning that some contexts tend to be consistently more ubiquitous in their speech than others. In turn, certain contexts were found to shape the
presence and absence of the Subjunctive, which suggests that the appearance of the mood is linked to the types of messages speakers wish to convey. This view of Subjunctive usage points to the fact that the analysis must include a much larger communicative context than the mere clause or sentence, or even the neighboring utterances. Instead, the whole interview should be deemed informative of speakers’ treatment of the Subjunctive mood. When examined in this manner, a research study on Subjunctive usage has a greater chance of reflecting the linguistic behavior of a Latino population as a whole.

In the present investigation, what stands out is that women (and particularly New York Raised women), New Yorkers (i.e. the New York Raised and Established Immigrants) and youth use Spanish differently from men, from the Newcomers (especially from Newcomer women), and from older informants. The availability of certain linguistic contexts and the occurrence of the Subjunctive mood overall and in specific linguistic contexts vary based on the socio-demographic characteristics cited above. A common hypothesis was offered to explain these distributional patterns.

First of all, female, New Yorker and younger Latinos are more exposed to the out-group of non-Latino whites as a result of their education, their jobs and their friends. That is, their out-group networks are denser than those of male, Newcomer, or older speakers, whose networks are still predominantly in-group ones. I have proposed here that female, New Yorker, and younger Latinos are using their greater connections to non-Spanish speakers as a way to display or to elevate their social status within their own communities in New York City. These speakers are (consciously or unconsciously) distinguishing themselves from their Spanish-speaking peers by means of their distinct treatment of the Subjunctive.
Accordingly, women, youth and the New Yorkers’ grammars of the Subjunctive could have undergone or be in the process of undergoing a form of simplification. However, this plausible instance of simplification is limited to specific linguistic contexts, and to particular groups of speakers. Furthermore, the slight expansion of the Indicative mood onto Subjunctive terrain is not attributed to the second generation’s supposed deficiencies in terms of acquisition, mood distinction, or avoidance. In fact, certain linguistic contexts are more available to New York Raised informants (particularly to the women) than to Newcomer participants (especially the women). This may be due to the fact that the linguistic contexts in question appear to convey notions of possibility, hope and desire, reflecting these speakers’ social aspirations. Thus, the increased presence of the Indicative mood represents a way for these less powerful (women and youth) or more integrated (the New Yorkers) groups to signal a real or desired social change.

Similar studies should be conducted in other bilingual U.S. cities in order to gauge whether these apparent linguistic innovations are unique to bilingual Spanish speakers in New York. It is entirely possible that New York City’s competitive culture and reputation as “the city of dreams” have given birth to a distinctive linguistic phenomenon.
Chapter 6: Conclusion

1. Contributions and applications

The current investigation has endeavored to insert itself into the long-standing discussions on the Spanish Subjunctive, and specifically on the more recent conversations on the Spanish Subjunctive in the Spanish of the United States. This study hopes to have contributed to the fields of variationist Sociolinguistics and Second Language Acquisition through its discernment of speakers’ patterns of Subjunctive and linguistic context usage; its extension of the sociolinguistic variable to include clause types; and its case for mood variability in all linguistic contexts and in the speech of all types of informants. Finally, this research is multifaceted not only because of its theoretical aspects, but also because of its applicability in the classroom, in terms of language pedagogy and the training of language teachers.\(^{53}\).

1.1 Contributions to the study of Spanish in the U.S.: Patterns of Subjunctive usage in New York City

This work has contributed to the study of Spanish in the United States by establishing that all speakers have consistent patterns of Subjunctive usage (Chapter 5). We speak of consistent patterns because we find that all of our informants’ Subjunctive behavior is consistent across contexts. On the one hand, there are speakers who exhibit a greater preference for the Subjunctive mood in all of their linguistic contexts; they are never inconsistent, that is, these individuals never prefer the Subjunctive in one linguistic context and disfavor it in another. On the other hand, there are speakers who show no

\(^{53}\) The term foreign language is purposely avoided, as Spanish is not a foreign language to Spanish heritage speakers, for whom Spanish is the home language (see definition further along, in section 1.4 of this chapter).
preference for either mood, in *all* of their linguistic contexts. These patterns illustrate the fact that Subjunctive usage is systematic for all speakers in the Otheguy-Zentella Corpus (OZC).

Furthermore, it turns out that the Subjunctive-inclined group does not correspond to the strict generational categories that it has been assigned in much of the literature (Bookhamer 2013; Fernández Pedraza 2014; Gutiérrez 1990; Lynch 1999; Mikulski 2010; Montrul 2009; Ocampo 1990; Silva-Corvalán 1994, among many others). Instead, we have found that this group is socio-demographically diverse. First, the divergence between the Subjunctive rates of New York Raised and Newcomer informants is much more limited than expected, as it only involves one linguistic context, the Protasis *Si* context (Chapter 4). This is a surprising finding, given the substantial role played by generational categories in the literature. Second, the Established Immigrants bear a greater resemblance to second-generation informants than to their first-generation peers, the Newcomers. Together the Established Immigrants and the New York Raised make up the *Subjunctive-disinclined* group, meaning that they show a relative preference for the Indicative mood (compared to the Newcomers). (My exploration of the Subjunctive patterns of the Established Immigrants of the OZC constitutes a relatively novel undertaking.) Third, other groups besides the Newcomers belong to the Subjunctive-inclined category, namely men and older informants, just as, in addition to the New Yorkers, women and younger informants belong to the Subjunctive-disinclined group. Finally, the suggestion in the literature (Labov 1990; Orozco 2007; Shin & Otheguy 2013) that these linguistic patterns would be statistically associated with socio-economic
status (SES) is misguided; no significant differences were found between Low and Middle SES informants (Chapter 4).

Just as with Subjunctive rates, group patterns became apparent when examining speakers’ Linguistic context availability. Before delving into the statistical analyses that would answer our research questions, we decided to explore the availability (i.e. presence) and non-availability (i.e. absence) of our nine linguistic contexts in the 142 speakers. This decision was based on the suspicion that not all linguistic contexts occur to the same degree in everyone’s speech. And, indeed, some linguistic contexts were found to be more popular (found in more speakers) than others. Thus, our dependent variable *Linguistic context availability* was fashioned in order to compare the availability and non-availability of different linguistic contexts in speech (the selection process for the different linguistic contexts is explained in Chapter 3). The patterns that were discovered by means of statistical analyses involve different linguistic contexts for different groups. Women tend to use one linguistic context to a lesser degree than men, while the New York Raised—and particularly New York Raised women—tend to use one linguistic context *more* than Newcomers, especially Newcomer women. These distinct Linguistic context availability patterns were credited to the relationship between message types and social conditions (Chapter 5).

To sum up, some of the same groups are Subjunctive-inclined *and* linguistic context-inclined (that is, a same linguistic context tends to occur more in their speech), while others behave differently with respect to the two. For instance, women’s attitude is similar with respect to Subjunctive mood (i.e. they are Subjunctive-disinclined) and Linguistic context availability (i.e. they are linguistic context-disinclined). On the other
hand, New York Raised consultants are \textit{not} as susceptible to the Subjunctive mood as the Newcomers, but are \textit{more} likely to have a particular linguistic context appear in their speech.

Research on Spanish in the U.S. has generally focuses on language contact to explain differences found in second generation speech (Otheguy & García 1993; Silva-Corvalán 1990, 1994, among many others). On the other hand, the effects of \textit{human contact} are rarely addressed, even though human contact admittedly goes hand in hand with language contact. We attributed the diverse patterns of Subjunctive discussed above to the desire for social change on the part of women, and to the New Yorkers’ and younger informants’ aspiration to higher social status and greater social advantages (Chapter 5). Within these groups, it could be that a number of individuals’ grammar of the Spanish Subjunctive is affected by the minor distinction that exists in English between the Indicative and Subjunctive moods.

The reduced Subjunctive rates found in women, New Yorkers, and the young could be viewed as an instance of simplification. However, the present study is not entirely committed to this theory, based on the fact that certain linguistic contexts tend to be \textit{more} available to the New York Raised, and particularly to New York Raised women, than to the Newcomers, and more specifically to Newcomer women. This finding casts doubt on Gutiérrez (1990)’s idea that the second generation avoids subordinate clauses in an effort to circumvent the Subjunctive mood. Furthermore, it is argued that the mood distinction hypothesis (e.g., Montrul 2009; Silva-Corvalán 1994, see Chapter 1), which concerns second generation grammar, does not hold. We have noted that the New York Raised draw on the Subjunctive mood to the same extent as the Newcomers, overall and
in all but one linguistic context. Thus, claims as to the presumed loss of mood distinction among the New Yorkers appear to be false.

Finally, many of the investigations on the Spanishes spoken in the U.S. have claimed that cross-generational differences point to language change (Gutiérrez 2003; Otheguy, Zentella, & Livert 2007; Shin & Otheguy 2013; Silva-Corvalán 1986, 1989, among many others). While acknowledging that this may very well be the case with respect to the Spanish Subjunctive in New York City, the current investigation takes a more restrained approach. It is difficult to gauge how men and women, and younger and older speakers have been employing the Subjunctive mood, as these groups are missing from much of the research on the Spanish Subjunctive in the U.S. Moreover, we fail to find a basis of comparison for the novel concept of Linguistic context availability. With little similar research to compare our results to, we cannot be certain that these group differences are recent. Just as gender and age power dynamics (discussed in Chapter 5) are not new, neither may be women and youth’s reduced Subjunctive usage relative to men and older speakers.

1.2 Applications of this study to variationist sociolinguistic research

All of the patterns described in the previous section attest to the fact that variability is anything but arbitrary. As maintained throughout this dissertation, speakers choose, consciously or subconsciously, to employ one or another mood depending on the message that they wish to express. Central to this theory is the idea that the speaker is always in charge of a verb’s mood (in any clause), rather than it being the clause that commands the appropriate mood, thereby overriding the speaker. Another way to put it is that the meaning conveyed by the speaker in each instance of use takes precedence
over a clause’s requisite mood --a concept that this study does not adhere to. In other words, the syntactic perspective (presented in Chapters 1 and 3), which emerged in the nineteenth century, is still inadvertently present in variationist sociolinguistic studies that appear, at first glance, to take a semantic/pragmatic approach to the treatment of the Spanish Subjunctive (e.g., Bookhamer 2013; Fernández Pedraza 2014; Lynch 1999; Ocampo 1990; Silva-Corvalán 1994). This syntactic view emerges through the treatment of obligatory/categorical linguistic contexts, in which the speaker is thought to have no other choice than to use a Subjunctive verb. The distinct messages that, for instance, *quiero que vienes* and *quiero que vengas* may be conveying, have not been considered. This idea can be generalized to other grammatical features that are often unknowingly interpreted through the prism of syntactic analyses (reflecting prescriptive grammar rules).

This notion led to our discussion of what constitutes a sociolinguistic variable (see Chapters 1 and 3). In this work, we have expanded the scope of sociolinguistic variables to include the availability of linguistic contexts, which represent categorical dependent variables in the logistic regressions (Chapter 4). As described in great detail in Chapter 3, for each linguistic context, the binary variable Linguistic context availability has two factors (or levels): Available and Not Available. Furthermore, the same independent variables (Gender, NYR, LARI, Age, SES, LARI*Female and NYR*Female) that, in the linear regressions, were relied on to predict groups’ higher and lower Subjunctive rates helped forecast which groups’ linguistic contexts tended to be, in the logistic regressions, more or less available. The current study has made clear that the variable Linguistic
context availability can operate, and does indeed operate, in the same way as any other sociolinguistic variable.

Yet, conventional linguists (e.g., Lavandera 1978) would likely contend that using the presence and absence of a linguistic context, meaning a clause (e.g., *Aunque quieran*) or a cluster of clauses (e.g., *Quiero que vengan* and *espero que vengan*), rather than the occurrence and non-occurrence of a single word (e.g., *vengan*), is unsuitable for a proper variationist analysis. We would counter that, however unusual, there is no theoretical motive for excluding strings of words from the range of possible sociolinguistic variables. Moreover, their inclusion did not cause any practical issues. As such, we hope to have demonstrated that the presence and absence of clauses can function as sociolinguistic variables in studies on variation.

We realized, though, that an essential methodological step had to be taken before we could begin our exploration of the distribution of forms (Subjunctive versus Indicative verbs) and the distribution of linguistic contexts (available versus not available) in our informants’ speech. I proceeded to ask myself what it would mean to find, for example, 91 percent of Subjunctive verbs and 9 percent of Indicative verbs in a given linguistic context. In addition to plausibly labeling such a linguistic context as obligatory (e.g., Bookhamer 2013), much of the literature would assume that Subjunctive usage in such a context represents the norm. It would follow that a rate of, for instance, 45 percent of Subjunctive verbs in that same linguistic context by an informant or group of informants could well be viewed, even if implicitly, as a deviation from the norm.

This imaginary scenario is in fact commonplace in the literature, with first generation speech (i.e. Subjunctive usage) embodying the norm, and second generation
speech implicitly or explicitly representing the aberration (e.g., Montrul 2009).

Comparing groups’ distinct rates (i.e. Subjunctive rates here) is informative of speaker tendencies within particular linguistic contexts. The knowledge that first generation informants have a significantly higher rate of Subjunctive verbs (e.g., 91 percent) than second generation informants (e.g., 45 percent) in a same linguistic context is valuable in that regard. However, if the linguistic context in question is absent from most speakers’ repertoire, that is, if it hardly ever occurs, then examining cross-generational differences within this context is not particularly useful.

For this reason, I began with a simple examination of the number of speakers that use (and fail to use) each linguistic context. In this way, if very few individuals said *quiero que + Subjunctive or Indicative verb*, then including this linguistic context’s Subjunctive verbs in my general analysis would not be informative of how most speakers treat the Subjunctive mood.

Generally speaking, this model can be applied to any linguistic feature (be it grammatical or phonological) in variation. Counting the number of consultants who use the sociolinguistic variable in a particular context (e.g., only five of 100 informants said *ser/estar pesado ‘to be heavy/boring/annoying’*), in order to determine whether or not to include that context (and thus the variables that appear in that context) in the analysis, is much more informative than simply exploring the variable’s rate (e.g., *ser* was used 70 percent of the time in *ser/estar pesado ‘to be numb’*). In my view, gauging *how many people say what* rather than just knowing about the *what* offers the bigger picture, and

---

54 Of course, investigations that are concerned with less common linguistic features and/or contexts are of interest as well. But just as with more prevalent linguistic features, both researcher and reader should be cognizant of the number/proportion of informants who use each of the contexts in which the sociolinguistic variable occurs (and fails to occur).
constitutes a vital step for any variationist study that aims to represent the speech of an entire community.

1.3 Contributions to Second Language Acquisition research

Now that we know that our New York Raised group’s Subjunctive rate is lower than that of the Newcomer group in just one linguistic context (i.e. the Protasis *Si* context, e.g. *Si vinieran, si hubieran venido*), the theory of incomplete acquisition (Montrul 2002, 2009, 2014; Polinsky 2006), which the second generation has been assailed with, appears irrelevant. Although it may very well be that many of the bilingual speakers who were born and/or raised in the U.S. are more English-dominant than Spanish-dominant, particularly with respect to the formal varieties exercised in school, we have shown that our New York Raised speakers are just as adept at employing the Subjunctive mood as any recent immigrant from Latin America. It is therefore not the case that our second generation speakers have an incompletely developed grammar of the Subjunctive. The claim by Montrul that we have already cited does not hold. It is not true that these speakers may have “missed the opportunity to develop productive use and written comprehension of more complex structures typically developed during the school-age period” (Montrul 2009: 265). If they had missed this opportunity, they would show lower Subjunctive rates than the Newcomers in a majority of linguistic contexts, as well as overall. But that is not the case.

The fact that the New York Raised and the Newcomers use the Subjunctive (and the Indicative) in the same way the majority of the time was a surprising finding. As explained in Chapters 1 and 3, the unusual methodology that the current investigation relies on surely accounts for its distinct results. Furthermore, it may be that
grammaticality judgment tests and fill in the blanks, which are common in the controlled environments prevalent in Second Language Acquisition (SLA) studies (e.g., Montrul, 2009), fail to pick up on the great amount of variability that is present not only in second generation grammar (of the Subjunctive) but also in that of the first generation. Indeed, just as with sociolinguistic research in bilingual settings, SLA research tends to focus more on the discourse of heritage speakers or heritage language learners (as the second generation tends to be called) than on language acquisition among recent immigrants. This is of course understandable, since the Spanish of heritage speakers (the minority language) has been in contact with English (the majority language) for a longer period of time, given these speakers’ upbringing in the U.S. Thus, it is often regarded as more worthy of analysis in this respect.

The SLA studies (e.g., Montrul 2009) that obtain very high rates of Subjunctive, often at or close to 100 percent among native speakers (labelled as such in these studies, and corresponding to our Newcomers) fail to examine authentic speech such as that found in sociolinguistic interviews. Yet, authentic and prescriptive grammar widely differ. The Newcomers who take these grammatically judgment and cloze tests may be recalling the grammar rules of the Spanish Subjunctive that they studied in school. In Latin American schools, Latin American children study Spanish grammar (Vaquera, 2005). The same is true of French children in France. I can still recall the rules of French grammar that I had to memorize, recite and apply in written exercises in my

---

55 Some sociolinguistic studies (e.g., Bookhamer 2013; Lynch 1999) also present very high Subjunctive rates (between 90 and 100 percent) among their first generation informants in several linguistic contexts. However, it is imperative for the reader to remember that the proportion of informants in whose speech these linguistic contexts appear is not made explicit in these investigations. In other words, these studies do present the number of informants that they begin with, e.g. 26 Newcomers in the case of Bookhamer (2013), but they then fail to specify the proportion of informants by linguistic context. Thus, if the linguistic context in question was barely used, the high Subjunctive rates may only belong to a handful of informants.
Parisian elementary and middle schools. I would not be surprised if the Newcomers recall them as well, when participating in these experiments. Yet, this does not mean that the Newcomers’ natural speech reflects the prescriptive rules that they may recall when faced with a written task. Indeed, we have observed through examples from the OZC that the speech of first generation informants often fails to abide by these prescriptive rules.

1.4 Applications to language pedagogy and teacher training: Spanish

Heritage Speakers as prospective Spanish teachers

In the years since I started training prospective foreign language teachers in the School of Education at the City University of New York’s College of Staten Island, nearly all of my future Spanish teachers have been self-proclaimed Spanish heritage speakers (SHS), that is, U.S.-born/raised second generation bilinguals. These Spanish majors enroll in the School of Education track from the World Languages Department with the intention of becoming secondary school Spanish teachers. In the Curriculum Development and Methods courses that I was fortunate to fully design and teach, my undergraduate and graduate students create their own lesson plans, which they practice teaching in our own classroom, in beginner college Spanish courses, and in the middle and high schools in the five boroughs of New York City where most will ultimately end up working as certified Spanish teachers.

What struck me the most when I began teaching these two courses was my SHS’ lack of metalinguistic awareness of prescriptive grammar rules and especially grammar

---

56 According to Valdés (2000: 1), heritage speakers were “raised in a home where a non-English language is spoken, who speak or merely understand the heritage language, and who are to some degree bilingual in English and the heritage language.”
terminology (e.g., subject pronoun). I had assumed that my SHS who were Spanish majors, and who, for the most part, had studied Spanish in high school alongside second language learners (L2L), would come to my courses equipped with some degree of metalinguistic knowledge (in and about any language). Yet, this belief revealed my ignorance of my SHS’ academic experiences with their home language. In college, nearly everyone had placed into the advanced Spanish courses on literature, cinema, theater and civilization, which rarely cover grammar. Although a few of my students had taken at least one Spanish Linguistics course, most showed little metalinguistic knowledge of any sort.

In her study on how metalinguistic knowledge affects the accurate production of the Spanish Subjunctive, Correa (2011) found that her L2L’ accuracy was positively correlated with their metalinguistic knowledge at three levels of instruction (beginner, intermediate and advanced), but that the same could not be said of the SHS, for whom there was no relationship between metalinguistic knowledge and accuracy. (In fact, Correa even finds a negative correlation between the two for the SHS, although this relationship was not statistically significant. In other words, metalinguistic knowledge is detrimental to SHS’ Subjunctive accuracy.) Even though her view of Subjunctive accuracy is debatable, as it seems to align with prescriptive grammar rules, Correa (2011) is not alone in believing that introducing grammar rules might be harmful to the many linguistically insecure SHS. She asserts, “[…] it has been suggested that teaching

---

57 Metalinguistic awareness has been defined as “that part of explicit knowledge that is potentially verbalizable” (Isemonger 2007: 113, cf. Correa 2011).
58 Correa (2011: 131) defines metalinguistic knowledge (MK) as “the ability demonstrated by participants to identify grammar terminology and ungrammatical sentences, as well as to provide grammar rules regardless of their exposure to explicit/implicit teaching.”
methodologies that require MK [metalinguistic knowledge] or over-emphasize the mastering of grammatical rules may be negative both in terms of performance and of self-confidence for this population (Beaudrie 2009; Anderson 2008)” (p.130). Shin & Hudgens Henderson (2017: 196) attribute this to the fact that “some of the grammatical rules in textbooks do not match the patterns of language use in real life.”

Consider the following rules obtained in a Spanish language textbook (A mí me encanta, p.78).59 The two linguistic contexts of interest are bolded for emphasis (and I have also added the glosses):

- To translate the future tense in a temporal subordinate clause, we use **cuando** ‘when’ followed by the present Subjunctive.
- To offer advice […], we use **aconsejar**60 que ‘to recommend’, pedir que ‘to ask to/request’, decir que ‘to tell’ followed by the present Subjunctive.

Now, consider what SHS students often write in the activities that they design:

(38) ¿Qué recomienda Clau que **lleven** cuando **van** a visitar Machu Picchu?

‘What does Clau recommend you bring when you will visit [the] Machu Picchu?’

59 I have translated both examples from French, as they were taken from a Spanish textbook published in France (for learners whose native language is French).
60 Aconsejar ‘to advise/recommend’ and recomendar ‘to recommend’ are close in meaning. Recomendar que, which is used in example (38), could have been included in this list.
61 As a side note, since I know the larger context, it is clear to me that SHS’ use of the periphrastic future in this sentence (**cuando van a visitar** ‘when you are going to visit’) is in fact a transfer from the habitual present tense in English, as in when you (go) visit. I think that they meant to write **cuando visitan** ‘when you visit’, as in What should one bring when one visits the Macchu Picchu? If this is indeed the message that the writers wished to convey, one can see how the Indicative mood (albeit in the present tense) would be more fitting than the Subjunctive. This is a great example of how a speaker/writer’s message drives mood choice (rather than the linguistic context at hand).
Following prescriptive grammar rules, example (38) includes two errors. The Indicative mood is used twice (*lleven* and *vayan*) in the place of the Subjunctive verbs *lleven* and *vayan* (or *visiten*), which are required (according to grammar rules) in both of these linguistic contexts (i.e. Recommendations introduced by, e.g., *recomendar* *que*, and the Temporal context, introduced by, e.g., *cuando*). If these so-called errors appeared in the writing of an L2L, I am certain that any Spanish teacher would correct them. I myself would expect to be corrected as a non-native Spanish speaker, and I definitely would have corrected these “errors” years ago when I taught middle and high school Spanish in France. And, I cannot say for certain that I wouldn’t now (although I would also ideally offer my secondary school students examples of sociolinguistic variation).

However, it is an entirely different story when it comes to heritage speakers, and particularly to heritage speakers who are studying to become Spanish teachers and who may sooner or later be faced with teaching prescriptive grammar rules. So the question that I have had to ask myself, as a teacher trainer and sociolinguist, is the following: *How do I validate the authentic grammar of SHS students while simultaneously touching on the prescriptive grammar rules that they need to learn in order to be able to explain them to their own students, even if only to their L2Ls (not necessarily to their own SHS)?*

---

62 Although I believe that I held prescriptive views of grammar at the time, I only had foreign language learners in my classes (I never had any heritage speakers of Spanish). Of course, my thinking has evolved since then, in that L2Ls would also benefit from being exposed to differences between authentic and prescriptive grammar, and from learning about where and how these differences originate (i.e. discussing language ideologies (del Valle 2014), linguistic prejudice (Shin & Hudgens Henderson 2017), etc.).

63 Ideally, L2 learners and heritage speakers would be in separate high school Spanish classes, whereby the Spanish heritage classes would resemble language arts courses, as proposed by Potowski (2005) and Samaniego & Pino (2000). However, most New York City high schools still mix these two populations together, even when Spanish heritage speakers are nearly as numerous as the L2 learners in a given Spanish class. Thus, if prescriptive grammar rules only target the L2 students, the teacher would have to differentiate instruction in the classroom.
There is much debate in the fields of Second Language Acquisition and Language Pedagogy around whether grammar should be taught *explicitly* or *implicitly* in foreign language classes, and about the specifics in each case (DeKeyser 2008; Ellis 2002; Ellis 2011; Lado, Bowden, Stafford, & Sanz 2014; Lee & Van Patten 2003; MacWhinney 1997; Spada & Lightbown 2008). While I have my students read about and discuss both perspectives, I share Lee & Van Patten (2003)’s belief that grammar should be at the service of communication (rather than the other way around) in language classes, as it is in real life. Besides, practicing and applying the rules of a language does not turn learners into fluent speakers of that language (Schmidt 1992). Accordingly, I prepare my prospective teachers to create lesson plans centered first and foremost on a communicative goal (e.g., students should be able to order food) that is tied to lexical, grammatical and cultural objectives. In this way, the grammar goal(s) that the teacher has in mind for the lesson are *implicitly* incorporated into activities, meaning that their students will use the new grammatical feature(s) without prior introduction through any sort of explanation (that is, no rules, paradigms or attention to forms). In order to achieve this, my student-teachers have to learn to contextualize and model the grammar that they expect their own students to use. Although this method has proven successful (as I have found that students *do* acquire grammatical features through contextualized implicit instruction), my prospective teachers also learn to teach the same grammatical features in an explicit manner (through a focus on form led by student discovery), in the last few minutes of their lessons, after they have been introduced implicitly throughout. In the end, my students are given the chance to exercise both approaches, and are prepared to accommodate different types of learners.
In addition to explicit grammar teaching, the sociolinguistic articles and
eamples from the present investigation that I share in class help my SHS gain
knowledge on how their own language functions. This, in turn, may actually make them
feel less linguistically insecure (i.e. more linguistically self-confident), according to
Mikulski (2006), Potowski (2002), and Shin & Hudgens Henderson (2018). This
happens if SHS learn to consider their utterances as equally or more acceptable than the
standardized forms found in language textbooks and grammar books. I firmly believe
that boosting SHS student-teachers’ linguistic confidence should go hand in hand with
enhancing their self-assurance as language teachers. This can be achieved by teaching
them to simplify grammar, in the way that prescriptive rules do, for the benefit of their L2
students. By learning to perceive language as a fascinating “object of study” (Shin &
Hudgens Henderson 2017: 196), SHS can become better-informed and more self-
confident teachers, who are capable of transmitting their newfound sociolinguistic and
ideological awareness to their own L2 and SHS students.

2. Limitations of the current study

2.1 No SES effect: Socio-economic status could be approached differently
In our study, we did not find any effect for Socio-economic status (SES), meaning that
speakers’ SES cannot be relied on to predict speakers’ overall Subjunctive rate, their
Subjunctive rate in any of the most popular linguistic contexts, or the availability of any
of the different linguistic contexts in their speech (Chapter 4). Yet, taking a different
approach with respect to this socio-demographic variable may lead to different results.

SES, which combines level of education and occupation, was examined as a
simple (non-interaction) variable, similar to most of the other independent variables in
the present study. But even though I did not find an SES effect, explorations of social class or socio-economic status (depending on the study) have obtained significant outcomes when participants’ gender has been taken into account. In monolingual settings, Labov (1990) suggests that interactions between sex and social class are needed “because the behavior of men and women in these various social groups has been found to be quite different […]” (p.221). In bilingual settings, Shin & Otheguy (2013: 431) find that national-group origin is a significant predictor of Spanish subject pronoun variability, and that national origin is tied to SES; in their words, the “hierarchy of change in pronoun rates in NYC mirrors a hierarchy of affluence”. The national groups’ different average socio-economic levels, themselves associated with their diverging social connections, help account for the variance in the use of pronouns. Still, in this study the variables Social class and Education were not found to significantly affect women’s pronoun rates. Moreover, interaction variables were not analyzed in this study.

In the present investigation, the *Women effect* (women’s comparatively lower Subjunctive rates overall and in the Modal context) discussed in Chapter 5 could be further informed if women’s socio-economic status were considered. Perhaps interesting subgroup differences would emerge, with Middle SES women showing lower Subjunctive rates than Low SES women, a hypothesis grounded in the linguistically conservative tendencies observed among poorer Latinos (Shin & Otheguy 2013: 441). Women’s socio-economic status might also affect their Linguistic context availability. Non-significant outcomes would also be informative, as they would lend support to our initial finding about SES. Thus, the inclusion in the regression analysis of the interaction
between SES and Female (in the form of the interaction variable SES*Female) could further enlighten our speakers’ treatment of the Subjunctive.

2.2 Analysis did not consider interviewer speech

The 6576 tokens (Subjunctive and Indicative verbs) derive from the speech of the interviewees, who were responding to questions posed by Latin American interviewers (see Chapter 3). An impressionistic examination of the interview questions suggests that interviewers tried to keep them short and simple. They appear to consist of main clauses, which usually involve the Indicative mood (Collentine 2010). Yet, if the interviewers did indeed draw on the Subjunctive, their usage of the mood may have primed that of the interviewees. Similarly, if the interviewers used a Subjunctive-inducing clause and drew on the Indicative mood, their usage of the Indicative may have induced the interviewees’ usage of the same mood. According to Ameri-Golestan (2012), structural priming (Bock 1986) “happens automatically and is not related to specific communication purposes […] (Levelt & Kelter 1982).” Although interviewer speech does not seem to include Subjunctive-inducing linguistic contexts, an in-depth analysis is in order if priming is considered a factor in mood choice.

3. Directions for future work

Additional inquiries emerged while pondering the various explanations for our results. These inquiries represent directions for future work, and are described one by one below.
3.1 Additional variables: Degree of Spanish use, English skills and Spanish skills

When discussing the social network hypothesis (Chapter 5), it was argued that those groups with reduced rates of Subjunctive (i.e. women, New Yorkers, and younger speakers) are more socially connected to Anglos than their counterparts (i.e. men, Newcomers and older speakers), whose social networks largely comprise Spanish speakers. The three groups’ greater contact with non-Spanish speakers, along with their desire for a higher social status (the New Yorkers) or greater sense of power (women and younger speakers) in the New York City pecking order, was said to account for their greater usage of the Indicative mood (Chapter 5).

The social status and social change hypotheses discussed in Chapter 5 would be further supported with additional information about speakers’ linguistic and social habits. Including socio-demographic variables already available in the OZC, such as Degree of Spanish use in general, Degree of Spanish use with friends, Degree of Spanish use with boss, English skills self-report, and Spanish skills self-report, might provide a fuller picture of our participants’ linguistic behavior. The more we know about how our bilinguals view their language skills, and the more we know about whom they associate with, the more we can infer about their Subjunctive usage. Zentella (2007) has a point when she adapts the traditional aphorism and says Dime con quién hablas y te diré quién

---

64 The socio-demographic variables Degree of Spanish use in general, Degree of Spanish use with boss, Degree of Spanish use with friends, English skills self-report and Spanish skills self-report (among others) have all been coded by O&Z (2012), as the participants’ personal information was collected through questionnaires that accompanied the sociolinguistic interviews for the OZC. O&Z (2012) coded these variables for their own study of Spanish subject pronouns.
eres ‘Tell me who you talk with and I’ll tell you who you are’.$^{65}$ This added knowledge could offer support for—or otherwise challenge—our hypotheses.

3.2 Exploring Subjunctive verb diversity

The idea that second generation speakers’ different (often simplified) grammar of Spanish is representative of a language change in progress is widespread (Shin & Otheguy 2013; Silva-Corvalán 1994, among many others). As outlined in Chapter 5, the present study does not exclude this theory, but takes a more measured stance. One way to gather further evidence for or against the notion of language change with respect to the Subjunctive would be to explore the diversity of Subjunctive verbs present in the speech of our New York Raised participants. A wide range of Subjunctive verbs, and a Subjunctive toolkit as varied as that of first generation consultants would cast doubt on this theory. In other words, the more speakers draw on a variety of Subjunctive verbs, the less likely they are to solely retain fixed or grammaticalized forms. Discourse markers or Idiomatic Usage, such as o sea ‘that is’ (with the unchanging Subjunctive verb sea ‘is’), which Bookhamer (2013) and Lynch (1999) include as a linguistic context (despite the lack of variability in both generations), were excluded from the present investigation. However, lo que is one of three clauses in our Modal clause linguistic context, and is often used as an idiomatic expression in Spanish, as in lo que sea ‘whatever it/this/that is’. Thus, it would be useful to pose the following research questions in an analysis of the diversity of Subjunctive verbs: (a) Which Subjunctive verbs are more and less frequent in speech? (b) How many informants use each Subjunctive verb, and how many informants use each type of Subjunctive verb (e.g.,

$^{65}$ This is a play on words on the expression Dime con quién andas y te diré quién eres ‘Tell me who you go about with and I’ll tell you who you are’ (Zentella (2007)’s own translation).
quiera, quisiera, haya querido, hubiera/hubiese querido would all be considered one type of verb)? (c) Who are the informants who have a larger and smaller variety of Subjunctive verbs? In other words, which socio-demographic variables shape speakers’ lexical (Subjunctive verb) variety? (d) Is it the case that our three Subjunctive-disinclined groups (i.e. women, the New Yorkers and youth), examined together, have a smaller variety of Subjunctive verbs relative to their Subjunctive-inclined peers (i.e. men, Newcomers and older participants)? Similar questions could be put forward with respect to the Subjunctive verbs in each of our nine linguistic contexts. The answers to some of these questions would shed light on the range of speakers’ Subjunctive repertoire, and would have implications for the theory of language change.

3.3 Adding verb tense to the exploration of Subjunctive verb diversity
Starting up where we left off in the previous section, an analysis of verb tense could also enlighten the discussion of Subjunctive diversity. Presumably, the greater the number of verb tenses drawn on in the Subjunctive (e.g., the present habe, the imperfect hablara/hablase, the present perfect haya hablado and the pluperfect hubiera/hubiese hablado), the better the handle a speaker has on the mood. For instance, Lynch (1999) found that all three of his generations employ all Subjunctive tenses. Similarly, Bookhamer (2013) discovered that the distribution of Subjunctive tenses is nearly identical in his Newcomer and New York Raised participants. A study of Subjunctive tense based on verb diversity could lend support, or otherwise challenge, the notion that second generation informants’ speech differs from that of the first generation. Finally, findings such as these would contribute to the debate on whether or not the Subjunctive
grammar of second generation speakers (and that of any other group) is undergoing a process of simplification.
Appendix A: Linguistic contexts

Although the examples below comprise only Subjunctive verbs, verbs in the Subjunctive and in the Indicative were explored in all linguistic contexts. In addition, the Spanish Subjunctive verbs are loosely translated in the glosses, as the meaning of the Subjunctive was not explored in this dissertation.

you = singular, formal form of address. There are two forms of address in Spanish: formal and informal

you = plural

1. Nine linguistic contexts used for overall Subjunctive rate, with glosses

• Volition context
e.g., Quiere que vengan, espera que vengan
‘S/he/you want(s) them/you to come, s/he/you hope(s) they/you come’

• Temporal context
e.g., Hasta que vengan, antes de que vengan, cuando vengan
‘Until they/you come, before they/you come, when they/you come’

• Protasis Si context
e.g., Si vinieran, si hubieran venido
‘If they/you came, if they/you had come’

• Hypothetical Como si context
e.g., Como si vinieran
‘As if they/you came’

• Apodosis Si context
e.g., Si..., quisiera ir, si..., hubiera querido ir
‘If..., s/he/you wanted to go, if..., s/he/you had wanted to go’

• Possibility context
e.g., Es posible que quieran, tal vez/quizá(s) quieran, a lo mejor quieran
‘It’s possible that they/you want, maybe they/you want’

• Concessive context
e.g., Aunque quieran
‘Even if they/you want’

• Uncertainty context
e.g., No creo que quieran, no sé si quieran
‘I don’t think they/you want, I don’t know if they/you want’
• **Modal context**  
e.g., Como quieran, lo que quieran, como que quieran  
‘Like they/you want, what(ever) they/you want, like/as they/you want’

2. **Four most popular linguistic contexts used for Subjunctive rate**

• **Modal context**  
e.g., Como quieran, lo que quieran, como que quieran  
‘Like they/you want, what(ever) they/you want, like/as they/you want’

• **Protasis Sí context**  
e.g., Si vinieran, si hubieran venido  
‘If they/you came, if they/you had come’

• **Temporal context**  
e.g., Hasta que vengan, antes de que vengan, cuando vengan  
‘Until they/you come, before they/you come, when they/you come’

• **Apodosis Sí context**  
e.g., Si..., quisiera ir, si..., hubiera querido ir  
‘If…, s/he/you wanted to go, if…, s/he/you had wanted to go’

3. **Four linguistic contexts used for Linguistic context availability**

• **Volition context**  
e.g., Quiere que vengan, espera que vengan  
‘S/he/you want(s) them/you to come, s/he/you hope(s) they/you to come’

• **Hypothetical Como si context**  
e.g., Como si vinieran  
‘As if they/you came’

• **Possibility context**  
e.g., Es posible que quieran, tal vez/quizás quieran, a lo mejor quieran  
‘It’s possible that they/you want, maybe they/you want’

• **Concessive context**  
e.g., Aunque quieran  
‘Even if they/you want’
Appendix B: Coding manual

The coding manual displays the variables and factors that were used in the statistical analyses, conducted in SPSS. It is divided into two parts. Part 1 includes the variables found in the verb file and in the informant file. The variables in the verb file were fashioned in the initial stages of this dissertation. A verb file includes one row per verb, whereas an informant file contains one row per informant (142 rows, in the present study). The informant file was created with much of the data from the verb file. (The verb file was not used in the statistical analyses, which is why we only present the variables). The informant file was exploited for all of the statistical tests in this study. Part 2 presents the factors that correspond to the informant file’s variables. An explanation in prose supplements the more complex variables and factors.

1. Variables

1.1 Verb file

1. Linguistic context
2. Linguistic environment
3. Verb number (from transcript)
4. Verb TMA (Tense Mood Aspect)
5. Mood choice (Subjunctive, Indicative, Conditional)

1.2 Informant file

Internal variables

- Variables 1 to 9: Total number of Subjunctive verbs in each of nine linguistic contexts
  1. Total number of Subjunctive verbs in Volition context
  2. Total number of Subjunctive verbs in Temporal context
  3. Total number of Subjunctive verbs in Protasis Si context
  4. Total number of Subjunctive verbs in Hypothetical Como si context
5. Total number of Subjunctive verbs in Apodosis Si context
6. Total number of Subjunctive verbs in Possibility context
7. Total number of Subjunctive verbs in Concessive context
8. Total number of Subjunctive verbs in Uncertainty context
9. Total number of Subjunctive verbs in Modal context

• **Variables 10 to 18: Total number of Indicative verbs in each of nine linguistic contexts**

10. Total number of Indicative verbs in Volition context
11. Total number of Indicative verbs in Temporal context
12. Total number of Indicative verbs in Protasis Si context
13. Total number of Indicative verbs in Hypothetical Como si context
14. Total number of Indicative verbs in Apodosis Si context
15. Total number of Indicative verbs in Possibility context
16. Total number of Indicative verbs in Concessive context
17. Total number of Indicative verbs in Uncertainty context
18. Total number of Indicative verbs in Modal context

• **Variables 19 to 22: Subjunctive rate in each of the four most popular linguistic contexts**

19. Subjunctive rate in Modal context
20. Subjunctive rate in Protasis Si context
21. Subjunctive rate in Temporal context
22. Subjunctive rate in Apodosis Si context

• **Variable 23: Subjunctive rate in nine linguistic contexts together**

23. Overall Subjunctive rate
• **Variables 24 to 27: Linguistic context availability of four linguistic contexts**

24. Availability of Volition context

25. Availability of Hypothetical *Como si* context

26. Availability of Possibility context

27. Availability of Concessive context

**External variables**

28. Informant’s identity number

29. Gender

30. Age

31. SES

32. Generation bivariate

33. LARI (Generation multivariate)

34. NYR (Generation multivariate)

35. LARI*Female

36. NYR*Female

2. **Factors (informant file)**

**Common values**

Variables with single-digit factor numbers:

7 Cannot decide
8 Not applicable
9 Missing data

Variables with double-digit factor numbers:

97 Cannot decide
98 Not applicable
99 Missing data
Internal variables

Variables 1 to 9: Total number of Subjunctive verbs in each of nine linguistic contexts
Give total number of Subjunctive verbs (in figures) in each of nine linguistic contexts

Variables 10 to 18: Total number of Indicative verbs in each of nine linguistic contexts
Give total number of Indicative verbs (in figures) in each of nine linguistic contexts

Variables 19 to 22: Subjunctive rate in each of the four most popular linguistic contexts
Give Subjunctive rate (percentage) in each of the four most popular linguistic contexts
Subjunctive rate ranges from 0 to 100 percent
When linguistic context not available, enter a period (so as not to be mistaken for 0 percent)

Variable 23: Overall Subjunctive rate
Give overall Subjunctive rate (percentage) of nine linguistic contexts taken together
Subjunctive rate ranges from 0 to 100 percent

Variables 24 to 27: Linguistic context availability in four linguistic contexts

24. Availability of Volition context
0 = Volition NOT AVAILABLE
1 = Volition AVAILABLE

25. Availability of Hypothetical Como si context
0 = Hypothetical Como si NOT AVAILABLE
1 = Hypothetical Como si AVAILABLE

26. Availability of Possibility context
0 = Possibility NOT AVAILABLE
1 = Possibility AVAILABLE

27. Availability of Concessive context
0 = Concessive NOT AVAILABLE
1 = Concessive AVAILABLE

Note: Recall from Chapter 3 that if one environment is available, the whole linguistic context is considered available.

External variables

28. Informant’s identity number
Enter numbers with no left zeros or letters
e.g., 002U = 2

211
29. Gender
0 = Male
1 = Female

30. Age
Give age in figures

31. SES
0 = Low SES
1 = Middle SES

**SES was computed in the following way** (by Otheguy & Zentella 2012):

- The respondent receives a number of points according to their educational level (attended -- not necessarily completed):

  1 point   - Elementary school
  2 points  - Secondary school
  3 points  - College
  4 points  - Graduate

- The respondent receives a number of points according to the category of occupation:

  1 point   - Unskilled
  2 points  - Skilled blue collar, clerical
  3 points  - Store owner, manager, white collar
  4 points  - Professional, business owner

For high school students, O&Z (2012:71) used high school for their education (2 points), and the occupational rating of the (highest rated) parent. For college students and all others for whom they did not have occupational scores (e.g. housewives, unemployed), they multiplied their education rating by 2. The result of this procedure is that all college students fall into Class C.

- Four constraints in O&Z (2012):

  1 = Level A: Informant scored 1-2 points
  2 = Level B: Informant scored 3-4 points
  3 = Level C: Informant scored 5-6 points
  4 = Level D: Informant scored 7-8 points
• The present investigation reduced O&Z’s number of constraints from four to two, to create a binary variable:

\[
0 = A + B = \text{Low SES} \\
1 = C + D = \text{Middle SES}
\]

*Note:* The only upper-class informant is in the Middle SES category.

32. **Generation bivariate**

\[
0 = \text{LARN (Latin American Raised Newcomers)} \\
1 = \text{LARI (Latin American Raised Established Immigrants)} \\
2 = \text{NYR (New York Raised)}
\]

33. **LARI (Generation multivariate)**

\[
0 = \text{Not LARI} \\
1 = \text{LARI}
\]

34. **NYR (Generation multivariate)**

\[
0 = \text{Not NYR} \\
1 = \text{NYR}
\]

35. **LARI*Female**

\[
0 = \text{Not LARI Female} \\
1 = \text{LARI Female}
\]

36. **NYR*Female**

\[
0 = \text{Not NYR Female} \\
1 = \text{NYR Female}
\]

• **Generation bivariate versus Generation multivariate:**

*Generation* was coded differently when used in bivariate and multivariate statistical tests. Bivariate statistics are less complex because only two variables are entered together. A Generation variable with three factors (LARN, LARI, and NYR) is not an issue. However, categorical variables with three levels or factors are difficult to examine when entered into a regression. For this reason, it appeared statistically sound to transform Generation into a binary categorical variable for the multivariate analyses. This was achieved by creating two dummy variables in its place. A dummy variable is an artificial variable created to represent an attribute with two or more distinct levels. Dummy coding a variable means representing each of its values by a separate dichotomous variable. These dummy variables only contain ones and zeroes (and sometimes missing values). Thus, heeding the advice of statisticians, the Newcomers category (i.e. LARN) was turned into the reference group, while the Established Immigrants (LARI) and the New York Raised (NYR) became the two dummy variables. This means that only the LARI and NYR variables appear in the regressions. The dummy variable LARI has two levels: Not LARI and LARI, and the dummy variable NYR has two levels: Not NYR and NYR. Although the LARN category is not visible in the regressions, this group’s results
can be inferred from those of the two other groups. In analysis, each dummy variable is compared with the reference group.

- **Interaction variables:**
The fact that both Generation and Gender are binary variables is especially helpful with respect to interaction variables, whose results are not easy to interpret in a regression. As shown in the coding manual above, the variable Gender has two factors, Male and Female. The present investigation hypothesizes that women’s Subjunctive usage will deviate from the norm (i.e. from what is expected), thus the factor called Female is the focus of the interaction variable. The two dummy variables, LARI and NYR, are introduced into two interaction variables (LARI*Female and NYR*Female). Each interaction variable has two factors, listed in the coding manual above. The interaction variable’s reference category is LARN*Female, that is, Female Newcomers. The statistical results pertaining to female Newcomers can be gathered from those of the two interaction variables.
References


