

The social correlates of lexical borrowing in Spanish in New York City

by

Rachel Marie Varra

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

2013

© 2013  
RACHEL MARIE VARRA  
All Rights Reserved

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

Ricardo Otheguy

---

---

Date

---

Chair of Examining Committee

Gita Martohardjono

---

---

Date

---

Executive Officer

José del Valle

Eva Fernandez

Arthur Spears

---

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

## Abstract

The social correlates of lexical borrowing in Spanish in New York

by

Rachel Varra

Adviser: Professor Ricardo Otheguy

This dissertation investigates lexical borrowing in Spanish in New York. English-origin lexical material was extracted from a stratified sample of 146 Spanish-speaking informants of different ages, national origins, classes, etc., living in New York City. ANOVAs and Pearson correlations determined whether lexical borrowing frequency and the type of vocabulary used (i.e. idiosyncratic and shared) were correlated with speaker traits. Results show that all speakers, regardless of their English proficiency or how long they have been in the U.S., borrow. However, borrowing frequency is best predicted by immigrant generation. Furthermore, borrowing rate may play a distinct role for first and second generation Spanish speakers. In the first generation, the middle class, those with more education, better proficiency in English and more Spanish confidence borrow most often. In the second generation, Puerto Ricans and those with more English confidence borrow most. Results for borrowing vocabulary suggest that the middle class is more responsible for introducing novel vocabulary to the Spanish-speaking community than the working class. Overall, though, both novel and shared vocabulary are integral components of speakers' borrowing inventories. Finally, this dissertation examines flagging (e.g. pauses, fillers, metalinguistic commentary) near other-language strings to determine whether flagging is

indicative of language awareness or linguistic disfluency. Results showed no support for the latter; but evidence for the former interpretation suggests that a simultaneous, albeit slight, process of deborrowing accompanies lexical borrowing in New York. The findings and the methodology of this dissertation contribute to several fields of language study. First, the definition of lexical borrowing used makes it germane to studying borrowing in situations of on-going, face-to-face contact in multidialectal communities. Furthermore, a corpus-based approach to differentiating between two types of bilingual speech phenomena, lexical borrowing and codeswitching, is offered. Second, findings show that borrowing in New York is not a 'deficit' behavior and that some aspects of identity (such as arrival age and class) cross-cut traditional (i.e. regional) characterizations of U.S. Latinos. Finally, findings for bilingual speech partially corroborate models describing the long-term outcomes of contact that predict that the quantity of contact features observable in a language is a function of the intensity of contact.

Para mis abuelos,

Quienes se atrevieron a tirar sus piedras al norte,  
y quienes engendraron un legado de mujeres apasionadas por la educación,  
dedicadas al servicio e intrépidas ante el trabajo duro

## **Acknowledgements**

Many individuals supported the completion of this work. The corpus on which the present investigation is based was created by Ricardo Otheguy and Ana Celia Zentella. Through their generosity and labor it has become available to their students, myself included. Their corpus was produced with the assistance of the National Science Foundation (grant number BCS 0004133). My appreciation extends to the informants of the corpus that not only gave their time, but whose candid stories made data collection a treat. I am grateful to Wilmer Rodriguez, Aldeny García, Carolina Chavez, Carmen Zúñiga, Juan Hernandez Espinosa, Edwin Ramos and Agustin Montalvo, who, among others, all volunteered to be Spanish dialect consultants. A very special thank you to Doctor Ramona Hernandez, Director of the Dominican Studies Institute (CUNY), and her team, who generously gave their time in lively discussion on Dominican migration in the U.S.

I extend my gratitude to my colleagues and friends, who graciously provided feedback on portions of this work. Thank you to Ashraf Nashed for being my loyal adversary and fierce ally, and for being there for all my beginnings and all my ends. Thank you to Agustina for gentle and broad-minded conversations. Thank you to Carolina for direction when I needed it. Very special thank you's go to Alexia and Berenice. To Berenice for the innumerable moments during many years in the city and after that I called on you and you answered. And to Alexia, for challenging me and your complete faith in friendship. To these formidable women (and man), I reiterate my thanks, because without them, this work, and my days, would have been less.

I feel fortunate when I reflect on the gifts of my family, which made research and writing a reality. To my father John Varra, thank you for conversation that enriches and inspires. To my stepmother Diane Varra, thank you for the hours you freed up and continued encouragement. To my brother, Daniel Varra, thank you for sanity and laughter. To my sister Michelle, thank you

for wise counsel. To Noemi and Eugene York, thank you for being my devoted and enthusiastic cheerleaders. And, finally, to Sonia Varra I write what I know pales in the immensity of the joyful sacrifice of the woman I call *Mutti*: Thank you.

A sincere thank you to my committee members, José del Valle, Eva Fernandez and Arthur Spears. Their early input strengthened the design and direction of my work. Their comments and critiques have improved its quality considerably. In particular, I would like to recognize, with much gratitude, the time, encouragement, flexibility and counsel of my advisor, Ricardo Otheguy. His insight, exacting standards and judiciously-placed remarks, which guided me and this work at every stage, were only the beginning. Through him I have been afforded an exemplary model of disciplined scholarship, eloquent and lucid locution and indefatigable commitment to students, from which I shall continue to benefit. To those that I have not explicitly named — those family, friends and professionals that complete the circle of generosity from which I have benefitted — a heartfelt thank you! Error or oversight that remains in this work are my own.

## Table of Contents

<b>Abstract</b> .....	iv
<b>Dedication</b> .....	vi
<b>Acknowledgements</b> .....	vii
<b>Notation and conventions used in this dissertation</b> .....	xix
<b>Chapter 1: Introduction</b>	
1. Introduction and research questions .....	1
2. Spanish in contact with English in New York City: A brief sketch.....	6
3. Lexical contact phenomena in Spanish in New York: A preview .....	9
4. Other terminology .....	12
<b>Chapter 2: Literature Review</b>	
1. The concept of lexical borrowing.....	14
2. The data and terminology of lexical contact phenomena.....	16
2.1 Lexical borrowing and codeswitching .....	17
2.2 A model of lexical borrowing and codeswitching as distinct phenomena.....	19
2.3 The Matrix Language Frame (MLF) model .....	20
2.4 Codemixing and government .....	22
2.5 Transfer and transversion .....	23
2.6 Criteria that have been used for classifying lexical borrowing and codeswitching .....	24
2.6.1 One-word donor language string .....	24
2.6.2 Corpus frequency .....	28
2.6.3 Recipient language phonology.....	29
2.6.4 Recipient language morphology .....	32
2.6.5 Summary .....	34
3. Lexical borrowing in language contact and bilingual speech studies .....	34
<b>Chapter 3: Methodology</b>	
1. Roadmap to this chapter .....	40
2. Informants and interviews .....	40
3. Independent variables and stratification of the Spanish in New York corpus .....	42
4. Data selection criteria and the formation of the English lexical borrowing database.....	43
4.1 Distinguishing between lexical borrowing and codeswitching.....	43
4.2 Result of contact in the U.S. ....	48
4.2.1 Preimmigration established loanwords .....	48
The precontact test .....	49
Exceptions.....	52
4.2.2 International loanwords .....	52
4.3 English-origin.....	53
4.3.1 Dialect contact loanwords.....	53
4.3.2 Synonymous bilingual homographs.....	55
The phonetics test .....	57
4.3.3 Etymologically foreign loanwords in English .....	62

4.4	Lexical pattern reproduction .....	63
4.4.1	Full articulation of a word.....	63
4.4.2	Interlingual coincidences .....	63
	The homonymy test.....	68
4.4.3	Proper noun phrases.....	69
4.5	General inclusions .....	72
4.6	General exclusions .....	73
4.6.1	Codeswitches .....	73
4.6.2	Edge of codeswitches.....	75
4.7	Summary .....	75
5.	Dependent variables .....	76
5.1	Discourse frequency measures .....	76
5.1.1	Lexical borrowing rate.....	76
5.1.2	Lexeme rate.....	77
5.2	Lexical innovation and reproduction measures: Nonshared and shared rate .....	80
6.	Analysis.....	81

**Chapter 4: Factors that condition English lexical borrowing frequency**

1.	Introduction .....	82
2.	The interpretation of social categories .....	83
3.	Overview of the corpus .....	85
4.	Immigrant generation, occupational class, level of education and daily Spanish use significantly condition lexical borrowing rate .....	86
4.1	The effect of arrival age and immigrant generation on borrowing frequency .....	86
4.2	The effect of occupational class on borrowing rate .....	91
4.3	The effect of level of education on borrowing rate .....	92
4.4	The effect of daily Spanish use on borrowing rate .....	93
4.5	Summary of the effect of immigrant generation, class, education and daily Spanish use on borrowing rate .....	94
5.	Discussion of immigrant generation and daily Spanish use as variables that condition borrowing rate .....	94
5.1	Interpreting the effect of immigrant generation on borrowing rate .....	94
5.2	Interpreting the effect of daily Spanish use on borrowing rate .....	96
6.	Immigrant generation and occupational class best predict borrowing rate in Spanish in New York City.....	98
7.	Variables that do not significantly condition borrowing rate in the Corpus of Spanish in New York City: Regional origin, sex, age, years in the U.S. ....	101
7.1	Ethnonational affiliation and regional origin as variables that do not significantly condition borrowing rate .....	101
7.2	Sex as a variable that does not significantly condition borrowing rate.....	103
7.3	Age as a variable that does not significantly condition borrowing rate .....	104
7.4	Years in the U.S. as a variable that does not significantly condition borrowing rate .....	106
7.5	Summary of variables that do not significantly condition borrowing rate.....	109
8.	Conclusion.....	110

## Chapter 5: Immigrant generations in focus

1. Introduction .....	112
2. The first immigrant generation.....	113
2.1 Occupational class, level of education, English proficiency, Spanish confidence and daily Spanish use condition borrowing rate in the first generation.....	114
2.2 Summary of variables that significantly condition lexical borrowing rate in the first generation .....	117
2.3 Discussion of occupational class, level of education, daily Spanish use, English proficiency and Spanish confidence in the first generation .....	117
2.3.1 Interpreting the effect of occupational class on borrowing rate in the first generation.....	117
2.3.2 Interpreting the effect of level of education on borrowing rate in the first generation.....	120
2.3.3 Interpreting the effect of English proficiency on borrowing rate in the first generation.....	122
2.3.4 Interpreting the effect of Spanish confidence on borrowing rate in the first generation.....	123
2.3.5 Interpreting the effect of daily Spanish use on borrowing rate in the first generation.....	124
2.4 Occupational class best predicts borrowing rate in the first immigrant generation ....	124
2.5 Variables that do not condition borrowing rate in the first generation: Regional origin, sex, age, arrival age and years in the U.S. ....	125
2.5.1 Sex, age and arrival age as variables that do not significantly condition borrowing rate in the first generation.....	125
2.5.2 Regional origin as a variable that does not significantly condition borrowing rate in the first generation .....	126
2.5.3 Years in the U.S. as a variable that does not significantly condition borrowing rate in the first generation .....	127
2.6 Summary of lexical borrowing in the first immigrant generation.....	130
3. The second immigrant generation .....	131
3.1 Regional origin, English confidence and daily Spanish use condition borrowing rate in the second generation .....	132
3.2 Summary of variables that significantly condition lexical borrowing rate in the second generation .....	134
3.3 Discussion of regional origin, English confidence and daily Spanish use in the second generation .....	135
3.3.1 Interpreting the effect of regional origin on borrowing rate in the second generation.....	135
3.3.2 Interpreting the effect of English confidence on borrowing rate in the second generation.....	142
3.3.3 Interpreting the effect of daily Spanish use on borrowing rate in the second generation.....	143

3.4	Regional origin and daily Spanish use best predict borrowing rate in the second immigrant generation .....	145
3.5	Variables that do not condition borrowing rate in the second generation: Sex, age, occupational class, level of education, ethnonational origin and Spanish proficiency .....	146
3.6	Summary of lexical borrowing in the second immigrant generation .....	147
4.	Conclusion.....	148

**Chapter 6: Innovation, reproduction and the dissemination of lexical borrowings in Spanish in New York City**

1.	Introduction and questions .....	150
2.	Shared and nonshared vocabulary in Spanish in New York .....	152
3.	The use of nonshared vocabulary: Lexical borrowing innovators .....	157
3.1	Variables that significantly condition nonshared rate: Occupational class, daily Spanish use and English skills.....	158
3.2	Discussion of class, daily Spanish use and English skills as variables that condition innovation in lexical borrowing .....	160
3.3	English skills best predict innovative borrowing .....	163
3.4	Discussion of variables that do not condition innovation in lexical borrowing .....	163
3.5	Innovative borrowing: Summary and conclusions .....	167
4.	The use of shared vocabulary: Lexical borrowing reproducers .....	167
4.1	Variables that significantly condition shared rate: Occupational class, daily Spanish use and English skills .....	168
4.2	Discussion of class, daily Spanish use and English skills as variables that condition reproduction in lexical borrowing .....	170
4.3	Daily Spanish use best predicts reproductive borrowing .....	173
4.4	Discussion of variables that do not condition reproductive lexical borrowing .....	174
4.5	Reproductive borrowing: Summary and conclusions .....	176
5.	On the dissemination of lexical borrowings.....	177
6.	Conclusion and review of questions.....	179

**Chapter 7: Deborrowing: The role and interpretation of flagged lexical transfers in Spanish in New York**

1.	Introduction .....	181
2.	Flagging in bilingual speech research .....	182
3.	From assumption to hypothesis .....	184
4.	The independent variable: Flagging rate .....	186
5.	The data and coding .....	187
5.1	Content borrowings: Nouns, verbs, adjectives and adverbs.....	188
5.1.1	Metalinguistic commentary .....	188
5.1.2	Fillers and filled pauses .....	190
5.1.3	False starts.....	192
5.1.4	Unfilled pauses.....	194

5.2	Functional words and expressions: Discourse markers, determiners, prepositions, coordinators, subordinators and conjunctions .....	198
5.2.1	Metalinguistic commentary .....	198
5.2.2	Fillers and filled pauses .....	199
5.2.3	False starts.....	200
5.3.4	Unfilled pauses.....	200
5.3	Summary of borrowing flags.....	200
6.	Results .....	201
6.1	The language membership hypotheses .....	201
6.1.1	Evidence in support of the language membership hypotheses .....	201
6.1.2	Results of analysis and discussion of hypothesis (1a): Flagging indicates awareness that an English lexical borrowing has a foreign origin.....	203
6.1.3	Results of analysis and discussion of hypothesis (1b): Flagging indicates that an English lexical borrowing is not part of the lexicon of Spanish in New York .....	205
6.1.4	Summary of findings for the language membership hypotheses .....	206
6.2	The disfluency hypotheses .....	206
6.2.1	Evidence in support of the disfluency hypotheses .....	206
6.2.2	Results of analysis for a disfluency hypothesis .....	208
6.2.3	Summary of results for a disfluency hypothesis .....	211
6.2.4	Discussion of hypothesis (2b): Flagging does not signal lack of proficiency in Spanish or English .....	211
6.3	Summary of conclusions .....	212
7.	Reinterpreting flagging for lexical borrowing in Spanish in New York.....	213
8.	Conclusion.....	215

## **Chapter 8: Conclusion**

1.	Contributions and applications.....	216
1.1	Contribution of the current study to Spanish in the U.S.: A portrait of lexical borrowing in Spanish in New York.....	216
1.2	Applications of the current study to language contact research.....	219
1.2.1	The concept of lexical borrowing .....	219
1.2.2	Some aspects of lexical borrowing are more tied to bilingual ability than others.....	219
1.2.3	Social determinants of lexical borrowing across contact situations .....	220
1.2.4	Some factors that promote a high density of permanent contact features in language also promote a high density of contact features in speech; others do not .....	224
1.3	Contributions of the current study to bilingual speech research: The classification of lexical contact phenomena .....	226
1.3.1	The distinction between codeswitching and borrowing can be made with respect to patterns of language use in a community .....	226
1.3.2	Flagging may not be a good diagnostic for classifying data.....	227
1.4	Applications of the current study to language pedagogy .....	227
1.5	Application of the current study to language ideology .....	228

2.	Limitations of the current study .....	229
2.1	Data selection: The shared lexicon of Spanish in New York may be underrepresented .....	229
2.2	Analysis did not control for variation due to interpersonal distance.....	230
3.	Future research .....	231
3.1	A comparative examination of several types of lexical and semantic contact data ....	231
3.2	Borrowability scales and how social and linguistic factors jointly contribute to patterns of contact features .....	231
3.3	Motivations for lexical borrowing.....	232
3.4	The morphosyntactic integration of lexical borrowings.....	233
3.5	The dissemination and sharedness of lexical borrowings .....	234
<b>Appendix A: Independent variables and sample stratification .....</b>		<b>235</b>
<b>Appendix B: Precontact test tool (sample) .....</b>		<b>240</b>
<b>Appendix C: Results of the homonymy test</b>		
1.	Data included in the study: Interlingual coincidences determined to be Spanicized English lexical borrowings .....	245
2.	Data excluded from the study: Interlingual coincidences determined to be semantic extensions .....	246
<b>Appendix D: Examples of lexical borrowing by part of speech category</b>		
1.	Nouns and noun phrases.....	247
1.1	Common nouns.....	247
1.2	Pronouns .....	249
1.3	Proper names and proper nouns .....	249
1.4	Exclusions .....	253
2.	Adjectives.....	254
2.1	Adjectives in isolation.....	254
2.2	Adjectives modified by other English words .....	255
2.3	Adjectives modifying English nouns .....	255
3.	Adverbs .....	255
3.1	Adverbs in isolation.....	255
3.2	Adverbs modifying English material .....	256
4.	Verbs .....	256
4.1	Verbs in isolation.....	256
4.2	Verbs with other English material.....	257
4.3	Verbs as set expressions .....	257
4.4	Exclusions .....	258
5.	Prepositions .....	259
5.1	Prepositions without an English complement .....	259
5.2	Analytic locutions without English objects.....	259
5.3	Prepositions in English collocations or idiomatic expressions .....	259
5.4	Exclusions .....	260

6. Determiners .....	261
6.1 Determiners in isolation .....	261
6.2 Possessive pronouns .....	261
6.3 Exclusions .....	261
7. Quantifiers: Numeric, generic and ordinal.....	262
8. Connectives: Conjunctions, coordinators and complementizers.....	262
9. Indexical demonstratives.....	264
10. Tag items .....	265
10.1 Tag items by function.....	265
10.2 Exclusions .....	268

**Appendix E: Supplemental tables**

1. Supplemental tables for Chapter 4 .....	269
2. Supplemental tables for Chapter 5 .....	270

<b>References</b> .....	274
-------------------------	-----

## List of Tables

Table 4.1:	Lexical borrowing in the Otheguy-Zentella Corpus of Spanish in New York City (OZC) .....	85
Table 4.2:	Lexical borrowing rate by immigrant generation-arrival age, All informants .....	87
Table 4.3:	Post hoc analysis of the effect of immigrant generation-arrival age on lexical borrowing rate, All informants .....	88
Table 4.4:	Lexical borrowing rate by immigrant generation, All informants (revised) .....	90
Table 4.5:	Lexical borrowing rate by occupational class, All informants .....	92
Table 4.6:	Lexical borrowing rate by level of education, All informants .....	92
Table 4.7:	Lexical borrowing rate by daily Spanish use, All informants .....	93
Table 4.8:	The ranked effects of immigrant generation, occupational class, level of education and daily Spanish use on lexical borrowing rates in Spanish in New York City, All informants .....	100
Table 4.9:	Lexical borrowing rate by years in the U.S., All informants .....	107
Table 4.10:	Lexical borrowing rate by years in the U.S., First generation .....	108
Table 5.1:	Lexical borrowing rate by occupational class, First generation .....	114
Table 5.2:	Lexical borrowing rate by level of education, First generation .....	115
Table 5.3:	Lexical borrowing rate by English proficiency, First generation .....	115
Table 5.4:	Lexical borrowing rate by Spanish confidence, First generation .....	116
Table 5.5:	Lexical borrowing rate by daily Spanish use, First generation .....	116
Table 5.6:	The ranked effects of occupational class, education level, English skills, Spanish confidence and daily Spanish use on lexical borrowing rates in the first generation .....	125
Table 5.7:	Lexical borrowing rate by years in the U.S., First generation .....	127
Table 5.8:	Lexical borrowing rate by regional origin, Second generation .....	132
Table 5.9:	Lexical borrowing rate by English confidence, Second generation .....	133
Table 5.10:	Lexical borrowing rate by daily Spanish use, Second generation .....	134
Table 5.11:	The ranked effects of regional origin, English confidence and daily Spanish use on lexical borrowing rates in the second generation .....	145
Table 6.1:	Shared and nonshared vocabulary in the OZC .....	154
Table 6.2:	Most frequent lexemes used to borrow in the OZC .....	155
Table 6.3:	Informants' preferences for shared or nonshared vocabulary .....	156
Table 6.4:	Nonshared rate by occupational class .....	159
Table 6.5:	Nonshared rate by daily Spanish use .....	159
Table 6.6:	Nonshared rate by English skills .....	160
Table 6.7:	The ranked effects of occupational class, English proficiency and daily Spanish use on nonshared rate .....	163
Table 6.8:	Shared rate by occupational class .....	168
Table 6.9:	Shared rate by English skills .....	169
Table 6.10:	Shared rate by daily Spanish use .....	169
Table 6.11:	The ranked effects of occupational class, English skills and daily Spanish use on shared rate .....	173
Table 7.1:	Percent of flagged and unflagged borrowings in the OZC .....	203
Table 7.2:	Corpus flagging rates for borrowings with different degrees of community acceptance .....	205
Table 7.3:	Informant flagging rate by immigrant generation .....	209

Table 7.4: Informant flagging rate by English skills .....	210
Table 7.5: Informant flagging rate by daily Spanish use.....	210
Table E.1: Cross-tabulation and Chi <sup>2</sup> for self-ascribed occupational class and socioeconomic status, All informants .....	269
Table E.2 Cross-tabulation of age cohort and immigrant generation .....	269
Table E.3 Cross-tabulation and Chi <sup>2</sup> for level of education and occupational class, First generation .....	270
Table E.4 Cross-tabulation and Chi <sup>2</sup> for level of education and occupational class, All informants.....	270
Table E.5 Cross-tabulation and Chi <sup>2</sup> for level of education and occupational class, Latinos age 19-64 in New York City, 2010.....	271
Table E.6: Tukey's post hoc: Ethnonational affiliation and borrowing rate, Second generation .....	271
Table E.7: Cross-tabulation and Chi <sup>2</sup> for ethnonational affiliation and English confidence, Second generation .....	272
Table E.8: Cross-tabulation and Chi <sup>2</sup> for national origin and English proficiency, First generation .....	272
Table E.9: Cross-tabulation and Chi <sup>2</sup> for ethnonational affiliation and English use, Second generation .....	273
Table E.10: Cross-tabulation and Chi <sup>2</sup> for national origin and English use, First generation .....	273

## **List of Figures**

Figure 4.1: Lexical borrowing rate by arrival age, All informants (outliers excluded).....	89
--	----

## Notation and conventions used in this dissertation

### In the running text

“ ”	quoting another’s words (long quotes are indented and blocked)
‘ ’	category labels (e.g. informants rated themselves as ‘poor’) meaning of a word or gloss
[ ]	IPA phonetic notation
<b>bold</b>	emphasis
<i>italics</i>	concepts or technical terms cites a word-form or lexeme
<i>ITALICIZED SMALL CAPS</i>	used for the citation form of a lexeme where the concepts of <i>word-form</i> and <i>lexeme</i> are juxtaposed (e.g. Ch. 3, section 4.4.2)

### Within quotations

Within quoted text, the following conventions are used:

[ ]	designates words not actually used by author/speaker; the material has been added for clarity and comprehension.
( )	designates material used by the author that should be ignored for ease of reading
(...)	material from the original quote has been omitted
<b>bold</b>	emphasis added

### In excerpts and (numbered) examples from informant interviews

Within excerpts and numbered examples, the following conventions are used:

...	material has been omitted from the excerpt
‘ ’	gloss / translation
xxx	untranscribable
excerpt text.. more text	informant pauses (undetermined length)
end of excerpt...	informant’s utterance continues
end of excerpt.	informant’s sentence ends or informant’s turn ends
<i>italics</i>	English-origin material (usually a lexical borrowing)
<b><i>bold italics</i></b>	distinguishes focal English-origin material from other nearby English material
<u>underlined</u>	draws attention to a feature of the excerpt

### Initialisms

LBD	lexical borrowing database
OZC	Otheguy-Zentella Corpus of Spanish in New York City

## Chapter 1: Introduction

### 1. Introduction and research questions

Lexical borrowing, the use in one language of words or phrases from another language, has long been a topic of central interest in linguistic research. Some of the early treatments catalogued the adoption of other-language neologisms found in languages in situations of contact (e.g. Espinosa 1917; Neumann 1938; Spencer 1947; Trager & Valdez 1937; Tsiapera 1964), while others examined the phonological and morphological adaptation of these borrowings to a recipient language (e.g. Acholonu, Penfield & Okezie 1980; Barkin 1980; Bowen 1975; Harshenin 1964a, 1964b; Holden 1976; Novotná 1967; Pfaff 1979). Recent linguistic treatments of lexical borrowing (and other lexical contact phenomena) have investigated the implications for bilingual language processing (e.g. Clyne 2003; Muysken 1995) and explored universal constraints on language mixture (e.g. Berk-Seligson 1986; Myers-Scotton 1997; Poplack 1980). Consideration of the extra-linguistic influences on lexical borrowing is found as early as Whitney (1881)<sup>1</sup> and Espinosa (1917). It was the work of Einar Haugen (1950, 1972) and Uriel Weinreich (1953), however, that emphasized that insight into contact-induced language change, of which lexical borrowing is just one manifestation, depends as much on identifying the social factors that encourage or inhibit the use of contact features as on linguistic constraints.<sup>2</sup> Accordingly, several

- 
- 1 Whitney writes: “Hence, wherever two tongues come in contact, each is liable to borrow something from the other; and more or less, according to wholly indeterminable circumstances: the measure and nature of the intercourse, the resources of the respective tongues, their degree of facilitating kinship or structural accordance, and so forth. And there are (...) few tongues in the world which are not to this extent mixed” (1881: 10).
  - 2 Weinreich writes: “To predict typical forms of interference from the sociolinguistic description of a bilingual community and a structural description of its languages is the ultimate goal of interference studies. Unfortunately this aim cannot be attained till the missing link – the correlations between characteristics of individual bilinguals and interference in their speech – is supplied” (1966: 86).

approaches to the study of language in bilingual societies, such as in anthropological linguistics,<sup>3</sup> the sociology of language<sup>4</sup> and discourse analysis,<sup>5</sup> have examined the discourse significance of and the macrosocial factors conditioning lexical contact phenomena. Most recently, attention has turned to the contribution of speaker identity to variation in lexical borrowing behavior.<sup>6</sup>

It has been shown that the extent to which individuals borrow varies according to their social class, place of residence (Eslami Rasekh, Ghoorchaei & Shomoossi 2008; Poplack, Sankoff & Miller 1988), sex (Poplack 1980), age (Eslami Rasekh et al. 2008; Ngom 2006), level of education (Eslami Rasekh et al. 2008; Sullivan 2008) and whether they are members of the contact language community or are learning the contact language (Thomason & Kaufman 1988).<sup>7</sup> Yet, the current body of work still exhibits some shortcomings. First, owing to a relative paucity of quantitative sociolinguistic investigations of lexical borrowing, we still lack a systematic understanding of how borrowings spread through a community and whether and how the permanent adoption of a borrowing into the lexicon of a community is conditioned by the social identity of the individuals that use them. Second, few quantitative studies employ data from sufficient numbers of participants so as to provide information on which individuals are most susceptible to borrowing. Finally, the concept of lexical borrowing, adopted from the

---

3 For example, Heller (1992) demonstrates the influence of macro-level social factors, such as institutional policies and group politics, on language choice in interaction. Ornstein (1976) addresses the contribution of sociocultural distance, need and motivation to lexical borrowing behavior.

4 For instance, Myers-Scotton's (1998) Markedness Model of codeswitching proposes that language choice in multilingual communities is one means by which speakers claim or reject sets of social rights and obligations. Ferguson (1959) and Fishman (1967) both highlight the impact of the domain of interaction on language choice.

5 On how the choice to switch to a different language is a tool for organizing talk and communicating meaning beyond the utterance level, see Gumperz (1976, 1982).

6 Although at least as early as 1962, Gumperz emphasized that understanding the dynamics of language change required investigation of variation in the use of language contact phenomena within a speech community (30).

7 Thomason and Kaufman (1988) have argued that for a language undergoing change, the domain likely to be most affected is a function of whether the language was the (community) second language or the (community) first language of the group of individuals introducing the changes. In general, when the language is the community first language, it is in a situation of *borrowing*, whereby lexicon, and only later syntax and phonology, would be affected. (However, see Winford (2003), where speaker agency and an individual's proficiency in the languages in contact are central for determining what gets borrowed.)

historical study of language change, is rarely defined in a way that makes it relevant to languages and speakers in contact situations.<sup>8</sup>

The current investigation addresses these problems and seeks to add to an understanding of the social correlates of lexical borrowing. We employ a quantitative sociolinguistic design to investigate variation in several aspects of borrowing behavior, such as its frequency in speech and its discourse treatment. To do so, this study uses data from 146 Spanish speakers from six ethnonational Latino groups in New York. The questions to be investigated are:

- (i) Does the frequency of lexical borrowing from English by Spanish speakers have correlates in the sociodemographic traits of the speakers? In other words, are characteristics such as age, sex or level of education (and others) related to the frequency of English lexical borrowings used by individuals? What does the constellation of significant correlative and non-correlative sociodemographic traits reveal lexical borrowing in Spanish in New York?
- (ii) How do speakers make use of idiosyncratic (used by one person) and widely-shared (used by several people) vocabulary? Is there a norm for the use of these types of vocabulary to which all or most Spanish speakers adhere?
- (iii) Is there evidence that the use of metalinguistic commentary, false starts, hesitations and fillers (i.e. flags) indicate that informants are aware of the language of origin of borrowings? Is there evidence that they use these discourse features because of lack of fluency in Spanish?

---

8 Haugen, for example, notes that few authors address the problem of distinguishing between “loans made before immigration, international words, and (...) bona-fide loans made during the period of inter-language contact” (1972: 106). In other words, scholars writing on borrowing and loanwords tend to group all donor-origin lexical material together, regardless of whether that material was picked up in the immigrant context or picked up prior to immigration. But distinguishing between the two types of donor-language material is important for, among other things, estimating the extent of linguistic change that results from cultural / language contact.

The main results of this investigation show that the most significant difference in borrowing frequency is immigrant generation (Chapter 4, section 6). The second generation, those born or raised in the U.S., borrow twice as often as the first generation (Chapter 4, section 4.1). Within the second generation, borrowing frequency is best predicted by an individual's regional ties and daily Spanish use (Chapter 5, section 3.4). Within the first generation, the best predictor of borrowing frequency is occupational class (Chapter 5, section 2.4). Middle class first generation informants borrow more than the working class. Notably, in the first generation, low or no proficiency in English does not prohibit borrowing. With respect to borrowing vocabulary, results show that using both widely-shared and idiosyncratic vocabulary are integral to how borrowing is done by the majority of informants (Chapter 6, sections 3.5 and 4.5). Finally, with respect to flagging, results suggest that flagging indicates that individuals are aware of the foreign-origin of borrowings (Chapter 7, section 6.1.3). There is no support, however, for the hypothesis that flagging indicates anything other than momentary disfluency, which is a feature of all spoken discourse (Chapter 7, section 6.2.4). In other words, when individuals flag borrowings, they are not doing so because they lack fluency in Spanish.

The questions, methodology and results of this investigation have applications in several fields of language study. First, although the results are consistent with perspectives that show lexical borrowing to be indexical of speaker identity in real time conversations, this study commences, as much sociovariationist work does, with an eye toward the future. In particular, it departs from the idea that the social identity of speakers constitutes a critical link between the variation we observe in speech and what the long-term outcomes of language contact will be (e.g. see Thomason & Kaufman 1988). In other words, to know what contact features may

become part of Spanish in New York, it is important to know who is using them.<sup>9</sup> Although the present investigation does not make claims about the permanence of particular lexical borrowings in Spanish in New York, or attempt a comprehensive theory of how individual speech behaviors relate to long-term changes, it contributes evidence for the building of such a theory. In particular, it supplies data that can be used for constructing an account of the social determinants of lexical change in diverse contact situations. Second, the use of a variationist framework, traditionally employed to study language change in monolingual communities, is shown to be compatible with studying language change in a multilingual context. As a result, this study makes a modest contribution toward the project of unifying theoretical approaches to language variation and language contact (e.g. see Milroy & Milroy 1985; Muysken 2000; among others). Third, results of this study challenge notions about the nature of language contact in general and lexical borrowing in particular. The results indicate that while the *locus of contact* may be the bilingual (Weinreich 1953/1966: 1), bilingual ability is not always the chief determinant of how much someone borrows. Results also show that monolinguals play an important role in propagating the use of lexical borrowings.<sup>10</sup> Yet, an equally vital factor for the dissemination of borrowings is the amount of time that individuals, whether bilingual or monolingual, spend using Spanish on a daily basis. Fourth, this study defines lexical borrowing in a way that makes the results of the investigation relevant to situations of on-going language contact. The methodology may be useful to other researchers that use bilingual, or so-called

---

9 As Eckert (2000) points out, the manifestations of personal style legitimize and sanction novel patterns of language use, such that they are reproduced by others and become established in the long run. Silva-Corvalán (1995: 4) seems to embrace a similar position when she notes that “in order to assess the stability of the [contact] features identified and the possibility that they may be passed on to new generations”, it is important to know who is using the feature, in particular their proficiency and how they use their languages in daily life.

10 Depending on how you define *lexical borrowing*, this may seem like a trivial conclusion. Given that lexical borrowings in this study **exclude** words that are part of (precontact) Spanish (Chapter 3, section 4.2.1), my assertion is that monolinguals appear to take up and use new or unfamiliar English-origin words quickly. See Chapter 5 (specifically, section 2.5.3) for results.

mixed, speech as data. Finally, this study supplies information on the linguistic and social situation of Spanish speakers in the U.S.

This dissertation is organized as follows. In the remainder of this chapter, I briefly describe the situation of Spanish in New York City, present a preliminary sketch of lexical borrowing and define other terminology. Chapter 2 reviews the literature on lexical contact phenomena. In Chapter 3, the research design and analysis techniques are described. Chapter 4 provides an overview of lexical borrowing in Spanish in New York. Chapter 5 examines borrowing for first and second generation Spanish speakers. Chapter 6 investigates whether there is a city-wide norm in the selection of borrowing vocabulary and who may be most responsible for introducing new borrowings and disseminating them throughout the community. Chapter 7 is a special chapter that investigates the role of flagging near borrowings. Chapter 8 discusses applications of the results, limitations of the study and directions for future work.

## **2. Spanish in contact with English in New York City: A brief sketch**

Spanish is the second most-spoken language in the U.S. and the number of speakers continues to rise (Bills 2005; Roca 2000). New York City, the most populous in the United States with around eight million people (U.S. Census 2010), is home to more Latinos than any other city in the U.S.<sup>11</sup> In 2010, over 2,338,000 Latinos<sup>12</sup> resided in New York, constituting about 28 percent of the city's population (U.S. Census Bureau 2012, Quick Facts New York City). The majority of Latinos in New York are natives to the U.S. (59%). The other 41 percent represent all 19 of the

---

11 Latinos comprise a larger proportion of the populations of Los Angeles (approximately 1,871,032 Latinos or 48.5 percent) and Miami (approximately 289,724 Latinos or 70 percent) (U.S. Census Bureau 2012, Quick Facts Los Angeles and Quick Facts Miami). Numerically, however, there are more Latinos in New York City.

12 Here and in the rest of this work, *Latino* is used to refer to individuals that claim ancestry to a predominantly Spanish-speaking Central or South American country, although *Latino* in more general usage refers to a person claiming ancestry with any Latin American country, Spanish-speaking or not.

officially Spanish-speaking Caribbean and Central and South American countries. Among the most numerous Latino heritages claimed in the city are Puerto Rican (32%), Dominican (26%), Mexican (13%), Ecuadorian (8%), Colombian (5%), Cuban (2%), Honduran (2%), Salvadorian (2%) and Peruvian (2%) (U.S. Census 2010).

Of the over two million Latinos over age five in New York City, 85 percent speak a language besides English at home (U.S. Census 2010). For over 99 percent of those, that language is Spanish. Without a doubt, Spanish in the U.S. is a home language. It is principally at home that individuals learn Spanish,<sup>13</sup> and it is with home interactions that it is most strongly associated. But Spanish is not relegated to home life or isolated neighborhoods. It can be heard and seen virtually anywhere one goes in the five boroughs. School-aged youngsters speak Spanish on the buses and trains en route between home and school. Preschoolers easily converse in Spanish with their parents and siblings. Store attendants and restaurant employees speak Spanish to each other and to patrons when appropriate. In the boroughs and neighborhoods where Latinos are concentrated, business signage can be exclusively in Spanish, while there is also no shortage of bilingual Spanish and English signage in areas less dense with Spanish speakers. Even Spanish-speaking tourists to the city unhesitatingly initiate conversations with strangers in Spanish.

Latino Spanish speakers are not just first generation immigrants, but their U.S.-born (second generation) offspring as well. Of the over one million U.S.-born Latinos over age five in New York City, 76 percent speak Spanish at home (compare to 97 percent of Latino first generation immigrants). In other words, Spanish is a fact of life in New York City and can be encountered in virtually any street, office or organization. While its continued use is no doubt aided by the

---

13 Note that no Census data are currently available on the extent to which Spanish may be spoken by individuals who nonetheless did not learn it at home. Questions on language use as formulated for the U.S. Census ask only whether the respondent “speak[s] a language other than English at home” (About Language Use webpage 2010).

on-going admittance of large numbers of Spanish-speaking immigrants into the U.S. each year (Veltman 1988), the presence of Spanish in the U.S. is not merely the residual of on-going immigration from Spanish-speaking nations: Spanish is being actively taught, used and maintained by the fully-fluent, and also English-speaking, second generation (Villa & Rivera-Mills 2009).

Spanish is not the only language spoken by Latino immigrants and their children. Many also speak English. About 93 percent of U.S.-born Latinos in New York speak English ‘well’ or ‘very well’,<sup>14</sup> while over half<sup>15</sup> of the foreign-born<sup>16</sup> over age five do (U.S. Census 2010). As a result of contact with English and English speakers as well as the high rates of bilingualism among both first and second generation Latinos (both in NYC and in the U.S.), Spanish throughout the U.S. has been noted to manifest several characteristics distinguishing it from Spanish spoken elsewhere, such as an elevated use of overt subject personal pronouns (e.g. Otheguy & Zentella 2012 for Spanish in New York), collocations modeled on English (e.g. Otheguy, García & Fernández 1989 for Cubans in New York), increased use of the present progressive, loss of the subjunctive (e.g. Silva-Corvalán 2002 for Spanish in Los Angeles) and, of course, the presence of English lexical items. It is with this latter, well-known phenomenon, that of English lexicon in Spanish, that the current investigation is concerned.

---

14 I arrive at the figure of “93 percent” as follows. There are 1,111,896 second generation Latinos age five or over in New York City. Of these, 59 percent speak English ‘very well’ and 10 percent speak English ‘well’ (59 + 10 = 69 percent.) I add another 24 percent to this for individuals that did not answer the question and who are, thus, assumed to speak only English at home (69 + 24 = 93 percent).

15 I arrive at the estimate of “over half” as follows. There are 904,763 first immigrant generation Latinos age five or older in NYC. Of these, 48 percent responded to the English ability question as ‘well’ or ‘very well’. Another three percent can be added to that figure to count first generation Latino immigrants that did not answer the question and who, thus, are thought to speak English at home (48 + 3 = 51 percent).

16 In this dissertation, *foreign-born* refers to anyone not born within one of the 50 U.S. states. Specifically, although Puerto Rico is a territory of the U.S. at the time of this writing, I classify individuals born in Puerto Rico as foreign-born.

### 3. Lexical borrowing in Spanish in New York: A preview

To facilitate discussion, a preliminary sketch of lexical borrowing is in order. The distinction between lexical borrowing and codeswitching is touched upon below and developed more fully in Chapters 2 and 3. Additional conditions on the definition of lexical borrowing and inclusion of lexical material in this study are found in Chapter 3. *Lexical borrowing* is defined as the reproduction in one language of a lexical pattern (i.e. minimally one *word-form*<sup>17</sup> plus meaning) from another language (Haugen 1969: 363). In this work, the language in which the form is reproduced will be referred to as the *recipient language*. The language from which the form is taken is called the *donor language*. In some cases, the lexical material may attain the status of *established loanword* (Poplack et al. 1988: 52). That is, it may become part of the lexicon of a recipient language, such that without knowledge of the word's history, one would be unable to distinguish it from any other native word of the recipient language. Before such time, a particular donor language string is hypothesized to undergo a process of social and linguistic modification (e.g. see Ngom 2006: 31), in which the lexical material is subject to variation in morphological and phonological realization, syntactic treatment and the extent to which it is employed to designate a given concept, both by different individuals and by the same individual on different occasions. The social and linguistic modification of donor language material is what is referred to here as the process of lexical borrowing. This process, theoretically begun the first time an individual uses a particular bit of donor language material in recipient language discourse, could also end with the eventual obsolescence of the donor language material for the recipient language

---

17 A *word-form* is a phonetic string that realizes at least one unbound morpheme. The expression *word-form* in the current work is intended to be synonymous with Lyons' (1977: 18) use of the same term, except that he explicitly defines it only with respect to written word-forms.

community. Examples (1)-(4), uttered by informants of this study, contain lexical borrowings (in italics).

- (1) No pero yo en *high school* era consejera y eso también me ayudó... 086P<sup>18</sup>  
'No but I in *high school* was a counselor and that also helped me...'
- (2) Entonces si *jitea* cualquiera de los números que está dentro de esas diez paginitas... 102P  
'So if he *hits* whichever of the numbers that's inside those ten little pages...'
- (3) La felicité... de haber manejado todo este *World Trade Center thing* muy bien, que fue muy flexible con todos... 324E  
'I congratulated her... for having managed all this *World Trade Center thing* very well, that she was very flexible with everyone...'
- (4) Pero llevé una clase de bilingüismo, se llama '*Bilingualism in Literature*' ... 350M  
'But I took a bilingualism class, it was called "*Bilingualism in Literature*" '

In examples (1) and (2) the words *high school* and *jitea* 'hits' are taken from English (the donor language) and used in a Spanish sentence (the recipient language). These are called single-word lexical borrowings. In (3) and (4) the strings *World Trade Center thing* and *Bilingualism in Literature* contain more than one English word. These are multiple-word lexical borrowings.

Lexical borrowings are distinguished from codeswitches, which are **not** examined in this study. A codeswitch is a clause-like donor language string, made up minimally of a finite verb plus another constituent of the verb's sentence. The constituent may be an NP, PP, AdjP, AdvP or another VP. (Such constructions are called clause-like since traditional grammar defines a clause as minimally containing a verb predicate and an overt or implied subject (Haegeman & Guéron 1999: 22)). So, for example, *it wasn't* in (5) would be an example of a codeswitch.

---

18 The number-letter combinations at the end of an excerpt are informant identifiers. "P" stands for Puerto Rican, "C" for Colombian, "D" for Dominican, "E" for Ecuadorian, "M" for Mexican and "U" for Cuban.

(5) Como *it wasn't* como tan formal... 331.1D

'Like *it wasn't* like so formal...'

In (5) the speaker uses the verb *wasn't* plus the subject NP argument *it*. The string *it wasn't* thus constitutes a codeswitch. The presence of a donor language verb does not by itself instantiate a codeswitch. Naturally, verbs can be borrowed. *Jitea* in example (2) is considered a lexical borrowing and **not** a codeswitch, as is *jangueando* in (6).

(6) Después de las dos ya está hecha tu tarea, *jangueando* con los amiguitos...

333D

'After two (o'clock) your homework is already done, *hanging [out]* with friends...'

In (6), *jangueando*, the present participle form of *janguear* 'to hang [out]', does not occur with any other English-language sentential constituent and is therefore classified as a borrowing.

This characterization of lexical borrowing includes lexical contact phenomena that would elsewhere be called codeswitches. For example, *World Trade Center thing* (example (3) and *Bilingualism in Literature* (example (4) would be characterized by Myers-Scotton (1997) as EL-islands and by Muysken (2000) as an insertional codeswitch. There are several reasons, however, for classifying these and similar strings as lexical borrowings, and not as codeswitches.

First, although lexical borrowing is typically restricted to one-word donor language strings, presumably because longer strings would imply either that the speaker is bilingual in the donor language or that two grammars are being accessed (both generally considered to be prerequisites for codeswitching), nothing prevents a monolingual speaker from reproducing longer strings of a foreign language despite not having the least knowledge of its structure (Muysken 1995: 189-190). For instance, monolingual English speakers can and do use complex phrases and entire sentences from another language, such as *je ne sais quoi*, *rendez-vous*, *ménage à trois*, *no hay*

*problema, hasta la vista* and *siéntete*. In fact, *rendezvous* and *menage a trois* are both found in English dictionaries (e.g. *Webster's Unabridged Dictionary* 2001), showing that even multiple-word strings are candidates for becoming established loanwords.

Second, and more importantly, Chapter 3 demonstrates that the distribution of English one-word strings, (non-clause-like) multiple-word strings and clause-like strings suggests that using multiple-word strings is, for our informants, more akin to using single words than to using clauses. Particularly, items in the corpus that we have characterized as single- and multiple-word lexical borrowings are used by virtually every informant of the corpus, whereas those we have called codeswitches are utilized by only a few informants (see Chapter 3, section 4.1 for more on this).

#### **4. Other terminology**

*Language contact situation* refers to situations where people are bilingual or are becoming bilingual (similar to Weinreich 1953/1966: 1), situations where monolinguals in one language live in proximity to and interact with speakers of another language and situations characterized by a combination of these. This interaction minimally includes periodic face-to-face communication, but also includes other modes and channels of communication such as written signs, the internet and telephone conversations.

A *contact feature* is regarded as a linguistic pattern, be it lexical, phonological, syntactic or morphological, from one language (X) that is reproduced in another language (Y) or while speaking language Y, such that at some precontact time, the feature was either non-existent or was used differently in language Y than it is at some post-contact time. This definition of contact feature is consonant with terms like *interference* (Weinreich 1953/1966: 1) and *transfer*, but avoids certain associations of these terms. For instance, *interference* suggests that a contact

feature was accidentally or unintentionally used, while *transfer* has occasionally been defined in a way that implies a more deliberate attempt to replicate patterns of one language in another (see, for example, Clyne 2003: 76). *Contact feature* avoids implications about the intentionality of the speaker in the reproduction of donor language patterns. The term *precontact*, when used to talk about the speech of someone that was born in the U.S., refers to a hypothetical situation where the person's parent(s) had not migrated and he was raised in the parental country of origin.

*Community first language (CL1)* and *community second language (CL2)* refer to a community's historical pattern of language use. For Spanish speakers in New York, Spanish is the community L1. The expressions *first language* and *second language* contained within these expressions are not meant to imply anything about the order of language acquisition or the bilingual competence of Spanish speakers in New York or the informants of this study. That is, the children of Spanish-speaking immigrants sometimes acquire Spanish and English simultaneously, resulting in their having two acquisitional L1s. Furthermore, even though Spanish may be the acquisitional L1 of children and adult migrants to the U.S., English can become their *dominant language* as they are educated in or work in predominantly English-speaking environments. That is, English may over time become the language in which an individual feels more able.

Finally, the expression *bilingual (mixed) speech* is used throughout this work. It refers to discourse that includes material from two languages, regardless of how many languages a person does or does not speak or his proficiency in those languages. That is, in this work, *bilingual speech* is rarely used to refer **exclusively** to discourse produced by fluent bilinguals.

## Chapter 2: Literature Review

### 1. The concept of lexical borrowing

The concept of lexical borrowing is rooted in the theoretical machinery of historical-comparative and sociolinguistics. In the former framework, lexical borrowings are identified in order to exclude words of foreign etymology from the inherited stock of native lexicon to which sound changes systematically apply. In the latter framework, the existence of lexical borrowings in a language may be taken to indicate the direction and degree of cultural contact. As the concept of lexical borrowing has made inroads into research on bilingual speech and synchronic analyses of language contact, there has arisen a tension in its characterization and the types of data to which it refers, owing in part to these origins.

On the one hand, the concept of lexical borrowing retains elements of its historical-comparative characterization: it refers to etymologically foreign elements that have become socially and linguistically incorporated into a recipient language lexicon so that they are indistinguishable from a stock of normally transmitted<sup>1</sup> lexical material. For instance, both Poplack (1987: 55; with Sankoff 1984: 99; with Sankoff & Miller 1988) and Muysken (1995: 190; with Kook & Vedder 1996: 489-450) have characterized lexical borrowings in this way. On the other hand, the concept also takes on a character of novelty, applying to foreign-origin lexical items recently introduced and easily identifiable in a recipient language. For instance, Espinosa's

---

1 Throughout this dissertation, the expressions *naturally transmitted* and *natural transmission* are used with respect to individual words of Spanish or English hypothesized to have come about through modes of language transmission that are free from abrupt change in form or meaning due to exogenous language influence. This usage reflects the spirit of *natural transmission* as defined by Thomason and Kaufman (1988: 146). In their work, however, *natural transmission* characterizes not single words, but whole languages and expresses the idea that the majority of a language's interrelated morphological, grammatical and lexical subsystems have remained intact over some period of time, *despite* evidence of language contact.

studies of English lexical borrowings in Spanish in the U.S. Southwest (e.g. see 1917) include lists, the very construction of which is predicated upon his ability to identify these words as new to Spanish. Haugen (1969: 384-386) asserts that the identification of an English lexical borrowing in U.S. Norwegian relied on the ability of the linguist to show that it did **not** exist in the speaker's variety of preimmigration Norwegian. In holding that already established English-origin loanwords would not qualify as lexical borrowings, Haugen seems to emphasize novelty as a necessary quality of lexical borrowings.

Distinguishing between foreign-origin words that have become part of the recipient language on the one hand, and, on the other, newer introductions is important for studies such as the present one that aim to examine the social factors that **precede and lead up to** the admittance of foreign-origin material into the lexicon of a recipient language. In principal, only newer introductions can be informative about this process.<sup>2</sup> Older incorporations cannot, since they are, as has been stated, already socially and linguistically part of the recipient language. Take, for example, socially and linguistically incorporated words, like *fútbol* 'football' or *clóse(t)* 'closet' from Puerto Rican Spanish.<sup>3</sup> Studying these can tell us about how Spanish speakers of the past have dealt with the introduction of unfamiliar sounds, like English /ʊ/ or /ɑ/. It cannot reveal the social variation that accompanied their incorporation into Spanish, since their use is no longer

---

2 Note that equating new introductions with socially and linguistically unincorporated lexical material is done for expositional simplicity. I recognize that some foreign-origin words may be taken up by speakers and considered part of their language after only a few months or years (see Myers-Scotton's (1997) *cultural borrowings*). In any case, I do not perceive a need to distinguish, within newer introductions, forms that have been incorporated into the lexicon of Spanish and those that have not. It is clear for me that in New York even phrases that with the greatest amount of social acceptance (as demonstrated by very frequent use), like *high school* and *landlord*, are not yet considered part of Spanish. This is demonstrated by the fact that the same speakers who use the English form also use its Spanish doublet, like *escuela secundaria* and *dueño*, to name the same concept.

3 That *clóse* 'closet' and *fútbol* 'football' have attained the status of native vocabulary in Puerto Rican Spanish is evidenced by their inclusion in national dictionaries, such as the *Tesoro Lexicográfico del Español de Puerto Rico* (2005).

conditioned by speakers' attitudes toward, exposure to or fluency in English, but rather by their need to name a particular concept.

The definition of lexical borrowing, in synchronic studies of situations of ongoing face-to-face contact, must shed some of the conceptual residue from its use in historical-comparative linguistics.<sup>4</sup> That is, lexical borrowing should not be equated, in theory or in practice, with foreign-origin material that has clearly become a part of the recipient language lexicon. Consequently, the techniques employed in this study to identify lexical borrowings explicitly attempt to exclude English-origin items known to be a part of Spanish prior to the migration of informants or their families to the U.S. Furthermore, the present study identifies borrowings not on the basis of the extent to which they resemble the recipient language, but upon the extent to which their origin in the donor language, English, can be supported. Operationalization of these criteria is provided in Chapter 3.

## **2. The data and terminology of lexical contact phenomena**

Students of language contact and bilingual discourse often note the variety of terminological and conceptual frameworks available for characterizing donor language words and expressions in a recipient language discourse.<sup>5</sup> In this section, I present just four, fairly prominent, frameworks. The presentation focuses on how the theoretical underpinnings of each are operationalized to

---

4 I do, however, agree with Haugen (1972: 104), who notes that there is no purely synchronic means to identify lexical borrowings.

5 What's more, the literature shows that different terms can be used to describe similar types of lexical contact phenomena, while a single term can be employed to designate different types of data. For instance, Muysken uses the term *codemixing* as a cover term for "all cases where lexical items and grammatical features from two languages appear in one sentence" (2000: 1). Myers-Scotton (1997) uses *codeswitching* to refer to roughly the same. On the other hand, *codeswitching* is used by Muysken to refer to a change in the main language of discourse. Muysken's definition of *codeswitching* is similar to the characterization of codeswitching proffered on at least one occasion by Poplack (1984: 102). Although, the data Poplack calls *intrasentential codeswitches* in her 1980 study seem better characterized with Muysken's definition of *codemixing* (i.e. as lexical items from two languages appearing in one sentence).

distinguish between two types of lexical contact phenomena: codeswitching and borrowing. Describing these frameworks serves two purposes. First, it provides the backdrop against which to contrast the method used in this study for distinguishing between lexical borrowing and codeswitching (Chapter 3). Second, it grounds discussion of results by demonstrating that what is called lexical borrowing in this study includes data classified under a different name in the work of others. I begin with an overview of codeswitching and borrowing.

## 2.1 Lexical borrowing and codeswitching

Recent scholarship on language contact and bilingual speech assumes a distinction between at least two different sorts of lexical phenomena: *lexical borrowing* and *codeswitching* (Woolford 1983; Poplack 1980). Few would argue that *average* and *meetings* in examples (1) and (2) represent a different sort of English language use than *she would buy them* in example (3). (Unless otherwise noted, all examples are from the Otheguy-Zentella Corpus of Spanish in New York, henceforth referred to as the OZC).<sup>6</sup>

- (1) Hice lo que tenía que hacer para pasar, tenía mis [sic] ochenta y cinco *average*. 329D  
'I did what I had to do to pass, I had my eighty-five *average*.'
- (2) Pero después, ya hubo los *meetings* de la escuela... 364E  
'But after, there were already the school *meetings*...'
- (3) Que se los (mis diseños) llevaría después que si ella le gustaba *she would buy them*. 012U  
'That she would take them (my designs) later, that if she liked it *she would buy them*.'

---

6 The number-letter combinations at the end of an excerpt are informant identifiers. "P" stands for Puerto Rican, "C" for Columbian, "D" for Dominican, "E" for Ecuadorian, "M" for Mexican and "U" for Cuban.

Researchers have exerted considerable effort in attempting to characterize the difference between these two phenomena. Typically, they are defined with respect to one another (Clyne 2003). Muysken (2000: 60) and Myers-Scotton (1997), for instance, subsume lexical borrowing under codeswitching, contending that the linguistic processes involved in borrowing of the sort in (1) and (2) are produced by the same processes underlying codeswitching. Alternatively, Poplack et al. (1988: 52-53) define lexical borrowing in contrast to codeswitching, noting that confusing the two would be detrimental for theories of bilingualism.<sup>7</sup>

Yet, what intuitively and theoretically are different phenomena have not been easily distinguished in practice. In an effort to tease them apart, authors have utilized several criteria. Length of the donor language string and morphological and phonological adaptation to the recipient language (Poplack 1980) have often been employed to make this distinction, as has an item's corpus frequency (Myers-Scotton 1997; Poplack et al. 1988; Poplack 1980: 598). In the presentation of authors' frameworks, their techniques for operationalizing the difference between codeswitching and lexical borrowing are discussed. For reasons subsequently made clear, these techniques were discarded for the present study. An alternate means for making the distinction is presented in Chapter 3 (section 4.1).

At this time, I would like to draw the reader's attention to the use of two terms that could be misleading. In the work of Poplack, the phrases *lexical borrowing* and *established loanword* **both** refer to lexical phenomena. Poplack uses *lexical borrowing* for words that are in the social and linguistic process of being adopted and *established loanword* to refer to words for which the

---

7 The work that I am citing does not explicitly say why confusing lexical borrowings and codeswitches is detrimental to a theory of bilingualism. I believe that the statement is based on the premise that borrowing is using one grammar and codeswitching is using two. Thus, if one were concerned with how bilinguals used two grammars together but included data that could not be classified as an example of using two grammars, one's theory would be flawed.

process has been completed. Other authors, however, use the term *lexical borrowing* to refer to both types of data.<sup>8</sup> I, for reasons elaborated upon in section 1 of this chapter, am concerned with only the first type of data: words and phrases that are in the process of being taken up (or not) by a recipient language community.

## 2.2 A model of lexical borrowing and codeswitching as distinct phenomena

Shana Poplack and her associates are major contributors to research on bilingual speech, encompassing work on both borrowing and codeswitching. Her work on borrowing (Poplack & Sankoff 1984; Poplack, Sankoff & Miller 1988) explores the linguistic and social aspects of how foreign-origin words gain or lose currency in communities. Poplack introduces a tri-partite terminology for lexical contact phenomena. *Established loanword* refers to a linguistically integrated foreign-origin word that is “widely used in the speech community, and (has) achieved a certain level of recognition or acceptance, if not normative approval” (Poplack et al. 1988: 52). *Lexical borrowing* “involves the incorporation of individual L2 words (...) into discourse of L1” (Poplack et al. 1988: 52). *Codeswitching* refers to a linguistic process of speaking one language and then another in discourse, without either language influencing the linguistic patterns of the other (Poplack 1980; Poplack & Sankoff 1984: 102). Codeswitching, for Poplack, can occur within a clause in constituent-sized, or smaller, chunks (called intrasentential codeswitching). It can also occur at clause boundaries (called intersentential codeswitching). In general, for Poplack, loanwords and lexical borrowings are necessarily single-word strings. Codeswitches,

---

8 Even Poplack herself has used the term *lexical borrowing* to refer both to already-incorporated words and words that are in the process of incorporation. For example, Poplack (1980) operationally defines *lexical borrowings* in a way that (partially) equates them with *established loanwords*; she required them to be linguistically adapted to recipient language patterns. To be fair, however, the focus of that study was codeswitching, not lexical borrowing. Excluding lexical material that was linguistically adapted to the recipient language, regardless of what the lexical material were called, was appropriate for her goals in that investigation.

while mainly referring to multiple-word foreign-origin strings, can also be single foreign-origin words.

Chief among the theoretical concerns found in several of Poplack's publications, then, is the differentiation between lexical borrowings and single-word codeswitches. Her massive study of English-origin words in Canadian French (Poplack et al. 1988) suggested that single words with indeterminate status in the community lexicon are structurally more similar to established loanwords<sup>9</sup> than to codeswitches. In particular, she and her coauthors found that when an English-origin word was used by more than ten speakers of their corpus (called *widespread*), it was very likely to be pronounced with French phonology.<sup>10</sup> They further found that 85 percent of all single English-origin words in their corpus were morphologically adapted to French. In other words, single English-origin words in French manifested two characteristics proper to established loanwords: morphological and (under certain conditions) phonological treatment according to recipient language patterns. Their results provide empirical support, on the basis of structural similarity, for the classification of single foreign-origin words as borrowings, as opposed to one-word codeswitches.<sup>11</sup>

### **2.3 The Matrix Language Frame (MLF) model**

Carol Myers-Scotton, principal architect of the Matrix Language Frame (MLF) model of codeswitching (1997), also recognizes a distinction between borrowing and codeswitching. She

---

9 *Established loanwords* were those that were well-attested in historical sources, such as French dictionaries.

10 Poplack and Sankoff (1984: 119-120) also found a similar relationship between corpus frequency and phonological integration for English loanwords in Puerto Rican Spanish in New York.

11 Poplack and Meechan (1998) summarize the morphological and syntactic behavior of single other-language words borrowed between typologically diverse language pairs. They likewise found that single other-language words tended to be morphologically and syntactically integrated into the recipient language, just as established loanwords, by definition, are. Poplack and Meechan interpret this as legitimizing the decision to consider single other-language words and established loanwords as part of a unified set of data, on the hypothesis that integration on the discourse/linguistic level implies integration on the community level. Poplack and Dion (2012) come to a similar conclusion

proposes that under certain circumstances the two phenomena are constrained by the same underlying processes. *Codeswitches* are multiple-word strings of a donor language (*embedded language* (EL) in Myers-Scotton's terminology) found in an otherwise recipient language discourse (*matrix language* (ML)). Borrowings, according to Myers-Scotton, come in two flavors. *Cultural borrowings* are foreign-origin words that fill a lexical gap in the speaker's or a language's lexicon. These types of borrowings are associated with abrupt *integration* into the recipient language, rendering them indistinguishable from any other native word. In having been integrated into the recipient language, cultural borrowings are used predictably and consistently to name that which they signify. *Core borrowings* express concepts for which speakers have a word in the recipient language. These are **not** integrated into the recipient language and must be retrieved for use via the donor language. They are, thus, controlled by the same linguistic production processes as codeswitches (Myers-Scotton 1997: 163).

In the MLF model, the notions of *integration* and *functional competition* are the characteristics that distinguish one type of borrowing from the other, as well as borrowing from codeswitching. Myers-Scotton operationalizes integration on the basis of corpus frequency of the donor language string. In particular, she reasons that since a cultural borrowing is systemically recipient language material (i.e. integrated), it can be employed freely in speech, wherever what it signifies must be named. This translates to higher rates of usage and, in turn, the higher occurrence rate of a cultural borrowing in a corpus. On the other hand, core borrowings and codeswitches, being donor language material and not integrated into the recipient language lexicon, functionally compete with recipient language forms to name concepts. Functional competition results in a situation where the appearance of a particular string to name a concept will be unpredictable and idiosyncratic in discourse, that is, infrequent. In the MLF model, then,

borrowings are single words that stand for new concepts (no matter how frequently or infrequently they occur in a corpus). Codeswitches are not only multiple-word strings from another language, but also single words that both stand for concepts already encoded in the recipient language and are infrequent.<sup>12</sup>

## 2.4 Codemixing and government

Pieter Muysken (2000) offers a typological synthesis of *codemixing* (his term for what has heretofore been referred to as codeswitching). Departing from the perspective that the use of two languages in a sentence is constrained by *governing*<sup>13</sup> relationships, he describes three distinct patterns of codemixing: insertions (single other-language words or entire constituents plugged into the corresponding grammatical slot in a recipient language sentence), alternations (the changing of one language and grammar to that of another language and grammar) and congruent lexicalization (stretches of discourse where the lexicon and grammatical frame come from two languages). Muysken maintains that codemixing and lexical borrowing instantiate different linguistic processes, distinguished by the level of language processing at which each occurs. He considers codemixing to be primarily a creative process that happens above the word-level. It involves the syntagmatic relationships between words (2000: 71). *Borrowing* par excellence occurs below the word-level via the combination of lexical formatives from two languages (1995: 190; 2000: 71-72).<sup>14</sup> Although lexical borrowing is generally not the focus of Muysken's analyses, in his earlier work (e.g. Di Sciullo, Muysken & Singh 1986), he and his co-authors

---

12 Myers-Scotton (1997: 204) picks three times as the cut-off point. If a word appears three or more times, it is a borrowing (core or cultural); if less than three, it is a codeswitch (unless it stands for a concept not in the recipient language, then it is a cultural borrowing).

13 The notion of *government* is used in Generative models of syntax (see, e.g. Haegeman 1994).

14 Nonetheless, Muysken (1995: 189-190) acknowledges that there are instances of mixed language below the word level that are more akin to codemixing in that they are creative and that there are cases of mixed language above the word level that are more akin to borrowing in that they are reproductive.

mention that borrowing could be distinguished from codemixing not only on the basis of the length of the other-language constituent (i.e. one-word versus multiple-word strings), but also on the basis of phonological nativization, speaker intuitions about an item's status, native word replacement (1986: 2) and frequency (1986: 15).

## 2.5 Transfer and transversion

In his comparative work on immigrant languages in Australia, Michael Clyne introduces a different terminology. A *transfer* (the product of the process of *transference*) is a “form, feature or construction (...) taken over by the speaker from another language, whatever the motives” (2003: 76). More particularly, a *lexical transfer* is the taking over of words from one language by speakers of another language. Although Clyne does not provide a precise operationalization of the concept, his examples show that lexical transfers may be single words or longer strings. They may be words regularly employed to refer to a particular concept, but they may also have a more ephemeral status.<sup>15</sup> Clyne defines *transference* in relation to *transversion*. Again, although he does not offer an explicit operationalization, he defines *transversion* as the “crossing over from one language to another” (Clyne 2003: 80). His examples (e.g. see page 75 of his 2003 work) suggest that a transversion minimally includes an entire clause of donor language material. This mirrors the distinction between codeswitching and lexical borrowing arrived at in this dissertation (Chapter 3).

---

15 Clyne's lexical transfers in part correspond to what are elsewhere called nonce borrowings, EL islands, intrasentential codeswitches and insertions, that is, constituent-sized other-language strings.

## **2.6 Criteria that have been used for classifying lexical borrowing and codeswitching**

In these frameworks, the distinction between lexical borrowing and codeswitching has been operationalized using several criteria, including string length, morphophonological adaptation and frequency. In the present study, these criteria posed problems. First, the justification for using them is founded on a characterization of lexical borrowings that equates them to recipient language lexical material. Doing this, in theory, limits the data to items that have already attained the status of established loanword. This, in turn, makes results uninformative about the intermediate stages of lexical borrowing in an on-going situation of language contact such as the one that occupies the present study. But these criteria also presented other difficulties, which will be discussed next.

### **2.6.1 One-word donor language string**

Typically, lexical borrowing refers to donor language material that is no more than a single word. A rationale for restricting lexical borrowings to one-word strings is most clearly articulated by Muysken, who considers the selection of single words to be mainly a reproductive behavior and the use of longer strings to be mainly a creative process (1995: 189-190).<sup>16</sup> However, the use of English by Spanish speakers in New York makes clear that even longer strings of English can exhibit a reproductive character. Consider examples (4)-(7), which contain multiple-word English constructions.

---

<sup>16</sup> In the works cited in the present study, Muysken does not quite spell out the logic. Reading into his statements, I believe his characterization of the difference between borrowing and codeswitching to be the conclusion of a syllogistic argument that might go something like this: Lexical borrowings can become established loanwords. Using established loanwords is just like using native recipient language vocabulary. Vocabulary selection is reproductive language behavior. Thus, the difference between lexical borrowing and codeswitching must be one of reproductive versus creative language use.

- (4) Lo que pasa es que nosotros estábamos en el *World Trade Center* el día que pasó todo, verdad... 325E  
 ‘What happens is that we were in the *World Trade Center* the day that everything happened, right...’
- (5) Si tiene una clase a las doce, le toca que leva.. levantarse a las ocho, salirse aquí de las nueve, coger el tren, coger el tren a la Cuarenta y dos, el (tren) Siete, entonces coger el (tren) Uno a la *South Ferry* esa... 311C  
 ‘If one has a class at twelve (o’clock), he has to get.. get up at eight, leave here at nine, get the train, get the train at Forty-second, the Seven (train), then get the One (train) to the *South Ferry*...’
- (6) Entonces fueron y me mandaron para.. para *Saint Patrick’s Military Academy*. 329D  
 ‘So they went and they sent me to.. to *Saint Patrick’s Military Academy*.’
- (7) Pero llevé una clase de bilingüismo, se llama *Bilingualism in Literature* y entonces conocí, eso tocaba con lingüística... 350M  
 ‘But I took a bilingualism class, it’s called *Bilingualism in Literature* and so I knew, that had to do with linguistics...’

Multiple-word strings like *World Trade Center*, *Saint Patrick’s Military Academy* and *Bilingualism in Literature* appear to display a degree of creative language processing in that they have constituent-internal English grammatical structure. Yet, they also seem to be of a highly formulaic nature. *World Trade Center* occurs five times in the corpus, used by at least three different speakers. *South Ferry*, *Saint Patrick’s* and *Bilingualism in Literature*, despite being used only once each, have at one time undoubtedly figured prominently in the daily routine of the speakers. Furthermore, each of these proper noun phrases appears to result from the insertion of donor language lexical material into a Spanish syntactic frame. This makes them qualitatively more similar to the use of *average* in (1) and *meetings* in (2), than to *she would buy them* in (3). Similar observations can be made for multiple-word strings like *liquor store* (8), *self-made woman* (9), *welfare check* (10) and *index cards* (11).

- (8) Pero buscarse las formas y que uno tiene que buscar para un *liquor store* no es todo más fácil... 329D  
 ‘But looking for the forms and that one has to look for a *liquor store* is not very easy...’
- (9) Ella ha avanzado. Es una *self-made woman* y... 324E  
 ‘She has advanced. She’s a *self-made woman* and...’
- (10) Es la meta de ella, la meta de una chica de dieciséis años, llegar a los dieciocho para recibir su propio *welfare check*... 370M  
 ‘That’s her goal, the goal of a sixteen-year-old girl, to get to eighteen in order to receive her own *welfare check*...’
- (11) Yo me acuerdo que la tabla periódica me las [sic] aprendí toda por medio de *index cards*... 300E  
 ‘I remember that I learned the entire periodic table by using *index cards*...’

*Liquor store* is used five separate times by informant 329D. *Self-made woman* is a formulaic expression of English (*Collins Online Collocation Dictionary*). *Welfare check*, as such, was used just once in the OZC, but *welfare* occurs ten other times from seven other informants. Finally, *index cards* is a phrase that clearly played a significant role in informant 300E’s educational years. Further problematizing the requirement that lexical borrowings be no more than a single word is the high rate of occurrence of multiple-word discourse markers both in the corpus as a whole and by individuals, which evidences their reproductive character.

- (12) Había un muchacho con el que yo estaba saliendo y un día voy yo a.. *I don’t know*.. era bien físico... 434P  
 ‘There was a boy who I was going out with and one day I go to.. *I don’t know*.. he was very physical...’
- (13) Solamente lo que quería era hacer su par de pesos y su plato de comida y *that’s it*, no pensaba en seguir progresando... 198P  
 ‘The only thing she wanted was to earn her money and her plate of food and *that’s it*, she wasn’t thinking about continuing to progress...’

- (14) Es que depende de uno mismo querer sobresalir pues *I mean* lo que hice en mi país pues.. trabajé y todo... 321E  
 ‘It’s that it depends on oneself to want to succeed well *I mean* what I did in my country was well.. I worked and everything...
- (15) La mamá me dice que yo soy el varón, *you know* y soy esto y lo o- .. yo soy el hombre de la casa... 311C  
 ‘The mom told me that I was the man, *you know* and I’m this and the oth-.. I am the man of the man of the house...’

As discourse markers, *I don’t know* occurs nine times in OZC, *that’s it* 16 times, *I mean* 34 times and *you know* 363 times. (There are 4507 total borrowings in the corpus.) Despite the implicit presence of English grammatical structure, multiple-word discourse markers provide the clearest example of the breakdown of the single-word criterion. They demonstrate that the generalization that multiple-word strings of donor language tend to be creative cannot be maintained, at least as regards the use of English by the Spanish speakers of the OZC. In the corpus overall, multiple-word English phrases of the sort illustrated above are instantiated 1,889 times, occurring almost as frequently as single-word English-origin strings (n=2,607).<sup>17</sup> Excluding multiple-word English strings from a study of lexical borrowing would miss a significant aspect of the process by which English-origin material potentially comes to be part of a New York Spanish lexicon.

---

17 The total number of borrowings in the corpus is 4507 (Chapter 6, Table 6.1). The number of single-word borrowings and multiple-word borrowings provided in this discussion do not add up to 4507, but to 4496 (a difference of 11). The reason for the discrepancy is that there were 11 borrowings where surrounding speech was not able to be transcribed. So, although the English-origin material was included as a borrowing, it was not certain whether the string was composed of a single word or of more than one word. For example, informant 005U says *Yo trabajaba para xxx company en Tampa*. ‘I was working for xxx *company* in Tampa’. *Company* is included as a borrowing in the database, but it is not certain if it is a single-word borrowing or a multiple-word borrowing.

### 2.6.2 Corpus frequency

In Myers-Scotton's (1997) work, the use of corpus frequency for identifying lexical borrowings rests on the idea that what is at stake in differentiating between borrowings and codeswitches is the difference between foreign-origin material that has become socially and lexically integrated into the recipient language and material that has not. Besides the arbitrariness of setting any particular number as the frequency cut-off (Myers-Scotton 1997: 204), the criterion of frequency presents logical and practical difficulties. First, while the characterization of codeswitching as the use of two or more functionally competing language systems virtually guarantees that a codeswitch would not be produced identically by two different speakers, the fact that a string has not been so-produced by two different speakers in a corpus cannot serve as positive proof that the string is a codeswitch (and thus, **not** a lexical borrowing). Myers-Scotton states: "It is not that a B [borrowed] form *must* recur to be a B form; it is that a CS [codeswitched] form *must not* recur in order to be a CS form" (1997: 204, original emphasis). My interpretation of this statement is that when a foreign-origin string **does** recur in a corpus, it is fairly certain that the string is not a codeswitch (making it classifiable as either a lexical borrowing or established loanword), but when a foreign-origin string **does not** recur, a pronouncement either way is uncertain. Second, although lexical borrowings<sup>18</sup> are, for Myers-Scotton, predictable bits of speech with an elevated probability of being uttered, it is the case, as Poplack and Sankoff (1984) noted, that

---

18 Given this characterization (i.e. that they have elevated chance of being uttered), I myself would not call the data to which Myers-Scotton refers *lexical borrowings*. I believe she has in mind lexical material that is socially and linguistically integrated into a recipient language. These data would, in Poplack's terms, be called *established loanwords*.

... to measure frequency (...) it is necessary to know not only the number of times an item occurred but also all the times it did not occur when it might have (...) this would involve identifying all semantic equivalents of every borrowed word (...) a formidable undertaking... (107)

And as in Poplack and Sankoff (1984), doing so is beyond the scope of the present work. Finally, excluding foreign-origin strings that have only been uttered once fails to allow that a norm for lexical borrowing might in fact be the use of nonce expressions. In fact, of 828 single-word borrowing types in the OZC, 50 percent (n=421) are only ever uttered one time.<sup>19</sup> Furthermore, results from Chapter 6 indicate that nonce borrowing is an integral part of how New York Spanish speakers in New York borrow.

### **2.6.3 Recipient language phonology**

The use of phonology to identify lexical borrowings in natural speech data is problematic for several reasons. First, doing so rests on the assumption that the distinction between lexical borrowing and codeswitching is a difference in the use of one grammar (i.e. lexical borrowing/loanwords) or two (i.e. codeswitching) or that lexical borrowings can, from the outset, be equated with recipient language vocabulary. But, as mentioned in section 1 of this chapter, where the appearance of foreign-origin strings represents the utilization of already-incorporated recipient language material, the data in principle cannot alone be informative about an on-going process of acquiring new lexical material, since by definition it is already a part of the recipient language.

A further difficulty in utilizing phonological adaptation is that it does not take heed of evidence that points to the highly variable linguistic realization of lexical borrowings between

---

<sup>19</sup> As a proportion of tokens uttered, the 421 single-word English-origin nonce items represent 16 percent of all single-word English borrowings uttered (n=2607).

the first instance of use and their eventual incorporation. Poplack et al. (1988) found that a single foreign-origin word used by fewer than five speakers had an equal tendency to be realized with either French or English phonology. In other words, although full phonological adaptation to the recipient language can be expected for foreign-origin words that have acquired social tender (operationalized as being used by ten or more speakers in a corpus), between the time a foreign-origin word is first borrowed and the time it attains a certain degree of social currency, its phonological adaptation to the recipient language is not yet assured.<sup>20</sup> The difficulty in using phonological adaptation lies in the asymmetrical implications relating linguistic processes to social processes.<sup>21</sup> Once a foreign-origin word becomes fully socially integrated into the recipient language, that is, once it becomes an established loanword, it will have been, by definition, morphophonologically adapted to the recipient language, and will likely be rendered so by any member of the community. But lack of phonological adaptation cannot disqualify a foreign-origin word from classification as a borrowing.<sup>22</sup> The word may have been recently introduced to the community or be rendered so for discourse or stylistic purposes.

Furthermore, using a phonological criterion to identify borrowings is less than ideal with respect to the OZC and to Spanish in New York. Anecdotal observations of Spanish speakers in New York indicate that second generation bilinguals typically have the ability to render both English and Spanish according to the phonological norms of either language in response to

---

20 Also, see Holden (1976), who shows for lexical borrowings in Russian that recipient language assimilation rates even vary for different segments within the same word.

21 Note that generalizing across speakers, Poplack et al. (1988) convincingly demonstrate that linguistic processes proceed hand-in-hand with social ones.

22 Poplack et al. (1988: 95) also recognize that the tendency for borrowings to be phonologically adapted as their social currency increases is true “only in a statistical sense”. In my opinion, this statement intimates that phonological adaptation may not be helpful for determining whether particular tokens are one-word codeswitches or borrowings. This is a position she explicitly espouses in Poplack and Dion (2012: 284).

situational conditions and discourse objectives.<sup>23</sup> Likewise, many first generation Spanish speakers, although tending to render English with Spanish phonology, are able to do so as well. To have made phonology a criterion for differentiating between lexical borrowing and codeswitching in this study would have required establishing a different set of conditions on phonological realization of English-origin strings depending on an informant's immigrant generation, age cohort, language history and the interviewer-interviewee relationship, among other things. Even were this feasible, evidence from the OZC shows that it would certainly have yielded mixed results. Different tokens of the “same” lexical borrowing are variably rendered by one and the same speaker. Consider examples (16) and (17), where informant 303P is talks about cable TV. A broad phonetic transcription follows the English-origin phrase.

- (16) Los medios de comunicación no eran tan eficientes, como el internet y *cable TV* ['kei bəl ti 'vi], y todas esas cosas... 303P  
'The means of communication were not so efficient, like internet and *cable TV*, and all those things...'
- (17) En aquel tiempo, no, porque no había cable tevé ['ka βle te 'βe]<sup>24</sup> ... 303P  
'At that time, no, because there was no cable TV...'

In example (16), 303P realizes *cable TV* with English phonology. In (17), she uses Spanish. In other words, this 36-year-old adult migrant (arrival age 24) ably renders the string using either

---

23 Note that Poplack and Sankoff (1984: 125-126) found evidence to the contrary: their Puerto Rican child Spanish-English bilinguals, who I take also to be ostensibly the second generation, tended to render English borrowings according to Spanish language phonology, like their parents. However, in their study, lexical borrowings were elicited, not from natural speech data. Thus, discourse and stylistic concerns that give rise to variation in phonological realization of donor language material would presumably not have been applicable to the task presented to their participants.

24 In fact, the token *cable tevé* from example (17) is **not** in the lexical borrowing database. This is because it was identified as a synonymous bilingual homograph, and the means operationalized in this study for determining whether synonymous bilingual homographs constituted a reproduction of donor language material in a recipient language discourse (i.e. the definition of lexical borrowing in this study) was phonological realization according to donor language patterns (see section 4.3.2 in Chapter 3).

Spanish or English phonology. The criterion of phonological adaptation would have excluded tokens produced using English phonology in response to discourse or pragmatic factors.

Finally, a phonological adaptation criterion would have created a significant drawback. It would have made a comparison of speakers' lexical borrowing inventories unreliable, since tokens of the same word, for instance *high school*, would be included in the corpus on one informant's pronunciation as [xaj e 'kul] but excluded on another's as ['hajs kul]. Thus, although several individuals may have the same English-origin word *high school* in their lexical repertoire, using phonological adaptation to exclude tokens would have obscured this fact. These considerations led to the discarding of phonological adaptation as a criterion for the identification of lexical borrowings (but see Chapter 3, where phonological realization was used to determine, for ambiguous strings, if an informant was reproducing an English-origin pattern).

#### **2.6.4 Recipient language morphology**

Poplack et al. (1988) suggest that morphological adaptation to the recipient language may also serve to distinguish lexical borrowings from one-word codeswitches. They found that plural English-origin words in French, regardless of their status as established loanwords or nonce borrowings, were likely to receive French affixation. They further found that, as the number of speakers who used a plural borrowing increased, so did the likelihood that it would be morphologically adapted to French. Eighty-five percent of plural nonce borrowings (plural English-origin nouns that appeared once in their corpus) received French affixation, while 98 percent of plural tokens that were used by 10 or more speakers received French affixation. Additionally, 100 percent of verbs received French verbal morphology. They interpret their results as suggesting that recently borrowed words (i.e. nonce borrowings) are more similar to established loanwords than to one-word codeswitches.

A morphological adaptation criterion, similar to other criteria thus far discussed, would have presented difficulties with respect to the OZC. First, contrary to Poplack et al.'s finding that over 85 percent of plural nouns have been adapted to the recipient language, morphological adaptation of English words to Spanish in our corpus appears to be very infrequent. Of the 276 plural single-word common nouns in this study,<sup>25</sup> 235 had overt plural lexical morphology (i.e. *-s*, *-es* or English ablaut as in *freshmen*). Of those, only 34 (14 percent) had unambiguous Spanish plural morphology. The other 200 (86 percent) had either English plural morphology or plural morphology that, depending on pronunciation of the root word, is ambiguous between Spanish and English morphological patterns. Furthermore, of 1,218 singular single-word common nouns, only 65 were realized with overt Spanish morphology, as in *babyito* or *fondución* 'foundation': a rate of only about 5 percent. The rate of morphological adaptation of single-word verbs (N=80) was higher: 82 percent. But taking these three categories of single-word strings together, the rate of morphological assimilation to Spanish was only 10 percent (165/1574). In other words, if morphological adaptation to Spanish were used to identify lexical borrowings in this study, 90 percent of single-word common nouns and verbs in the OZC would have been excluded.

I propose that the logic Myers-Scotton (1997: 204) articulated with respect to the frequency criterion also applies to morphological adaptation. It is not so much that borrowings must be morphologically adapted as it is that codeswitches are never fully adapted to the morphology and phonology of the recipient language. The implication, then, is that while morphologically adapted words may be classified as lexical borrowings or established loanwords, a lack of morphological adaptation gives no indication of whether a word is better classified as a borrowing or a codeswitch.

---

25 This does not include nine inherently plural English borrowings like *sports* and *blues* (musical genre).

### 2.6.5 Summary

This section described how the distinction between lexical borrowing and codeswitching is operationalized in the literature, variously using length of string, corpus frequency, morphological adaptation to the recipient language and phonological adaptation to the recipient language. For reasons both theoretical and practical, none of these criteria are suitable for the present study. Nonetheless, maintaining a distinction between borrowing and codeswitching is desirable, even if difficult to make in practice. The strategy used in this study to make the distinction is detailed in the following chapter, Chapter 3 (section 4.1). Before Chapter 3, however, a synopsis of the ways that lexical borrowing frequency has been interpreted in the literature is provided next.

### 3. Lexical borrowing in language contact and bilingual speech studies

This section presents several ways that lexical borrowing has been interpreted in quantitative investigations of lexical borrowing,<sup>26</sup> including as an indication of a speaker's language processing, as a marker of social group affiliations and as the reflection of a community's history of language and cultural contact, among other things.

Ngom (2006) and Matus-Mendoza (2002), for instance, posit that lexical borrowing frequency can be a *sociolinguistic index*. That is, more or less borrowing may “point(er) to speakers' social identities” (Gal & Irvine 1995: 973). For instance, Ngom explored the frequency of single-word lexical borrowings from English, French, Arabic, Pulaar and Spanish in Wolof among 200 individuals in Saint-Louise, Senegal. He found that younger individuals used more borrowings from English (than from Arabic, for example) and argued that English had become a

---

26 Note that this discussion cites several studies in which the data have been classified as intrasentential codeswitching. These findings of these studies are nonetheless relevant since their data is similar to what is called lexical borrowing in the present investigation.

symbol of urban youth and modern culture.<sup>27</sup> Matus-Mendoza (2002) found that migrant workers from Moraleón, Mexico used more English lexical items than the upper class (who had high levels of education and professional degrees). She proposed that the use of English lexical borrowings by Moraleonians is a class marker. Specifically, she hypothesized that migrant workers compensated for their lack of educational or professional credentials by using English to signal their increased buying power and, thus, elevate their status.<sup>28</sup>

Frequency has also been linked to the dissemination of lexical borrowings. For instance, in a study of borrowing by Punjabi speakers in Pakistan, Sullivan (2008) found that the least educated borrowed most. She hypothesized that this was because the least educated individuals had less access to the community prestige languages, Urdu and English. Lack of access, she reasoned, resulted in one's inability to filter out borrowings from foreign sources. This, in turn, she says, indicated that the less educated were responsible for the spread of borrowings. In an investigation of English and Hebrew borrowings in Arabic, Amara and Spolsky (1986: 51) found that those who worked outside their village, as professionals or construction workers, supplied a greater number of borrowings than students or the unemployed. They concluded that these two groups were the "conduits for the diffusion of the foreign words" (1986: 52).

Borrowing frequency has been interpreted as an indication of individuals' network affiliations. Poplack et al. (1988: 76) found that men residing in one neighborhood borrowed more frequently than women. This was due to the fact that the men habitually worked in a

---

27 Further, Ngom (2006: 45-46) indicates that the use of English borrowings has undergone a process of *recursivity* (Gal & Irvine 1995: 974). That is, the association of English with modernity has been projected inwardly to create distinctions between different groups of youth: those that are modern, on the one hand, and those that are conservative, on the other. Its association with modernity has also been projected outward, to distinguish between the old and the young.

28 In their study of Canadian Francophones, Poplack et al. (1988: 76) found that social class membership best predicted borrowing frequency: working class individuals borrowed more than the middle class. Although they refrained from explicitly interpreting borrowing frequency as marker of social class membership, it is tempting to see their results as additional support for the idea that borrowing frequency can be a sociolinguistic index.

neighboring community, where borrowing was more frequent. Ngom (2006: 103) found that the lack of a statistically significant difference between men and women's use of French and Arabic borrowings probably reflected the fact that, in Saint Louis, women are highly involved in French and Arabic educational institutions, contexts that are typically dominated by men in the rest of Senegal. Mendieta's (1999) study of lexical borrowing in Spanish in the U.S. may provide additional support for the importance of one's community. She found that Puerto Ricans used more borrowings than Cubans or Mexicans (1999: 91-94). But Mendieta's ethnolinguistic groups each lived in a different U.S. city: Her Puerto Ricans lived in Perth Amboy, New Jersey, her Cubans in Miami, Florida, and her Mexicans in San Antonio, Texas. In effect, her findings suggest that different areas of the country establish independent local norms for borrowing.

Borrowing frequency may also vary with community exposure to the donor language. Thomason and Kaufman note that "greater intensity of contact in general means more borrowing (...) The factors that promote greater intensity of contact (...) are: length of time—enough time for bilingualism to develop..." (1988: 72). In other words, communities with a long history of contact with a donor language culture or with a sufficient number of bilinguals will evidence a greater number of contact features (lexical, morphological and syntactic). Thomason and Kaufman are, of course, referring to linguistic features that have become entrenched in a language. But it may be reasonable to suppose that before such time, community language speakers may demonstrate greater or lesser quantities of contact features in their speech in accord with each subgroup's history of contact with the donor language culture or in accord with the number of bilinguals among the community language group. It should be born in mind, however, that bilingualism may not need to be extensive within a community for lexical borrowing to

occur. As Thomason and Kaufman state: “lexical borrowing frequently takes place without widespread bilingualism” (1988: 37).<sup>29</sup>

Borrowing frequency has also been connected to an individual’s *exposure* to the donor language. Eslami Rasekh et al. (2008) found that, among Turkman speakers in Iran, borrowing from Persian was more frequent among the more educated and city-dwellers. This was, they concluded, a function of the fact that highly educated and city-dwelling Turkman speakers had more contact with Persian, the socioeconomically dominant language.

Intuitively, borrowing frequency should also reflect differences in speakers’ *proficiency* in a donor language. Indeed, it is taken as axiomatic in studies of language contact that bilinguals are “the locus of language contact” (Weinreich 1953/1966: 1), the channel through which foreign-origin features are introduced into a recipient language. Yet, studies have revealed that borrowing frequency may be related to donor language proficiency only marginally (e.g. see Poplack et al. 1988: 76) or not at all (Vedder, Kook & Muysken 1996, cited in Muysken, Kook & Vedder 1996: 486).<sup>30</sup>

Borrowing frequency may be associated with recipient language proficiency. Vedder et al. (1996, cited in Muysken et al. 1996: 486) found that the use of Dutch (the donor language) in Papiamentu speech was greater among children because of their limited proficiency in Papiamentu. Additionally, the literature everywhere reminds that a primary motivation for

---

29 Similarly, Weinreich notes: “lexical borrowing is less restricted to the bilingual portion of a language community than phonic or grammatical interference” (1966: 56).

30 More often, it is how speakers combine donor and recipient resources in discourse that is related to donor language proficiency. For instance, Poplack (1988: 606-607) found that Spanish speakers who had equally high proficiency in English tended to use English within a clause. Those who were less proficient in the donor language (English) than in the recipient (Spanish) tended to do so at clause peripheries. Note that Poplack interprets her results not as a function of donor language proficiency, but as a function of the relative strength of speakers’ abilities in both languages. The association of (relative) language proficiency with patterns of combining foreign-origin material, however, is still under debate. For instance, Berk-Seligson (1986) did not find any relationship between proficiency and the type of language mixture done by Spanish-Hebrew bilinguals in Jerusalem.

lexical borrowing is *referential need* (e.g. Mendieta 1999: 43; Weinreich 1966: 56, 58). Referential need often refers to the fact that the recipient language lacks a word for something (a.k.a. *semantic / cultural gap*). It may also, however, refer to a situation where individuals lack a word due to forgetting it (*language attrition*) or never having learned it (*incomplete acquisition*).

In the case where community language speakers are bilingual, borrowing frequency may also be an indication of the relative *activation* or *inhibition* of a bilingual's languages (Paradis 1997: 342-343). Grosjean (1997a: 234) found that French-English speakers inserted English borrowings and codeswitches into French discourse more often when both of their languages were activated (in Grosjean's (1997a, 1997b) terminology, they were in *bilingual mode*) than when one language was activated and the other inhibited (i.e. *monolingual mode*). Likewise, Lipski (2005, 2008: 235-239) suggests that the frequent use of *so* and other discourse markers<sup>31</sup> by U.S. Spanish speakers may reflect the fact that their language monitoring mechanism is "circumscribed by a meta-level based on key English discourse delimiters" (2008: 239). This is, he speculates, is probably due to the fact that they "spend more time each day speaking English than Spanish" (2005: 12). In other words, regular activation of the donor language results in a greater tendency for it to "pop through" while speaking Spanish.

Finally, borrowing frequency may be indicative of speaker attitudes toward the donor and recipient languages. Weinreich suggests that borrowing may be less frequent for speakers that have strong feelings of loyalty to the recipient language or a negative perception of the donor language (1966: 65). On the other hand, borrowing may be more frequent among those that associate the donor language with positive prestige (Weinreich 1966: 59-60, 65).

---

31 Note that Lipski (2005: 239) does not classify English discourse markers as borrowings, but leans toward calling them instances of *momentary codeswitching*.

In sum, the foregoing discussion has offered an overview of several ways that lexical borrowing frequency been used and interpreted in the literature. Lexical borrowing has, for example, been used as evidence to support models of language contact and bilingual speech. Its occurrence has also been explained with recourse to concepts from speech accommodation and social identity theory, language ideology and variationist sociolinguistics. Likewise, these frameworks will provide the conceptual basis for interpreting the results of the present investigation (see Chapter 4, section 1). Next, however, the design of the study is presented.

## **Chapter 3: Methodology**

### **1. Outline of this chapter**

This chapter describes the research design of the present study. It begins with how informants for this study were located, as well as the structure and content of the interviews (section 2) from which data were obtained. It is followed by a description of the independent variables, the variables by which informants were primarily classified in order to produce a balanced sample of New York Spanish speakers (section 3). Section 4 describes in detail the procedures implemented in order to select and classify English-origin strings as lexical borrowings. The dependent measures are defined and illustrated in section 5. A description of how analysis was carried out is provided in the final section of this chapter (section 6).

### **2. Informants and interviews**

The English lexical borrowing database (LBD) was created using 140 interviews that comprise the Otheguy-Zentella Corpus of Spoken Spanish in New York, supplemented by six additional interviews that were gathered separately.<sup>1</sup> This corpus, a total of 146 interviews henceforth referred to as the OZC, is a stratified subset of over 300 Spanish language interviews collected between 2000 and 2005 (Otheguy & Zentella 2012: 27). The OZC contains over 130 hours of recorded conversation.

Informants of the OZC were obtained using a snowballing technique (Oppenheim 1992: 43) that began with the friends, families and acquaintances of the research team. All informants spoke Spanish during the interview and identified as belonging to one of the following

---

1 The corpus was developed with funding by National Science Foundation grant number 0004133.

ethnonational groups: Puerto Rican, Dominican, Mexican, Cuban, Colombian or Ecuadorian. The letters P, D, M, U, C and E following an informant identification number indicate, respectively, the ethnonational group with which the informant identified (e.g. 005U is a participant identifying as Cuban).

The OZC was compiled using Labovian sociolinguistic interviewing techniques (Labov 1963, 1966). 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 (Otheguy & Zentella 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 origin as the informant. So Dominican informants spoke with Dominican interviewers, Mexican informants with Mexican interviewers, and so on.<sup>2</sup> Occasionally, two or three informants participated in a single interview. Finally, interviews were loosely structured. To ensure comparability between interviews, all informants were prompted to talk on similar topics including personal stories and opinions related but not limited to: 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. After each interview, a two-part oral questionnaire was administered. 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 Columbia for ‘birthday cake’). The second part gathered demographic information

---

2 In terms of lexical selection, Zentella (1990) suggests that lexical borrowing might be promoted when an individual’s Spanish word to name a concept has restricted regional currency and using it marks him as a speaker of a stigmatized variety of Spanish. Having an interviewer that speaks the same variety of Spanish as the informant reduces the likelihood of this occurring.

on each informant. Only the first, conversational, part of the interview was used to obtain lexical borrowing data.

Although informants were aware that they were being interviewed because they spoke Spanish, no explicit attempt was made to force or suggest that Spanish be the only language admissible in the conversation. If informants asked about using another language (i.e. English), they were told that it was okay. The interviews were tape-recorded and later transcribed by the interviewer or a fellow interview-conducting project member. 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. The number of words uttered by informants during the conversational part of the interview ranged from 1,150 to 13,500.<sup>3</sup>

### **3. Independent variables and stratification of the Spanish in New York corpus**

The 146 informants of the OZC are stratified<sup>4</sup> according to five principal socio-demographic traits: (1) ethnonational affiliation, (2) Latin American region of origin, (3) sex, (4) age of arrival to the U.S. and (4) occupational class. An attempt was made to ensure sufficient numbers of informants across factor groups for the following secondary variables: (6) age, (7) years in the U.S., (8) level of education, (9) English skills, (10) Spanish skills and (11) Spanish daily use. These are considered secondary variables both with respect to their expected importance to linguistic variation in Spanish in New York City, as well as with respect to the sometimes uneven distribution of informants across the factor groups of each variable. In all, these eleven

---

3 Note that there is no significant correlation between borrowing rates (see section 5.1 for how this is calculated) and the wordiness of an informant ( $r=0.35$ ,  $p<.683$ ). To be expected, however, longer interviews produced a greater number of borrowings per interview ( $r=0.42$ ,  $p<.001$ ).

4 For more on the selection of interviews and stratification of the OZC, see Otheguy and Zentella (2012: 22-23, 25-27).

traits are the independent variables of the study. A description of the variables and the distribution of informants within factor groups is provided in Appendix A.

#### **4. Data selection criteria and the formation of the English lexical borrowing database**

In this study, lexical borrowing has been defined as *the reproduction in one language of lexical material (minimally one word-form<sup>5</sup> plus meaning) from another language* (Chapter 1, section 3; Haugen 1969: 363). In other words, what is required for the identification of a lexical borrowing is a demonstration (1) that a speaker is speaking Spanish, (2) that lexical material used while speaking Spanish has origins in the donor language (English) or the donor community's use of it, (3) that the lexical material was not used in the home country variety of Spanish spoken by the informant<sup>6</sup> and (4) that informants could not have made up the lexical material independently. In what follows, the methods to ensure that each of these conditions has been met are described. Section 4.1 addresses condition (1), section 4.2 addresses condition (3), sections 4.3 and 4.4 address condition (2). And as concerns condition (4), it will be assumed, following Haugen (1969), that lexical material that meets conditions (2) and (3) also meets condition (4).<sup>7</sup>

##### **4.1 Distinguishing between lexical borrowing and codeswitching**

Chapter 2 discussed why several means often used to distinguish between borrowing and codeswitching were inadequate, theoretically unsound or impractical. Despite the difficulty of making the distinction in practice, I maintain that lexical borrowing and codeswitching in fact

---

5 A *word-form* is a phonetic string that realizes at least one unbounded morpheme. The expression *word-form* in the current work is intended to be synonymous with Lyons' (1977: 18) use of the same term, except that he explicitly defines it only with respect to written word-forms.

6 For informants born in the U.S., condition (3) would be that the lexical material was not used in the variety of Spanish spoken by his parents.

7 Haugen writes that "under strong cultural pressure from English", informants are "not likely to have made up new linguistic items which parallel exactly the English forms except as an imitation of those English forms" (1969: 384).

represent two distinct types of contact behavior. Furthermore, I believe that the difference consists mainly in determining when someone is speaking Spanish and when someone is speaking English (condition 1 of the identification of lexical borrowings). If this is true, then, among a sample of diverse informants, such as is used in this study, not all informants could or would engage in both types of behavior. In particular, codeswitching, in instantiating essentially English discourse, should be more limited in its distribution among informants, who were informed that they were being interviewed because they spoke Spanish. Thus, to locate a dividing line between borrowing and codeswitching, a distributional analysis of English strings of varying lengths and syntactic compositions was undertaken. To carry out this analysis, absolutely **all** English-origin words, constituents, phrases, clauses and sentences were extracted from informant interviews. These strings can be characterized as follows:

- (i) **Established loanwords:** etymologically English words that are part of the lexicon of the recipient language (confirmed by their presence in general or national dictionaries of Spanish). By definition, they are fully morphophonologically adapted to the recipient language. E.g. *lobi* ‘lobby’ in Puerto Rican Spanish, *queique* ‘cake’ in Cuban Spanish.
- (ii) **Single words** that may or may not show morphological or phonological adaptation to Spanish. Their status in the lexicon of the community language was, at the time of data collection, indeterminate. E.g. *janguear* ‘to hang [out]’ and *nursery*.
- (iii) **(Multiple-word) Phrases:** Strings containing two or more words, not exceeding one constituent of the type: noun phrase, adjective phrase or discourse marker. E.g. *los Twin Towers, cap and gown, reading and writing, every word, senior citizen, above average, mentally ill, pretty much, I don’t know, oh my God!* These could also be complex (2-part) verbs, like *break into*.
- (iv) **Clause-like strings:** Strings consisting of a verb plus some other constituent of the verb’s sentence (such as an object NP, subject NP or PP). The string may contain any number of words, as long as it did not comprise a full sentence (subject + predicate). E.g. *this is, she’s a, going back to school, release their earnings, trust you*<sup>8</sup>
- (v) **Prepositional phrases:** English preposition + English noun phrase. E.g. *down the block, on Broadway, on Sundays, for shock value*
- (vi) **Sentences:** Strings consisting minimally of an entire sentence. Often, they were in fact several sentences long, uttered sequentially without little or no intervening Spanish language material. E.g. “Y ahora mismo yo no puedo vivir sin ellos. *Which is ok. That’s not a bad life.* Pero después de dos días...” 434P<sup>9</sup>

All English-origin established loanwords (type i) were excluded on the grounds that, since these data are already part of a Spanish lexicon, they were irrelevant to a study of the social variation that **accompanies** the process of incorporation of foreign-origin material into lexicon (for how established loanwords in Spanish were identified, see sections 4.2.1 and 4.2.2). Single words (type ii) were considered the quintessential example of lexical borrowing. And sentences

---

8 In traditional grammar, the last three examples (*going back to school, release their earnings, trust you*) would not be considered clauses, since clauses are defined as minimally containing a verb predicate and an overt or implied subject (Haegeman & Guéron 1999: 22).

9 ‘And right now I can’t live without them. *Which is okay. That’s not a bad life.* But after two days...’

(type vi), which are essentially changes in the language of discourse, were considered the quintessential example of codeswitching. Both of these latter types, single words and sentences, served as points of comparison for how other English words and phrases patterned among informants.

Distributional analysis revealed that of 146 informants, 139 (95 percent) use English in their interviews (i.e. at least one instance of types ii – vi). All 139 informants use single words of English (type ii) at least one time (but usually more than once). On the other hand, only 46 informants (30 percent) use full sentences of English (type vi). In other words, using single words of English, i.e. borrowing, is very common in the OZC. Codeswitching to English is not. It now remains to see whether data types (iii–v) were speech behaviors that co-occurred only among those that switched (suggesting they are more similar to switching) or whether they also co-occurred among those that borrowed.

Of the 139 informants that use single words, 126 (90 percent) also use phrases (type iii). Said differently, the use of single words implies to a great extent the use of phrases. This suggests that the use of English phrases instantiates a behavior similar to using single words. Turning now to clauses (type iv), it turns out that only 30 informants (22 percent of those that used some English) use this type of string. That is, if an informant uses single words, there is only a 22 percent chance that he also uses clauses. Further, all but four of the 30 individuals (i.e. 86 percent) that use clauses also use full sentences. In other words, the set of clause users overlaps to a greater extent with codeswitchers than it does with borrowers. This supports the conclusion that the use of English clauses instantiates a behavior indeed very similar, and in terms of language processing perhaps identical to, speaking English. Finally we turn to prepositional phrases (type v). PPs were used by the fewest number of speakers (n=16), all but three of whom

also use *both* clauses (type iv) as well as changed discourse language (type vi). This suggests that prepositional phrases are also best classified as codeswitches for informants of the OZC.

If corpus distribution is considered insufficient as means of classifying data and, in particular, if it is considered a weak justification for classifying multiple-word strings as lexical borrowings, it should be noted that the general outline of significant results in Chapters 4 and 5 still obtains when statistical analysis is performed with respect to only single-word English-origin strings (i.e. data of type ii) in the OZC.<sup>10</sup>

In sum, the corpus-wide distribution of English strings of varying lengths and syntactic composition led to identifying lexical borrowings *as the reproduction in Spanish of single-words and (non-clausal, non-prepositional) phrases from English*, i.e. data of types (i)-(iii). Codeswitches, not in this study, designate English lexical material of types (iv)-(vi).<sup>11</sup> As a result, lexical borrowings in this study include data that have elsewhere been classified as intrasentential codeswitching (Poplack 1980), EL islands and core borrowings (Myers-Scotton 1997), insertional codeswitching (Muysken 2000) and lexical transfer (Clyne 2003).

The remainder of section 3 describes the strategies employed to satisfy the conditions (2) and (3) for identifying English lexical borrowings. These strategies directly address difficulties for identifying lexical borrowings that were enumerated by Haugen (1969: Chapter 15, 1972: Chapter 4).

---

10 Prepositional phrases of English were **included** in the database only if they were part of proper noun phrases (e.g. *Fashion Institute of Technology*, collocations (e.g. *coast to coast*) or discourse markers (e.g. *of course*). This is because such phrases are believed to represent lexemes (i.e. vocabulary units). See section 4.1.2 for more on lexemes and how a phrase was determined to be a collocation. See Appendix D (section 5) for examples of single prepositions and analytic locutions that were included in the database as lexical borrowings.

11 Note that results in Chapters 6 and 7 have not been checked to see if they still obtain in a design where only single-word borrowings are taken as data.

## 4.2 Result of contact in the U.S.

Lexical borrowings, as defined in this study, result from exposure to English **in the U.S.** In other words, the English-origin lexical item should be reasonably presumed to be available for the informant's use due the circumstance of the informant's or his family's having come to reside in the U.S.<sup>12</sup> It makes no difference whether a speaker uses an English word because he cannot remember or never learned a Spanish way to say something (as might be the case with someone born in the U.S.) or for some other reason. Said another way, in order to be included in the corpus, English-origin lexical material should not have been part of the lexicon of the Spanish spoken by informants before arrival to the U.S. or in the Spanish spoken by their non-immigrating counterparts. That is, it should not be part of a *precontact* lexicon of Spanish (see condition 3). Because such *precontact loanwords*, like *queique* 'cake' in Cuban Spanish, have not come about due to contact with English **in the U.S.**, they were excluded from the database using the precontact test. This test is described in the following section with respect to two types of precontact items: *preimmigration loanwords* and *international loanwords* (Haugen 1950).

### 4.2.1 Preimmigration established loanwords

A borrowing was determined to be a preimmigration loanword (or, simply, part of preimmigration Spanish) if it was considered by a native speaker of Spanish to be used in the

---

12 It goes without saying that determining which items were U.S.-contact items, that is, which items would **not** have been part of a speaker's linguistic inventory had he not come to live in the U.S. involves best-guessing most of the time. Even with native speakers and dictionaries at one's disposal, the information provided by these sources is necessarily incomplete or biased. It also goes without saying that determining which items are of English origin depends on the comparison of hypothetical, idealized versions of both English and Spanish. Neither circumstance or the methods used to address classificational difficulties undermine the validity of this study, since the assumptions and guesswork involved are both common and required for any linguistic inquiry to take place when detailed accounts of all the language varieties being compared do not exist, which is the case for all but a few language varieties of the world.

country of affiliation.<sup>13</sup> If it was, all uttered tokens of that precontact item were excluded from the database. For example, the standard word in Puerto Rico meaning ‘apartment’ is *apartamento*.<sup>14</sup> Thus, every token of *apartamento* uttered by a Puerto Rican immigrant or the child of a Puerto Rican immigrant was excluded. (An exception to this procedure occurred if a preimmigration item was uttered using English phonology. See “Exclusions” below and section 4.3.2.)

**The precontact test.** To determine whether an (potentially) English-origin word was used in the informant’s country of origin, one or two native Spanish speakers born in Puerto Rico, the Dominican Republic, Cuba, Colombia, Ecuador and Mexico were recruited to report on whether specific English-origin words were used in the informant’s home country. Each enlisted reporter (called a *precontact source*) was presented with from 65-70 words in sentences.<sup>15</sup> The words and their attendant meanings were obtained from the context of their use in OZC interviews. The sentences were attempted to maximize transparency of the meaning of the possible English-origin word, and were checked for clarity by a native speaker of Spanish. Appendix B provides a sample data collection tool.

This means of establishing whether an expression was a preimmigration item, suffers some shortcomings. Most notably, geographical variation in the Spanish vocabulary of the home country complicated the task of identifying preimmigration loanwords, since a single precontact source was unlikely to know with 100 percent certainty whether a given word was used in his or her home country. For instance, in Mexico, it is clearly the case that some English-origin words

---

13 In other words, to be included in the lexical borrowing database (LBD), a borrowing should **not** have been in an immigrant’s lexical repertoire prior to arriving in the U.S. (see Poplack, Zentz & Dion 2012: 204) and should not have been likely to be in the lexical repertoire of a second generation informant had his parents never have migrated.

14 According to the *Diccionario de la lengua española*, the Real Academia Española considers *apartamento* to be etymologically Italian (from *appartamento*) (<http://lema.rae.es/drae/?val=apartamento>).

15 The precontact text for Puerto Rico contained 98 items.

are common in the everyday speech of Spanish-speakers living near the U.S.-Mexico border (e.g. *nice* in: "...muy buena gente.. muy *nice*..." 349M), while the same words may never be heard by those in central Mexico, that is, until they migrate to the U.S.

Likewise, highly stratified social classes in the home country, where stratification was also associated with differential exposure to English, in some cases led to difficulty determining whether a word was a precontact item or not. For instance, a Colombian precontact source informed me that *campus* would be a common English loanword in Colombia, but only among the upper classes (who had greater access to a university education). Colombians from a lower class or who had not attended university would not be likely to know it. An additional factor complicating this task was that the status of an English-origin word in the Spanish of the home country may have changed between the time that informants of the corpus were interviewed (from 2000-2005) and the time that precontact sources were reporting on a lexical item. For instance, a Puerto Rican source reported that *pamper* 'diaper' had not been used in Puerto Rico 11 years prior (when data for this study was being collected), but that it was commonly used in Puerto Rico at the time he was providing feedback on potential precontact words, in 2011. Thus, the subjectivity of making the precontact determination was twofold: the source's limited knowledge of geographical, historical and social variation in the use of words and my own subjective determination of how to utilize the information provided by the source.<sup>16</sup>

---

16 The number of these difficult cases that ended up being included in the database was just a handful. Concretely, I faced the question of whether to exclude the word *pamper* 'diaper' when used by a Puerto Rican informant or the word *campus* when used by a Colombian. I decided to include *pamper* on the grounds that at the time it was uttered by the Puerto Rican it probably was not part of the Spanish lexicon in Puerto Rico. With respect to *campus/campuses/campo/campos* '(university) campus(es)', it was excluded regardless of the class status of the Colombian informant (although *campus/campo* was used by only one Colombian informant, 194C). This was under the hypothesis that a lower class Colombian may have learned the word through contact with upper class Colombians or other Spanish speakers for whom *campus* is part of their homeland Spanish. See section 4.3.1 on how the influence of dialect contact was controlled in the current study.

Nevertheless, short of a full-scale study of established English loanwords in national varieties of these Spanishes, using a native national to help make the determination of whether an English-origin lexical item is a U.S.-contact item or not provides insight not achievable through other methods.<sup>17</sup> For example, if only a standard dictionary of Spanish were used to identify preimmigration items, all of the following words (which were not, for example, in the *Diccionario de la RAE* with the meanings listed) would be included in the database<sup>18</sup>: *actualmente* (discourse marker-like) (**M**, C, D, U) ‘in fact, actually’, *aplicación* (**C**, **P**) ‘solicitation for entry into an organization, *aplicar* (**C**, **E**, **P**, D, M, U) ‘to submit materials as a means to ask for a job’, *elementaria, elemental* (**C**, **P**, E, M) ‘education for ages 5-12’, *envolverse, envuelto* (**P**, M) ‘to become involved / be involved’, *ganga* (**D**, **P**, C, M) ‘street gang’ and *proyecto* (**M**, **P**, C, D) ‘a concentration of living quarters for low-income families’. But all of these words were in fact identified by precontact sources as part of Spanish in the home country. That is, their use in Spanish in New York City is not evidence of language change due to contact in the immigrant context. Using a standard dictionary of Spanish would have incorrectly led to classifying these items as having arisen due to contact with English in New York. Using a precontact source, on the other hand, has allowed these words to be identified as preimmigration loanwords and **excluded** from the study. This procedure has helped to avoid an exaggeration of the extent to which Spanish in New York City is influenced by contact with English.

---

17 For instance, in pilot studies of this investigation, a regional dictionary of each informant’s affiliated national Spanish was used to determine if etymologically English words were part of an immigrant’s preimmigration lexicon or not. This method was functional, but dictionaries tend to filter out perceived foreign influences and also fail to inform about extremely geographically or contextually localized use of English in the home country. For example, *(la) high* ‘high class’ is not in the *Diccionario del Español de Cuba* (2000), but was reported by my Cuban precontact source to be used by himself and other Cubans in Cuba. Thus, *(la) high* ‘high class’ was excluded from the database.

18 Letters following a word indicate the ethnonational affiliation of the informants that uttered it. Letters in boldface indicate that the precontact source reported that it was used in his or her home country.

**Exceptions.** It has been noted that the same donor language lexical item may be borrowed more than once at different times in a recipient language's history (Pyles & Algeo 1993: 297). For instance, the word *chief*, a French-origin, established loanword in English was borrowed during the fourteenth century; English borrowed the naturally transmitted offspring of *chief*'s etymon in French again, as *chef*, in the nineteenth century. Likewise English *gentle* and *jaunty* are two reflexes of the same French word *gentil*, borrowed at different times. In recognition, then, of historical circumstances that give rise to such doublets and given the immediacy and strength of cultural pressure on Spanish-speakers in the U.S., it was desirable to allow for the possibility that lexical items determined to be precontact items of Spanish could be reborrowed from English. Thus, a methodological trapdoor on the condition that lexical material **not** be used in the informant's home country Spanish (condition 3) was put in place. A lexical item which has, via the precontact test, been identified as a preimmigration (or international) item can be rescued from exclusion if it was realized with English phonology (section 4.3.2).

#### **4.2.2 International loanwords**

Some English-origin lexical items are part of an interlingual or international lexicon, such as words like *music* and *museum*, which Haugen (1972: 101) notes were certainly part of the home country Norwegian of his U.S. Norwegian immigrants.<sup>19</sup> Similarly, some English-origin words from this corpus, including *club*, *sandwich*, *drama*, *jazz*, *Internet* and *taxi*, are common to several European and South American languages, including Spanish. English-origin international loanwords were included as lexical borrowings or excluded as precontact items based, first, on

---

<sup>19</sup> Although, Haugen contends that the pronunciation that his Norwegian speakers gave these words almost certainly indicated that they were probably not in common use among his immigrant informants before immigration.

the precontact test (section 4.2.1) and, second, as suggested by Haugen (1972: 101-102), based on their phonetic realization (section 4.3.2).<sup>20</sup>

### 4.3 English-origin

Identifying English-origin words and expressions in Spanish-language discourse in New York is, in the majority of cases, fairly straightforward for those with sufficient knowledge of Spanish and English. However, certain (potentially) English-origin words in the transcripts posed a challenge, among them were bilingual homographs (e.g. the grapheme sequence *d-o-c-t-o-r* used in Spanish and English to spell the word for ‘doctor’) and words that were determined to be preimmigration loanwords in one variety of Spanish, but not in another. (These latter items, when used by informants for whom they were not determined to be precontact items, may be evidence of Spanish dialect contact, rather than English contact.) The methods and rationale used to classify words of both types are discussed next.

#### 4.3.1 Dialect contact loanwords

Since this study is concerned with the influence of English on Spanish, it was desirable to limit borrowings to English-origin strings resulting from contact with English and, therefore, exclude items that may have come about through speakers’ contact with other varieties of Spanish. In other words, an English-origin lexical item should be available to an informant either because of personal exposure to or because of the exposure of his speech community to English or English speakers, and **not** because of exposure to (speakers of) other varieties of Spanish. Including only

---

20 Besides common nouns like *museum* and *jazz*, another notable type of international loanword was single-word proper place names, such as of continents (*Africa*), states/territories (*Florida*), boroughs (*Queens*) and neighborhoods (*Bedford*). Consistency in methodological application would have dictated that these types of words should have been subject, as international and preimmigration items were, to the phonetics test (section 4.3.2). However, they were so numerous (over 1400 were used in the OZC) that phonetic evaluation would have been too time-consuming at this juncture. All single-word proper place names were excluded from the database unless they met the conditions enumerated in section 4.4.3.

English-origin lexical items that result from exposure to English ensures that generalizations made with respect to lexical borrowing in Spanish in New York City are attributable to English language contact.

It may be difficult to determine whether an etymologically English lexical item has come to be available for an informant's use because of his or his community's exposure to English or because of exposure to other Spanishes. In fact, in the highly multilingual New York environment, it is unrealistic to expect that the use of any English-origin phrase would be due entirely to exposure to either just English or just Spanish. For instance, the Puerto Rican Spanish word for 'apartment' is *apartamento*. In Mexico, it is typically *departamento*. When a New York Mexican uses the word *apartamento* 'apartment', it may be that the Mexican's use is made possible by his knowledge and acquiescence to norms of Puerto Rican Spanish or by his knowledge and morphophonological adapting of English *apartment*. (A potentially etymologically English lexical item, like *apartamento* uttered by a New York Mexican, would theoretically be included as an English-contact lexical borrowing if it resulted from the latter scenario, but not from the former.) More likely, a Mexican's use of *apartamento* is encouraged due to a combination of these situations.<sup>21</sup>

In an effort to include in the database only lexical items that resulted from contact with English and not those of etymologically English origin that may have come to be used because of an informant's contact with other varieties Spanish, all instances of preimmigration loanwords (as determined by the precontact text) were excluded from the database, regardless of the ethnonational identification of the informant using it. For instance, because *apartamento* 'apartment' is a preimmigration word of Puerto Rican Spanish, not only were all instances of

---

21 Of course, these are just two possibilities of several, since a Mexican may also have been prompted to use *apartment(o)* through interaction with speakers of other immigrant languages in the city.

*apartamento* uttered by Puerto Ricans excluded, but all instances of *apartamento* were excluded regardless of which country the informant had ties to. So, despite the fact that the standard word for ‘apartment’ in Mexico is *departamento* (and **not** *apartamento*), when *apartamento* was used by a Mexican, it was not included in the LBD. There were approximately 200 tokens of these either-language-or-dialect-contact loanwords that were excluded from the database.<sup>22</sup> (Excluding these items makes the LBD a more conservative representation of the extent of English lexical influence on Spanish in New York.) An exception to the protocol of excluding these items occurred when they were realized using an English phonetic pattern. Some examples of dialect contact items for which a token was occasionally pronounced with English phonology, and which were thus included in the LBD, were: *email* (10x), *internet* (5x), *kindergarten* (12x), *record* (2x), *shock* (9x),<sup>23</sup> *volleyball* (2x), *basketball* (5x), *football* (4x),<sup>24</sup> *baseball* (7x), *bar* (1x), *manager* (5x), *disco* (2x) and *ghetto* (2x).<sup>25</sup>

### 4.3.2 Synonymous bilingual homographs

English has a large stock of now fully linguistically- and socially-integrated vocabulary of Romance origin. In some cases, the linguistic processes acting on these words since their introduction to English and acting also on the naturally transmitted descendants of their etymons in Spanish, in conjunction with orthographic conventions in each language, have resulted in *synonymous bilingual homographs*: words that are spelled identically in Spanish and English and have roughly the same meaning. For instance, the word for ‘doctor’ is spelled identically in

---

22 Compare this number with the total number of borrowing tokens in the database: 4507. Had the dialect-or-contact loanwords been included, the total number of borrowing tokens would have exceeded 4700. In that case, they would have represented about four percent of the total borrowing database.

23 *Shock* also occurred in the borrowing *cultural shock* (326E).

24 *Football/fútbol* is included in the database when it is pronounced with Spanish phonology but refers to the game of U.S. football. There is one such token in the database.

25 These words accounted for 66 borrowing tokens, or one and a half percent of the lexical borrowing database, which contained 4507 borrowings.

English and Spanish. All synonymous bilingual homographs were extracted from the OZC and included in or excluded from the LBD based on their phonological realization. If a synonymous bilingual homograph was uttered with English stress placement and/or phonetic realizations (see description of the phonetics test in this section for specifics), the informant was considered to be importing English lexical patterns. The item was thus classified as a borrowing and included in the database. If the synonymous bilingual homograph failed to meet phonological criteria, the grounds for establishing that the lexical material was English-origin (condition (2) for lexical borrowing identification) were considered insufficient, and the word in question was excluded from the database.<sup>26,27</sup>

The use of phonetic realization to ascertain whether synonymous bilingual homographs were English-origin is a proxy in the absence of certainty about whether they in fact constitute a reproduction of English lexical material in Spanish. It must be highlighted, in any case, that bilingual homographs represent, by definition, lexical patterns that are also part of a Spanish language lexicon. Thus, synonymous bilingual homographs uttered with English phonetic

---

26 As may be recalled (Chapter 2, section 2.6.3), I consider recipient language phonological adaptation as a criterion for the identification of borrowings (versus codeswitches) to be undesirably limiting for synchronic studies of on-going contact. Note that the use of phonological realization in this study is not for distinguishing between lexical borrowing and codeswitching, but for deciding whether synonymous bilingual homographs (and also preimmigration and international loanwords) instantiate reproductions of English-origin (or U.S.-contact) lexical material in the recipient language or not. Thus, as utilized for lexical borrowing identification in this study, phonological realization is employed in a way distinct from that prescribed by Poplack, for example. Here, the presence of **donor** language phonological patterns is used to justify the classification of a foreign-origin string as a lexical borrowing, whereas typically scholars would stipulate that it is **recipient** language phonology that supports the classification of a foreign-origin string as a lexical borrowing.

27 In employing a phonetic means of evaluating interlingual cognates, I am not claiming that the speaker is accessing an exclusively English *mental lexicon* (Levelt 1989). (Some words may be registered simultaneously in an English and in a Spanish lexicon (Clyne 2003: 165).) The phonological component of one language may be active during speech produced with the syntactic and lexical components of another language (see, for instance, Muysken 2000: 252-259). For example, a speaker could say [p<sup>h</sup>uε do a blar esp<sup>h</sup>an yo] (English phonology with Spanish lexicon and syntax) or [ay εs pik in gli] (Spanish phonology with English lexicon and syntax). (In everyday terms, this is referred to as “having an accent”.) Thus, just because a word, like *puedo* ‘I can’, is rendered using English phonology ([p<sup>h</sup>uε do]) does not mean that *puedo* is part of the speaker’s English lexicon. This indicates that phonetic realization does not necessarily reflect the membership of a word in one mental lexicon of one language or another.

patterns, and which are thus included in the database of English lexical borrowings, are exceptions to the condition that English lexical borrowings in this study should **not** be part of the precontact Spanish of the informant or in the variety of Spanish spoken by an informant's home country counterparts (i.e. they were exceptions to condition (3)). In this way, the phonetics test is a trapdoor on condition (3) of the definition of lexical borrowings in this study, that lexical material not be used in the home country variety of Spanish spoken by the informant. Words that are able to meet condition (2) (i.e. reproduction of English word-form(s) + meaning) but fail to meet condition (3) can escape exclusion using the phonetics test.

**The phonetics test.** When met with a single word that was a synonymous bilingual homograph (*doctor*) in a transcription, phonetic realization was used to approximate whether the informant was (mostly) importing English lexical material, in which case the item would be included in the study. The phonetics test was also used as a means for determining whether preimmigration loanwords (*manager* in Cuban Spanish) and international loanwords (*jazz*) were, nonetheless, instances of U.S.-contact items (see Haugen 1972: 101). For expositional purposes, these three types of words will be referred to as *interlingual cognates*.

The phonetics test is a truncated methodology for facilitating the assessment of interlingual cognates as English-origin, U.S.-contact items. The mechanism of the phonetics test operates to include an interlingual cognate item when evidence of English phonological realization is present, rather than to confirm exclusion of these forms if evidence of Spanish phonological realization were present. With respect to preimmigration loanwords, this mechanism allows for the possibility that languages may borrow the same word from the same language at different times in their history, and that these borrowed words reflect the phonological patterns of the donor language at the time of the borrowing (consider, for instance, the established French-origin

loanwords *chief* and *chef* in English). With respect to bilingual homographs, this mechanism is simply a methodological means for determining whether lexical items can be considered reproductions of English-origin material (in which case they are included) or not (in which case they are excluded).<sup>28</sup>

The phonetics test uses a hierarchy of segmental and suprasegmental features. Stress placement is given primacy over vowel realization, which in turn is given primacy over consonant realization. This order of consideration reflects the relative ease of hearing these features on audio recordings. Stress placement is more easily audible than vowel realization, and vowel allophonic distinctions are more easily audible than consonantal ones.

The first criterion used for deciding whether an interlingual cognate would be included as a U.S.-contact English lexical borrowing was syllable stress. If an English versus Spanish realization of the word predicted different stress location, stress placement in the informant's uttered word was the sole criterion for inclusion or exclusion from the database, regardless of the realization of the phonological segments of the word.<sup>29</sup> For example, excerpts (1) and (2) were uttered by informant 065P.

- (1) Y al momento llega el *Doctor* ['dok tor] Rossen, bueno, dice... 065P  
'And at that moment *Doctor* Rossen arrives, well, he says...'
- (2) Le puso la medicina, entonces porqué el (a) *Doctor* ['dok tor] Rossen no puso ese (b) doctor [dok 'tor] a atenderme... 065P  
'He gave him the medicine, so why didn't *Doctor* Rossen have that doctor attend me...'

---

28 It is assumed in the case of bilingual homographs that, in a discourse in which Spanish is primarily activated (Grosjean 1997a; Paradis 1997), the use of English phonological patterns by an informant instantiates a more deliberate attempt to reproduce English lexical material.

29 Primacy was given to stress location because, as mentioned, segment realization is under the control of some speakers (altering it, for instance, depending on the discourse meaning or (perceived) proficiency of an interlocutor) and because evidence also shows segment realization to be an unreliable indicator of language membership in early stages of borrowing to loanword transition, where the data of this corpus reside (Chapter 2, section 1).

A Spanish pronunciation of *doctor* places stress on the second syllable, while an English one predicts stress on the first. Based on the audio recording of the interview, it was found that in example (1) informant 065P places stress on the first syllable of *doctor*, conforming to English stress placement on this word. It was thus included in the database (regardless of the fact that the informant also realized all vowels and consonants according to Spanish phonetic patterns). In (2), she also uses the English stress pattern. Token (2a) of *doctor* was thus included in the database. In the third instance of *doctor* (example 2b), the informant locates the word stress in the expected position for Spanish, and so this token of *doctor* is not included in the corpus.

When English and Spanish realization of a word predicts stress on the same syllable, the realization of phonological segments was used to determine whether or not to include it in the LBD. If all vowels were phonetically English, it was included in the study, regardless of the pronunciation of consonants. For instance, in (3), the word *rock* gets primary stress in English and Spanish. Because the speaker in (3) realized ‘o’ as [ɑ], as in English (and not [o] as would be expected for a Spanish pronunciation), it was included in the corpus.

- (3) A mí me gusta más *rock* [ɹɑk], pero me gustan muchos diferentes estilos de música... 013U  
‘I like *rock*, but I like many different styles of music...’
- (4) [Mis niños] no viven en *clubs* [klʌbs], nada, nada. Son sanos todos... 347M  
‘[My kids] don’t live in *clubs*, nothing of that sort. They are good clean [kids] all of them...’

Likewise, in (4), *clubs* was uttered with an English mid central lax vowel [ʌ] (instead of the tense high back vowel [u] expected in a Spanish realization of the word). Because of this, and despite the realization of the consonantal plural allomorph as a Spanish /s/ (i.e. a voiceless

alveolar fricative [s]) as opposed to English's voiced [z] plural allomorph after a voiced final consonant, it was included in the database.

When both (i) Spanish and English require stress to fall on the same syllable and (ii) not all vowels are realized with English allophones, a third strategy, utilizing consonant realization, was employed to determine whether an interlingual cognate would escape exclusion. A tally was done of consonantal segments (or segment positions) of the word where an English realization or a Spanish realization of it would differ (what I term a *consonantal conflict site*). If a greater number of these conflict segments were realized as English consonantal allophones, the word was included in the database; if a greater number of conflict segments were realized in Spanish, the word was excluded. Although there are numerous allophonic differences between the realizations of mutually-shared Spanish and English consonantal phonemes, only a handful were selected as conflict sites.<sup>30</sup> Phonemes (plus one morphophonemic context) qualifying as conflict sites included:

- /R/<sup>31</sup> (U.S. English retroflex [ɹ]; Spanish alveolar tap [ɾ], trill [r], uvular [ʀ], etc.)
- /L/ (U.S. English [ɫ]; Spanish non-velarized [l])
- /S/+C word initially (English [s]; Spanish [ɛs])
- Syllable-initial voiceless stops /P/, /T/, /K/ (U.S. English [p<sup>h</sup>], [t<sup>h</sup>], [k<sup>h</sup>]; Spanish unaspirated [p], [t], [k])
- Syllable-initial voiced alveolar affricate spelled 'j' in English (U.S. English [dʒ]; Spanish [x], [h], [j])
- plural allomorph /s/ following a voiced C or V (U.S. English [z]; Spanish [ɛs] or [s])

---

30 There are a number of reasons that I selected only a few phonemes and environments to be conflict sites. First, time constraints prevented comprehensive phonetic analysis of words; second, some types of segments are easier to hear on an audio recording than others and; three, my accuracy in making more subtle phonetic distinctions by ear has not been tested.

31 For the purpose of this exposition, small caps between forward slashes (such as /R/, /L/ and /S/) are meant to serve as a bilingual notation. Each should be understood as a sort of bilingual archiphoneme that could be realized by distinct sets of allophones (or phonemes) in each language. In the case of /R/, I am positing that U.S. English /ɹ/ may be associated with two different phonemes in Spanish: tap /ɾ/ and trill /r/.

An example of a context for which conflict sites were relied upon to determine whether an interlingual cognate could be included in the database is the word *rock* in example (5).

- (5) Entrevistador: ¿Qué te gusta?  
Entrevistado: Ciertas partes del *rock* [ɾok]. Tú sabes. 005U  
‘Interviewer: What do you like?  
Informant: Certain parts of *rock*. You know.’

Stress placement for *rock* is the same in Spanish and English. In addition, the informant does not realize the vowel as English [ɑ]. Since, there is only one conflict site in this word (i.e. the /R/), its realization determined whether this international loanword could nonetheless be included in the LBD. In (5) the informant realizes /R/ according to English norms of pronunciation, as [ɾ]. It was therefore included in the database. If, for some reason, there was a tie in the number of conflict sites realized in English and Spanish, the word was excluded from the database.

- (6) A mí me gusta más [k<sup>h</sup>in dɛr] porque pues son.. todavía son niñitos y son inocentes... 310C  
‘I like *kinder*[*garten*] most because well they’re.. they’re still little and they’re innocent...’

*Kinder* in example (6) contains two conflict sites: /K/ and /R/. The first is realized according to English phonetic constraints on syllable initial voiceless stops: as aspirated [k<sup>h</sup>]. The /R/ is realized according to Spanish norms as [r]. In that there was an equal number of conflict sites with Spanish pronunciation and English pronunciation, the token *kinder* uttered by 310C was excluded from the study.

In total, phonetic realization was the sole criterion<sup>32</sup> for the inclusion of approximately 326 bilingual homographs and interlingual loanwords (e.g. *CD* [si 'dij], *cable* ['kej bəl], *camping* [k<sup>h</sup>am piŋ], *DVD* [di vi 'di], *email* ['i: mejl], *GED* [dʒi 'di], *golf* [gɒlf], *graffitti* [grɪə 'fi ri], *hip*

32 In one case, the case of *final* (in “mi primer *final* [fi 'nal]...” 349M), a bilingual homograph was also determined to be an *interlingual coincidence* and its inclusion in the database hinged on the homonymy test (section 4.4.2).

*hop* [xɪp xɑp], *hockey* ['xo k<sup>h</sup>ij], *I.D.* [aj 'di], *Internet* ['ɪn teɪ nɛt], *jacket* ['dʒæ kɛt], *jacuzzi* (dʒə 'ku zi), *jazz* [dʒæz], *kindergarten* ['kin dɛɪ ɡaɪd ɹ̩], *lukemia* [lu 'ki mjə], *minor* ['maɪ nɛr], *punk* [p<sup>h</sup>ʌŋk], *reggae* ['rej geɪ], *record* ['ɪɛ koɪ], *X Box* ['ɛks baks]). The majority of the 326 lexical borrowings had clear English pronunciations. Only 60 of these (just over one percent of the total database; several of which are given as examples above) had mixed Spanish-English phonology and were admitted as English lexical borrowings using the procedure described above.

### 4.3.3 Etymologically foreign loanwords in English

The criterion that lexical borrowings be English-origin does not preclude the possibility of including lexical items that have their etymological origin in languages other than English. As long as a lexical borrowing was considered a word of English and was used in consonance with its English meaning, it was included in the database. *Graffiti* in example (7) is an example of a word with non-English etymology that was considered a lexical borrowing in this study.

- (7) Todo por *graffiti*, y por escribir un nombre en la pared... 432P  
 'Everything for *graffiti*, and for writing a name on the wall...'

The word *graffitti* in (7) is etymologically Italian. It is hypothetically possible that the speaker learned the word via contact with Italian or Italian speakers. However, when such words are used with a meaning consistent with that of the word in U.S. English, it is posited, given the immediacy of cultural pressure in an immigrant language situation, that they came to be used in a person's Spanish as a result of contact with speakers of English in the U.S. Thus, *graffiti* and similar words of English that themselves have non-English origins are included in the database.

#### 4.4 Lexical pattern reproduction

As here defined, lexical borrowing requires the reproduction of at least one donor language word-form plus meaning pair. This section primarily addresses one major challenge to the task of establishing that every instance of a borrowing represented an instance of lexical pattern reproduction: namely the occurrence of interlingual coincidences (section 4.4.2).

##### 4.4.1 Full articulation of a word

Lexical borrowings must consist of at least one fully-articulated word. If a speaker does not finish articulating a word, it is not included in the database. For example, in (8), *receptionist* is within the envelope of study, but the partial articulation *recep* is not.

- (8) Lo que se llama la recep... la rec... ay Dios mío... *receptionist*... 206U  
'What is called the recep... la rec... oh my god... *receptionist*...'

##### 4.4.2 Interlingual coincidences

The Spanicization of English lexical material is common in Spanish in New York. That is, English-origin lexical items are frequently to partially or wholly reflect the phonological, morphological and/or syntactic patterns of Spanish. For instance, *rufo* 'roof', *boila* 'boiler', *jai-eskúl* 'high school' and *cuora* 'quarter' are all typical renderings of English-origin lexicon when used in Spanish discourse. In addition, English has a great number of words of Latinate origin. The Spanicization of these Romance-origin English words can result in English lexical borrowings that are phonologically identical to the naturally transmitted descendants of their etymons in Spanish. For instance, the word *carpeta* when used by a New York Spanish speaker often means 'carpet'. *Carpeta* is also a standard Spanish word, but its meaning is 'folder'. Likewise the word *realización* 'attainment' is part of standard Spanish lexicon. When used by

Spanish speakers in the U.S., it often has a meaning more similar to English *realization* ‘awareness’ than to standard Spanish *realización*. The question is whether a particular word-form, such as *carpeta* or *realización* (a) is an inherited recipient language form whose meaning has been extended to accommodate a meaning of a cognate word in the donor language (i.e. [direct] contact-induced semantic extension<sup>33</sup>) or (b) is donor language lexical material (word-form+meaning) that has been transferred and adapted to the phonology (and/or morphology) of Spanish.<sup>34</sup> Word-forms whose manner of having come to be is uncertain in this way will be called *interlingual coincidences* (Haugen 1950: 102-103, 1953/69: 400).

In other words, the main question being posed at this point is **not** whether words like *realización* ‘awareness’ have come to be used due to language contact or whether such usages have origin in patterns of lexical use in English (in all likelihood both are true given strong Anglo cultural and linguistic pressures on Spanish-speakers in the U.S. and given the establishment through the precontact test that such word-form-meaning combinations were not used in the home country variety of Spanish spoken by the informant). Rather, the question is whether such uses instantiate the reproduction in Spanish of English lexical material or only English semantic patterns. Making the distinction is central to this study since, as here defined,

---

33 I assume for now, in the absence of lexicographic data on community-wide usages of such words, that complete semantic transfer (see footnote 34) has not occurred.

34 From the perspective of bilingual language processing, several more distinctions could be made to explain the presence of Spanish-English interlingual coincidences like *carpeta* and *realización*. For instance, to take just two examples, these forms could hypothetically be semantic extensions that occurred independent of contact with another language or its speakers that only happens to be semantically similar to a cognate form in the donor language (i.e. *non-contact-induced semantic extension*). Else, these forms could be *semantic transfers* (Clyne 2003: 77). That is, the meaning of English *carpet* and *realization* may be associated with and transferred to similar phonological strings that are already part of the Spanish lexicon. In other words, it could be the case that *realización* ‘realization / awareness’ comes to be used because the speaker knows the word *realización* exists in Spanish, and substitutes the English meaning for the Spanish one. The difference between a *semantic extension due to contact* and a *semantic transfer* is a difference in quantity, not in kind. That is, a semantic extension due to contact would add a sense from an English word to the senses already expressed by a Spanish word-form. A semantic transfer would replace all senses of the inherited Spanish form with the senses associated with the English form. Making the distinction would be best made via a lexicographic study of the use of words in the community language.

lexical borrowing entails the reproduction of a donor language word-form + meaning pair, while semantic extension (under contact) would involve only the transfer or reproduction of donor language meaning patterns. Furthermore, attempting to operationalize a distinction is important, seeing as the inclusion of all interlingual coincidences as lexical borrowings would have resulted in the overstatement of the extent of English lexical influence in Spanish (Otheguy 2012: 227). On the other hand, excluding these items *carte blanche* from the database would have ignored valuable data that informs about the nature of lexical borrowing in New York.<sup>35</sup>

From the perspective of the recipient language,<sup>36</sup> Spanish, the question amounts to asking whether a change introduced into Spanish with the admittance of an interlingual coincidence to its lexicon would result in a polysemous Spanish word (i.e. whether it represents a single lexeme with several different senses), or rather whether it is accidentally homonymous with an inherited Spanish word. To decide, then, whether an interlingual coincidence, such as *realización*, should be included in or discarded from this study, the homonymy test was developed. Before describing how it was applied, a brief discussion of homonymy is in order.

Homonymous words are, according to Lyons (1977: 21-22, 206), words that are identical in *form*<sup>37</sup> but that represent distinct *lexemes*.<sup>38</sup> For a word-form to represent two distinct lexemes, it

---

35 For example, if future work with this database inquired about the rate of morphophonological incorporation of lexical borrowings, the exclusion of borrowings that just happen to be Spanicized would give the impression that morphophonological incorporation occurred less often than it did, or fail to reveal that, perhaps, the morphological adaptation of lexical borrowings to recipient language patterns occurs most frequently when the adaptation produces a word-form that patterns after already-existent recipient language word-forms.

36 Making the distinction with respect to the idealized notions of the community language, and not, say, from the perspective of bilingual language processing, is necessary since the linguistic evidence available (i.e. morphosyntactic traits of a word form and its meaning in context) is insufficient to do so. The reason is that more processing possibilities are available than linguistic and semantic features for distinguishing them. In other words, multiple processes map onto a fixed set of linguistic resources with the result that one form-meaning combination could be the result of several distinct processes. Although the processes by definition entail a certain arrangement of linguistic features, the linguistic features do not mutually or exclusively entail one particular process.

37 As defined by Lyons (1977: 18-19), a *word-form* is a sequence of symbols separated by a space. In this dissertation, *word-form* designates a sequence of sounds that minimally realizes one unbounded morpheme.

is not enough for the form to have two different meanings.<sup>39</sup> For example, *EYE* ‘organ of sight’ is signifies something distinct from *EYE* ‘hole in a needle’. Yet, these two meanings of *eye* are related through the idea of *hole in the middle of something*. They are thus considered to be two meanings of the same lexeme.<sup>40</sup> That is, having different meanings is not a sufficient basis on which to establish homonymy. Rather for word-forms with different meanings to be considered homonyms, the meanings must be unrelated.<sup>41</sup>

Semantic relatedness is difficult to define and its correlation with speakers’ judgments is problematic<sup>42</sup> (Lyons 1977: 550-552). For instance, the basis on which a speaker could claim that two meanings are semantically related may differ in both degree and kind for different pairs of words. For instance, I would wager that most Spanish-speakers would **not** think that the meaning of *CARPETA* ‘rug’ is related to that of *CARPETA* ‘folder’, even though the *denotatum*<sup>43</sup> of both words share the feature of shape; they are both flattish objects (i.e. when used in typical fashion, they appear far more wide and long than they are tall). Yet the same individuals (if also

---

This latter definition is intended to reflect Lyons’ use of the term, except that he explicitly defines it with respect to written word-forms only.

38 As defined by Lyons (1977: 19), a lexeme is an abstract group of word forms with the same denotational/cognitive meaning. Note that Lyons reserves the use of single quotes ( ‘ ’ ) to indicate the citation form of a lexeme. In the current work, the citation form of a lexeme is typically provided in *italicized lower case* fonts (as are word-forms). *ITALICIZED SMALL CAPS* are used for the citation form of a lexeme only when *lexeme* is being defined or exemplified in exposition.

39 Lyons uses double quotes ( “ ” ) to indicate the meaning of a word-form or lexeme. In the present study, single quotes ( ‘ ’ ) are used for meanings (and glosses).

40 That the form *eye* with both of these meanings represents a single lexeme is reflected in the work of dictionary makers who do not give *eye* two entries, but rather give it a single entry with the various meanings of *eye* enumerated within the entry.

41 When the meaning of identical word-forms are taken to be related, the word-form is said to be polysemous.

42 For instance, although most dictionaries give *ear* ‘external protrusion of most mammals that aids in hearing’ and *ear* ‘fruiting spike of grain’ different entries (indicating that the word-form *ear* is homonymous and stands for two unique lexemes that have two unrelated meanings), Lyons notes that some English speakers have been known to see a connection between the two meanings of *ear*.

43 For the current discussion, *denotatum* is “the class of objects, properties, etc., to which to which an expression correctly applies” (Lyons 1977: 207). For example, the denotatum of *blue* is the class of all blue things. The denotatum of *dog* is a particular class of animals. Denotation, in other words, is “the relationship between a linguistic entity and something outside the language-system” (Lyons 1977: 210).

speakers of English) would likely say that the meanings of *EYE* ‘organ of sight’ and *EYE* ‘the hole in a needle’ **are** related, and this, clearly by virtue of the shape of the denotata (and probably also its location in the object).

Lyons notes that endeavors to define semantic relatedness may be in vain and, indeed, despite utilizing the concept of semantic relatedness, I do not attempt a comprehensive definition here. Instead, I stipulate the basis upon which semantic relatedness can and cannot be established. Semantic relatedness of two word forms cannot be established on the basis of the ontological nature of denotata, such as being an object, an idea, an action or a quality. That is, for the purposes of this discussion, it is not sufficient to claim that the meaning of *BANK* ‘container to hold money’ and *BANK* ‘ridge of earth on the side of a river’ are semantically related because they both denote objects; or that [sow] ‘to plant seeds’ and [sow] ‘to fasten together with thread’ are related because they both denote activities. Rather, relatedness of meaning can be established on relationships of antonymy (e.g. hot~cold), causality, hyperonymy,<sup>44</sup> location, temporality and function. In addition, I consider two meanings related if they can be related by processes noted to be the basis of semantic change, such as metonymy,<sup>45</sup> generalization / specialization, pejoration / amelioration, intensification / weakening and metaphorical extension (see Pyles & Algeo 1993: 237-251 for descriptions of these processes).

---

44 For example, ‘red’ is a hypernym of ‘scarlet’. Also see Smead (2000: 292).

45 Metonymy is a transfer or change of meaning of a word facilitated by the metalinguistic context in which it is used. For example, in 1300, English *hearse* meant ‘candle framework, hung over a coffin’. The meaning then shifted to ‘temporary canopy hung over a coffin’. By 1650, *hearse* had been registered with the meaning ‘vehicle for transporting the dead’. Since lexicographers generally see all these meanings as related (i.e. there is only one entry for *hearse* in the dictionary (Dictionary.com)), the relatedness of the meanings must be the funereal context of use (perhaps in addition to the functionality of the object with respect to a coffin). Another example of metonymy is demonstrated by the shift in meaning of the word *foot*, originally meaning ‘lower extremity of a human’. It later came to acquire the meaning ‘approximate length of a human foot’ through metonymy, that is, because of the use of the human foot to make measurements (example from Pyles & Algeo 1993: 242-243).

**The homonymy test.** To determine whether a word form, like *carpeta* ‘rug’ in Spanish in New York, is better characterized as an extension of *CARPETA* ‘folder’ or rather should be considered a novel addition to the lexicon of Spanish (i.e. that it is *homonymous*<sup>46</sup>), a two-step protocol is employed. First, it is asked whether *carpeta* ‘rug’ and *carpeta* ‘folder’ are semantically related. Because ‘folder’ and ‘rug’ are not semantically related in the ways above specified, it is posited that *carpeta* ‘folder’ and *carpeta* ‘rug’ qualify as homonymous word-forms. After establishing that a (potentially) Spanicized English lexical borrowing **is not semantically related** to the meaning of its phonological cognate form in Spanish, it must be further established that *carpeta* ‘rug’ has come to be used as a result of contact with English. The second step of the protocol therefore asks whether the meaning of word-form as used by an informant **is semantically related** to the meaning of the cognate form in U.S. English. That is: Is the meaning of *carpeta* ‘rug’ semantically related to the meaning of its English cognate form, *carpet* ‘carpet’? In this case, the answer is ‘yes’. Thus, the word *carpeta* ‘rug’ is included into the LBD as an example of a U.S.-contact English lexical borrowing. Other interlingual coincidences determined to be Spanicized English lexical borrowings (and which are thus included in the lexical borrowing database) are in Appendix C.

By means of this protocol, several interlingual coincidences were excluded from the database. For example, the word *colegio* is used by corpus informants to mean ‘university’. If *colegio* ‘university’ is subjected to the procedure just outlined, its meaning is found to be both (i)

---

46 Note that in Lyons’ discussion, the homonymous word-forms in question are taken to be lexemes of a language, recognizable by any “native speaker” (1977: 551) as words of that language. Although I attempt to operationalize the homonymy criterion in order to distinguish between spanicized English lexical borrowings and historically inherited Spanish words that have shifted or changed in meaning, I am not making claims about the status of Spanicized English lexical borrowings in the mental lexicon of individual Spanish speakers or in a dictionary for any variety of Spanish. That is, while it may be safe to posit that the meanings of words determined to be Spanicized English lexical borrowings are not semantic extensions of historically inherited cognate forms, I do not claim that these items are themselves listed in a Spanish lexicon or the mental lexicon of the speakers who use them.

semantically related to the inherited Spanish meaning of *colegio* ‘K-12 school’ or ‘private/catholic school’ (through their function of both being an ‘institution for education’) **and** (ii) semantically related to the English cognate word *college* ‘university’. Due to inability to demonstrate that *colegio* ‘university’ is **not** an extension of Spanish’s historically inherited *colegio* ‘private/catholic school’ (because of answering ‘yes’ to question (i) of the protocol), it cannot be convincingly argued to have resulted from the Spanicization of an English lexical borrowing. It is therefore excluded from the database. Appendix C (section 2) contains interlingual coincidences that were excluded from the corpus due to failing the homonymy test.

#### 4.4.3 Proper noun phrases

The language membership of proper names is a contentious issue (Lyons 1977: 222). Clyne (2003) suggests that proper names may belong to the lexicon of multiple languages simultaneously.<sup>47</sup> Consequently, in order to meet the condition that their presence in Spanish was an example of reproduction of English material (condition 2), proper names had to meet additional criteria before they could be included in the LBD.<sup>48</sup> Particularly, they are included when a translation equivalent is either (a) known to be used by New York Spanish speakers (9)-(13), (b) part of a general or standard lexicon of Spanish<sup>49</sup> (9)-(13), (c) easily rendered into Spanish, as in the case of acronyms (14) or personal titles (19)-(20), or (d) when the proper name is a compound English word (15).

---

47 It is for this reason that Clyne (2003: 165) identifies them as a *trigger* in bilingual speech. That is, because proper nouns simultaneously belong to multiple lexicon of plurilingual speakers, they are able to *trigger* the switching of a discourse language in mid-thought or mid-clause.

48 As a species of international loanword, single-word proper place names, such as names of continents (*Africa*), countries (*Zimbabwe*), states/territories (*Florida*), boroughs (*Manhattan*, *Queens*) and neighborhoods (*Bedford*) should have been subject to the phonetics test (section 4.3.2). However, they were so numerous (over 1400 were used in the OZC) that phonetic evaluation would have been too time-consuming at this juncture. All single-word proper place names were excluded from the database unless they met the conditions enumerated in this section.

49 The translation equivalent of an English-origin proper noun is considered to be part of the general lexicon of Spanish if it is in *Diccionario de la RAE*.

Examples (9) to (15) contain single-word proper nouns that meet at least one of these conditions and are, thus, included in the database.

- (9) Una vez, bueno, eso fue en un.. tiempo de *Christmas*, de.. así, de navidad, que hicieron un.. hicieron una rifa... 025C  
'Once, well, that was in a.. time of *Christmas*, of.. yeah, of Christmas, and they did a.. they did a raffle...'
- (10) Y, y al otro día, nosotros llegamos en una víspera de.. de (a) *Easter*.. en la misma semana de (b) *Easter*... 118D  
'And, and on the next day, we arrived during the eve of.. of *Easter*.. during the same week as *Easter*...'
- (11) Me dice mi mamá ¿Y a dónde tú vas? Y yo le dije "Ay, mami, a la fiesta de *Halloween*." Dice "¿Tú me pedistes permiso?"... 183U  
'My mom said to me "And where are you going?" And I told her "Ay, mom, to the *Halloween* party. She said "Did you ask me permission?"...'
- (12) Ella se graduó en *February* del próximo año porque tenía que hacer las clases que no pasó... 365E  
'She graducate in *February* of the following year because [she] had to do the classes she didn't pass...'
- (13) Aunque a veces también como el *Spanglish* se empieza a meter en tu idioma... 340M  
'Although sometimes too like *Spanglish* begins to appear in your language...'
- (14) Por ejemplo este año el grupo, hay muchos estudiantes que no.. que no estudiaron en *NYU* [en wai ju]... 092P  
'For example this year the group, there are many students that don't.. that don't study in *NYU*...'
- (15) Uno tenía que ñagotarse de espalda y bajar como *Spiderman*... 432P  
'One had crouch his back and go down like *Spiderman*...'

Among New York Spanish speakers, *Christmas* (9) is often rendered as (*las*) *navidad(es)* (also the standard Spanish) and *Easter* (10) as *Pascua*. *Halloween* (11) can and is sometimes rendered as *Día de los muertos*. *February* (12) is often rendered *Febrero* (also in standard Spanish) and *Spanglish* (13) as *Espanglish*. *NYU* [en wai ju] (14) could have been rendered as [enei grie ga u]

(although for whatever reason, initialisms are rarely converted to Spanish (see Teschner 1972: 1069-1076)). *Spiderman* (15), because it is a compound consisting of the common English nouns *spider* and *man*, was included in the database.

Excerpts (16)-(20) contain multiword proper nouns that meet at least one of the above-mentioned criteria and that are also in the LBD.

- (16) ¡Bueno viví en *New Jersey*! Pero me gustaba mucho el ambiente... 351M  
'Well, I lived in *New Jersey*! But I liked how the atmosphere was very much...'
- (17) No, él vive en *Long Island*... 325E  
'No, he lives on *Long Island*...'
- (18) Solamente [había] un autobús venía directo, aquí a *Port Authority*... 346M  
'[There was] only one bus came direct, here to *Port Authority*...'
- (19) Em... otra persona eh de, de cuando de eh de chiquito, *President Kennedy* también, cuando lo mataron, eh eso me afectó mucho... 417P  
'Uhm... another person uh from, from when, when uh I was little, *President Kennedy* too, when they killed him, uh that affected me a lot...'
- (20) El profesor era.. era *Mister K---*, todavía me acuerdo, era tan difícil el profesor... 300E  
'The professor was.. was *Mister K---*, I still remember, he was so difficult that professor...'

Because *New Jersey*, *Long Island* and *Port Authority* in examples (16)-(18) each contain at least one English common noun or adjective (i.e. *island/long*, *new*, *authority/port*) that could be easily rendered in Spanish, each is counted as a lexical borrowing.<sup>50</sup> Because *President* (19) and *Mister K---* (20) are easily translatable into Spanish as *Presidente* and *Señor*, *President Kennedy* and *Mister K---* are included in the database.

---

50 Note that although *New Jersey* is sometimes rendered as *Nueva Jersey*, resulting in its also meeting the criterion that the proper name is known to be Spanicized by Spanish speaker in New York. For whatever reason, community standards seem to dictate that *Long Island* and *Port Authority* are not ever rendered as *Isla Larga* or *Autoridad de Puertas*, respectively.

#### 4.5 General inclusions

Any noun, adjective, adverb, pronoun, etc. that meets the criteria outlined above is included if it occurs at the edge of a codeswitch and plays a role in the structure of an adjacent Spanish clause. For instance, in example (21), *seafood* is included in the corpus because it is a constituent of the syntactic structure of the previous Spanish-language clause *ellos tienen mucho seafood*.

- (21) So ellos siempre tienen mucho **seafood** you know shrimps, and little fish and little scallops, and things like that, *scampi*, it's more of a.... 333D  
'So they always have a lot of **seafood** you know...'

On the other hand, the strings (i) *you know*, (ii) *shrimps*, (iii) *and little fish and little scallops* (iv) *and things like that* and (v) *scampi* are not arguments or adjuncts of the *tener* clause. Rather they are like descriptive afterthoughts about *tener*'s direct object *seafood*. Since they are not part of the syntactic structure of the previous clause, the codeswitch that begins with the subject-verb combination *it's* is considered to bleed out to its left and right, encompassing all adjacent English-origin material until it meets a word that **is** part of the syntactic structure of an adjacent Spanish-language clause.

Consider also example (22).

- (22) Él vendía.. él trabajaba en una.. en una tienda de beepers, de **beepers and cell phones, telecommunications, you know electronics store, he was the manager there...** 333D  
'He used to sell.. he used to work in a.. in a *beeper* store, for **beepers and cell phones, telecommunications, you know, electronics store, he was the manager there...**'

In example (22), *he was the manager there* is a codeswitch. The codeswitch grows to encompass all English adjacent to it, both before and after, until either (i) meeting a Spanish verb + argument string or (ii) meeting an English-origin constituent that is clearly part of the syntactic

structure of an adjacent Spanish clause. In the case of example (22), *beepers and cell phones* is the object of the preposition *de* that is subordinate and descriptive of the NP *una tienda* ‘a store’ (which itself is the object of the Spanish preposition *en* ‘in’ which modifies *trabajaba* ‘he used to work’). *Beepers and cell phones* is thus counted as a lexical borrowing. However, *telecommunications, you know* and *electronics store* are not syntactically linked to preceding *trabajaba* clause, so they are considered to be part of the codeswitch and excluded.

## 4.6 General exclusions

### 4.6.1 Codeswitches

Codeswitches are **not** examined in this study. Any English-language string that includes a verb plus some other constituent of the verb’s clause is considered a codeswitch to English (see section 4.1). Excerpts (23) and (24) contain examples of codeswitching in the OZC.

- (23) Si ella me compra aunque sea un diseño, *she’s giving me exposure*. 012U  
‘If she buys even [just] one design from me, *she’s giving me exposure*.’
- (24) Las compañías tienen que este.. *release their earnings*, precisamos eso... 324E  
‘The companies have to uh.. *release their earnings*, we demand it...’

In (23), the speaker begins a sentence with a conditional clause in Spanish (*Si ella me compra aunque sea un diseño*) and then utters the main clause in English. In (24), after having spoken the subject and main verb of the sentence in Spanish (*las compañías tienen que* ‘the companies have to’), the speaker utters the infinitive verb phrase *release their earnings* in English and follows with more information in Spanish (*precisamos eso*). Both *she’s giving me exposure* and *release their earnings* are examples of codeswitches. (The former is often called an intersentential codeswitch while the latter is often called an intrasentential codeswitch.)

In identifying codeswitches, it makes no difference whether the string is completely or partially articulated. If the speaker changes her mind in the middle of the codeswitch, it still qualifies as such.

- (25) Muchos de los muchachos estos días están interesados en, en solamente em, como, llamamos un *good time*, and *they don*.. un pasa tiempo... 228D  
'Many of the boys these days are interested in, in only, uhm, like we say a *good time*, and *they don*.. a past time...'

In (25) the speaker uses the English phrase *good time* and continues in English for a bit uttering *and they don* before returning to Spanish. The string *they don* is a codeswitch because it contains an NP constituent, the subject *they*, and part of the verb, *don*. It does not matter that the speaker did not finish uttering the word *don't* or the rest of the verb: the fact that a string contains at least part of a finite verb and part of another constituent from that verb's sentence makes it a codeswitch.

In addition, all-English prepositional phrases (P+NP) are considered codeswitches, based upon the distribution of these strings among informants of the corpus (see section 4.1). So, for instance, *for about two hours* in (26) is excluded from the database.

- (26) Me quedo ahí *for about two hours*... 333D  
'I stay there for about two hours...'

- (27) Para nada, y ah.. con la mayoría la gente eran Latinos, casi todos. Habían cuantos gente americana, *like*<sup>51</sup> *Irish*<sup>52</sup>. 365E  
'Not at all, and ah.. with the majority, the people were Latinos, almost all of them. There used to be a number of Americans, *like Irish*.'

---

51 *Like* is considered a preposition when 'similar to' can be sensibly substituted for it in an utterance. On all other occasions *like* is considered a discourse marker (see Appendix D, section 10 for examples of discourse markers).

52 Proper adjectives that stand in for noun phrases are considered nouns in this study (Appendix D, section 1 for examples).

In example (27), the use of an English preposition plus English object NP classifies the string *like Irish* as a codeswitch, and it is excluded from the database.

#### 4.6.2 Edge of codeswitches

A string that is identified as a codeswitch, in turn, subsumes all adjacent same-language material as part of the codeswitch (unless, as mentioned in section 4.6, it meets a constituent that is part of an adjacent Spanish-language clause). So in (25), because the discourse marker *you know* plays no syntactic role in the previous Spanish-language clause (headed by *tienen* ‘they have’), it is considered part of the codeswitch *it’s more of a*. In general, discourse markers at the edge of codeswitches are excluded from the LBD, even though under other circumstances they qualify as borrowings (see Appendix D, section 10 for examples of discourse markers in this study).

#### 4.7 Summary

Expressions that met all of the above criteria were found to be part of the following lexical categories: common nouns and noun phrases, proper nouns, simple and complex verbs, adjective and adjective phrases, proper adjectives, tag phrases, adverbs, quantifiers, quantifier phrases, conjunctions, possessives, determiners and demonstratives. Although the definitions and criteria used to decide whether data fell within or outside the envelope of study suffer some shortcomings,<sup>53</sup> it should be stressed that definitions and criteria were not decided on apriori but selected only after a review of **all** segments of English found in the speech of the informants of

---

53 For example, some may find it strange to have excluded short phrases such as *release their earnings*, while including others like *World Trade Center thing* alongside items like *high school*, *laundry* and *landlord*. One justification for taking the data set I have lies with the different diffusion of these two types of strings among corpus informants (see section 4.1). The distributional analysis justifies the examination of *World Trade Center thing* and *high school* together as a single data set (that excludes *release their earnings*, i.e. a clause-like string).

this study. Appendix D provides additional examples of the words and phrases that are included in this study.

## 5. Dependent variables

The measures used to capture lexical borrowing behavior can be grouped into two categories: discourse frequency measures and lexical sharedness measures. These are described next.

### 5.1 Discourse frequency measures

#### 5.1.1 Lexical borrowing rate

The principal dependent variable used to gauge the frequency with which an informant borrowed from English is called the lexical borrowing rate. The lexical borrowing rate is a token-level measure. It counts how many times (per 1000 words of speech) the speaker utilizes a borrowing while speaking Spanish. One borrowing occurs each time a speaker uses a string of English, whether that string is composed of one word or several. For example, even though *double major* in example (28) has two words, it counts as one borrowing.

- (28) Yo estoy haciendo un *double major*... 340M  
'I am doing a *double major*...'

Each time a speaker utters *double major*, it counts as one more instance of a borrowing. Other excerpts from the database that include lexical borrowings are found in examples (29)-(31).

- (29) No, sí es cerca de *Williams Bridge*. 220P  
'No, yes it's close to *Williams[burg] Bridge*.'
- (30) Porque trabajaba yo *part-time* en esa época. 315M  
'Because I was working *part-time* during that period of time.'
- (31) Si tú estás en una colina bien alta, tú puedes ver *coast to coast*... 333D  
'If you are on a very high hill, you can see *coast to coast*...'

In (29) the entire string *Williams Bridge* is one lexical borrowing. Likewise, *part-time* is one lexical borrowing and so is *coast to coast*. Each and every time a speaker dips into English by uttering one of these phrases, one is added to the total count of his lexical borrowings. His total number of borrowings is then divided by the total number of words he uttered during his interview. So, for instance, if the four borrowings cited above (*double major*, *Williams Bridge*, *part-time* and *coast to coast*) were uttered by a single informant that uttered 570 words<sup>54</sup> during his interview, his borrowing rate would be  $\frac{4}{570} = .007$ , or seven borrowings per 1000 words of speech, written as 7/k.

### 5.1.2 Lexeme rate

The lexeme rate as defined in the present study is similar to what in other works is called a type rate (e.g. see Lyons 1977: 20). It indicates the number of different lexemes (per 1000 words of discourse) used by an informant in borrowing. A *lexeme* is akin to what is referred when it is said that two different forms, such as *give* and *gave*, are instances of the same vocabulary word. More precisely, a lexeme is an abstract group of forms with the same meaning (Lyons 1977: 19). To illustrate, consider examples (32) and (33).

(32) Que se llama el *principal's list*... 201U  
 ‘Which is called the *principal's list*...’

(33) Para hacer eso, yo voy a tratar (...) tratar ahí par.. por.. por, eh.. la posición de *teaching principal*. 329D  
 ‘To do that, I’m going to try (...) try there ff.. for.. for, uh.. the position of *teaching principal*.’

---

<sup>54</sup> For calculating this variables, a word is a string of graphemes separated on either side by a space. It is arrived at by using the word count function of Microsoft Word ®.

In (32), the informant uses one borrowing: *principal's list*. It is composed of two lexemes (i.e. two vocabulary words): *principal's* and *list*. In (164), 329D borrows *teaching principal*. *Teaching principal* is one borrowing, also containing two lexemes: *TEACHING* and *PRINCIPAL*.

Suppose, now, that examples (32) and (33) had been uttered by the same informant. It would be said that the informant used two borrowings: *principal's list* and *teaching principal*. These two borrowings are composed of four words, but only three lexemes: *PRINCIPAL*, *LIST* and *TEACHING*. That is, *principal's* and *principal* are two tokens of the same lexeme. Together they add one to the informant's lexeme count. If examples (32) and (33) were the only borrowings uttered in a 200 word interview, the informant's borrowing rate would be  $\frac{2}{200}$ , or 0.01, or 10 borrowings per 1000 words (written as 10/k). Her lexeme rate would be  $\frac{3}{200}$ , or 15/k.

Forms counted as instances of the same lexeme are nouns inflected for number and possession, regardless of whether the inflection is Spanish or English. As indicated above, *principal's* in example (32) and *principal* in example (33) are two instances of the same lexeme. Forms used as verbs, regardless of tense or aspect inflections and regardless of whether the inflection is English or Spanish, are also instances of the same lexeme. For example, *janguear*, *jangeando* and *hanging* are three instances of the same lexeme. In addition, a shortened form is considered part of the same lexeme as that of its non-shortened version. For example *high* 'high school', instantiates the same lexeme as *high school* or *jai eskul*. But *(la) high* 'high class' represents a distinct lexeme. Similarly, *super* 'superintendent' and *superintendent* are two tokens of the same lexeme *SUPERINTENDENT*. But *super* [su pəɪ] 'very', if used, would instantiate a distinct lexeme.

The grammatical category of a form is somewhat taken into account when counting lexemes (following Lyons 1977: 558). For instance, *man* used as a noun and *man* used as a discourse

marker represent two different lexemes. Likewise *like* used as a discourse marker, verb, preposition, etc., would each be considered instances of different lexemes. Not identified as distinct lexemes are forms used as either nouns or as adjectives. For instance, the form *World Trade Center* in (34), where it is used as a noun, is an instantiation of the same lexeme as *World Trade Center* in (35), where it is an adjective.

(34) Fui al *World Trade Center* como dos veces... 330D  
'I went to the *World Trade Center* about two times'

(35) También la felicité... de haber manejado todo este *World Trade Center thing* muy bien, que fue muy flexible con todos... 324E  
'I also congratulated her... for having managed this whole *World Trade Center thing* very well, cause [she] was very flexible with everyone...'

Examples (34) and (35) show that not all lexemes are instantiated by one-word forms (e.g. as were *PRINCIPAL*, *LIST* and *TEACHING*). In some cases, an English string, although containing several identifiable smaller words of English, is considered to represent a single lexeme (e.g. *WORD TRADE CENTER*).<sup>55</sup> This happens when the string is a proper noun, collocation or discourse marker. Consider example (36).

(36) No, sí es cerca de *Williams Bridge*. 220P  
'No, yes it's close to *Williams[burg] Bridge*.'

In (36), *Williams Bridge* is a proper noun. It is hypothesized that the entire string represents a vocabulary item distinct from either *bridge* or *Williams[burg]*. So, although it contains the words *Williams* and *bridge*, the whole phrase is counted as one lexeme. In general, proper noun phrases, of whatever length, are considered to instantiate a single lexeme. To determine whether

---

55 This characterization of lexemes is in accord with Lyons, who says "Roughly speaking, we can say that lexemes are the words **and phrases** that a dictionary would list under a separate entry" (1977: 23, my emphasis).

other multiple-word strings are lexemes, the dictionary test was used. If a multiple-word string is listed in a dictionary of standard English or if it is in a dictionary of English collocations, it is considered a lexeme. Examples (30) and (31) are repeated as (37) and (38).

(37) Porque trabajaba yo *part-time* en esa época. 315M  
'Because I was working *part-time* at that time.'

(38) Si tú estás en una colina bien alta, tú puedes ver *coast to coast*... 333D  
'If you are on a very high hill, you can see *coast to coast*...'

*Part-time* is a single entry in *Webster's English dictionary* (2001) and *coast to coast* is listed in *Collins Collocation Dictionary Online*. Each string, then, despite containing more than one word, is considered one lexeme. Lexeme rate is not a measure used often in this study, but it provides the basis for corpus frequency rates, which are described next.

## 5.2 Lexical innovation and reproduction measures: Nonshared and Shared rate

Lexical borrowings in this study are coded for their corpus frequency. That is, each lexeme used in an informant's borrowings is coded according to the number of informants that use it. Two corpus frequency measures are examined in this study: shared lexeme rate and nonshared lexeme rate. When a lexeme is used by five or more informants, it is considered a *shared lexeme*. A *nonshared lexeme* is used by just a single informant. The shared lexeme rate (*shared rate*, for short) is calculated by dividing the number of shared lexemes by the total number of lexemes an informant uses. For example, suppose an informant used the lexical borrowings: *high school*, *you know*, *negative attitude* and *big picture*. In these four borrowings, there are six lexemes: *HIGH SCHOOL*, *YOU KNOW*, *ATTITUDE*, *NEGATIVE*, *PICTURE* and *BIG*. Since *YOU KNOW* and *HIGH SCHOOL* are shared lexemes (they are used by more than five speakers), this informant's shared rate would be  $\frac{2}{6}$ , or 0.33, or 33 percent. Nonshared lexeme rate (*nonshared rate* for short) is calculated in the

same way: by dividing the number of nonshared lexemes by the total number of lexemes. Again using the above example, the lexemes *ATTITUDE* and *NEGATIVE* are used by only one informant. Thus, the nonshared rate for this hypothetical informant is also  $\frac{2}{6}$ , or 33 percent. Since some lexemes are not classified as shared or nonshared (e.g. *BIG* and *PICTURE*<sup>56</sup>), an informant's shared and nonshared rate are not, strictly speaking, inverses of each other. These terms are reviewed in Chapter 6, where they are used in analysis.

## 6. Analysis

Measures of lexical borrowing behavior (i.e. lexical borrowing rate, lexeme rate, nonshared lexeme rate, shared lexeme rate and, to be defined in Chapter 7, flagging rate) were calculated for each informant. Exploratory analysis of informants' borrowing rates was done in order to exclude statistical outliers. One-way analyses of variance (ANOVAs) were used to determine the relationship between the dependent measures of lexical borrowing and categorical independent variables like sex and occupational class. Pearson correlations were used to determine the relationship between the dependent measures and continuous independent variables, like age. Linear regression analysis revealed the extent to which predictive variables contributed to the observed variance in lexical borrowing behavior. Finally, to help interpret results, cross-tabulations and Chi<sup>2</sup> analyses were used to determine the relationship between two or more independent variables, such as between arrival age and English skills. For all statistical tests, the probability value was set to  $p < .05$ , although results approaching the probability value cut-off, falling between .05 and .1, are also presented.

---

56 *Big* and *picture* are classified as *periodic* lexemes: they are uttered by at least two (but less than five) speakers.

## Chapter 4: Factors that condition English lexical borrowing frequency

### 1. Introduction

This chapter is the first of four providing evidence on intra-community variation in lexical borrowing behavior in Spanish in New York City. This chapter and the next present the social factors that condition the relative frequency of borrowing. The chapter begins, in section 2, with a brief discussion of how results are interpreted. In section 3, an overview of borrowing in the OZC is presented. The results of a sociovariationist analysis are detailed in sections 4 through 7. Note, however, that in the current chapter the language proficiency variables, English skills and Spanish skills, are not examined.<sup>1</sup> The results of this chapter show that, although nearly everybody does it, borrowing is not done in the same way by all individuals. Different groups borrow to different extents. That is, the extent to which an individual borrows appears to be influenced by factors such as the age of migration, occupational class and daily Spanish use. In particular, results show that those that use Spanish infrequently borrow most. However, the strongest overall predictor of borrowing frequency is immigrant generation, where the second generation borrows, on average, more than twice as often as the first. The marked difference between the first and second generation suggests that each generation represents a distinct subcommunity with respect to borrowing frequency. Results in Chapter 5 further support this by

---

1 Language proficiency variables are not examined in this chapter, where the results of the whole corpus are reported, because the interpretation of skill levels differs depending on who is reporting. For instance, when a native Latin American Spanish speaker (particularly, a teen or adult migrant) reports that his Spanish is ‘good’ or ‘passable’, it is taken as an indication not of competence or proficiency in Spanish, but as an expression of confidence in his Spanish vis-à-vis an idealized or standard version of Spanish (see Chapter 5, section 2.3.4). When, on the other hand, an informant born or raised in the U.S. reports his Spanish skill level, it is interpreted as an indication of his competence in Spanish (see Chapter 5, section 3.3.2). Language proficiency variables are explored in Chapter 5, where variables conditioning borrowing rate are analyzed for the first and second immigrant generation separately.

showing that borrowing in each generation is conditioned by a unique set of conditioning variables.

## **2. The interpretation of social categories**

The appropriate theorization of social categories has, for some time, been one of general concern and interest to sociolinguists (Eckert 2002; Milroy & Gordon 2003: 224; Lavandera 1978: 182). A recent trend in variationist research is to provide local and performance-based justifications for the interpretation given to social categories found to correlate with linguistic variables. These justifications are made possible through detailed ethnographic observation of an individual's in-group and out-group interactions, patterns of dress, places where he spends time, opinions and aspirations. These interpretive practices reconceptualize the role of social categories within sociovariationist theory. Social categories are not apriori givens. Rather, they are local constructions arising (in part) from the allocation of available linguistic devices and the performance of identity by the individuals that use them (Eckert 2002: 4).<sup>2</sup> In keeping with this trend of appropriate theorization of social categories, and in recognizing that any one sociodemographic variable may be simultaneously linked with several cultural, historical and linguistic realities (Eckert 2008), the independent variables of the present study are not assigned a single pre-determined value. Instead, social categories, and in turn the linguistic feature that covaries with them, are interpreted with recourse to concepts from five frameworks:

---

2 For example, in Eckert's (2000) study of burnout and jock talk in a Michigan high school, the meaning of the variable realization of /ay/ within the high school is in part created by the extreme raising of it by individuals such as Judy who, through flamboyant dress, makeup and demeanor, locates herself as a benchmark of burnout practice. In other words, the association of raised /ay/ with burnout identity is created via the social performance of those that use it most, the self-ascribed burnouts.

- language contact (e.g. *intensity of contact, exposure, agentivity*)
- variationist sociolinguistics (e.g. *prestige, change from above, network membership*)
- speech accommodation and social identity theory (e.g. *ethnic identity, in-group/out-group interactions*).
- bilingual language processing and language proficiency status (e.g. *language activation, language attrition, bilingual mode*)
- language ideology (e.g. *indexicality, iconicity, recursivity*)

The suitability of a particular concept for interpretation will be grounded in the present work with respect to five types of information: the constellation of other sociodemographic variables significantly related to the linguistic variable under discussion, the constellation of other non-significantly related sociodemographic variables, corpus-internal demographic trends, census-based demographic trends for Latinos in New York City and examples of borrowings from the OZC. For example, if it were found that sex was significantly related to how much a person borrowed from English, this would not necessarily be interpreted as indicative of gendered speech patterns or a change occurring in the language of the speech community. Sex differences might be better interpreted as differences in the social networks of men and women or perhaps a difference in their exposure to English. The network explanation might be favored, for example, if it were found that, in addition to sex, borrowing frequency varied by occupational class, or that, in New York, Latino men were more often employed outside of the home. The account in terms of exposure to English might be favored if it were found that, in addition to sex, borrowing frequency was correlated with English proficiency (but not class) and that one sex demonstrated higher proficiencies in English than the other. It is in this way, then, through the use of the constellation of significantly correlated variables, non-correlated variables and sociodemographic trends, that the interpretation of social categories is grounded.

### 3. Overview of the corpus

On the whole, lexical borrowing in Spanish discourse is an infrequent occurrence. Because it can be difficult to grasp the idea that someone borrows, for example, 0.7 times per 100 words, borrowing rates are presented as *per 1000 words of interviewee discourse*, indicated by “/k”. Table 4.1 shows the average lexical borrowing rate for the corpus as a whole.

**Table 4.1**  
**Lexical borrowing in the OZC**

<b>Total excluding outliers<sup>3</sup> (n=138)</b>	
<b>Count</b>	3426
<b>Rate</b>	5.5/k

Total participant words: 621,915

In over 620,000 words of interviewee discourse, informants borrowed from English a total of 3426 times, or at a rate of five and a half times per 1000 words of speech. To conceptualize this another way, if the speech of each informant were typed as a separate chapter in the same font as appears on this page and then compiled into one single-spaced document, there would, on average, be five or six English lexical borrowings on each page. Of course, this corpus average of 5.5/k provides only an approximation of lexical borrowing in New York City. If we were to flip through our 622-page<sup>4</sup> document we would notice that, in some chapters, no borrowings occur,<sup>5</sup> while in others, borrowings occur as many as 15 or 20 times per page.<sup>6</sup> Closer

3 This total excludes borrowings made by eight informants whose borrowing rates were statistical outliers. (Statistical outliers were borrowing rates exceeding 20/k.) These eight informants are also excluded from analyses presented in the rest of this chapter and for analyses presented in Chapter 5. For the actual number of borrowings in the LBD, see Table 7.1 in Chapter 7.

4 I estimate that about 1000 words of Spanish fit onto one typed, single-spaced, Times New Roman, 12 font size page (with 1" margins). Thus, 621,915 words would take up about 622 pages.

5 There are seven informants that do not borrow at all in their interviews.

examination of the distribution of borrowings in our document would further reveal that the concentration of borrowings in different portions of the document corresponds to whose chapter we were reading. Said differently, analysis reveals that borrowing rates diverge from the corpus average according to the sociodemographic traits of informants. That is, members of certain Spanish-speaking subcommunities in New York City borrow more (or less) than others. The following sections begin to detail these findings.

#### **4. Immigrant generation, occupational class, level of education and daily Spanish use significantly condition lexical borrowing rate**

Borrowing frequency in the OZC significantly varies with immigrant generation ( $F(1, 136) = 45.88, p < .001$ ), occupational class ( $F(1, 131) = 9.87, p < .002$ ), level of education ( $F(1, 135) = 4.70, p < .032$ ) and daily Spanish use ( $F(1, 135) = 19.07, p < .001$ ). The results of ANOVA for each of these variables follow.

##### **4.1 The effect of arrival age and immigrant generation on borrowing rate**

The most pronounced distinctions in borrowing frequency in the present investigation are tied to the immigrant generation of the speaker. Immigrant generation is determined by a person's arrival age to the U.S. Those that are born in the U.S. or that arrived before the age of three are considered the second generation. Those that arrived at or after age twenty are considered the first generation. In between, there is the one-and-a-half generation: those that arrived in the U.S. anywhere from age four to nineteen. This group is further divided into two subgroups: those that arrived as children (from age four to twelve) and those that arrived as teens (from age thirteen to

---

6 One informant borrows as frequently as 40 times per 1000 words. The present analysis, however, does not include this informant since his borrowing rate (and that of seven other individuals) is a statistical outlier.

nineteen). It is important to remember that a second generation informant is not necessarily younger than a first generation one. The ages of informants in the second generation range from age 12 to 63. Ages of first generation informants range from twenty-three to eighty.

An ANOVA of borrowing rates<sup>7</sup> by arrival age produces significant results (Table 4.2).<sup>8</sup>

**Table 4.2**  
**Lexical borrowing rate by immigrant generation-arrival age, All informants**

	Mean/k	SD
<b>Second generation</b> (n=24) (arrived ≤ 3)	9.1	5.1
<b>Child arrivers</b> (n=17) (arrived age 4-12)	8.6	4.9
<b>Teen arrivers</b> (n=26) (arrived age 13-19)	3.8	3.2
<b>First generation</b> (n=71) (arrived age 20+)	4.0	3.3
<b>Total</b> (n=138)	5.4	4.5
$F(3, 134) = 15.18$ $p < .001$		

As shown in Table 4.2, the second generation borrows most often, around nine times per 1000 words. The first generation borrows only about four times per 1000 words. Child arrivers borrow 8.6 times per 1000 words, while teen arrivers do so about 3.8 times. These results are highly

7 For all analyses in this chapter and subsequent ones, it is the case that where borrowing rate was significantly related to an independent variable, mean *unit rate*, i.e. the number of syntactically independent word-like pieces in a borrowing, was also significant. (Think of a *unit* like a token of a lexeme.) That is, groups that used more borrowings than another group, also used more individual units of English to borrow. Given the way that unit rate and borrowing rate were calculated (over total interview words), this may seem like an obvious result. What it indicates however, is that informants did not diverge greatly from one another in the average number of units they used in a borrowing. Borrowings averaged between 1 and 1.5 units in length (keeping in mind that a unit could be more than one word in the case of proper nouns, discourse markers and tag expressions). This fact, however, does not preclude the possibility of finding that certain speakers preferred single-word borrowings, while others preferred multiple-word borrowings. Single- and multiple-word borrowings are not examined in the present investigation.

8 A Pearson correlation for arrival age (i.e. immigrant generation in its raw form) and borrowing rate is also significant ( $r(136) = -0.41$ ,  $p < .001$ ).

significant, with  $p < .001$ , and the variation in borrowing rates between groups being more than 15 times of that within groups ( $F(3, 134) = 15.18$ ).

From the group means in Table 4.2, it is clear that, in terms of borrowing rate, child arrivers pattern with the second generation, while teen arrivers pattern with the first generation. A Tukey's test (Table 4.3) confirms that, indeed, there are in reality only two factors with respect to lexical borrowing frequency and the generation-arrival age complex: those who arrived at or before age twelve, and those who arrived later.

**Table 4.3**  
**Post hoc analysis of the effect of immigrant generation-arrival age on lexical borrowing rate, All informants**

		Mean diff	p
<b>Second generation</b> (arrived $\leq 3$ )	Child arriver	0.5	.979
	Teen arriver	5.3*	<b>.001</b>
	First generation	5.0*	<b>.001</b>
<b>Child arriver</b> (arrived age 4-12)	Teen arriver	4.9*	<b>.001</b>
	First generation	4.5*	<b>.001</b>
<b>Teen arriver</b> (arrived age 13-19)	First generation	-0.3	.991

Table 4.3 shows that borrowing rates of the second generation differ from those of teen arrivers and the first generation ( $p < .001$  in both cases). On the other hand, the borrowing rates of second generation informants do not differ at all from those of child arrivers ( $p < .979$ ). Additionally, borrowing rates for the first generation are significantly different from those of child arrivers ( $p < .001$ ), but not from teen arrivers ( $p < .991$ ).

A scatterplot of borrowing rates by arrival age provides the graphic representation of these facts (Figure 4.1).

**Figure 4.1**  
**Lexical borrowing rate by arrival age, All informants (outliers excluded)**

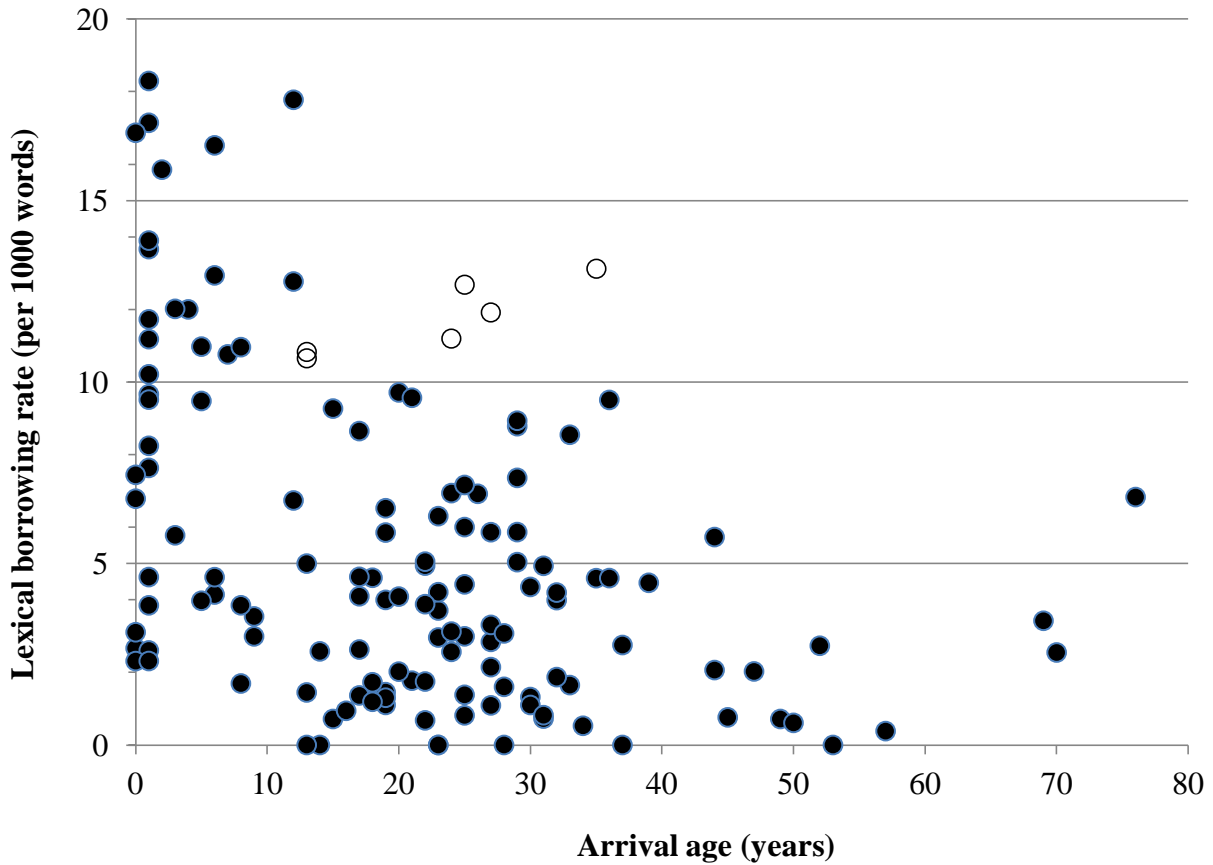


Figure 4.1 shows that as arrival age increases, the borrowing decreases. In particular, those born in the U.S. or who came as infants (dots on or near the y-axis) borrow as little as two times per 1000 words, or as much as 18 times. After arrival age twelve, almost all speakers never borrow more than 10 times per 1000 words, and there are a handful ( $n=7$ )<sup>9</sup> that never do so at all. The

<sup>9</sup> There appears to be only six markers at  $y=0$  because two informants that do not borrow at all have the same arrival age of twenty-three years old.

figure further shows that those arriving to the U.S. from age four to twelve clearly pattern with those that arrived prior to age three or who were born in the U.S. That is, child arrivers, like the second generation, all use some borrowings (captured in that there are no informants between age four and twelve with a borrowing rate of 0/k) and may use up to as many as 18/k. Those that arrive from age thirteen on (n=97) all tend to borrow from English no more than 10 times per 1000 words of discourse. (The exceptions to this are depicted by hollow markers.) Furthermore, it can be seen from Figure 4.1 that those arriving after age 40, all tend to borrow five times or less per 1000 spoken words. However, it has not been possible to identify these informants as an independent arrival age group on statistical grounds due to how few informants fall into this age range (n=12). To summarize, child arrivers are similar to the U.S.-born in that both groups are more likely to borrow at rates exceeding 10/k and show the same degree of within-group variability in their borrowing rates. Teen arrivers pattern with the second generation in that both groups are more internally homogenous and rarely borrow from English more that 10 times per 1000 words. Occasionally teen and adult arrivers borrow nothing, but, on average, they borrow at a rate of about 4/k. Given these findings, another ANOVA was performed using a two-factored immigrant generation-arrival age variable (Table 4.4).

**Table 4.4**  
**Lexical borrowing rate by immigrant generation, All informants (revised)**

	Mean/k	SD
<b>2<sup>nd</sup> generation + Child arrivers</b> (n=41)	8.8	5.0
<b>1<sup>st</sup> generation + Teen arrivers</b> (n=97)	4.0	3.3
<b>Total</b> (n=138)	5.4	4.5

$$F(1, 136) = 45.88 \quad p < .001$$

Table 4.4 shows that, when combined, second generation and child arrivers together average about nine borrowings per 1000 words. The first generation and teen arrivers, on the other hand, borrow at less than half that rate: only 4/k. That this two-way arrival age division is a powerful one is shown by the enormous F value ( $F(1, 136) = 45.88$ ). Given these findings, I will, for the sake of expositional convenience, use the term *second (immigrant) generation* to refer to those both that were born in the U.S. and that arrived up to and including age twelve. The term *first (immigrant) generation* designates those who arrived to the U.S. after age twelve.<sup>10</sup>

#### **4.2 The effect of occupational class on borrowing rate**

Results of analysis in the present study show that occupational class significantly conditions lexical borrowing frequency ( $p < .002$ ). In this study, occupational class is a self-ascribed variable. That is, informants were asked whether they considered themselves to be members of the high, middle or working class. Informants were also classified according to a more objective measure of socioeconomic status (SES), which was based on level of income and parents' level of education (see Appendix A). However, self-ascribed class produced more robust results with respect to borrowing rate than did SES.<sup>11</sup> Only results of the ANOVA for (self-ascribed) occupational are presented (Table 4.5).

---

10 The individuals comprising the first and second generation in this study are different from those comprising the first and second generation in Otheguy, Zentella & Livert (2007) and in Otheguy & Zentella (2012), whose corpus was utilized for this study. The basic difference lies in whether 19 individuals that arrived from ages four to twelve are considered to be first or second generation. For Otheguy and Zentella they are first generation, in the present chapter and in Chapter 5, their lexical borrowing behavior aligns them with the second immigrant generation.

11 ANOVA results for SES in the corpus as a whole were  $F(1, 133) = 8.07, p < .005$ . Self-ascribed occupational class and SES are correlated with each other ( $Chi^2(3, N=138) = 22.94, p < .001$ ). See Appendix E, Table E.1.

**Table 4.5**  
**Lexical borrowing rate by occupational class, All informants**

	Mean/k	SD
<b>Middle class</b> (n=69)	6.5	4.4
<b>Working class</b> (n=64)	4.2	4.2
<b>Total</b> (n=133)	5.4	4.4
<i>F</i> (1, 131) = 9.87		<i>p</i> < .002

As seen in Table 4.5, middle class Spanish speakers on average borrow from English over six times per 1000 words of discourse, while working class speakers do so about four times. This class-based difference in borrowing rate is robust, as demonstrated by the F value, 9.87.<sup>12</sup>

### 4.3 The effect of level of education on borrowing rate

Education level also significantly influences borrowing rate ( $p < .032$ ) (Table 4.6).

**Table 4.6**  
**Lexical borrowing rate by level of education, All informants**

	Mean/k	SD
<b>College or graduate</b> (n=114)	6.2	4.5
<b>High school or less</b> (n=23)	4.6	4.1
<b>Total</b> (n=137)	5.5	4.4
<i>F</i> (1, 135) = 4.70		<i>p</i> < .032

As Table 4.6 shows, those with less than a high school education borrow least, averaging about four and a half borrowings per 1000 words. Those with a college education borrow more often,

<sup>12</sup> Data in Chapter 5 show that the significance of occupation class for borrowing rates in corpus overall is probably due to the significance of class for the first, but not second, immigrant generation. In other words, class does not significantly condition borrowing in the second generation. However, results for class in the second generation do show that the overall trend still obtains. That is, middle class second generation informants borrow more ( $M=12.0/k$ ) than the working class ( $M=9.2/k$ ).

averaging six borrowings. The finding is fairly robust, as indicated by an F value of almost five ( $F(1, 135) = 4.70$ ).<sup>13</sup>

#### 4.4 The effect of daily Spanish use on borrowing rate

The extent to which informants use Spanish on a daily basis also significantly conditions borrowing rates in the OZC ( $p < .001$ ). ANOVA results for daily Spanish use are in Table 4.7.

**Table 4.7**  
**Lexical borrowing rate by daily Spanish use, All informants**

	Mean/k	SD
<b>Infrequent Spanish use</b> (n=48)	7.6	5.2
<b>Regular Spanish use</b> (n=89)	4.3	3.5
<b>Total</b> (n=137)	5.5	4.4
$F(1, 135) = 19.07$ $p < .001$		

Table 4.7 shows that those that use Spanish on a daily basis borrow from English less often (4.3/k) than those that rarely or never use it (7.6/k). The distinction in borrowing frequency between those that use Spanish infrequently and those that use it regularly everyday is well-defined as indicated by the F value of 19.

13 As will be seen in Chapter 5, the effect of education on borrowing rates operates primarily within the first generation. The effect of education on the corpus as a whole is due, it seems, to the fact that first generation informants account for a large proportion of corpus informants (of 146 total informants, first generation = 99). However, results for education in the second generation do show that, although not significant ( $F(1, 45) = 0.86$ ,  $p < .359$ ), the overall trend still obtains. Those with a college or graduate education borrow more ( $M = 11.4/k$ ) than those who have not been formally educated beyond high school ( $M = 9.4/k$ ).

#### **4.5 Summary of the effect of immigrant generation, class, education and daily Spanish use on borrowing rate<sup>14</sup>**

The results of ANOVA for borrowing rates have shown that immigrant generation, education, class and daily Spanish use all influence the extent to which Spanish speakers borrow. In particular, those of the second generation borrow more than the first generation; the middle class borrows more than the working class; those with more years of formal education borrow more and; those that use Spanish less frequently on a daily basis borrow more. In this chapter, only results for immigrant generation and daily Spanish use will be discussed. This is because the effect of occupational class and education appears to function primarily within the first generation. Its apparent effect on borrowing in the corpus as a whole is probably due to the fact that the first generation (n=98) account for 67 percent informants (of n<sup>total</sup>=146). The interpretations of class and education are given in Chapter 5 (sections 2.3.1 and 2.3.2).

### **5. Discussion of immigrant generation and daily Spanish use as variables that condition borrowing rate**

#### **5.1 Interpreting the effect of immigrant generation on borrowing rate**

Immigrant generation has been found to be a central determinant for both grammatical and lexical aspects of language contact. Both Weinreich (1966: 94) and Haugen (1972: 88) noted the importance of immigrant generation in determining the quantity and phonetic shape of foreign

---

14 In summary and conclusion sections, results will be reported as generalizable to the population of Spanish speakers in New York. Newton and Rudestam (1999: 53-57) note that for results of inferential statistical procedures to be generalizable to a larger population, the sample must have been randomly generated. As described in Chapter 3 (section 1), sampling for the present investigation was done using a snowballing technique. This may arguably limit the extent to which results reported herein are generalizable to the Spanish-speaking population in New York at large. Nevertheless, the use of a snowballing technique for sampling purposes does not necessarily obviate the ability to obtain valid insight into a larger population. Several linguistic studies employing a snowballing technique (and stratified sampling, such as used in the OZC) have obtained valuable results (Labov 2001: 38-39 and footnotes 9 and 10). Furthermore, the nature of the results obtained in this study (particularly those for immigrant generation and Spanish language use) suggest they are probably accurate reflections of patterns of lexical borrowing for the greater Spanish-speaking population of New York City. So discussion proceeds in terms of New York City-wide generalizations.

lexical material in a recipient language. More recently, Otheguy and Zentella (2012: 88) show with respect to overt pronoun usage that those raised in the U.S. use more overt pronouns than those raised in Latin America. While not the main thrust of her 2002 work, Silva-Corvalán (2002: 169) addresses some lexical influences from English and finds that second and third generation Spanish speakers in Los Angeles tend to use different English-influenced lexico-semantic formations in their Spanish than do the first generation.<sup>15</sup> While all of her generations used single-word loans (171), only the second and third generation used lexico-syntactic calques (186). Likewise, Otheguy, García and Fernandez (1989) found that second generation Cubans in West New York used both loanwords and semantic calques to a greater extent than first generation Cubans.

Given that immigrant generation is closely tied to proficiency in the donor language (English),<sup>16,17</sup> and the long-recognized axiom that bilinguals are “the locus of language contact” (Weinreich 1953/1966: 1), these results are perhaps unsurprising. The more bilingual second generation borrows more often than the first generation. Yet these findings do more than validate the obvious. They provide additional empirical support, on the level of speech, for theoretical models that predict that the more *intense* the contact (see Thomason & Kaufman 1988: 72), the greater the number of contact features that will be evidenced in a language. With regard to the present study, however, the notion of intensity of contact refers not just to the relative

---

15 Silva-Corvalán (1994) does not use the term *immigrant generation*, but simply designates a ‘group 1’, including informants born in Mexico, a ‘group 2’, for those born in the U.S., and a ‘group 3’ for those born in the U.S. with at least one parent who would be classified as a group 2 member. She does not state why she divides her informants by ‘groups’ instead of immigrant generations. However, I will note that the term *immigrant generation* may not be felicitous in the Southwest where the family of many Spanish speakers did not migrate to the U.S., but simply found themselves within U.S. borders after Mexico’s 1848 cessation of Arizona, California, New Mexico and Colorado to the U.S.

16 In the OZC, only 37 percent (n=36) of first generation informants rate themselves as ‘good’ or ‘excellent’ in English; of the second generation, 90 percent (n=43) do.

17 In the OZC, only 38 percent (n=37) of first generation informants rate themselves as ‘good’ or ‘excellent’ in **both** English and Spanish; the second generation, 65 percent (n=31) do.

socioeconomic prestige or power of donor and recipient languages or the length of contact between groups, but to a group's overall competence in the donor language or their degree of proficiency in both languages.

## **5.2 Interpreting the effect of daily Spanish use on borrowing rate**

Results showed that those that use Spanish on a daily basis borrow at a rate of 4/k, while those that use Spanish less often borrow at a rate of 7/k. This fact could be described in one of two ways. It may be said that regular use of Spanish is a **deterrent** to lexical borrowing. Or else it may be said that not using Spanish often on a daily basis **stimulates** or **encourages** lexical borrowing. The latter of these descriptions is preferable. This is because the mean borrowing rate of those that **do** use Spanish regularly (4/k) is closer to the corpus average (5.5/k) than that of the infrequent Spanish users (7.6/k). That is, what appears to be occurring is that, while using Spanish daily does not seem to stop individuals from borrowing (their rate of borrowing is close to average), non-regular use of Spanish leads to an increased tendency to borrow from English.

These facts could be accounted for in several ways. It may be that, in immigrant language contexts, lack of use of Spanish leads to *attrition* of Spanish, causing individuals to more often rely on English words and phrases to express themselves. An alternative account rests on the idea of *language activation* (Paradis 1997: 342-343). In New York, lesser use of Spanish on a daily basis implies more use English. In other words, it may be that those that do not use Spanish daily borrow more because English is more activated for them, resulting in quicker access to English phrases, of which they are able to avail themselves when they know their interlocutor is bilingual (as was the case in for our OZC interviewers).

An interpretation of findings for daily Spanish use that appeals to language attrition could be supported if it were found, for instance, that those that use Spanish infrequently struggle more to

remember a Spanish word or phrase than those that use Spanish frequently. Particularly, discovering that infrequent Spanish users flagged their borrowings to a greater extent could support an interpretation of daily Spanish use as language attrition.<sup>18</sup> Additionally, an attrition interpretation may be supported if it were found that those that use Spanish less have a greater tendency to restate their English borrowings in Spanish. In other words, if individuals that use Spanish infrequently borrow more because they are forgetting or have forgotten how to render something in Spanish, they might also be more inclined to restate an English borrowing in Spanish, after the use of an English borrowing has bought them time to recover a comparable Spanish equivalent. An analysis of flagging and recasting among informants, however, shows that neither the flagging nor the recasting of borrowings in Spanish occurs more (nor less) for those that use Spanish infrequently. Both those that use Spanish regularly and those that do not flag borrowings to the same extent ( $M=12\%$ ,  $F(1, 136) = 0.39$ ,  $p<.535$ ).<sup>19</sup> And both groups also recast their borrowings in Spanish to the same extent ( $M=6\%$ ,  $F(1, 137) = 0.17$ ,  $p<.683$ ).

On the other hand, daily Spanish use could be interpreted as an indication that English is more available than Spanish for those that use Spanish less on a daily basis. This position could be supported if it were found, for instance, that those that use Spanish less frequently opt to borrow from English even when referring to concepts that have a Spanish translation equivalent that is in common use in New York. At the moment, an exacting analysis of this type has not been carried out. However, perusal of the nonshared vocabulary of infrequent Spanish users suggests that a language activation account alone may not be a sufficient interpretation of the

---

18 Results of Chapter 7 suggest that flagging in the OZC indicates that informants are aware that something uttered is not part of Spanish. Furthermore, no support could be found for the idea that flagging indicates a lack / loss of proficiency in Spanish.

19 The analysis of flagging in Chapter 7 does show that flagging differs based on daily Spanish use. However, the difference is not between those that use Spanish infrequently and those that use Spanish regularly, the factors for which an interpretation is sought in the current discussion. Rather, flagging rate varies significantly between those that use Spanish, even a little on a daily basis, and those that **do not use Spanish at all** in their daily lives.

daily Spanish use variable. Although infrequent Spanish users certainly borrow even when naming things that are easily and typically rendered in Spanish (e.g. *sixth grade, swimming, t-shirt, turkeys, stain*), their nonshared borrowings<sup>20</sup> are dominated (more than among regular Spanish users) by phrases that may be unfamiliar to these speakers in Spanish or else that are more unwieldy to render in Spanish than in English. For example, the borrowing inventory of infrequent Spanish users includes: *Amish people, annual report, anesthesiologist, award night, beginner freshman, blackbooks, boardwalk, book signing, daycare center, food runner, food stamps, French shutters, moldings, mission statement, out of business, open-minded, paperwork, profit sharing, snowflake, spring break, steak house, track stars, trump card* and *word processing*. Many of the items in fact are collocations in English or else designate concepts and index situations that could be perceived as specifically pertaining to the New York City sociocultural context. It seems, then, that, although an account cannot not yet be definitive, daily Spanish use may indicate the extent to which speakers are involved with U.S. institutions or oriented toward Anglo cultural phenomena. That is, the greater use of borrowings by infrequent Spanish users demonstrates the greater degree of involvement that these speakers have in U.S. life. Of course, this interpretation is not necessarily inconsistent with interpretations that appeal to language attrition or language activation.

## **6. Immigrant generation and occupational class best predict borrowing rate in Spanish in New York City**

Lexical borrowing frequency in Spanish in New York is conditioned by an individual's immigrant generation, his class membership, his level of education and the extent to which he

---

20 I examine the stock of nonshared borrowings because this is the stock that, by definition, informants of the OZC do not share among each other, which, as a result, can provide insight into the reasons why infrequent and frequent Spanish users differ from one another in their borrowing frequency.

uses Spanish on a daily basis. In order to ascertain which of these four factors best accounts for differences in lexical borrowing frequency, these variables were entered into a multiple linear regression analysis.

In addition to meeting typical assumptions for carrying out parametric tests (Newton & Rudestam 1999: Chapter 6),<sup>21</sup> a valid regression must also avoid the problem of colinearity, that is, the existence of a high degree of correlation between two independent variables (Chen, Ender, Mitchel & Wells 2003: Chapter 2). When colinearity between independent variables in a regression exists, the coefficients for those variables become unstable and the standard error for the coefficients becomes inflated, thus increasing the chances that a real statistical relationship present in the data fails to be revealed.<sup>22</sup> Fortunately, statistical procedures allow the investigator to determine the degree of colinearity that exists between the independent variables (i.e. predictor variables) and thus make a determination as to whether the colinearity may be too high. In a linear regression, the investigator may request the *tolerance level* for each variable. *Tolerance*, a value between 0 and 1, is “an indication of the percent of variance in the predictor that cannot be accounted for by the other predictors” (Chen et al. 2003: Chapter 2). A low tolerance level for an independent variable (i.e. it is close to zero) indicates that only a small percentage of the variance in that variable cannot be accounted for by variance in other variables.

---

21 Newton and Rudestam note that linear regression requires that independent variables entered into the regression both be continuously distributed (1999: 248) and demonstrate a linear effect on the dependent variable (i.e. lexical borrowing rate) (1999: 128). Level of education and daily Spanish use are continuous variables that only appear as ordinal measurements at the manifest level (Newton & Rudestam 1999: 182). But the variables of immigrant generation and occupational class are categorical. In order to simultaneously assess the relative contribution of each of these variables on lexical borrowing rate, dummy coding has been used (1999: 187) for immigrant generation and occupational class to approximate an informant’s having “more” or “less” immigrant status, on the one hand, or class standing, on the other.

22 The standard error is used to predict the closeness of the sample to the population (also referred to as *standard deviation of the sampling distribution* or *random sampling error*) and is part of what determines the probability level for the regression. A higher standard error would result in a smaller t value for a given level of degrees of freedom, and could potentially mean the failing to report the existence of a significant relationship (a Type II error).

Hence, the smaller the tolerance for a variable, the more redundant that variable is with respect to other variables in the model. Chen et al. (2003) suggest that tolerances less than .10 indicate that there is too much redundancy in a variable of a regression model and, thus, that the results of a regression have likely been compromised. In the present study, a more stringent criterion is adopted. Only regression models wherein the tolerance values of independent variables are no less than .60 are reported. That is, for a regression model to be reported in the present investigation, 60 percent of the variance in the variables entered into the model must be unaccounted for by the variance in other variables of the model.

Results from the regression including immigrant generation, class, education and daily Spanish use are displayed in Table 4.8.

**Table 4.8**  
**The ranked effects of immigrant generation, occupational class, level of education and daily Spanish use on lexical borrowing rates in Spanish in New York City, All informants**

	Standardized beta coefficient	p	Tolerance
<b>Immigrant generation</b>	.43*	<b>.001</b>	.81
<b>Occupational class</b>	.19*	<b>.016</b>	.87
<b>Daily Spanish use</b>	- .13	.125	.80
<b>Level of education</b>	.08	.287	.86
$R^2 = .33$ $F(4, 127) = 15.25$ $p < .001$			

Table 4.8 shows that the tolerance values of each of the variables entered into the regression are .80 or higher. That is, more than 80 percent of the variance in the variables of the model is independent of the variance accounted for by the other variables. Table 4.8 also shows that a model including the conditioning variables of immigrant generation, class, education and daily Spanish use significantly accounts for one-third of the variance in lexical borrowing rate in the

OZC corpus ( $R^2=.33$ ,  $p<.001$ ). Table 4.8 further reveals that of these four variables, immigrant generation ( $\beta=.43$ ,  $p<.001$ ) and social class ( $\beta=.19$ ,  $p<.016$ ) together best account for differences in borrowing rate among informants. The importance of immigrant generation, as revealed by its large beta value of .43, suggests that the first generation and the second comprise two distinct subgroups of Spanish speakers with respect to borrowing frequency. As a result, borrowing trends within these two groups will be examined separately in Chapter 5.

## **7. Variables that do not significantly condition borrowing rate in the Corpus of Spanish in New York City: Regional origin, sex, age and years in the U.S.**

### **7.1 Ethnonational affiliation and regional origin as variables that do not significantly condition borrowing rate**

Spanish-speakers in New York are aware of stereotypes about the characteristics of regional Spanishes. Comments made by informants in their interviews reveal that the locus of this variation is perceived to be not just in phonological realization, but also in lexical choices and in the use of English while speaking Spanish (see Chapter 5, section 3.3.1 for examples). Prior to analysis, it was hypothesized that Puerto Ricans' relatively longer and more intense history of contact with Anglo culture of the U.S. would result in a situation where Puerto Ricans would have an established pattern of borrowing from English that would differ from that of other Spanish speakers. In particular, the fact that Puerto Ricans (i) are the only Spanish-speaking migrants that have citizen's rights, (ii) are the largest group of Spanish speakers in New York, (iii) have been present in substantial numbers in New York the longest and (iv) have a greater degree of exposure to English in the home country led to the hypothesis that Puerto Ricans would borrow more than other Spanish-speaking groups. Likewise, due to U.S.'s position as a host country for Cuban refugees over several decades, and as a correlate to their relatively more successful economic and political integration into U.S. life (Otheguy, García & Roca 2000: 17-

170), it was anticipated that Cubans would borrow more from English than other groups. Finally, the confluence of these two facts, it was thought would, in turn, lead to finding that Caribbean speakers would borrow more than individuals with ties to the Latin American mainland.

Results for the corpus as a whole, however, have not supported these hypotheses. Ethnonational affiliation (a six-factored variable) failed to significantly correlate with lexical borrowing rates in the corpus as a whole ( $F(5, 132) = 1.28, p < .276$ ), as did regional origin (rate  $F(1, 136) = 0.99, p < .322$ ). In other words, regardless both of historically longer and more involved relations between the U.S. and Puerto Rico, on the one hand, and Cuban migrants, on the other, and contrary to popular opinions that differences in ethnonational affiliation are played out on the stage of lexical borrowing, it appears that the hypothesis that speakers of particular regional varieties of Spanish would more freely utilize English while speaking Spanish cannot thus far be supported. However, results in Chapter 5 (sections 3.2 and 3.3.1) indicate that ethnic or regional differences do condition borrowing frequency in the second generation.

Besides failing to support anticipated results, these findings also superficially appear to be at a variance with Mendieta's (1999) investigation of English loanwords in U.S. Spanish. Mendieta (1999: 91) found that Puerto Ricans used English borrowings during their sociolinguistic interviews about two times as often as the Mexican informants and three and a half times as often as Cuban informants. Put another way, Puerto Ricans were responsible for more than half the borrowing tokens in her corpus, while Mexicans provided about 26 percent and Cubans 17 percent of the borrowings. However, it should be emphasized that Mendieta's results were obtained from ethnolinguistic groups from different parts of the U.S. Her Puerto Rican informants were New Yorkers, her Cubans were from Florida, and her Mexicans lived in Chicago. Thus, her results may not so much reflect an ethnolinguistic division of speakers, but

regional variation in the extent to which English is used in different Spanish-speaking communities in the U.S.

## 7.2 Sex as a variable that does not significantly condition borrowing rate

Sex ( $n^{\text{males}}=70$ ,  $n^{\text{females}}=76$ ) was also anticipated to be a dimension along which borrowing rates would vary. In Poplack's (1980) study of codeswitching among Harlem-residing Puerto Ricans, women were found to use intrasentential codeswitches (some of that data appear to be similar to data here called *lexical borrowing*) more than men. Yet in the present study, sex is not significantly correlated with borrowing rate ( $F(1, 136) = 0.56$ ,  $p < .442$ ). Men and women both averaged between five and six borrowings per 1000 words of discourse. The reason for the difference in findings may be because of how the dependent variable was calculated in Poplack's study. The rate of use of intrasentential codeswitches in her study was calculated as a proportion of all codeswitches, both intrasentential and intersentential. And, her category of intersentential switches includes sentence-length foreign-origin strings, strings that have been excluded from the current investigation. In other words, her findings report the ratio between what have, in the present investigation, been called borrowings and codeswitches. Reinterpreting her findings in the terminology of the present study, then, it might be said that women preferred borrowing to codeswitching, and men preferred codeswitching to borrowing. A comparative analysis of borrowing and codeswitching in the present study has not been carried out. However, preliminary calculations of codeswitches in the OZC suggests that findings would still conflict with those of Poplack (1980). Of all the English strings used by women, about 77 percent are borrowings. But of all English strings uttered by men, about 87 percent are borrowings. In other words, both male and female informants of the OZC seem to prefer as borrowing (as here defined) to codeswitching.

This negative finding with respect to sex corroborates the findings of other lexical borrowing studies. Neither Mendieta (1999: 96-8) nor Eslami Rasekh et al. (2008) found any definitive correlation between an informant's sex and the amount of a donor language used in discourse. And, while Poplack et al. (1988) did find that French-speaking women in one neighborhood (Hull) used fewer English loanwords than did men from the same neighborhood, this finding was reinterpreted as a neighborhood effect,<sup>23</sup> and not a reflection of gendered language. Although biological sex and gender are not isomorphically mapped to one another (Milroy & Gordon 2003: 100), the failure to find a significant difference in average borrowing rates of men and women suggests that whatever gendered speech patterns may be present among Spanish speakers in New York City, they are not reflected in the frequency with which borrowings are employed.

### **7.3 Age as a variable that does not significantly condition borrowing rate**

In sociolinguistic investigations of lexical borrowing, age often emerges as a variable that conditions borrowing frequency (Eslami Rasekh et al. 2008; Matus-Mendoza 2002; Ngom 2006).<sup>24</sup> In the places where these studies were carried out, the donor language is associated with modernity. Older individuals borrowed less than the young, presumably because they may have been focussed on maintaining tradition and culture and, thus, avoided using the donor language. Age was **not** anticipated to be significantly correlated with borrowing frequency in the present

---

23 The sex difference was interpreted as a neighborhood effect due to the fact that (i) there were no sex differences present in any of the other three neighborhoods they studied and (ii) the Hull men's higher loanword use approximated elevated rates of loanword use characteristic of an adjacent Ottawa neighborhood, where over half of the Hull men worked.

24 For instance, Ngom (2006) found that younger and older age groups differed in the extent to which they used Arabic and English (but not French) loan words in their Wolof speech. Older people used more Arabic loanwords than the young, and younger people used more English than the older people. He explains: "... while Arabic loans are used as markers of group membership or religious erudition among older people, English words are used as markers of membership to the trendy urban youth" (2006: 98). Likewise Eslami Rasekh et al. (2008) found that their younger group of Turkman speakers in Iran borrowed more words from Persian into their Turkman than the older group. They explained this fact by saying that older speakers wanted to keep their Turkman identity more than younger speakers.

investigation. This is because, in New York, the donor language, English, is strongly associated with economic advantage (as evidence by metalinguistic comments by informants in the OZC<sup>25</sup>) among both older and younger Spanish speakers. Furthermore, younger Spanish speakers, whether foreign-born or born in the U.S., are, more often than older individuals, bilingual in English.<sup>26</sup> They, therefore, need not borrow from English to connect themselves with modernity or demonstrate their potential for economic success, but merely switch to speaking it.

In accord with anticipated results, analysis showed that borrowing frequency among Spanish speakers in New York is not conditioned by the age of the informant. At first blush, a Pearson correlation ( $r(136) = -.26, p < .002$ ) and ANOVA ( $F(3, 134) = 3.21, p < .025$ ) appeared to indicate that the age of the informant was inversely correlated to informants' borrowing rate, such that as age increased, borrowing decreased. However, further examination of the data revealed that the effect for age was only apparent. The apparent significance of age is due to a sampling imbalance. Concentrated among the young of the corpus, especially among teens, are individuals

---

25 As just one example of comments about the relationship between knowing English and economic advancement, informant 325E says:

No pensaba que iba a hacer tanta la presión [de conocer el inglés], y nunca pensé en ver a tantas personas que venían acá y se daban cuenta de que lo que primero que tenían que hacer es aprender el idioma porque sino, sienten muchas frustraciones o se sienten atados o todo lo que saben.. Si eres un profesional que viene a este país y no sabe el idioma no importa [RISA] no importa que hayas sino presidente de tu compañía allá que aquí eso no importa, pero si no hablas [el inglés], nadie, ósea no eres nadie...

'I never thought that the pressure (to know English) would be so much, and I never thought I would see so many people that come here and realize that the first thing they needed to do is learn the language because if not, they experience many frustrations or they feel stuck, or everything they know... If you're a professional who comes to this country and you don't know the language, it doesn't matter [LAUGHTER], it doesn't matter if you've been president of your own company over there because here that doesn't matter, but if you don't speak [English], nobody, that is, you're nobody...'

26 Of course, the vast majority (93 percent) of U.S.-born Latinos in New York City are fluent in English or speak it 'well' or 'very well' (U.S. Census 2010). Among the foreign-born, a larger proportion of young speakers (age twelve to thirty-four) are fluent in English (57 percent) than those age thirty-five or older (47 percent) (U.S. Census 2010).

of the second generation,<sup>27</sup> which is a distinction highly predictive of lexical borrowing behavior (see sections 4.1 and 6 of this chapter). When the second generation<sup>28</sup> is removed from the sample, the effect of age on borrowing rate disappears, both in the Pearson's ( $r(95) = -.13$ ,  $p < .211$ ) and in the ANOVA ( $F(3, 93) = 0.96$ ,  $p < .414$ ). The effect of age on borrowing frequency in the OZC, then, is only apparent; it is a product of the distribution of immigrant generations among age cohorts in the sample.<sup>29</sup>

#### **7.4 Years in the U.S. as a variable that does not significantly condition borrowing rate**

Years in the U.S., even though it is a non-significant variable in a statistical sense, merits special attention. It was hypothesized that the longer that informants were in New York, the more often they would borrow from English. This was based on the belief that increased time in the U.S. would mean increased exposure to English and, therefore, a corresponding tendency to draw on English while speaking Spanish.<sup>30</sup> In fact, when questioned about who is most likely to “mix” English with Spanish, informants of the OZC sometimes indicated that they believed that long-time residents were the most prone (see Chapter 5, section 2.5.3 for an example of one such comment). Results of analysis show that borrowing rates initially appear to be correlated to the

---

27 See stratification of sample by age and immigrant generation in Appendix E (Table E.2).

28 See section 4.1 for why child arrivers are grouped with the New York born as the *second generation*.

29 That it is immigrant generation producing an apparent effect of age on borrowing rate (and not the reverse) is confirmed by the fact that when younger informants are removed from the sample (age 25 or less), immigrant generation significantly conditions borrowing rate ( $F(1, 94) = 25.35$ ,  $p < .001$ ). However, and as discussed in section 7.3, the opposite is not true; when the second generation is removed from the sample, there is no effect of age on borrowing rate (ANOVA:  $F(3, 93) = 0.96$ ,  $p < .414$ ; Pearson:  $r(95) = -.13$ ,  $p < .211$ )

30 Although there is no direct measure of *exposure* to English, results of a Pearson correlation show that there is in fact a trend in the corpus for informants' *proficiencias* in English to increase with time spent in the U.S. ( $r(145) = 0.32$ ,  $p < .001$ ). However, this trend disappears, as does the apparent correlation of years in the U.S. with borrowing rate, when second generation informants are removed from the sample ( $r(97) = -0.10$ ,  $p < .313$ ). In other words, for first generation informants, longer residence in the U.S. does not appear to lead to better skills in English. Nonetheless, I maintain, as a matter of common sense, that cumulative exposure to English certainly must increase over the course of a person's life in the U.S.

years an informant has spent in the U.S. ( $p < .044$ ), such that the longer he is in the U.S., the more likely he is to utilize English borrowings (Table 4.9).<sup>31</sup>

**Table 4.9**  
**Lexical borrowing rate by years in U.S., All informants**

	Mean	SD
<b>Recent</b> (n=17) (0-2 yrs in U.S.)	4.1	3.6
<b>Long</b> (n=56) (3-15 yrs in U.S.)	4.6	3.9
<b>Established</b> (n=65) (16+ yrs in U.S.)	6.4	4.9
<b>Total</b> (n=138)	5.4	4.5
	$F(2, 135) = 3.20$	$p < .044$

Table 4.9 appears to indicate that, as individuals spend greater amounts of time in the U.S., the amount that they borrow from English increases. However, as with age, when the second generation is removed from the sample (n=41), the effect of years in the U.S. on borrowing rate for the first generation (i.e. arrival after age twelve) disappears for both the Pearson's ( $r(95) = -.07, p < .505$ ) and the ANOVA ( $F(2, 94) = 0.32, p < .730$ ) (Table 4.10).<sup>32</sup>

31 Similar results were obtained from a Pearson's test ( $r(136) = 0.21, p < .012$ ).

32 It appears to be the case that immigrant generation produces an apparent effect for years in the U.S. and not that years in the U.S. produces an apparent effect for immigrant generation. This is supported by the finding that when established immigrants (years in the U.S. for 16+ years) are removed from the sample, immigrant generation significantly conditions borrowing rate ( $F(1, 71) = 10.42, p < .002$ ). (In this case, however, there are only six informants in the second generation compared to 67 in the first.) As presented in the exposition, the reverse is not true; when the second generation is removed from the sample, years in the U.S. does not significantly condition borrowing rate (ANOVA  $F(2, 94) = 0.32, p < .730$ ; Pearson  $r(95) = -0.07, p < .505$ ).

**Table 4.10**  
**Lexical borrowing rate by years in U.S., First generation**<sup>33</sup>

	Mean/k	SD
<b>Recent</b> (n=16) (0-2 yrs in U.S.)	4.3	3.6
<b>Long</b> (n=50) (3-15 yrs in U.S.)	4.1	3.3
<b>Established</b> (n=31) (16+ yrs in U.S.)	3.6	3.0
<b>Total</b> (n=97)	4.0	3.3
$F(2, 94) = 0.32$	$p < .730$	

Table 4.10 shows that, for the Latin American immigrants of the OZC, lexical borrowing neither increases or decreases with time spent in the U.S. ( $p < .730$ ). Rather, Latin American immigrants, from those that have only just arrived in the U.S. to those that have been in the U.S. for over fifteen years, are a fairly homogenous group with respect to how often they take from English while speaking Spanish. In other words, borrowing from English requires very little, perhaps no, exposure to English. What may be happening is that, upon arrival, immigrants acquire a set of borrowings that are proper to the New York City context, just as any Latin American, Spanish-speaking traveler to Spain would quickly acquire a local vocabulary to facilitate communication in their new environment.<sup>34</sup> Whereas the Puerto Rican Spanish speaker in Spain would adopt *naranja* for his native *china* ‘orange’ or *coche* for his native *carro* ‘car’, a Spanish-speaking migrant to the U.S. might simply be adopting *subway* in place of *metro* and *buildin* for his native

33 Other divisions of years in the U.S. factor groups were explored. None resulted in significant differences once the second generation was removed from the sample.

34 This finding and its interpretation are consonant with Silva-Corvalán’s observation with respect to both lexicosyntactic calquing and loanwords in LA that “regular inferential processes allow (...) speakers and out-group members to draw the necessary semantic connections” in order to arrive at the meaning of previously unfamiliar words and phrases in each other’s speech (2002: 186).

*edificio* ‘building’. A more detailed look at the borrowing behavior of recent arrivals is provided in Chapter 5 (section 2.5.1).

### **7.5 Summary of variables that do not significantly condition borrowing rate**

Ethnonational origin and regional origin of informants, contrary to expectations, were not related to significant differences in lexical borrowing frequency among Spanish speakers in New York. This finding is a powerful indication that, despite the not uncommon perception that certain types of Spanish (such as Puerto Rican and Dominican) evidence a greater degree lexical influence from English than other regional varieties of Spanish (such as Colombian Spanish), on the whole, speakers of regional varieties in New York City do not ostensibly differ with respect to how frequently they borrow from English. (But see Chapter 5, sections 2.5.2, 3.1 and 3.3.1 for refinements and exceptions to this finding).

In addition, age did not significantly influence borrowing. An apparent significant correlation between lexical borrowing frequency and age was, in fact, an artifact of the distribution of informants among the age categories. Namely, second immigrant generation informants, who are a clearly-defined subgroup Spanish speakers in the city, were mostly distributed within the younger age categories, creating the illusion that younger individuals borrowed more. When second immigrant generation informants are removed from the sample, age is not correlated with borrowing frequency.

Finally, expectations that years in the U.S. would be related to lexical borrowing rates were not confirmed. It was thought that with increased time in the U.S., individuals’ exposure to both English and conversation in Spanish as done in the city would lead to their becoming accustomed to borrowing, and would result in their borrowing from English more often. Like age, an initial look at the data seemed to confirm this belief. However, when the second

generation is removed from the sample, years in the U.S. is not significantly correlated with borrowing frequency. This findings is notable since it indicates that once persons over the age of twelve emigrate to the U.S., they are immediately able to utilize English words and phrases in their Spanish. The findings for years in the U.S. as it pertains to borrowing frequency are examined in greater depth in Chapter 5 (section 2.5.3).

## **8. Conclusion**

The data presented in this chapter suggest that lexical borrowing behavior in Spanish in New York City is not the same for all individuals. The extent to which lexical borrowing is likely in a Spanish discourse is conditioned by an individual's membership in subgroups of the Spanish-speaking population. Borrowing rates are most significantly impacted by an informant's immigrant generation. Particularly, they are highest among those that are born or raised in New York: frequently between 10/k and 20/k. Those that come as adults (from age twenty on) are a more homogenous group. Adult immigrants rarely use more than 10 borrowings per 1000 words; in a small number of cases (n=7), they do not borrow at all. Additionally, it was discovered that child arrivers borrow as frequently as the second generation and that teen arrivers pattern similarly to the first generation. This latter finding led to a decision to identify child arrivers as part of the second generation and teen arrivers as part of the first with respect to borrowing frequency. The robustness of the difference in borrowing rates of the first and second generation suggests that each represents a distinct subcommunity with respect to borrowing frequency.

It was also found that the extent to which individuals are accustomed to using Spanish on a daily basis influences borrowing. In particular, borrowing is more frequent among those that do not often use Spanish on a daily basis. An examination of the borrowings of both the regular and infrequent Spanish users suggested that infrequent Spanish users utilize far more borrowings that

name things perceived to be specific to the New York sociocultural context. It was proposed that the extent of individuals' use of Spanish for daily activities is an indication of the extent to which they are oriented to life in the U.S.

Additionally, analysis revealed that borrowing frequency is influenced by occupational class and education level. The effect of these variables on borrowing rates in the corpus as a whole is likely to the fact that they are mostly influential in the first generation (n=98), which happens to represent a large portion of the entire corpus. As a result, the interpretation of these variables is provided when the first generation is examined by itself in Chapter 5 (sections 2.3.1 and 2.3.2).

Negative findings, particularly with respect to years in the U.S., also tell the story of lexical borrowing in the city. Borrowing frequency does not appear to increase over the course of a person's life or sojourn in a host country. Rather, it is a phenomenon to which new immigrants adjust immediately: the 20-year-old immigrant who arrived six months ago from Ecuador could use just as many English lexical borrowings as the 40-year-old who arrived from Puerto Rico 20 years ago. This state of affairs speaks of immediate and complete acquiescence to local norms of Spanish use. Although, as will be seen in Chapter 7, there may be a subtle reversal of this trend instigated by greater familiarity with English. The borrowing trends of recent immigrants are examined further in Chapter 5 (section 2.5.3).

## **Chapter 5: Immigrant generations in focus**

### **1. Introduction**

In the previous chapter, it was shown that borrowing frequency in the OZC as a whole is conditioned by an individual's immigrant generation, occupational class, level of education and daily use of Spanish. Of these four conditioning variables, immigrant generation is the overall strongest predictor of lexical borrowing frequency in Spanish in New York City. The very different borrowing rates of the first immigrant generation (arrival age  $\geq 13$ ) and those of the second immigrant generation (born in U.S. or arrival age  $\leq 12$ ) led to the conclusion that, with respect to lexical borrowing, each generation is a distinct subcommunity of Spanish speakers. This chapter examines each immigrant generation separately, exploring whether the factors that condition lexical borrowing in the corpus as a whole also condition it within each subcommunity. Furthermore, this chapter presents results for two variables whose effect on borrowing frequency has not yet been examined: the variables of English skills and Spanish skills.

Results of analysis show that borrowing frequency within each immigrant generation is influenced by a distinct set of conditioning factors. For first generation immigrants, lexical borrowing is affected by class and education and, additionally, by an individual's confidence in his Spanish and proficiency in English. Results suggest that while having more proficiency in English encourages borrowing, English proficiency is not a prerequisite for borrowing; those with little or no practical proficiency in English also borrow. In addition, results show that all immigrants, whether long-time residents in the city or recently arrived, use lexical borrowings to the same extent. This result suggests that using English lexical borrowings is part of a process of rapid acclimation both to the new cultural environment of New York City and to what may be

perceived as its local Spanish lexicon. Of the variables that condition borrowing in the first generation, occupational class comes closest to being the best overall predictor of borrowing rate. In the second generation, on the other hand, regional origin, and the degree to which informants use Spanish on a daily basis best predict borrowing frequency. It is proposed that lexical borrowing is the domain in which ethnic distinctions are maintained for second-generation Spanish speakers. Overall, evidence suggests that lexical borrowing is not a deficit behavior. In the first generation, it is most frequent among those with more education, those of higher class standing and with better English skills. In the second generation, borrowing is more frequent for those with more English confidence and is **not** connected with an informant's proficiency in Spanish.

## **2. The first immigrant generation**

In Chapter 4, it was shown that teen arrivers to the U.S. (arriving from age thirteen to nineteen) borrow similarly to adult migrants (those arriving from age twenty on). It was decided that *first immigrant generation* would refer to both adult and teen migrants. The expression continues, in the present chapter, to refer to these same individuals: those that arrived at or after age thirteen.

In this section, it will be seen that borrowing in the first generation is, as for the corpus as a whole, conditioned by occupational class, level of education and daily Spanish use. In other words the pattern of significant and non-significant findings in the first generation is the same as what is found for the corpus as a whole. This is likely because first generation informants make up a larger proportion of the corpus than second generation informants. In addition, results for daily Spanish use also condition borrowing frequency in the first generation. Since this variable was interpreted in Chapter 4, the discussion of these findings will be a restatement of conclusions presented in that chapter. Furthermore, this chapter elaborates on results for years in the U.S. are

further elaborated upon here. Finally, results for English proficiency and Spanish confidence, which have not thus far been presented, will be also be addressed.

## 2.1 Occupational class, level of education, English proficiency, Spanish confidence and daily Spanish use condition borrowing rate in the first generation

The results of an ANOVA show that class conditions borrowing rate in the first immigrant generation ( $p < .003$ ). Table 5.1 provides the details.

**Table 5.1**  
**Lexical borrowing rate by occupational class, First generation**

	Mean/k	SD
<b>Middle class</b> (n=47)	5.0	3.5
<b>Working class</b> (n=47)	3.0	2.8
<b>Total</b> (n=94)	4.0	3.3
$F(1, 92) = 9.35$		$p < .003$

Table 5.1 shows that middle class first generation immigrants borrow almost five times per 1000 words. The working class borrows less: about 3/k.<sup>1</sup> As can be seen by the F value, the difference in borrowing between the middle class and the working class is strong. Variation in borrowing rates between the groups is over nine times the amount of variation within the groups.

In addition, as with the entire corpus, one's level of educational attainment conditions borrowing in the first generation ( $p < .046$ ) (Table 5.2).

<sup>1</sup> The perhaps more objective variable socioeconomic status (SES), likewise, conditions lexical borrowing frequency in the first generation. As with (self-designated) occupational class, those with higher socioeconomic status borrowed more (4.7/k) than those with a lower status (3.0/k) ( $F(1, 92) = 6.91, p < .010$ ). I report on occupational class because results are more robust, as indicated by the higher F value (9.35).

**Table 5.2**  
**Lexical borrowing rate by level of education,<sup>2</sup> First generation**

	Mean/k	SD
<b>College or graduate</b> (n=49)	4.7	3.5
<b>High school or less</b> (n=47)	3.3	2.8
<b>Total</b> (n=96)	4.0	3.3
$F(1, 94) = 4.08 \quad p < .046$		

Table 5.2 shows that those with no more than a high school education borrow less often (3.3/k) than those that attend college or graduate school (4.7/k).

English proficiency<sup>3</sup> also has a statistically significant impact on borrowing frequency in the first generation ( $p < .008$ ) (Table 5.3).

**Table 5.3**  
**Lexical borrowing rate by English proficiency, First generation**

	Mean/k	SD
<b>Excellent</b> (n=11)	6.4	3.8
<b>Non-excellent</b> (n=85)	3.7	3.0
<b>Total</b> (n=96)	4.0	3.3
$F(1, 94) = 7.24 \quad p < .008$		

First generation informants that rated their proficiency in English highest borrow more (6.4/k) than those that rated it as less than excellent (3.7/k). The distinction between those that have excellent English skills and those that do not is pronounced, as indicated by the F value of 7.24.

- 
- 2 Other configurations of the education variable, including the elementary-educated versus those with at least some secondary education ( $F(1, 94) = 2.34, p < .129$ ) and the four-factor construction of the variable (elementary, secondary, college, graduate) ( $F(3, 92) = 1.87, p < .140$ ) were not significant. It was, however, the case that the configuration of graduates versus those with less than a graduate education was marginally significant ( $F(1, 94) = 3.60, p < .061$ ). I report the ‘high school or less’ versus ‘college or graduate’ finding because it is more robust ( $F(1, 94) = 4.08$ ).
  - 3 Informants rated themselves on a four-point scale using the categories: ‘poor’, ‘passable’ ‘good’ and ‘excellent’.

Informants' confidence in their Spanish also significantly conditions borrowing frequency ( $p < .053$ ) (Table 5.4).

**Table 5.4**  
**Lexical borrowing rate by Spanish confidence, First generation**

	Mean/k	SD
<b>Most confidence</b> (n=54)	4.5	3.4
<b>Less confidence</b> (n=42)	3.2	3.0
<b>Total</b> (n=96)	4.0	3.3
$F(1, 94) = 8.84$		$p < .053$

Table 5.4 demonstrates that those with more confidence in their Spanish borrow more (4.5/k) than those with less (3.2/k).<sup>4</sup> As with English skills, these findings are fairly pronounced, as indicated by the F value, 8.84.

Finally, daily Spanish use influences borrowing rates in the first generation ( $p < .072$ ) (Table 5.5).

**Table 5.5**  
**Lexical borrowing rate by daily Spanish use, First generation**

	Mean/k	SD
<b>Less Spanish use</b> (n=54)	4.5	3.6
<b>Most Spanish use</b> (n=42)	3.3	2.7
<b>Total</b> (n=96)	4.0	3.3
$F(1, 94) = 3.30$		$p < .072$

4 As with English, informants rated their abilities in Spanish on a four-point scale. Given that all first generation informants, having been raised in Latin America until the age of 12, at least, are from a competence point of view completely fluent in Spanish, informants' Spanish self-rating is interpreted as a reflection of their confidence in their Spanish vis-à-vis prescriptivist notions of correctness, rather than an indication of their proficiency in Spanish. This interpretation of the Spanish self-rating for first generation informants follows Otheguy & Zentella (2012: 97-98).

Table 5.5 shows that daily Spanish use marginally influences borrowing frequency in the first generation ( $p < .072$ ). The results reflect those for the corpus overall (Chapter 4, sections 4.4 and 5.2). That is, first generation individuals that use Spanish less for daily activities borrow more often (4.5/k) than those that use Spanish most (3.3/k).

## **2.2 Summary of variables that significantly condition lexical borrowing rate in the first generation**

Class, education, English proficiency, Spanish confidence and daily Spanish use all condition borrowing frequency among first generation immigrants. In particular, those of the middle class, with more education, better proficiency in English, more confidence in Spanish and those that use Spanish less all borrow more than others.

## **2.3 Discussion of occupational class, level of education, daily Spanish use, English proficiency and Spanish confidence in the first generation**

### **2.3.1 Interpreting the effect of occupational class on borrowing rate in the first generation**

It was found that middle class informants of the first generation borrow more than the working class. In several studies, class membership conditioned the extent to which individuals use lexical borrowings from a donor language. Poplack et al. (1988) discovered that, among French speakers in Canada, working class individuals used more English loanwords than the middle class. Likewise, Matus-Mendoza (2002) found that working class migrants used English borrowings to a much greater extent than individuals with professional degrees and higher levels of education. She hypothesized that the use of English lexical borrowings by Spanish-speakers from Moroleón, Mexico was a class marker.

Similarly, occupational class significantly conditions borrowing frequency among first generation Spanish speakers in New York City. Yet, the direction of the influence of class on

borrowing rates in the present investigation is opposite of that found in Poplack et al. (1988) and Matus-Mendoza (2002). In their studies, the working class borrowed more than the middle or upper classes. In New York, however, the direction of the influence of social class is reversed. It is the upper class that borrows more. The reason for the reversal is most likely due to differences in the relative social prestige of the donor language, English, vis-à-vis French in Francophone Canada, Spanish in Moroleón and Spanish in New York City. In Canada, for instance, following the nationalist Francophone movement in the 1960s, French has become the language of the mobilization and is a “major key to upward social mobility” especially in areas along the English-French border (Heller 1992: 131) (precisely where Poplack et al.’s study was carried out). Poplack et al.’s results could be said to reflect typical sociolinguistic class-based patterns where the prestige-conscious upper classes avoid the use of non-prestige or low-prestige forms (i.e. English borrowings); else, that the working class, in defiance to prestige conscientiousness, openly utilizes stigmatized or non-prestige forms. Although one may infer from Matus-Mendoza’s exposition that in Moroleón, Mexico the local prestige of English is a contentious issue,<sup>5</sup> the use of English loanwords by Moroleonians appears also to be interpretable in accord with traditional class-based accounts of linguistic variation whereby the lower classes utilize non-prestige features of language to a greater extent than those of the upper classes.

On the other hand, in New York, English is 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. Thus,

---

5 Matus-Mendoza (2002: 334) notes that for some migrants, knowing English and using English-origin loanwords confers a degree of prestige, in that it is an indication of the migrant’s purchasing power. Conversely, some Moroleonian nonmigrants criticize returning migrants as troublemakers. One intuits from her discussion that for these nonmigrant Moroleonians, the external markers of the migrant worker, including his use of English, are evaluated disparagingly.

the prestige conscious and those with aspirations of upward social mobility, such as the middle class, would be inclined to more often demonstrate their command of English, in part, it appears, through the use of English lexical borrowings. What is noteworthy, however, is that, in New York, despite associations of English with social advancement, and despite it being commonly reported that most Spanish speakers “mix languages”<sup>6</sup> (of which lexical borrowing as defined in this study is but one manifestation), there is a perception among Spanish speakers that using English while speaking Spanish may not be as expected of individuals of higher social standing. Consider a comment made by one informant of the OZC in example (1).

- (1) Entrevistadora: Okay ¿Tú crees que la alternancia, la mezcla de los dos idiomas ocurre porque los que hacen eso saben los dos idiomas bien o [porque] no saben los dos idiomas bien?  
 Entrevistado: Sí, yo creo que depende del caso y de qué tipo de mezcla. La gente culta también mezcla dos idiomas, o se.. se.. se.. se utiliza mucho una frase en francés... 092P  
 ‘Interviewer: Okay, do you think the alternation, the mixing of two languages occurs because those that do it know the two languages well or [because they] don’t know the two languages well?  
 Interviewee: Yes, I think that [it] depends on the case and the type of mixture. Refined people also mix two languages, or a phrase in French is.. is.. is often used..’

In example (1), informant 92P states that individuals with a higher social status (i.e. *gente culta*) “also” mix languages. The *también* ‘also’ in her comment appears to make her statement function as a clarification of a generally held, but opposite, perception (one that she is observant enough to see is false): that those of a high social status are generally thought to **not** mix languages. Notice, however, how she distances the association of *gente culta* ‘refined people’

---

6 Consider, for example, this informant’s acknowledgment of the pervasiveness of language mixing:

Yo creo que todo el mundo los mezcla, todo.. todo el que hable inglés y español lo mezcla, ¿verdad? 005U  
 ‘I think that everyone mixes them, every.. everyone that speaks English and Spanish mixes it, right?’

with mixing English into their Spanish, by adding that what they mix in is **French**. Yet, the pattern of results uncovered in this study, that is, the empirical data reflecting what Spanish speakers **actually do** with English, conflict with these descriptions of what Spanish speakers **say** is done with English. Clearly, then, what is at hand is a language behavior with covert prestige (Labov 2001: 215). In other words, when pressed, Spanish speakers in New York will impart to their interlocutors the *monoglossic* wisdom (del Valle & Stheeman 2004: 10) of their host and parent nations (i.e. that languages should be kept separate), all the while orienting to a distinct set of norms in interaction.

### **2.3.2 Interpreting the effect of level of education on borrowing rate in the first generation**

Educational attainment has been correlated with lexical borrowing in several studies. Matus-Mendoza found that among Mexicans who are former and current migrant workers to the U.S., those with fewer years of education “favor the use of English in their speech”, and that those who have attained higher levels of education progressively use English to a lesser extent or not at all (2002: 333). She explains this as an occupational class effect: those migrant workers with post secondary education tended to have higher occupational class standing after leaving migrant work (as lawyers, engineers, etc.).<sup>7</sup> Likewise, Sullivan (2008) examined borrowings from Urdu and English in Punjabi in Lahore, Pakistan. She found that the less education individuals had, the more likely they were to use borrowings from Urdu and English. She concluded that those with less education were primarily responsible for the spread of borrowed lexical items among Punjabi speakers. A corollary to her conclusion was that the least-educated had less access to

---

7 Matus-Mendoza conjectures that after individuals obtain a higher class standing they feel that they “no longer need to separate from the rest of the city” and that they therefore do not wish to emphasize their separateness by using more English lexical borrowings (2002: 333-334). On the other hand, migrant workers that continued in this line of work tended to socialize only with those in similar types of jobs. This latter group uses English lexical borrowings to mark their non-localness when back in their hometown of Moreleón, Mexico.

community prestige languages, Urdu and English, and thus less practice at recognizing and filtering out elements foreign to their native Punjabi. Thus, for Sullivan, level of education was related not only to donor language exposure but also to language processing. On the other hand, Eslami Rasekh et al. (2008) found that more-educated Turkman speakers in Iran borrowed from Persian more frequently than the less-educated. They also interpreted level of education to be an indication of the degree to which speakers were exposed to the socially dominant language, in their case, Persian.

Similar to the above-mentioned studies, lexical borrowing has, in the current investigation, been linked to level of education. Yet, as with findings for class, the effect of education level on borrowing in the present study contradicts with what has been found by others. Whereas lack of education may be associated with greater borrowing in Mexico and Pakistan, in New York City, it is Spanish speakers with more education that borrow more. This finding is not interpreted, as it was for Sullivan (2008) and Eslami Rasekh et al. (2008), as reflecting Spanish speakers' exposure to the donor language, English. This is because the classroom is not the primary means by which Spanish-speakers are exposed to English in New York.<sup>8</sup> Rather, level of education in New York is tied to social class standing. In particular, 70 percent of first generation informants with no more than a secondary education identify themselves as working class, while 70 percent of those with a college education or beyond claim middle class status ( $Chi^2(1, N=95) = 16.01$ ,

---

8 However, it is true that, in the first generation, a larger percentage of those with college or graduate education have probably been exposed to English as a language of instruction (approximately 38 percent) than those with education to the high school level or less (just four percent or  $n=2$ ). This is also true for OZC informants overall, where about 62 percent of those with a college or graduate education have probably been exposed to English in the classroom, as compared to just 28 percent of those with less than a high school education.

$p < .001$ ).<sup>9</sup> In other words, the significance of education on borrowing rate is believed to be a reflection of the more powerful influence of class on borrowing (see section 6).

The reversal of the class/education trend among the first generation in New York City, as compared to Moroleón, is probably because the location of the more-educated within the local socioeconomic hierarchy in each place differs. In Moroleón, those with more education likely represent the most prestigious socioeconomic strata within their local communities. Borrowing from English, a behavior linked to one's potential for economic advancement and social betterment, earns educated Moroleonians no additional advantage either in how others perceive them or in the types of employment they have access to. In New York, more-educated and middle class, first generation, Spanish speakers occupy a more central position within the socioeconomic hierarchy of the larger New York City context. Borrowing from English, then, may simultaneously reflect their orientation toward the larger socioeconomic hierarchy of the city as well as their ability or potential ability to move in more prestigious circles (similar to the working class and less-educated of Moroleón).

### **2.3.3 Interpreting the effect of English proficiency on borrowing rate in the first generation**

It is a well-established axiom of contact studies that bilinguals are the channel through which influences from one language flow to another (e.g. see Haugen 1969: 4). It was, thus, expected that borrowing would be more frequent among those with higher English proficiency. Results from Table 5.3 confirm this expectation: first generation Spanish speakers with better English skills borrow more than those with lesser proficiency in English.

---

9 The distribution of first generation informants by education level and occupational class can be found in Appendix E (Table E.3). This significant positive relationship between occupational class and education is also true of informants of the OZC overall ( $Chi^2(1, N=141) = 19.48$ ,  $p < .001$ ; Table E.4) as well as for working Latinos in New York City (Appendix E, Table E.5) (U.S. Census 2010).

However, while high proficiency in English encourages greater frequency of borrowing, it is not only the English-proficient that borrow. About one-third (n=30) of the first generation has little or no ability in English, according to their own self ratings.<sup>10</sup> Of those, only four do not borrow at all. That is, 86 percent of what are, for all practical purposes, monolingual Spanish speakers borrow at least once during their interview and 66 percent (n=20) had borrowing rates between 1/k and 12/k. Said differently, this group (n=30), those with the least knowledge of English, averaged about 3.2 borrowings per 1000 words of discourse. In other words, although better proficiency in English seems to grease the channel through which lexical influence on Spanish in the U.S. flows, lack of proficiency in English does not appear, in itself, to be a barrier to lexical borrowing. This interpretation of results is further bolstered by an examination of group means with respect to the total mean (4.0/k). From Table 5.3, it can be seen that those with excellent English proficiency borrow at a rate farther from the group mean (a difference of +2.4/k) than those with less proficiency (a difference of -0.3/k). That is, having excellent proficiency in English **encourages** lexical borrowing.

#### **2.3.4 Interpreting the effect of Spanish confidence on borrowing rate in the first generation**

At the time of data collection, informants were asked to rate their skills in Spanish on a four-point scale as *pobre*, *pasable*, *bueno* or *excelente* (i.e. ‘poor, passable, good, excellent’). First generation informants, having been raised in Spanish-dominant societies, were all fluent speakers of Spanish. It was believed, therefore (following Otheguy & Zentella 2012: 93-94), that speakers’ self-rating in Spanish was a reflection not of their proficiency in Spanish, but of whether they felt that their Spanish resembled some notion of good or proper Spanish. In other

---

<sup>10</sup> On the four-point scale (*pobre—pasable—bueno—excelente*), they gave themselves the lowest rating ‘poor’.

words, for the first generation, the Spanish skills variable is thought to be a reflection of a person's confidence in or insecurity about his language abilities.

Table 5.4 showed that those with more confidence in their Spanish borrow more than those with less. In comparing factor means to that of the first generation overall (4.0/k), it can be seen that those with less confidence are farther from the overall mean (-0.8/k) than those with more confidence (+0.5/k). Findings for Spanish confidence are therefore stated in the following way: having less confidence in Spanish **inhibits** borrowing. Perhaps in perceiving themselves to have poor Spanish (despite being fluent in it), these first generation informants avoid borrowing to prevent negative perceptions of their speech habits.

### **2.3.5 Interpreting the effect of daily Spanish use on borrowing rate in the first generation**

It was found that, as in the corpus overall (Chapter 4, section 4.4), daily Spanish use conditions borrowing frequency in the first generation. As in Chapter 4 (section 5.2), daily Spanish use is interpreted as an indication of individuals' investment in life in the U.S. First generation Spanish speakers who are more oriented toward life in the U.S. (by necessity or desire) borrow more.

## **2.4 Occupational class best predicts borrowing rate in the first immigrant generation**

To determine which of the five conditioning variables (occupational class, educational attainment, English skills, Spanish confidence and daily Spanish use) best accounts for borrowing frequency in the first generation, a regression analysis was conducted. The results are in Table 5.6.

**Table 5.6**  
**The ranked effects of occupational class, level of education, English skills, Spanish confidence and daily Spanish use on lexical borrowing rates in the first generation**

	Standardized beta coefficient	p	Tolerance
<b>Occupational class</b>	.20	<b>.088</b>	.71
<b>English skills</b>	.17	.136	.80
<b>Spanish confidence</b>	.11	.320	.78
<b>Educational attainment</b>	.05	.653	.71
<b>Daily Spanish use</b>	.01	.949	.76
$R^2 = .147$ $F(5, 86) = 2.96$ $p < .016$			

Table 5.6 indicates that a model including the five variables significantly predicts only a small amount of the variance in the borrowing rates of the first generation: about 15 percent ( $R^2=.147$ ,  $p<.016$ ). No one factor is found to be the main predictor of borrowing rate at the  $p<.05$  level. However, of these five variables, occupational class appears to be the best predictor of borrowing frequency in the first generation ( $beta=.20$ ,  $p<.088$ ).

## **2.5 Variables that do not condition borrowing rate in the first generation: Regional origin, sex, age, arrival age and years in the U.S.**

### **2.5.1 Sex, age and arrival age as variables that do not significantly condition borrowing rate in the first generation**

Analysis reveals that sex does not play a role in borrowing frequency among first generation informants ( $F(1, 95) = 2.43$ ,  $p<.122$ ). Likewise, the age at which first generation informants arrive to the U.S. has no impact on the amount of borrowing he does ( $r(97) = -.11$ ,  $p<.307$ ). Furthermore, as detailed in Chapter 4 (section 7.3), when the first generation is considered in isolation, an informant's age is not significantly related to differences in borrowing frequency ( $F(3, 93) = 0.96$ ,  $p<.414$ ). Said differently, results for age show that for first generation

individuals, a 65-year-old migrant and a 20-year-old migrant can be expected to borrow to the same extent.

### **2.5.2 Regional origin as a variable that does not significantly condition borrowing rate in the first generation**

Neither ethnonational affiliation nor regional origin significantly influences borrowing rates in the first generation ( $n=97$ ) (ethnonational affiliation  $F(5, 91) = 1.38, p < .236$ ; regional origin  $F(1, 95) = 2.55, p < .114$ ), despite expectations to the contrary (Chapter 4, section 7.1). One explanation for this finding could be in a modified version of the doctrine of the *first effective settlement* (Labov 2001, citing Zelinsky 1992: 13). The first effective settlement doctrine states that characteristics of the linguistic system of the first group of settlers to an area establishes a pattern “that dominates the history of the speech community” (Labov 2001: 45). Modifying this definition for immigrant contact situations and with respect to lexical borrowing, it may be that the first groups of Spanish-speaking migrants to New York City established a pattern for lexical borrowing that is replicated in the speech patterns of immigrants of later vintages. In other words, if Caribbean settlers, particularly Puerto Ricans, established a pattern of lexical borrowing that later immigrants have taken up, then there would be no difference in borrowing rates for Caribbean speakers and Mainlanders, who tend to be of later immigrant vintages. This seems a satisfactory account of the fact that no difference in borrowing rates by regional origin or ethnonational affiliation has been found in the corpus as a whole<sup>11</sup> or among the first generation. In effect, Spanish-speaking migrants of the first generation, regardless of origin, their country’s history of contact with English and the length of time their communities have been in the city, all borrow to the same extent. This principle, however, does not characterize borrowing

---

11 Corpus results of an ANOVA for region were  $F(1, 136) = 0.99, p < .322$ . The ANOVA results for ethnonational affiliation were  $F(5, 132) = 1.28, p < .276$ . (See Chapter 4, section 7.1).

in the second generation, where ethnic identity appears may play a part in conditioning borrowing rate (sections 3.1, 3.3.1 and 3.4 of this chapter).

### 2.5.3 Years in the U.S. as a variable that does not significantly condition borrowing rate in the first generation

Results presented in Chapter 4 (section 7.4) show that lexical borrowing among first generation immigrants, on average, neither decreases nor increases with time spent in the U.S. ( $p < .730$ ). ANOVA results are reproduced here in Table 5.7.

**Table 5.7**  
**Lexical borrowing rate by years in U.S., First generation**

	Mean/k	SD
<b>Recent</b> (n=16) (0-2 yrs in U.S.)	4.3	3.6
<b>Long</b> (n=50) (3-15 yrs in U.S.)	4.1	3.3
<b>Established</b> (n=31) (16+ yrs in U.S.)	3.6	3.0
<b>Total</b> (n=97)	4.0	3.3

$F(2, 94) = 0.32$        $p < .730$

Table 5.7 shows that years in the U.S. does not contribute to differences in borrowing frequency in the first generation ( $F(2, 94) = 0.32, p < .730$ ).<sup>12</sup> In other words, even recent immigrants take from English as often as someone who has been in the U.S. for decades. These results are unexpected and contradict beliefs about borrowing among Spanish speakers in the city. To wit, when asked why mixing Spanish and English happens, several informants of the OZC suggested that it was due to the amount of time one lives in the city (2).

<sup>12</sup> Non-significant results were also obtained from a Pearson correlation ( $r(95) = -0.07, p < .505$ ).

- (2) Entrevistadora: ¿Por qué razón [mezclas el inglés y el español]?  
Entrevistada: Bueno, porque.. tantos años viviendo aquí, en este país y la costumbre de tener que hablar los dos, hay ocasiones que uno lo dice sin darse cuenta. 112D
- ‘Interviewer: For what reason [do you mix English and Spanish]?  
Interviewee: Well, because.. so many years living here, in this country and the habit of having to speak both, there are times when one does it without realizing.’

The finding that recent migrants borrow as frequently as long-established residents suggests several, but not mutually exclusive, accounts of borrowing behavior in the first generation. It may suggest that newly arriving immigrants are faced with a great number of novel cultural artifacts and experiences, which are most expediently referred to by adopting their English names.<sup>13</sup> It could also indicate that English lexical borrowings are perceived by recent migrants not as English per se, but as novel, local (Spanish) vocabulary that must be quickly integrated into their speech if they desire efficient communication. Alternatively, it may be that, upon arrival, recent migrants perceive a discourse pattern in Spanish that is characterized by using English every so often. They then mimic this pattern by inserting what English they do know into speech as frequently as established migrants.

Deciding which of these scenarios best accounts for the behavior of recent migrants requires comparing their borrowing inventory to that of longer-time residents. For instance, if recent arrivals draw from a smaller inventory of vocabulary to borrow, it could suggest that what they acquire upon arriving to the U.S. is a discourse pattern. In other words, it may be that recent migrants pick up a few local (English-origin) words and drop them into their Spanish about as often as long-time residents, who may have a more varied inventory, do. If the borrowing

---

13 Numerous studies suggest that referential need is perhaps the strongest motivation for borrowing (e.g. see Neumann 1938; Ornstein 1976). Weinreich writes, “Lexical borrowing of this type [i.e. to name new things] can be described as a result of the fact that using ready-made designations is more economical than describing things afresh. Few users of language are poets.” (1966: 57).

inventory of recent migrants were not only smaller, but found to be a subset of the inventory used by long-time residents, it would strengthen the impression that recent migrants are quickly adopting what they perceive to be a local Spanish lexicon. On the other hand, if their inventory is distinct from that of long-time residents, then rapid lexical innovation in the face of great cultural change may best explain the borrowing behavior of the recently arrived.

An analysis of lexeme rates and the sharedness of borrowing inventories indicates that a combination of the latter two scenarios may describe how new arrivals borrow. First, an ANOVA showed no difference in the number of lexemes used by recent arrivals (in the U.S.  $\leq 3$  years) and immigrants in the U.S. for longer ( $F(3, 93) = 0.46, p < .710$ ). Both recent arrivals and established immigrants use about three or four distinct lexemes<sup>14</sup> for every 1000 words of discourse. In other words, recent arrivals do not appear to be merely mimicking a discourse pattern for speaking Spanish in New York,<sup>15</sup> but quickly acquiring a substantial, new inventory of English-origin words. An analysis of how that new inventory overlaps with others' is also informative. Of 244 lexemes used by recent arrivals, 126 of them (52 percent) are used by informants that are either long-time residents or U.S.-born. In other words, 52 percent of the borrowing inventory of newcomers is shared with other informants of the corpus. The other 48 percent of their inventory is unique to them.<sup>16</sup> Thus, it seems that recent arrivals both promptly

---

14 For more on lexemes, see Chapter 3 (section 4.1.2).

15 In other words, while recent immigrants may in fact perceive a norm for how often to use English lexical borrowings in speech, and implement that norm, this is not all they are doing.

16 This set of lexemes includes *Columbus Day*, *AT&T*, *Brooklyn Museum of Art*, *Cross Bronx*, *crawfish*, *Down's Syndrome*, *drive-through*, *factory*, *fellowship*, *Forty-second [Street]*, *I-95*, *infomercial*, *matter of fact*, *Miami Beach*, *Napa Valley*, *popcorn*, *ponchar*, *proud*, *realtor*, *slogan*, *Savoy Inn*, *stripper* and *website*. Several of these, such as *Forty-second*, *drive-through*, and *I-95*, are probably in regular use throughout New York or among individuals in certain contexts, but were simply not used by another informant of the corpus. This indicates that even the supposedly unique vocabulary items of recent migrants are probably acquired as part of speaking Spanish in New York City. See Chapter 6 (section 3.4) for more on the nonshared lexemes of recent arrivals.

acquire an inventory of borrowings in use among long-time residents, as well as augment their inventory with borrowings pertaining to their particular experiences in the U.S.

## **2.6 Summary of lexical borrowing in the first immigrant generation**

Borrowing rates for the first generation are not influenced by sex, age or arrival age. Notably, national and regional origin are also not significantly predictive of differences in borrowing rate, despite different histories of contact with English and Anglo-American institutions. It was proposed that an adapted version of the doctrine of the *first effective settlement* (Zelinsky 1992) may provide an explanation for this unanticipated result. Further, years in the U.S. is also not significantly related to differences in borrowing frequency. This latter finding is noteworthy for several reasons. First, it indicates that even as Spanish speakers reside in the English-speaking environment for years, presumably adding to their English knowledge along the way, there is no ostensible increase in how often they borrow. In other words, borrowing from English is more or less stable throughout an immigrant's time in the U.S. Second, the fact that recent arrivals borrow just as much as long-established residents suggests that lexical borrowing, while thought to be the provenance of bilinguals, does not require much exposure to or proficiency in English. Rather, as indicated by an inventory analysis, newly-arrived immigrants in part acquire a stock of words already in use by established-residents, much like a foreigner learning the locally appropriate foreign words for things upon living in a new country. At the same time, they rapidly augment their inventory with the English words and phrases descriptive of their particular experiences in a new place.

On the other hand, borrowing rates in the first generation are influenced by occupational class, level of education, English skills, Spanish confidence and daily use of Spanish. In particular, borrowing from English is more frequent among the middle class and those with

college or graduate educations. These individuals use between one and a half and two times as many borrowings from English as those with fewer years of schooling and the working class. These results indicate that, contrary to popular beliefs about who is most susceptible to lexical borrowing (e.g. the lower classes or the less educated), borrowing in Spanish in New York City is in fact more common among those with more education and higher social standing. It is further speculated that borrowing in the first generation may be a marker of covert prestige as well as an indication of an individual's ability or potential ability to navigate relationships and employment in more prestigious circles in the larger New York City context.

Findings for English proficiency and Spanish confidence were both expected and surprising. As expected, those with better English skills borrow at a higher than average rate compared to the first generation overall. Of note, though, was that most of those with little or no proficiency in English also borrow, again suggesting substantive knowledge of English is not a prerequisite for lexical borrowing. It has also been seen that those that are less confident in Spanish borrow at rates that appear suppressed with respect to the group mean. This finding suggests that perhaps a sense of linguistic confidence is a cornerstone of lexical borrowing among immigrant groups. Of these four influencing factors, however, class ( $\beta=.20, p<.088$ ) came closest to being a significant predictor within the five-variable regression model. These results should be tempered, however, by the fact that the model accounts for only about 15 percent of the variance in borrowing rate in the first generation ( $R^2=.147, p<.016$ ).

### **3. The second generation**

In the previous chapter, analyses showed that those that arrived to the U.S. as children (i.e. from age four to twelve) have borrowing rates that are similar to those born in the U.S. As a consequence, it was decided that, for expositional simplicity, the expression *second immigrant*

*generation* would refer not only to the U.S.-born, but also to child arrivers. In the present section, the expression *second (immigrant) generation* continues to be used with reference to informants either born in New York or who came at or before age twelve.<sup>17</sup>

### 3.1 Regional origin, English confidence and daily Spanish use condition borrowing rate in the second generation

Regional origin conditions borrowing rate among second generation informants ( $p < .023$ ). Table 5.8 displays the results of an ANOVA.

**Table 5.8**  
**Lexical borrowing rate by regional origin, Second generation**

	Mean/k	SD
<b>Caribbean</b> (n=24)	12.8	7.4
<b>Mainland</b> (n=23)	8.4	5.2
<b>Total</b> (n=47)	10.7	6.7
$F(1, 45) = 5.55$		$p < .023$

Lexical borrowing is highest for second generation informants that have ties to Caribbean nations (i.e. Puerto Rico, the Dominican Republic and Cuba). Caribbean informants borrow about 13/k, while informants with ties to countries of the Latin American mainland (i.e. Mexico, Colombia and Ecuador) do so less often, about 8/k.

<sup>17</sup> However, in Chapter 7, *second immigrant generation* will only refer to those born in New York and those arriving at age three or before. This is because with respect to flagging (and other aspects of lexical borrowing not included in this dissertation), child arrivers do not pattern with the U.S.-born (see Chapter 7, section 6.2.2).

Borrowing rates in the second generation are also significantly conditioned by an informant's confidence in English<sup>18</sup> ( $p < .050$ ) (Table 5.9).

**Table 5.9**  
**Lexical borrowing rate by English confidence,<sup>19</sup> Second generation**

	Mean/k	SD
<b>More confidence</b> (n=33)	12.0	7.2
<b>Less confidence</b> (n=14)	7.7	4.5
<b>Total</b> (n=47)	10.7	6.8
$F(1, 45) = 4.06$		$p < .050$

Particularly, those with less confidence in their English borrow less (7.7/k) than those that have the most confidence in their English abilities (12/k).

18 The English confidence variable was obtained from informant's self-rating of their English skills on a four-point scale (*poor—passable—good—excellent*). Otheguy and Zentella (2012: 98) find evidence in support of the position that the English skills variable is in fact a good indication of informant's proficiency in English. However, in line with their interpretation of Spanish proficiency as an indication of linguistic insecurity for the first generation, I have chosen to interpret the English skills variable for the second generation also as an indication of linguistic confidence or insecurity. I have done this because I have assumed that all second generation informants are probably fluent in English. My support for this assumption comes from two sorts of facts. First, 32 of 47 second generation informants (68 percent) rated gave themselves a better rating in English and another 13 (28 percent) gave themselves the same rating in Spanish and English. Only three (six percent) rated their Spanish higher than their English. Furthermore, 38 of 47 second generation informants (81 percent) have, as far as can be established, received all of their formal education in English. The other nine (19 percent), it appears, have had no less than five years of formal schooling in English. These latter facts on language of education were not obtained directly from informants, but from an analysis of the second generation informant's arrival age, age at the time of the interview and the highest level of education attained.

19 Other arrangements of the factor groups including a division between those that rated themselves as 'poor' or 'passable' on the one hand, and those that rated themselves as 'good' or 'excellent' in English, on the other, were not significant at the  $p < .05$  level.

Finally, daily Spanish use also influences borrowing rates ( $p < .016$ ) in the second generation. These results are in Table 5.10.

**Table 5.10**  
**Lexical borrowing rate by daily Spanish use, Second generation**

	Mean/k	SD
<b>Infrequent Spanish use</b> (n=34)	12.2	6.9
<b>Regular Spanish use</b> (n=13)	6.9	4.6
<b>Total</b> (n=47)	10.7	6.8
$F(1, 45) = 6.23 \quad p < .016$		

Table 5.10 reveals that second generation informants that use Spanish regularly (6.9/k) borrow about half as often as those that use Spanish infrequently (12.2/k).

### **3.2 Summary of variables that significantly condition lexical borrowing rate in the second generation**

Lexical borrowing in the second immigrant generation is conditioned by a distinct set of variables than in the first generation. Specifically, the region of Latin America to which an informant has ties is significantly related to borrowing rate in the second generation, but not in the first generation. Those with ties to Caribbean nations borrow more than those with ties to Latin American mainland nations. In addition, those with the most confidence in their English borrow more than those with less confidence. One variable that influences borrowing frequency in both the first and second generation is daily Spanish use. In both, those that use Spanish infrequently in the day to day borrow more than those that use Spanish regularly.

### 3.3 Discussion of regional origin, English confidence and daily Spanish use in the second generation

#### 3.3.1 Interpreting the effect of regional origin on borrowing rate in the second generation

To understand what may be occurring in the second generation with respect to the variable of region, let us consider their borrowing rates in terms of ethnonational affiliation, a similar but more detailed description of an individual's regional ties:

Puerto Ricans (n=8):	15.3/k
Dominicans (n=9):	12.7/k
Ecuadorians (n=6):	10.5/k
Cubans (n=7):	10.3/k
Colombians (n=9):	8.6/k
Mexicans (n=8):	6.6/k

Mexican second-generation informants borrow less often than other ethnonational groups (6.6/k). Ecuadorians and Cubans borrow at rates close to the mean borrowing rate in the second generation: an average of 10 times per 1000 words. Finally, Puerto Rican and Dominican second generation informants borrow more than any other group (15/k and 13/k, respectively). Although results for ethnonational group are not significant at the  $p < .05$  level ( $F(5, 41) = 1.81, p < .133$ ), a post hoc analysis shows that the difference between Puerto Ricans and Mexicans approaches statistical significance ( $p < .098$ ) (for these results, see Table E.6 in Appendix E). Said another way, there is a 90 percent chance that the difference in group means for Puerto Ricans and Mexicans in this sample is true of Puerto Ricans and Mexicans in the population.

Within the second generation, then, the results of statistical analysis confirm anecdotal commentary about the groups most involved in the so-called mixing of Spanish and English, of which borrowing as defined in this study may be considered one manifestation. Consider

comments given by just four informants in the questionnaire portion of interviews when asked who they think mixes Spanish and English.

- (3) Entrevistadora: ¿En qué consiste el español de los puertorriqueños? ¿O cómo se distinguen? (...)  
Entrevistado: [Los que son criados acá] lo mezclan más con el inglés... 158C  
‘Interviewer: What are the characteristics of Puerto Rican Spanish? Or how are they different?  
Interviewee: [Those that are raised here] mix it with English more...’
- (4) Entrevistado: ¿Y tú conoces a otras personas que lo hacen [que mezclan el inglés y el español], otros grupos? (...)  
Entrevistada: Eh.. bueno los puertorriqueños que yo conozco, y los dominicanos que yo conozco... 010U  
‘Interviewer: And do you know other people that do it [that mix Spanish and English], other groups?  
Interviewee: Uh.. well the Puerto Ricans that I know, and the Dominicans that I know...’
- (5) Entrevistadora: ¿Usted conoce unas personas que hablen el inglés mezclado y el español?  
Entrevistada: Sí, lo puertorriqueño... 118D  
‘Interviewer: Do you know people that speak mixed English and Spanish?  
Interviewee: Yes, Puerto Ricans...’
- (6) Entrevistadora: ¿Conoces a otras personas que lo hacen [mezclan el inglés y el español]? (...)  
Entrevistada: Em, puertorriqueñas y ecuatorianas... 180C  
‘Interviewer: Do you know other people that do it [mix English and Spanish]?  
Interviewee: Uhm, Puerto Ricans and Ecuadorians...’

As can be seen from excerpts (3)-(6), the groups most often associated with language mixing, Puerto Ricans, Dominicans and Ecuadorians, are precisely those with the highest borrowing rates in the second generation. In other words, the perception among New York Latinos that language mixing is more common among some ethnonational groups than others is confirmed by results of this investigation, but only as regards the second generation.

It should not escape attention that the borrowing rate of each ethnonational group, from highest to lowest, roughly corresponds to periods of migrations of these groups to New York City, from earliest to latest. For instance, one of the first major movements of Puerto Ricans to the city occurred after the passing of the Jones-Shafroth Act of 1917, which granted U.S. citizenship to Puerto Ricans. Dominicans came in substantial numbers in the 1960s and 70s, following the fall of the dictator Trujillo in the Dominican Republic, although they were most likely present in New York since the 1930s.<sup>20</sup> The migration of Ecuadorians, Cubans and Colombians to the city also rose significantly in the 1960s and 1970s. Finally, economic crisis in Mexico in the 1980s led to large numbers of Mexicans migrating to New York at that time. The fact that borrowing rates mirror this relative ordering seems more than coincidental. The variable of ethnonational affiliation may be interpretable as an indication of the length of contact that each of these communities has had in the U.S. That is, it may be that individuals associated with particular migrant communities borrow more because their communities have, over time, been able to acquire a larger inventory of borrowings from which members draw while speaking Spanish. If this is the case, it would be expected that all members of those communities, including first generation migrants, would be prone to borrow more. However, analysis does not reveal this to be the case. Borrowing frequency in the first generation is not conditioned by national affiliation ( $F(5, 91) = 1.39, p < .236$ ) nor region ( $F(1, 95) = 2.55, p < .114$ ) (see section 2.5.2 of this chapter).

It may be hypothesized, nonetheless, that although the trend reflecting length of community contact is not significant in the first generation, the trend begins in that generation and

---

20 Ramona Hernandez, Director of the Dominican Studies Institute, and her team have found evidence of the passage of 5,000 Dominicans through Ellis Island between the years 1892 and 1924 (personal communication, 3 July 2013). Although it is not yet certain where these Dominican migrants settled, I speculate that many stayed in the city.

subsequently becomes more focused (and approaches statistically significant difference) in the second generation. If this were the case, it would be expected that first generation migrants of the Caribbean would borrow more than mainlanders. It would further be expected that, although not significant, the relative ranking of first generation ethnonational groups by borrowing rate would anticipate the ranking demonstrated within the second generation (i.e. that Puerto Ricans would have the highest average, followed by Dominicans and Ecuadorians, and that Mexicans would have the lowest group average). It turns out that first generation Caribbean informants do borrow more than mainlanders. Caribbean informants borrow about 4.5/k and mainlanders do about 3.4/k, although the difference is not statistically significant ( $F(1, 95) = 2.55, p < .114$ ) (section 2.5.2 of this chapter). However, the ranking of ethnonational group in the first generation does not resemble that of the second generation:

Puerto Ricans (n=8):	5.7/k
Mexicans (n=8):	4.0/k
Cubans (n=7):	3.9/k
Dominicans (n=9):	3.8/k
Colombians (n=9):	3.1/k
Ecuadorians (n=6):	3.1/k

In the first generation, Puerto Ricans borrow most. They are followed by Mexicans, Cubans and Dominicans. Finally, Colombians and Ecuadorians of the first generation borrow least of all. This ranking by national group in the first generation is not just different from the ranking in the second generation, it is **very** different. Although Puerto Ricans and Colombians in both generations maintain the same relative ranking, the rankings of other groups diverge considerably. Dominicans and Ecuadorians are two of the three highest-ranking groups in the second generation, but are two of the three lowest-ranking groups in the first generation.

Mexicans and Cubans are two of the three lowest-ranking groups in the second generation but are two of the three highest-ranking groups in the first generation.

In other words, there is little evidence in the current sample that supports an interpretation of the ethnonational affiliation variable as indicative of length of community contact. For such an interpretation to be supported, the borrowing pattern demonstrated by second generation informants should also be present for first generation informants. Although Caribbean first-generation informants do borrow more than mainlanders, the ranking of national groups between the generations is so different that it makes regional origin and ethnonational affiliation questionable proxies for length of community contact. Particularly, Dominicans, what may be one of the earlier-arriving groups, have one of the lowest borrowing rates in the first generation. Furthermore, Mexicans, the group that settled in New York in substantial numbers the latest, have the second-highest borrowing rate in the first generation. The only data that makes a length of contact interpretation tenable is the fact (a) that Puerto Ricans, the group with the longest history of presence in the U.S., have the highest borrowing rates in both the first and second generation and (b) Colombians, one of the later arriving groups, have one of the lowest borrowing rates in both generations. But two groups of six are not sufficient evidence for a length of contact interpretation of region findings.<sup>21</sup> Furthermore, analyses using other measures,

---

21 A “length of group contact” interpretation may be tenable if exceptions to the ranking could be explained with information on the settlement patterns of migrating ethnonational groups. For instance, Dominicans of the first generation may have lower than expected borrowing rates because they did not constitute a well-defined ethnic group until they arrived in more substantial numbers in the 1960s and 1970s. And if this is so, it may be that Dominicans of the second generation have higher rates only because they tend to settle and work among Puerto Ricans more than first generation immigrants do. I have not researched historical patterns of settlement and interaction between Latino groups in New York. So, I am unable to offer explanations of this sort.

such as lexeme rate, have turned up little evidence in support of a length of contact interpretation.<sup>22,23</sup>

The effect of regional origin on borrowing rates in the second generation may be due to the fact that Puerto Ricans, a Caribbean-origin group, borrow so much more often than other groups. The fact that they do in both generations may be due, not to the fact that this group has been present in New York longer, but because Puerto Ricans have a unique position as Spanish speakers in New York. They are the only group automatically granted U.S. citizenship.

To account for the findings, then, it is proposed that regional origin, more particularly ethnonational affiliation, should be interpreted as ethnic identity.<sup>24</sup> That is, lexical borrowing

---

22 An analysis was also carried out with respect to lexeme rate with similar findings. The effect of region on lexeme rate among first generation informants is not significant ( $F(1, 95) = 2.25, p < .137$ ), although Caribbean informants on average do appear to use more unique vocabulary words while borrowing ( $M=3.2/k$ ) than mainlanders ( $M=2.5/k$ ). The effect of ethnonational group on lexeme rate in the first generation approaches significance ( $F(5, 91) = 1.92, p < .098$ ). However, the results by ethnonational group are (in order): Puerto Ricans ( $M=4.2/k$ ), Mexicans ( $M=3.3/k$ ), Dominicans ( $M=2.7/k$ ), Cubans ( $M=2.7/k$ ), Ecuadorians ( $M=2.2/k$ ) and Colombians ( $M=1.8/k$ ). (Tukey's post hoc analysis shows that the difference between Puerto Ricans and Colombians approaches significance:  $p < .078$ .) In other words, the ranking of ethnonational groups by lexeme rate, again, does not correspond to the ranking in the second generation.

23 It was thought that a distinction between the Caribbean and mainland informants of the first generation might be detectable in a difference in the actual number of shared lexemes used by each group. It was found, however, that first generation Caribbean informants, on the one hand, and mainland informants, on the other, use approximately the same number of shared lexemes. Caribbean informants use 84 different lexemes and mainlanders use 71. (Sixty-one lexemes of the lexemes used by each group are also used by the other group.) In other words, the actual inventory of lexemes used by each regional grouping in the first generation does not substantially differ either. Analysis of the inventory of shared lexemes of second generation Caribbean and mainlanders, however, shows that their inventories differ more pronouncedly than those of the first generation. Second generation Caribbean informants use 63 shared lexemes, while mainlanders use only 38. (Thirty-three of each groups lexemes are used by the other group.)

24 Other explanations for the fact that Puerto Ricans borrow most in both generations were also considered. I thought that their higher borrowing rates were perhaps reflecting the fact that they either use English more on a daily basis or are more proficient in English than other groups. It is true that a large proportion of Puerto Ricans in both generations are proficient or confident in English. Yet, this is also true of groups that have lower borrowing rates. For instance, 100 percent ( $n=8$ ) of second-generation Puerto Ricans rate their English highly, whereas between 55 and 85 percent of other second generation national groups do. Among the highest of these other groups were Cubans; 70 percent of Cubans ( $n=5$ ) have high confidence in their English (Appendix E, Table E.7). But Cubans have only middle-of-the-range borrowing rates ( $M=10.3/k$ ). In the first generation, 53 percent of Puerto Ricans ( $n=9$ ) are proficient in English (Appendix E, Table E.8). The group with the next largest proportion of English-proficient individuals are Ecuadorians. Forty-one percent ( $n=7$ ) of Ecuadorians are proficient in English. But Ecuadorians have one of the lowest borrowing rates in the first generation. Additionally, Cubans have the lowest concentration of English-proficient individuals (12 percent or  $n=2$ ), but are among the most frequent borrowers in the first generation. As regards English use, in the second generation,

may take on the function of indexing the ethnic affiliation of second generation Spanish speakers. Ethnic identity is important to the first generation as well. First generation individuals, however, arrive to the U.S. using a set of linguistic features that index their regional origin and ethnic identity, such as a regional lexicon (e.g. *chiringa* or *cometa* ‘kite’; *pantallas*, *pendientes* or *aretes* ‘earrings’). But, as recent evidence shows, differences in both the grammatical and lexical domains of Spanish appear to be leveling out in the second generation (see Otheguy & Zentella 2012 for this evidence with respect to the use of the overt subject personal pronoun; Zentella 1992 for this evidence with respect to Spanish lexical vocabulary). It is proposed, then, that it is with more consciously controllable linguistic features, such as English lexicon, that second generation Spanish speakers, particularly Puerto Ricans, in part maintain or emphasize ethnic distinctions. This interpretation, that borrowing may be a marker of ethnic identity in the second generation, is consistent with Poplack’s (1987: 67, 71) observation that bilingualism (and in turn the use of English origin lexical material) is emblematic of Puerto Rican identity in New York City. More specifically, results of the current study suggest that, as an index, speech mixture, and in particular lexical borrowing, may show signs of *recursivity* (Irvine & Gal 2000). That is, the difference that it marks between New York Puerto Ricans, on the one hand, and island Puerto Ricans and New York Anglophones, on the other, may be projected **within** the Spanish-speaking community, to mark differences between Spanish speakers with ties to different nations.

---

88 percent of Puerto Ricans (n=7) use mostly English in their daily lives (Appendix E, Table E.9). Eighty-nine percent of Colombians (n=8) do as well. Yet, second-generation Colombians have very low borrowing rates. In the first generation, 77 percent of Puerto Ricans (n=13) are frequent English users (Appendix E, Table E.10). A majority of Ecuadorians (59 percent, n=10) also use mostly English for daily activities. Ecuadorians, nevertheless, have some of the lowest borrowing rates in the first generation. In sum, then, the fact that Puerto Ricans borrow most in both generations cannot be attributed directly to the fact that a large proportion of Puerto Ricans of both generations are proficient in, confident in or frequent users of English. Other national groups of which the same could be said, nonetheless, borrow infrequently.

### 3.3.2 Interpreting the effect of English confidence on borrowing rate in the second generation

At the data collection phase of this study, informants were asked to rate their skills in English on a four-point scale as ‘poor’, ‘passable’, ‘good’ or ‘excellent’. It is hypothesized that among second generation informants, this variable reflects not an informant’s proficiency in English, but whether he feels his English resembles some idea of good, proper or correct English.<sup>25,26</sup> This is because second generation informants are all fluent in English. Even those that came during later years of their childhood, from eight to twelve years old, had all been in the U.S. for at least ten years at the time of their interview and had received at least six years of formal education in the U.S. Analysis of the four-factored English confidence variable shows that it was most powerful as a two-factored variable: those with the most confidence in their English and those with less.

Analysis revealed that, in the second generation, those with less confidence borrow less frequently than those with the most confidence. Furthermore, the average borrowing rate of those with less confidence is farther from the total mean (a difference of -3.0/k) than that of the highest confidence group (a difference of +1.3/k). This fact, in conjunction with the fact that all second generation informants borrow, suggests that lack of confidence in one’s language abilities inhibits borrowing. Perhaps those that have less confidence attempt to forestall negative perceptions of their speech by avoiding the use of foreign-origin words in their Spanish.

---

25 Also explored was the idea that the English skills variable, for these fluent English speakers, may be indicative of one’s language preference or cultural orientation. In other words, I postulated that when a second generation informant gives himself a low rating for English skills, it may be a move to align himself more with Spanish or to identify as a Spanish speaker. If that were the case, then it would not be surprising to find that informants that rated themselves poorly in English (i.e. manifesting a strong allegiance to Spanish) rated themselves better in Spanish. However, this allegiance interpretation also could not be supported, since among the eight second generation informants that rated themselves poorly in English, three actually give themselves an even lower rating in Spanish, four evaluated their Spanish to be at the same level as their English and only one rated their Spanish as superior to their English. The interpretation of the English skills variable as linguistic confidence or linguistic insecurity for second generation informants seems the most tenable.

26 Similarly, Otheguy and Zentella (2012: 97-98) interpret the Spanish skills variable among the first generation as an indication of linguistic confidence and not an indication of an informant’s proficiency in Spanish.

### 3.3.3 Interpreting the effect of daily Spanish use on borrowing rate in the second generation

It was seen that frequent Spanish users borrow less frequently than those that do use Spanish infrequently for daily activities. In Chapter 4 (section 5.2), two interpretations of this finding were considered, but evidence to support them was not forthcoming or complete. For instance, it was proposed that daily Spanish use could be interpreted as *language attrition*. That is, it was thought that borrowing was more frequent for those that use Spanish less because perhaps these informants were forgetting Spanish words and phrases for expressing particular concepts, which in turn led to reliance on English for expressing those concepts. It was further proposed that if that were the case, these informants would at first make an effort to use Spanish before deciding on an English expression and that, as a result, their borrowings would be flagged or recast in Spanish more frequently. That is, their borrowings should occur with hesitations, false starts and metalinguistic commentary to a greater extent than those that do use Spanish regularly. Else, their borrowings should be rephrased in Spanish to a greater extent. But evidence of more frequent recasting or flagging by those that use Spanish infrequently was not found.<sup>27</sup>

An alternative account of daily Spanish use findings relied on the notion of *language activation*. It was hypothesized that if those that use Spanish infrequently borrow more simply because English is more active for them (thus allowing quicker access to English words and phrases to express themselves), their borrowing inventory should include at least as many words that are easily rendered in Spanish as are not. However, an examination of the inventory of infrequent Spanish users showed that their borrowing inventory is in fact dominated by phrases

---

27 It is also not the case that there is any difference in flagging or recasting in the second generation. In the second generation, both those that use Spanish regularly and those that use it infrequently flag their borrowings to the same extent ( $M=15\%$ ,  $F(1, 46) = 0.03$ ,  $p<.875$ ) and both groups recast their borrowings in Spanish to the same extent ( $M=6\%$ ,  $F(1, 46) = 0.02$ ,  $p<.892$ ).

that are collocations in English or that may require wordier expressions in Spanish to communicate them.

It is also conceivable that the effect of daily Spanish use on borrowing frequency may be interpreted differently for different types of speakers. For instance, although it cannot be posited that daily Spanish use is at all related to Spanish proficiency or incomplete acquisition for the first generation, it may be that, for second-generation Spanish speakers, this is the case. That is, it is conceivable that second-generation Spanish speakers that do not use Spanish on a daily basis do not do so **because** they never acquired the resources in Spanish to express certain concepts. However, such an interpretation must be discarded. First, among second-generation Spanish speakers, daily Spanish use and Spanish skills are not significantly correlated ( $r(47) = .16$ ,  $p < .286$ ). Second, Spanish proficiency does not significantly condition borrowing in the second generation (see section 2.1). Third, all second generation informants demonstrated competence in Spanish by having produced lengthy interviews in Spanish. And, finally, if it were the case that second generation informants that use Spanish infrequently had more gaps in their lexicon for which they needed to resort to English, it would be expected that they would be less able than those that use Spanish frequently to render their borrowings in Spanish. An analysis of recasting, however, shows this is not the case. Both groups of second generation informants, those that use Spanish regularly and those that do not, recast their English borrowings in Spanish to the same extent ( $M=6\%$ ,  $F(1, 46) = 0.02$ ,  $p < .892$ ). In other words, even if language attrition or incomplete acquisition is a motivation for borrowing, it does not appear to capture the difference in borrowing between those that use Spanish regularly and those that do not. In light of these facts, then, results for daily Spanish use in the second generation are interpreted similarly to their interpretation for the corpus as a whole (Chapter 4, section 5.2) and for the first generation

(section 2.3.5 of this chapter). It is proposed that for second and first generation Spanish speakers alike, daily Spanish use is an indication of the extent to which individuals are invested in life in the U.S. That is, infrequent users of Spanish, by choice or obligation, use English more often because they are more oriented toward U.S. cultural institutions. This greater involvement leads to a propensity to also use English words and phrases to name the concepts and ideas central to their life circumstances in the U.S., rather than attempt to render these concepts in Spanish.

### 3.4 Regional origin and daily Spanish use best predict borrowing in the second immigrant generation

To determine which of the three significant conditioning variables, regional origin, English confidence and daily Spanish use, best predicts borrowing in the second generation, a regression analysis was performed (Table 5.11).

**Table 5.11**  
**The ranked effects of regional origin, English confidence and daily Spanish use on lexical borrowing rates in the second generation**

	Standardized beta coefficient	p	Tolerance
<b>Daily Spanish use</b>	- .37*	<b>.007</b>	.96
<b>Regional origin</b>	.35*	<b>.016</b>	.94
<b>English confidence</b>	.18	.192	.93
$R^2 = .295$		$F(3, 43) = 5.99$	$p < .002$

Table 5.11 indicates that a model of lexical borrowing that includes these three variables accounts for 30 percent of the variability found in borrowing rates of the second generation ( $R^2=.295$ ,  $p<.002$ ). Of the three, daily Spanish use ( $beta=-.37$ ,  $p<.007$ ) and regional origin ( $beta=.35$ ,  $p<.016$ ) are the best predictors of borrowing frequency. In addition, it is fairly certain

that these results have not been compromised by the problem of colinearity.<sup>28</sup> Tolerance values for these variables are all greater than .90. That is, at least 90 percent of the variability associated with each variable cannot be attributed to other variables in the model.

### **3.5 Variables that do not condition borrowing rate in the second generation: Sex, age, occupational class, level of education and Spanish proficiency**

Neither an informant's sex ( $F(1, 45) = 0.21, p < .674$ ) nor age ( $F(3, 43) = 1.45, p < .242; r(39) = -.50, p < .755$ ) significantly conditions the borrowing rates of second generation informants. Furthermore, occupational class ( $F(1, 43) = 1.80, p < .187$ ) and educational attainment ( $F(1, 45) = 0.86, p < .359$ ) are not significantly correlated to borrowing frequency. The fact that class and education have not been found to influence borrowing among the second generation indicates that the effect of these variables on the corpus as a whole is due mainly to the fact that they are influential among the Latin American immigrants, who comprise a larger proportion of participants in the corpus ( $n=98$ ), and for whom these findings are fairly robust (class  $F(1, 92) = 9.35, p < .003$ ; education  $F(1, 94) = 4.08, p < .046$ ; see sections 2.1, 2.3.1 and 2.3.2). On the other hand, it should be noted that, within the second generation, middle class informants, nonetheless, borrow more (11.9/k) than the working class (9.2/k) and those with more education (i.e. college and graduate school) borrow more (11.4/k) than those with a high school education or less (9.4/k). In other words, these differences reflect borrowing trends in the corpus overall. The fact that the differences were not statistically significant may perhaps be because there are too few informants ( $n=47$ ) in the second generation.

Prior to analysis, it was thought that Spanish proficiency would significantly impact the extent to which second generation informants borrow. It was thought that Spanish speakers that

---

28 See Chapter 4, section 6 for a discussion of colinearity.

rated themselves as less proficient would do so because they perceived gaps in their knowledge of Spanish, in particular with respect to their knowledge of Spanish vocabulary. It was, therefore, thought that those with lower proficiency ratings in Spanish would borrow from English more often than those that rated themselves as more capable in Spanish. However, it was found that Spanish proficiency **does not** significantly condition borrowing frequency among the second generation ( $F(1, 45) = 1.31, p < .258$ ). That is, second-generation Spanish speakers that rate themselves as excellent in Spanish borrow as much as second generation informants that rate their Spanish as 'poor' or only 'passable'.

### **3.6 Summary of lexical borrowing in the second immigrant generation**

This section has revealed that sex, age and arrival age play no role in borrowing frequency in the second immigrant generation ( $n=48$ ). Men and women borrow to the same extent, as do the old and the young. Likewise, those born in the U.S. and those that came to the U.S. as children (from four to twelve years old) borrow to the same extent. Additionally, it was seen that class and level of education do not significantly influence borrowing in the second generation. However, it is the case that the trend found in the corpus overall still obtains: those of higher occupational classes and with more education borrow more. Furthermore, Spanish proficiency plays no role in borrowing rate differences. That is, individuals with better skills in Spanish borrow as much as those that rate themselves as less proficient.

On the other hand, borrowing in the second generation is significantly correlated with regional origin, English skills and daily Spanish use. Particularly, those with low confidence in English borrow less, while those that use Spanish infrequently borrow more. This latter finding was hypothesized to indicate, **not** that second generation informants lack resources in Spanish, but that daily Spanish use is an indication of the extent that individuals are involved and oriented

toward life in the U.S. Furthermore, Puerto Ricans and Dominicans borrow most (about 15/k and 13/k, respectively) and Mexicans borrow least often (about 7/k). This finding partially corroborates lay perceptions about who is most responsible for so-called language mixing, since this correlation is only evidenced in the second generation. As a result, and in light of evidence that suggests that grammatical differences in the Spanish of the first generation are disappearing in the second, it was hypothesized that, for the second generation, lexical borrowing may be a means to express ethnic differences. Of the three influencing variables, regional origin and daily Spanish use best account for variation in borrowing rates.

#### **4. Conclusion**

This chapter examined borrowing frequency within the first generation and second generation. Results and their interpretations have led to several conclusions about lexical borrowing among Spanish speakers in New York City. First, not only is lexical borrowing a sensitive sociolinguistic index, but it may serve different functions for different groups of individuals. This is supported by the fact that borrowing frequency is conditioned by a distinct set of variables within each immigrant generation. In the first generation, borrowing frequency seems to be a marker of class status. In the second, it may operate as a marker of ethnic identity.

Second, borrowing is not a deficit behavior. This conclusion is supported by the fact that borrowing in the first generation is more frequent among those of higher class status, more education and better English skills, and among those that have more confidence in their language abilities in both generations. It is further supported by the discovery that borrowing could **not** be connected with lack of proficiency in Spanish or English in either the first or second generation.

Third, among first generation immigrants, lexical borrowing from English may be part of a community-wide norm for speaking Spanish in New York City. This was supported by the fact

that among first generation informants, there is no distinction in borrowing rate for informants of different regional and national origins nor between immigrants that have only recently arrived and those that are long-time residents. In other words, regardless of the length of his sojourn, whether one year or ten, Spanish-speaking migrants use English borrowings to the same extent. That borrowing is part of speaking Spanish in New York City is further supported by finding that half of the vocabulary that recent immigrants use in borrowing is used by long-time residents. The other half, although not shared among other informants in the corpus, appears to consist of names for words or concepts that would nonetheless be in common use among those who share the same employment, education or living experiences as recent migrants.

Fourth, proficiency in English facilitates borrowing, but proficiency in English is not required for borrowing to occur. This was supported by the fact that even monolingual Spanish speakers borrow.

Fifth, in New York City, borrowing is for the linguistically secure. For both generations, borrowing is less frequent among those that have a diminished sense of their linguistic abilities. That is, borrowing appears to be inhibited for first generation Spanish speakers that have a low opinion of their Spanish and second generation individuals that have a low opinion of their English, despite being fluent in these languages. It is proposed that the linguistically insecure refrain from extensive borrowing, a behavior that is overtly stigmatized in the city, in order to avoid negative perceptions of their speech.

Finally, in both generations, borrowing may be an indication of how invested in life in the U.S. an individual is. This was supported, in Chapter 4 (section 5.2), when it was seen that infrequent Spanish users (i.e. those that use English more) use a large quantity of borrowings that may be perceived to name things particular to the sociocultural context of New York City.

## **Chapter 6: Innovation, reproduction and the dissemination of lexical borrowings in Spanish in New York City**

### **1. Introduction and questions**

This chapter investigates evidence for a city-wide norm in the selection of borrowing vocabulary or whether the tendency to use innovative and reproductive vocabulary is restricted to subsets of Spanish speakers in New York. It is posited that the corpus frequency of a word reflects its degree of social incorporation in the Spanish-speaking community at large. This supposition is based on the findings of Poplack et al. (1988). They found that widespread items, that is, items used by several participants in their corpus, tended to have a long history of attestation in Canadian or international French.<sup>1</sup> In the current investigation, it is likewise proposed that vocabulary items shared by many informants of the OZC have gained a degree of acceptance in Spanish in New York and that those used by only a single informant are more recently introduced in the community. In turn, it is supposed that the individuals that use this latter type of borrowing vocabulary are most responsible for introducing English-origin words into Spanish. That is, they are borrowing innovators. The questions investigated in this chapter are:

---

1 The length of history of attestation is determined by a borrowing's having been recorded in written sources, such as dictionaries (Poplack et al. 1988: 52).

- (i) How do informants make use of shared and nonshared vocabulary in borrowing? More specifically:
  - a. Is it the case that shared vocabulary makes up a greater proportion of informants' borrowing inventories than nonshared vocabulary? This may indicate a tendency to borrow with respect to a community-accepted repertoire of borrowings.
  - b. Do informants tend to use nonshared items in equally high proportions as shared items? This may indicate that the introduction of new vocabulary is an acceptable norm for borrowing in New York City.
  - c. Which groups appear to be most responsible for introducing novel vocabulary when borrowing?
  - d. Which groups appear to be the most responsible for the dissemination of lexical borrowings throughout the community?

The results of the current analysis show that the use of novel vocabulary is part of a community-accepted pattern for borrowing in the city. Novel vocabulary accounts for over 70 percent of borrowings in the OZC. Additionally, nonshared and shared vocabulary represent approximately equal portions of individuals' borrowing inventories. This indicates that lexical borrowing in New York is very personalized (in that each person, in part, uses unique borrowings) as well as evidences cohesion (in that each person draws also on a stock of established vocabulary). Furthermore, the analysis reveals that the same variables (i.e. Spanish daily use, English skills and occupational class) condition the use of both shared and nonshared vocabulary. It should be recognized that the measures for shared and nonshared vocabulary are not inverses of each other. Thus, this pattern of results was not a foregone conclusion.<sup>2</sup> In some

---

<sup>2</sup> See Chapter 3 (section 4.2) for more on how these rates were calculated.

cases, however, a group's low shared rate is a function of the fact that they use so much more nonshared vocabulary (e.g. see findings for class in section 4.2 of this chapter).

## 2. Shared and nonshared vocabulary in Spanish in New York

The unit of analysis in this chapter is the *lexeme*, not the borrowing. Two types of lexemes are examined in this chapter: shared and nonshared lexemes.<sup>3</sup> *Shared lexemes* are words and phrases that are used by five or more speakers. *Nonshared lexemes* are words used by only one informant. In addition, a third type of lexeme was identified; *periodic lexemes* are used by two, three or four informants. Periodic lexemes will **not** be examined in this chapter. As a review of the discussion from Chapter 3 (sections 4.1.2 and 4.2), lexemes can be thought of as the vocabulary words that informants draw upon in order to borrow. More specifically, a *lexeme* is a group of word forms with a related meaning (Lyons 1977: 19). Thus, the borrowings *janguear* 'to hang [out]', *janguées*, *hanging*, *janguéo* and *janguéamos* are five instantiations of a single lexeme. We may refer to the lexeme using a citation form such as *JANGUEAR*. If the lexeme *JANGUEAR* were instantiated in the speech of only a single informant, that is, if *janguear*, *janguées*, *hanging*, *janguéo* and *janguéamos* were all uttered by a single informant and no other instantiation of *JANGUEAR* appeared in the OZC, *JANGUEAR* would be classified as a nonshared lexeme or vocabulary word. If the borrowing tokens *janguées*, *janguear*, *hanging*, *janguéo* and *janguéamos* were used by five different informants, the lexeme *JANGUEAR* would be classified as a shared lexeme.

---

3 In this investigation, *shared* vocabulary is akin to Poplack et al.'s (1988) categories of *widespread* (used by ten or more speakers) and *recurrent* borrowings (used more than ten times, but not necessarily by as many speakers). Nonshared vocabulary roughly corresponds to Poplack et al.'s (1988) *idiosyncratic* and *nonce* categories. (*Nonce* words are used only one time in the corpus. *Idiosyncratic* words are used more than once, but by only one speaker.) Note, however, that the operational definition of shared vocabulary used in this dissertation does not correspond to Poplack et al.'s (1988) definitions. This is because Poplack et al.'s (1988: 49) database was five times as large (over 19,500 tokens) as the current one (4,507 tokens). The definitions use in the current study have been adjusted for the smaller size of our lexical borrowing database.

In the foregoing discussion, the term *lexeme* has been illustrated using a vocabulary word, that is the verb *JANGUEAR*, for which inflectional forms are numerous. Note, however, that some lexemes are instantiated by a much smaller set of inflectional forms – perhaps only one. For instance, lexemes that perform functional roles in a sentence, such as conjunctions (e.g. *and*) or discourse markers (e.g. *however*), are typically instantiated with only a single, invariant form.

Typically, each word in a lexical borrowing is considered to instantiate one lexeme. Thus, the borrowing *negative attitude* contains two lexemes: *NEGATIVE* and *ATTITUDE*. However, it should be clear that, in the present research, multiple-word proper nouns, collocations and discourse markers are **not** decomposed into smaller words when counting lexemes. That is, borrowings like *New York*, *I don't know* or *cap and gown*, for example, instantiate one lexeme each. Chapter 3 (section 4.2) further elaborates on the means used to decide whether a borrowing is (or is not) decomposed into smaller words.

When counting word frequency, authors often talk in terms of *tokens* and *types*. However, Lyons notes that, sometimes, when authors talk about types, what they have in mind are, in fact, lexemes (Lyons 1977: 20). To illustrate, a frequency count of the borrowings *principal*, *principal* and *principals* could be summed up by saying that a speaker has used three word tokens, **one** type and **one** lexeme. For Lyons, however, the concept of *type* applies to word-forms, that is, written (or, as in the present investigation, spoken) strings. For Lyons, then, if a person uttered *principal*, *principal* and *principals*, they would have used three word tokens, **two** word types (*principal* and *principals*) and **one** lexeme (*PRINCIPAL*). In this investigation, I utilize Lyons' terminology. Thus, in the overview of shared and nonshared vocabulary in the OZC supplied in Table 6.1, findings are presented in terms of (word) tokens and lexemes. The reader should freely substitute the term *type* for *lexeme*, if he is used to talking about frequency counts in those

terms. Be aware that, in this chapter, I use the word *type* to mean *corpus type*, that is, whether a lexeme is classified as shared or nonshared.

**Table 6.1**  
**Shared and nonshared vocabulary in the OZC**

	Nonshared		Shared		Total*	
	n	%	n	%	n	%
<b>Tokens</b>	1592	<b>32</b>	2244	<b>45</b>	5041	<b>100</b>
<b>Lexemes</b>	1261	<b>73</b>	92	<b>5</b>	1737	<b>100</b>

\* Percent totals do not add up to 100 percent because the table does not include all corpus types. In addition to shared and nonshared, some borrowing vocabulary is classified as *periodic* (i.e. vocabulary used by two, three or four informants). Periodic vocabulary is not examined in this chapter.

Table 6.1 shows that borrowings in the OZC are composed of 5041 word tokens and 1737 lexemes. That is, an inventory of 1737 words and expressions are instantiated 5041 times by 146 informants. Of these 5041 word tokens, almost half (45 percent) are shared among five or more informants. Another substantial portion of tokens, 32 percent, are vocabulary words used by a single speaker. The trend in terms of tokens is toward a slightly greater use of shared items. Table 6.2 provides a list of some of the most frequent lexemes borrowed. Note that, for the rest of this chapter, the citation form of a lexeme is given in *lower case italics*, and not in *ITALICIZED SMALL CAPS*.<sup>4</sup>

4 *ITALICIZED SMALL CAPS* are used for the citation form of a lexeme only when *lexeme* is being defined or exemplified in exposition, such as in Chapter 3 (section 4.4.2) and section 2 of this chapter. The rest of the time, the citation form of a lexeme is typically provided in *italicized lower case* fonts (as are word-forms). Note that Lyons (1977) uses single quotes ( ‘ ’ ) to indicate the citation form of a lexeme.

**Table 6.2**  
**Most frequent lexemes used to borrow in the OZC**

<b>lexeme</b>	<b>n</b>	<b>lexeme</b>	<b>n</b>	<b>lexeme</b>	<b>n</b>
<i>you know</i> .....	363	<i>anyway</i> .....	20	<i>tax</i> .....	14
<i>so</i> .....	269	<i>baby</i> .....	20	<i>GED</i> .....	13
<i>yeah</i> .....	134	<i>bill</i> .....	20	<i>major</i> .....	13
<i>high school</i> .....	126	<i>building</i> .....	20	<i>well (DM)</i> .....	13
<i>New York</i> .....	64	<i>superintendent</i> ....	20	<i>bachelor</i> .....	12
<i>CD</i> .....	47	<i>weekend</i> .....	19	<i>basement</i> .....	12
<i>New Jersey</i> .....	47	<i>business</i> .....	18	<i>Fort Washington</i> .	12
<i>wow</i> .....	40	<i>master[ 's]</i> .....	18	<i>kindergarten</i> .....	12
<i>college</i> .....	38	<i>City College</i> .....	17	<i>social</i> .....	12
<i>I mean</i> .....	34	<i>like (DM)</i> .....	17	<i>tour</i> .....	12
<i>man (discourse marker)</i> ....	30	<i>mall</i> .....	17	<i>truck</i> .....	12
<i>restaurant</i> .....	30	<i>boss</i> .....	16	<i>year</i> .....	12
<i>the</i> .....	30	<i>right</i> .....	16	<i>break</i> .....	11
<i>Long Island</i> .....	29	<i>that's it</i> .....	16	<i>downtown</i> .....	11
<i>(oh) my God</i> .....	27	<i>Central Park</i> .....	15	<i>email</i> .....	11
<i>Queens College</i> .....	25	<i>Nuyorrican</i> .....	15	<i>ESL</i> .....	11
<i>janguear</i> .....	24	<i>ticket</i> .....	15	<i>Halloween</i> .....	11
<i>and</i> .....	22	<i>man (n)</i> .....	14	<i>Hunter College</i> ....	11
<i>junior</i> .....	22	<i>Play Station</i> .....	14	<i>welfare</i> .....	11
<i>subway</i> .....	21	<i>school</i> .....	14		

As can be seen from Table 6.2, the most frequent individual borrowings are the discourse markers *you know*, *so* and *yeah*, each with over one hundred tokens a piece. The next most frequent lexeme is *high school*, with 126 occurrences, followed by *New York*, with 64 occurrences. It is of note that, of the top ten most frequently-uttered lexemes, half are discourse markers. This seems to present a challenge the current wisdom that nouns are the most frequently borrowed lexical category. It appears to be the case that, in aggregate, nouns outnumber other lexical categories. It may also be the case that the lexemes that eventually become part of a recipient language lexicon tend to nouns. However, the generalization does not hold for individuals tokens uttered in speech. As Table 6.2 demonstrates, as individual borrowings go, a given discourse marker is likely to be far more frequent than any given noun.

Returning to Table 6.1, when borrowing vocabulary is presented as an inventory of unique words and expressions, namely, in terms of lexemes, shared vocabulary represents a significantly reduced proportion of the inventory. Of 1737 different words used in the OZC, 73 percent is used by only a single informant. Only five percent of lexemes is shared among informants.<sup>5</sup> The importance of nonshared vocabulary for borrowing is also apparent in the borrowing inventories of individual informants. Table 6.3 shows the number of informants that use shared and nonshared vocabulary.

**Table 6.3**  
**Informants' preferences for shared or nonshared vocabulary**

		<b>n</b>	<b>%</b>
<b>Majority nonshared</b>	<i>Only nonshared</i>	6	4
	More nonshared than shared	66	45
<b>Majority shared</b>	<i>Only shared</i>	16	11
	More shared than nonshared	35	24
<b>Equal proportion shared &amp; nonshared</b>		16	11
<b>No borrowings</b>		7	5
Total informants		146	100

Table 6.3 illustrates that, of the 146 informants, only seven do not borrow at all (second row from the bottom). The table also shows that just 15 percent of informants (n=22) uses **only** shared or **only** nonshared vocabulary to borrow (4% + 11% = 22%). The majority of informants

5 It may occur to the reader that this results for shared vocabulary stem, in part, from the definition of lexical borrowing used in this study. That is, it was required that lexical borrowings **not** be present in heritage or precontact varieties of Spanish (i.e. words that may have been a part of Spanish prior to contact in the U.S. were excluded). Were precontact items and international loanwords included in this study, they would likely have evidenced higher frequency of use and would have most likely been classified as shared vocabulary. As a result, shared vocabulary might have represented a larger portion of the lexical borrowing database. However, should all precontact items have been included in the database (adding perhaps 2000 more borrowing tokens), the large margin of difference between shared and nonshared lexemes indicates that, in terms of the borrowing inventory of the OZC, there would still have been a substantial portion of nonshared vocabulary words.

uses **both** shared and nonshared vocabulary (45% + 24% + 11% = 80 percent). There are, however, slightly more informants that prefer nonshared vocabulary (4% + 45% = 49 percent) than those that prefer shared vocabulary (11% + 24% = 35 percent). It now remains to be seen whether individuals' personal traits or membership in particular groups affects their preference for shared or nonshared vocabulary.

In the rest of this chapter, the presentation and organization of results differ from the previous two chapters. Significant findings are presented and discussed first. This is followed by the results of regression analysis. Nonsignificant results are discussed last. The discussion of results relies primarily on Poplack et al. (1988), whose work is the conceptual basis of the present chapter. It should also be noted, however, that in one important way, this study diverges from that of Poplack et al. Whereas Poplack et al. included English-origin words that had become part of the lexicon of international or Canadian French (1988: 52-55), the present study has spent considerable effort excluding such items (see Chapter 3, section 4.2). Furthermore, Poplack et al. examined only single words of English origin (1988: 53). In the present study multiple-word English-origin strings are included as borrowings (Chapter 3, section 4.1). These differences in data selection are what most likely give rise to findings of this study that diverge with respect to theirs (see section 4.2, for example).

### **3. The use of nonshared vocabulary: Lexical borrowing innovators**

A rate for nonshared vocabulary was calculated for each informant. An informant's nonshared rate represents the proportion of vocabulary words he uses that are unique to him. More precisely, it is the proportion of *lexemes* the informant uses in borrowing that he and no other informant uses. It is given as a percentage of all lexemes used. The results of ANOVA analysis, presented below, show whether the proportion of nonshared lexemes used by one subgroup

differs from the proportion of nonshared lexemes used by another. In interpreting results, one subgroup's statistically higher nonshared rate is taken to indicate that it has, comparatively speaking, a greater tendency for innovation in borrowing. Furthermore, groups that are seen to use a larger proportion of nonshared vocabulary are discussed as if they are most responsible for the introduction of novel English-origin vocabulary into Spanish. However, it should be noted that the identification of nonshared vocabulary only approximates what would be the inventory of truly novel introductions into the Spanish-speaking community of New York City. This is because, although some English vocabulary words may not be frequent in the OZC, they may nonetheless be commonly used for certain topics of conversation that did not come up frequently during OZC interviews. Or, some English vocabulary may be commonly used by groups of Spanish speakers that were not well-represented in the OZC (e.g. individuals in a specific profession).<sup>6</sup>

### **3.1 Variables that significantly condition nonshared rate: Occupational class, daily Spanish use and English skills**

For the corpus as a whole, nonshared rate is significantly influenced by occupational class ( $F(1, 132) = 4.12, p < .044$ ), daily Spanish use ( $F(1, 136) = 7.62, p < .007$ ) and English skills ( $F(1, 136) = 8.91, p < .003$ ).<sup>7</sup>

---

6 For examples of lexemes coded as nonshared that are probably not novel introductions into Spanish, see section 3.4 (this chapter), where non-significant results for years in the U.S. are discussed.

7 When all informants are taken together, nonshared rate is not significantly conditioned by ethnonational affiliation ( $F(5, 133) = 0.65, p < .662$ ), region ( $F(1, 137) = 0.30, p < .585$ ), sex ( $F(1, 137) = 1.52, p < .219$ ), age ( $F(3, 135) = 0.34, p < .768$ ;  $r(138) = 0.05, p < .575$ ), arrival age ( $F(3, 135) = 0.95, p < .418$ ;  $r(138) = -0.01, p < .920$ ), immigrant generation ( $F(2, 136) = 1.43, p < .243$ ), years in the U.S. ( $F(2, 88) = 0.25, p < .776$ ;  $r(138) = 0.04, p < .669$ ), level of education ( $F(1, 136) = 0.98, p < .322$ ) or Spanish skills ( $F(1, 136) = 2.48, p < .117$ ).

Table 6.4 shows ANOVA results for class.

**Table 6.4**  
**Nonshared rate by occupational class**

	Mean (%)	SD
<b>Middle class</b> (n=73)	38.7	18.7
<b>Working class</b> (n=61)	31.5	22.3
<b>Total</b> (n=134)	35.5	20.7
$F(1, 132) = 4.12$		$p < .044$

Table 6.4 shows that, on average, 35 percent of informants' borrowing vocabulary is nonshared. The table also indicates that middle class informants use a greater proportion of nonshared vocabulary (39 percent) than the working class (31 percent).

Results for daily Spanish use are in Table 6.5.

**Table 6.5**  
**Nonshared rate by daily Spanish use**

	Mean (%)	SD
<b>Infrequent Spanish use</b> (n=55)	41.5	18.3
<b>Regular Spanish use</b> (n=83)	31.8	21.1
<b>Total</b> (n=138)	35.7	20.5
$F(1, 136) = 7.62$		$p < .007$

Table 6.5 shows that Spanish speakers that do not often use Spanish in their daily lives utilize nonshared vocabulary to a greater extent (41 percent) than regular Spanish users (32 percent).

The effect of English skills on the use of innovatory vocabulary in borrowing is detailed in Table 6.6, which provides the results of an ANOVA.

**Table 6.6**  
**Nonshared rate by English skills**

	Mean (%)	SD
<b>Excellent</b> (n=46)	42.8	15.6
<b>Non-excellent</b> (n=92)	32.1	21.8
<b>Total</b> (n=138)	35.7	20.5
$F(1, 136) = 8.91$ $p < .003$		

Table 6.6 shows that the inventory of excellent English speakers contains a larger proportion of nonshared words (43 percent) than that of non-excellent English speakers (32 percent).

### **3.2 Discussion of class, daily Spanish use and English skills as variables that condition innovation in lexical borrowing**

Middle class informants use more nonshared lexemes while borrowing than the working class. If nonshared items approximate the stock of what are truly novel introductions into Spanish, it may be said that middle class Spanish speakers have a greater propensity than the working class to introduce new borrowings. In other words, lexical borrowing of the most noticeable sort, that is, the introduction of novel English-origin vocabulary, is most advanced among those of higher social classes. In fact, a comparison of the inventory of nonshared items used by the middle class and working class seems to confirm this impression. The inventory of the middle class (n=896) is almost three times the size of that used by the working class (n=326). These findings bolster conclusions reached in the previous chapter. In particular, they strengthen the impression that, not only is borrowing by Spanish speakers in New York **not** a deficit behavior, but that it is a behavior licensed most for those of privileged classes. These results do not reflect those of

Poplack et al., who found that French-speaking Canadians of the working and the middle class do not differ with respect to the types of borrowings they use. That is, middle class and working class Canadian Francophones borrow approximately equal proportions of nonshared vocabulary (*nonce vocabulary* in their parlance). However, our findings for class **do** parallel what Labov (2001: 360-363) has found for the introduction and spread of phonological change in Philadelphia. In particular, Labov says that although it may be impossible to locate the innovators of change (i.e. those that use a linguistic form for the first time), they are likely to be “marginal” individuals “with high communication scores” (362). That is, innovators communicate with many members of a group while also maintaining a high number of contacts with people outside of the group. The middle class of this study seem to match this profile; in general, they are more likely than the working class to use English on a daily basis and, thus, to interact with English speakers.<sup>8</sup>

English skills is also positively correlated to the use of nonshared vocabulary. That is, those that rated their English better use a larger than average proportion of innovative vocabulary. Results for English skills are interpreted differently for first and second generation informants (as was done in Chapter 5). In the first generation, the English skills measure is a straightforward indicator of informants’ proficiency in English. The results of Table 6.6, with respect to the first generation, then, corroborate findings from Poplack et al. (1988). They found that, in Canada, it was the highly bilingual speakers that used more *nonce* borrowings (1988: 84). Thus, although Chapter 5 (section 2.4) revealed that English proficiency was not the **primary** factor conditioning how often an immigrant borrows, it is a significant predictor of the extent to which

---

8 While 85 percent (n=63) of middle class informants appear to use at least some use of English on a daily basis (that is, they report lesser degree of Spanish use on a daily basis), only 53 percent (n=34) of working class informants do.

novel words are introduced while borrowing. Said differently, lexical borrowing innovators in the first generation are Spanish speakers that have high proficiency in English, who are, in other words, more bilingual.

For the second generation, the English skills variable cannot be interpreted as indicative of proficiency since they all were raised in the U.S. and all had **at least** 10 years of formal schooling in English. As in Chapter 5 (section 3.1 and 3.3.2), this variable is interpreted, for the second generation, as an indication of linguistic confidence or insecurity.<sup>9</sup> In light of this, results in Table 6.6 show that second generation informants who downplay their linguistic abilities or who feel less optimistic about them are also slightly less inclined to risk using vocabulary that may be new or unfamiliar to their interlocutor. Conversely, results show that those with the most confidence more often play the role of innovators in borrowing.

Finally, infrequent Spanish users also tend, to a greater extent than regular Spanish users, to be borrowing innovators. In other words (and assuming that nonshared vocabulary is a good proxy for truly novel introductions), English-origin innovations in Spanish are introduced by those that use Spanish least often, that is, those that use English most. In sum, findings up to this point show that familiarity with and daily use of English, although not prerequisites for borrowing, are important for promoting borrowing, in terms of both frequency of use (Chapter 5, sections 2.3.3 and 2.3.5, also 3.3.3) and propensity for innovation (Table 6.5).

---

9 In all reality, there is probably a correlation between confidence and some types of linguistic competence, such as mastery of sophisticated vocabulary. Still, even if this is the case, it would be hasty to interpret English skills results for the second generation in terms of English competence / proficiency due to the nature of lexical borrowings in the LBD. The English vocabulary in the LBD is extremely common, accessible to elementary school children. That is, both the vocabulary called shared and that called nonshared are of a basic sort, requiring no special or advanced study to be learned or used by anyone raised in New York City. So, even if confidence does imply competence, there is no necessary reason for there to be a significant correlation between nonshared vocabulary and English competence for completely fluent English speakers. Clearly, then, an interpretation of English skills as English proficiency for the second generation must be put aside.

### 3.3 English skills best predict innovative borrowing

Results of a regression including the variables of English skills, daily Spanish use and occupational class are in Table 6.7.

**Table 6.7**  
**The ranked effects of occupational class, English proficiency**  
**and daily Spanish use on nonshared rate**

	Standardized beta coefficient	p	Tolerance
<b>English skills</b>	.18	<b>.073</b>	.74
<b>Daily Spanish use</b>	- .13	.167	.76
<b>Occupational class</b>	.12	.166	.94
$R^2 = .101$ $F(3, 129) = 4.81$ $p < .003$			

Table 6.7 shows that, together, the three variables account for 10 percent of variation in the use of novel vocabulary among OZC informants ( $R^2=.101$ ,  $p<.003$ ). Although the model as a whole is significant ( $p<.003$ ), only English skills significantly predicts (or nearly so) variation in the nonshared rate ( $beta=.18$ ,  $p<.073$ ). That is, one's confidence or proficiency in English is the strongest indication of how much novel English vocabulary one is likely to introduce in Spanish.

### 3.4 Discussion of variables that do not condition innovation in lexical borrowing

The use of nonshared vocabulary is not influenced by immigrant generation, years in the U.S., education level, age, sex or Spanish skills. The pronounced distinction in borrowing frequency according to an informant's immigrant status (Chapter 4) has led to positing two distinct subgroups of Spanish speaker in New York City (Chapter 5): the first generation and the second. It was expected that differences in borrowing behavior according to immigrant generation would be apparent in all aspects of borrowing behavior, including one's preference for novel

vocabulary. In particular, given that second generation informants are generally more proficient in English, it was expected that they would use a larger proportion of novel vocabulary. ANOVA results, however, do not show this to be the case (arrival age  $F(3, 135) = 0.95, p < .418$ ;  $r(138) = -.01, p < .920$ ; immigrant generation  $F(2, 136) = 1.43, p < .243$ ). Innovative vocabulary is used to the same extent by both first and second generation informants (about 35 percent of a speaker's inventory).<sup>10</sup> This fact suggests that, despite the sharp stratification of first and second generation informants with respect to borrowing frequency, both groups conform to a pattern of borrowing that consists of utilizing nonce lexemes for about a third of their borrowing vocabulary.

Ethnonational affiliation ( $F(5, 133) = 0.65, p < .662$ ) and region ( $F(1, 137) = 0.30, p < .585$ ) also do not condition nonshared rate. That is, Puerto Ricans, Mexicans, Dominicans, etc. all borrow using about equal proportions of novel vocabulary while borrowing.

Nonshared rates are not significantly correlated with an informant's years in the U.S. ( $F(2, 88) = 0.95, p < .909$ ;  $r(138) = .04, p < .669$ ). This finding contradicted expectations. It was thought that the longer one was in the U.S., the more exposure to English one would have and, thus, the greater proportion of novel vocabulary he would use. This is, however, not the case. Regardless of whether an individual has only just arrived, or whether he has been in the U.S. for decades, first generation Spanish speakers use the same proportion of nonshared (and as will be seen, shared) vocabulary. This, on hindsight, is perhaps reasonable given that as soon as recent arrivals find a job or attend school, they are faced with vocabulary that is English and that is, likely, particular to the context they find themselves in. In fact, several items of nonshared vocabulary

---

10 However, an ANOVA for overall lexeme rate shows that second generation informants also use a more varied stock of lexemes in their interviews than first generation informants ( $F(1, 121) = 47.90, p < .001$ ). The second generation averaged eight different lexemes per thousand words of discourse ( $M=8/k, n=29$ ); the first generation averaged three ( $M=3/k, n=98$ ). Thus, if 35 percent of the inventory of a second generation informant is nonshared vocabulary, this represents a greater actual number of novel vocabulary items than are used by the first generation.

used by recent arrivals are probably not, in fact, novel introductions in Spanish, but just happen to be used by only a single informant of the OZC.<sup>11</sup> For instance, examples of vocabulary used by those with less than two years in the city that were classified as nonshared include: *AT&T*, *Atlantic City*, *Basic* (computer programming language), *big wheel* (child's toy), *Brooklyn Bridge*, *CBS*, *chef*, *Chinatown*, *Clueless* (a movie), *Columbus Day*, *Cross Bronx* (name of a New York highway), *Down's Syndrome*, *drive-through*, *excuse me*, *Forty-two* (in reference to 42<sup>nd</sup> Street), *Golden Girls* (a TV show), *HBO*, *I-95* (immigration paperwork), *IRS*, *layaway*, *massage*, *nickels*, *pedicure*, *popcorn*, *realtor*, *shiftear*, *slogan*, *stand-by*, *TECU University*, *website* and *White Castle* (fast food restaurant). These words are probably regularly used in specific contexts, like work or school, in which recent arrivals find themselves.

Several studies have found education level to be an important determinant not only of borrowing frequency (Eslami Rasekh et al. 2008; Sullivan 2008), but also of the types of borrowings used. For instance, Poplack et al. (1988: 87) found that speakers with secondary education preferred unattested forms. This was due, they hypothesized, to the fact that those with more education have greater exposure to English. Education does not significantly influence the use of nonshared vocabulary in the present study ( $F(1, 136) = 0.98, p < .322$ ). Those with few years of formal schooling and those with college and graduate education use novel vocabulary to the same extent. This may be due to the fact that, where education is related to the amount or type of borrowings used, the recipient language is the primary vehicle of public discourse and

---

11 This is not true of all words coded as nonshared in this study. Other nonshared lexemes appear to be spontaneous introductions by the individual. For example, *between* (002U), *brass* (423P), *bungalows* (434P), *garbage can* (401P), *capitalism* (201U), *challenge* (417P), *championship* (329D), *choreography* (331.1D), *coast to coast* (333D), [apartment] *complex* (401P), *mid-life crisis* (417P), *decline* (326E), *deep inside* (434P), *different* (331.1D), *dinner* (384E), *diversity* (333D), *exciting* (330D), *finally* (401P) and *figure out* (333D). All of these, and many others, are probably not in regular use in any specific context. Furthermore, it is difficult to attribute their use to "not knowing" the Spanish word or phrase since each is easily rendered in Spanish for the native speaker of Spanish.

exposure to the donor language is associated primarily with educational contexts, as a language or topic of instruction. In New York City, the donor language, English, is the de facto language of public transaction. Formal education is just one context of several where exposure to it can be obtained. It may be, then, that education level plays a role in proficiency-related aspects of borrowing, such as the introduction of novel vocabulary, when education itself is a principal source of exposure to the donor language. The fact that education level is positively correlated with borrowing frequency in the first generation (Chapter 5, sections 2.1 and 2.3.2) does not necessarily conflict with this conclusion. This is because evidence from this investigation suggests that borrowing frequency is not, for informants of the OZC, a proficiency-dependent aspect of borrowing.

Several studies have found a relationship between age and borrowing (e.g. Ngom 2006; Poplack et al. 1988). For instance, Poplack et al. (1988) found a weak, but definite, tendency for the young to use novel (i.e. *nonce* in their terminology) borrowings. In New York, there appears to be no such relationship ( $F(3, 135) = 0.34, p < .768$ ;  $r(138) = .05, p < .575$ ). That is, younger individuals introduce as many novel borrowings as older Spanish speakers. It may be, however, that there is a qualitative difference in the nonshared vocabulary used by the young and old. For instance, Poplack et al. (1988: 87-88) found that older individuals used many nonce forms too, but that theirs were archaizing, rather than novel forms. A closer look at the nonshared vocabulary of the old and the young of this study is warranted, but has not yet been undertaken at this juncture.<sup>12</sup>

In addition, sex has no influence on nonshared rates ( $F(1, 137) = 1.52, p < .219$ ). This finding corroborates Poplack et al. (1988), where no significant differences in men and women's use of

---

12 The comparison of novel forms of older individuals and the young requires verification of the dates of attestation of the English-origin forms they use, a task not yet undertaken due to its time-consuming nature.

novel vocabulary were found.<sup>13</sup> Finally, despite the fact that having more confidence in one's Spanish is related to higher borrowing frequencies in the first generation (Chapter 5, sections 2.1 and 2.3.4), there is no relationship between proficiency or confidence in Spanish and the extent to which individuals select novel vocabulary to borrow ( $F(1, 136) = 2.48, p < .117$ ).

### **3.5 Innovative borrowing: Summary and conclusions**

We have seen that, on average, 35 percent of the lexemes that Spanish speakers use to borrow are unique to each speaker. That is, over a third of each informant's borrowing inventory consists of words and phrases not in wide use throughout the Spanish-speaking community in New York City.<sup>14</sup> Additionally, it was seen that innovation in borrowing is conditioned by class and, moreover, by a handful of speaker traits that all relate to familiarity with, confidence in, exposure to or knowledge of English. Borrowing innovators in New York seem to be middle class individuals. Furthermore, they are, regardless of the age at which they arrived, their sex, age or immigrant generation, are Spanish speakers that are more proficient in English, that have more confidence in English and that use English more for daily activities.

### **4. The use of shared vocabulary: Lexical borrowing reproducers**

A rate for shared vocabulary was calculated for each informant. An informant's shared rate is the proportion of the vocabulary he used that is used by five or more informants of the OZC. As with the nonshared rate, shared rate is reported as a percentage of all *lexemes* an informant used.<sup>15</sup> A group's preference for shared vocabulary, that is, words that have gained currency in the

---

13 However, Poplack et al. found a non-significant trend where women used more nonce vocabulary than men (1988: 89), as well as a significant difference between men and women in one neighborhood only, where women used less novel vocabulary than men (1988: 76).

14 Although, as mentioned in section 3.4 of this chapter, a selection of what have been identified as nonshared borrowings may, in fact, be commonly used among specific groups or in specific contexts.

15 For more on this, see section 2 in this chapter. See Chapter 3 (section 4.2) for a full account.

Spanish-speaking community, is interpreted as a tendency for reproductive borrowing behavior. It is not posited that reproductive borrowers are necessarily the principal disseminators of borrowing vocabulary since, as explained in section 5 of this chapter, disseminators must also be those that use Spanish regularly.

#### 4.1 Variables that significantly condition shared rate: Occupational class, daily Spanish use and English skills

In the OZC, shared rate is significantly influenced by an informant's occupational class ( $F(1, 132) = 3.96, p < .049$ ), English skills ( $F(1, 136) = 6.91, p < .010$ ) and daily Spanish use ( $F(1, 136) = 10.98, p < .001$ ).<sup>16</sup> These are the same three variables that condition the use of nonshared vocabulary (section 3.1). However, it should be borne in mind that, since shared and nonshared rates are not inverses of each other (see Chapter 3, section 4.2), the fact that these variables condition both shared and nonshared rates was not an inevitable outcome.

Results for occupational class are in Table 6.8.

**Table 6.8**  
**Shared rate by occupational class**

	Mean (%)	SD
<b>Working class</b> (n=61)	36.7	22.3
<b>Middle class</b> (n=73)	30.2	15.1
<b>Total</b> (n=134)	33.1	18.9
$F(1, 132) = 3.96$		$p < .049$

<sup>16</sup> Shared rate for all informants does not vary with ethnonational affiliation ( $F(5, 133) = 0.78, p < .563$ ), region ( $F(1, 137) = 0.00, p < .980$ ), sex ( $F(1, 137) = 0.80, p < .372$ ), age ( $F(3, 135) = 1.89, p < .133; r(138) = -.09, p < .314$ ), arrival age ( $F(3, 135) = 1.11, p < .349; r(138) = .01, p < .905$ ), immigrant generation ( $F(2, 136) = 1.63, p < .201$ ), years in the U.S. ( $F(2, 88) = 0.25, p < .775; r(138) = -.11, p < .191$ ), level of education ( $F(1, 136) = 2.15, p < .145$ ) or Spanish skills ( $F(1, 136) = 0.36, p < .550$ ).

Table 6.8 shows that, on average, 33 percent of informants' borrowing vocabulary consists of shared items. Additionally, it can be seen that the working class tends to use a slightly greater proportion of shared vocabulary (37 percent) than the middle class (30 percent).

Table 6.9, results for English skills, shows that those that rated their English abilities as poorer opt for shared vocabulary more than those that rated themselves better.

**Table 6.9**  
**Shared rate by English skills**

	Mean (%)	SD
<b>Less good</b> (n=59)	38.1	20.8
<b>Better</b> (n=79)	29.9	16.1
<b>Total</b> (n=138)	33.4	18.6

$$F(1, 136) = 6.91 \quad p < .010$$

In particular, Table 6.9 shows that 38 percent of the inventory of those with lower self-ratings in English are shared items. Only 30 percent of the inventory of those with higher self-ratings in English are shared vocabulary items.

Details for daily Spanish use are in Table 6.10.

**Table 6.10**  
**Shared rate by daily Spanish use**

	Mean (%)	SD
<b>Most Spanish use</b> (n=41)	41.2	22.4
<b>Less Spanish use</b> (n=97)	30.1	15.8
<b>Total</b> (n=138)	33.4	18.6

$$F(1, 136) = 10.98 \quad p < .001$$

In Table 6.10, it can be seen that, just as those that use Spanish less on a daily basis incorporate more nonshared vocabulary (Table 6.5), those that use Spanish most utilize a greater proportion of shared words (41 percent). Infrequent users of Spanish employ shared items for about 30 percent of their borrowing inventory. The effect of daily Spanish use on shared rate is strong. The F value (10.98) indicates that the variation in shared rate between the groups is 11 times larger than the variation within each group.

#### **4.2 Discussion of class, daily Spanish use and English skills as variables that condition reproduction in lexical borrowing**

On average, 33 percent of each informant's borrowing vocabulary is shared. This is not very different from the use of nonshared vocabulary, which, on average, accounts for 35 percent of each informant's inventory.<sup>17</sup> This result differs from that of Poplack et al., who found that widespread borrowings "constitute a large proportion of the average speakers' [sic] borrowed vocabulary" (1988: 83), namely, between 50 and 65 percent (1988: 79).<sup>18</sup> In New York City, however, it appears that novel **and** shared vocabulary are both integral components of a city-wide pattern of borrowing. The difference between Poplack et al.'s findings and those of the present investigation most likely results because of how borrowings were defined in each study. As noted at the end of section 2, Poplack et al.'s study included borrowings that had been attested in Canadian and international French dictionaries (1988: 52). In other words, several of

---

17 The other third of informants' borrowing inventory consists of *periodic lexemes*, that is, vocabulary that was used by from two to four different informants. As mentioned at the beginning of section 2, these types of lexemes are not examined in this dissertation.

18 The difference between Poplack et al.'s findings and those of this study might be attributable to the fact that their investigation included many loanwords that were already part of French lexicon in Canada or an international lexicon of French (1988: 94) and that, consequently, appeared frequently in their corpus. In the present investigation, however, borrowings that were established in a preimmigration Spanish lexicon were excluded (Chapter 3, section 4.2.1) (also see footnote 5 in this chapter). It is likely that had such items been included, they would have accordingly increased the shared rates of informants. Even were this the case, however, it is clear that both novel and shared vocabulary constitute a substantial portion of the corpus as a whole (Table 6.1), and a consistent proportion of the vocabulary utilized by individuals (Table 6.2).

their borrowings had already been established in the lexicon of French at the time of their study. In the present investigation, on the other hand, these kinds of words, that is English-origin, *established loanwords* in Spanish, were carefully excluded from the data set (as detailed in Chapter 3, section 4.2.1). What the findings of Poplack et al. appear to demonstrate, then, is not that Francophones of their study draw upon a local repertoire of English borrowings, but that they speak French like Francophones the world over: using established French lexicon where the message of their discourse demands it. Should these now-French words have been excluded from Poplack et al.'s study, it is possible that they would have found lower overall rates for the use of widespread items. Perhaps their widespread rates would even have been in line with the average shared rate of the current study.

It has also been seen that, just as the middle class appears to be more responsible for the introduction of novel vocabulary while borrowing (Table 6.4), the working class tends, to a greater extent, to use English words and expressions that are already in wide use in the city. That is, they tend to be reproductive borrowers. The finding that the working class uses a higher proportion of shared lexemes appears to be a consequence of the fact that the middle class opts to a greater extent for innovative vocabulary (section 3.2).<sup>19</sup> Together, these findings suggest that the flow of lexical borrowings among Spanish speakers in New York begins with the middle class, a selection of whose borrowings are taken up by the working class. In fact, it turns out that the 80 shared lexemes used by the working class are virtually a perfect subset of the lexemes used by the middle class. Specifically, 99 percent (n=79) of the shared inventory of the working class is also used by middle class, whose shared inventory is about 10 percent larger (n=88).

---

19 I believe that results for shared rate and class are a reflex of results for nonshared rate because both the working and middle class use the same proportion of periodic lexemes (which are not examined in this dissertation). That is, the results of ANOVA for *periodic rate* are  $F(1, 132) = 0.09, p < .759$ .

In addition, English skills influence the extent to which Spanish speakers select shared vocabulary for borrowing. This finding appears to be a reflex of findings for English skills and nonshared rate.<sup>20</sup> As before, this variable must be interpreted differently for first generation and second generation Spanish speakers.

For the second generation, results in section 4.1 show that those with less confidence in English opt for shared items more often than those with more confidence. For the first generation, for whom English skills is an indicator of English proficiency, the less proficient select lexemes that are in wide circulation in Spanish to a greater extent than the more proficient. In other words, less proficiency and less confidence in English is associated with the reproductive use of borrowings. This finding accords with that of Poplack et al., who found that “there is a greater tendency of loanwords to circulate between the lower-proficiency [in English] groups” (1988: 93). Further, the finding supports conclusions reached in the previous chapter: certain aspects of borrowing are less integrally tied to bilingual proficiency than others. For instance, in the previous chapter, it was found that although English proficiency does condition borrowing frequency in the first generation (Chapter 5, sections 2.1 and 2.3.3), it is not a principal predictor (Chapter 5, section 2.4). Likewise, it seems that, while the introduction of novel English vocabulary is best predicted by one’s proficiency in English, the use of widely shared vocabulary is, like overall frequency, less dependent on an individual’s mastery of the donor language. This conclusion also corroborates Poplack et al. They stated that despite claims that most bilinguals tend to be not only the “importers” of innovation but also the “agents of

---

20 I believe that results for shared rate and English skills are a reflex of results for nonshared rate and English skills because both those with higher English skills ratings and those lower English skills ratings use the same proportion of periodic lexemes. That is, the results of ANOVA for periodic rate are  $F(1, 136) = 0.24, p < .625$ .

loanword diffusion”, their data suggest that the main role of the highly bilingual is that of borrowing innovators (Poplack et al. 1988: 84).

Finally, it was found that regular Spanish users opt for reproductive borrowings to a greater extent than infrequent Spanish users. This finding reflects the fact that, in interacting with other New York City Spanish speakers more often, regular Spanish users are more exposed to the English-origin vocabulary in wide circulation in Spanish and, consequently, are more prone to using that vocabulary. This fact suggests that using shared vocabulary is not, for the individual speaker, an act of “using English” but of using local patterns for speaking Spanish.

### 4.3 Daily Spanish use best predicts reproductive borrowing

The three variables that condition the use of shared vocabulary – class, English skills and daily Spanish use – were entered into a regression. The results are in Table 6.11.

**Table 6.11**  
**The ranked effects of occupational class, English skills**  
**and daily Spanish use on shared rate**

	<b>Standardized beta coefficient</b>	<b>p</b>	<b>Tolerance</b>
<b>Daily Spanish use</b>	.21*	<b>.054</b>	.64
<b>Occupational class</b>	.09	.365	.78
<b>English skills</b>	-.05	.635	.59
$R^2 = .08$		$F(3, 129) = 3.87$	$p < .011$

Table 6.11 shows that together, English skills, class and daily Spanish use significantly account for just eight percent of the variance in reproductive borrowing behavior ( $R^2=.08$ ,  $p<.011$ ). Of these variables, daily Spanish use ( $beta=.21$ ,  $p<.054$ ) best predicts the use of shared vocabulary.

In other words, how often one uses Spanish is the best overall indicator of one's tendency to select English-origin material that already has wide circulation in Spanish in New York.

#### **4.4 Discussion of variables that do not condition reproductive lexical borrowing**

Neither immigrant generation, regional origin, sex, education, years in the U.S., age nor Spanish skills significantly influences the use of shared vocabulary. Given the fact that, on the whole, the first generation is less proficient in English than the second, it would not have been surprising to find that they favored shared vocabulary. That is, it was expected that the first generation would opt to use items that require little or no proficiency in or familiarity with English. This is, however, not the case. As with innovative borrowings, the first and second generation use reproductive borrowings to relatively the same extent (arrival age  $F(3, 135) = 1.11, p < .349$ ;  $r(138) = .01, p < .905$ ; immigrant generation  $F(2, 136) = 1.63, p < .201$ ). This adds support to the idea that underlying frequency differences is an established pattern of borrowing in New York, characterized by equal proportions of both nonshared and shared vocabulary (section 4.2 of this chapter).

Ethnonational affiliation ( $F(5, 133) = 0.78, p < .563$ ) and regional origin ( $F(1, 137) = 0.00, p < .980$ ) also do not influence shared rate. Spanish speakers, regardless of whether they are Cuban or Puerto Rican, from the Latin American mainland or the Caribbean, use shared vocabulary to the same extent. Likewise, whether one is a man or a woman is not related to shared rate ( $F(1, 137) = 0.80, p < .372$ ). That is, men and women Spanish speakers in New York City tend to draw upon shared vocabulary to the same extent. These results conflict with Poplack et al.'s findings. On one hand, Poplack et al. report the trend that "men, and not women, favor established [similar to *shared*] loanwords" (1988: 89). (This trend, however, did not reach statistical significance.) On the other hand, they also found that women of the less bilingual

neighborhood (Hull) favored widespread loanwords (1988: 77) at rates that significantly differed from men.

Furthermore, education, although important to borrowing rate in the first generation (Chapter 5, sections 2.1 and 2.3.2), does not condition the use of shared vocabulary ( $F(1, 136) = 2.15, p < .145$ ). In other words, individuals with little formal education and those with a lot use approximately the same proportion of shared vocabulary. This finding does not support the results of Poplack et al. (1988: 87), who found that established loanwords were “concentrated among (Francophone Canadian) speakers with primary schooling or less”, which they furthermore interpreted in terms of exposure to English. The findings of the present study also do not corroborate Sullivan (2008). She found that less-educated Punjabi speakers, who borrowed most, were more responsible for disseminating borrowings. The results of the present study contrast with those two studies most likely because whereas education may be a principal or primary means of exposure to the donor language in those contexts, such is not the case in New York City, where the donor language is the local majority language (section 3.4) and exposure to it is not relegated to educational contexts.

But if English exposure is the key to accounting for the type of vocabulary that informants use, then it is striking that years in the U.S. does not influence the use of reproductive vocabulary ( $F(2, 88) = 0.25, p < .776; r(138) = -.11, p < .191$ ). It was expected that new arrivals, in having had less exposure to English than long-established immigrants, would use more borrowings that are widely-shared. Despite the lack of a significant correlation between shared rate and the number of years immigrants have resided in the city, it may nevertheless be that recent arrivals use vocabulary with some social currency. An examination of the nonce vocabulary of recent arrivals

(section 3.4 of this chapter lists examples) shows that, although they may not be widely used in New York, they may be common in recent immigrants' particular spheres of interaction.

Again, age plays no role in the types of borrowings used ( $F(3, 135) = 1.89, p < .133; r(138) = -.09, p < .314$ ). Both the young and the old use shared vocabulary to the same extent. Finally, neither proficiency nor confidence in Spanish conditions the use of shared vocabulary ( $F(1, 136) = 0.36, p < .550$ ). Second generation individuals who are very proficient in Spanish use the same proportion of shared vocabulary as those who are less proficient. In the first generation, individuals that have low confidence and those with high confidence in Spanish use shared vocabulary to the same extent.

#### **4.5 Reproductive borrowing: Summary and conclusions**

On average, 33 percent of informants' borrowing vocabulary is shared among several informants. This percentage, that is, one-third of speakers' inventories, matches that found for innovative borrowings (section 3.1). The selection of reproductive borrowings is conditioned by class, English skills and daily Spanish use. The working class uses a greater proportion of shared vocabulary than the middle class. Those with lower English skills ratings employ a greater proportion of shared vocabulary. These findings are thought to result from the fact that the middle class, on the one hand, and those with better English skills ratings, on the other, tend to play important roles as innovators in borrowing. Their elevated use of nonce and idiosyncratic vocabulary forces down the proportion of shared vocabulary they appear to use.

The strongest predictor of the use of shared vocabulary, however, is daily Spanish use. Regular Spanish users employ a greater proportion of vocabulary with some social tender. This reflects the fact that regular Spanish users are more exposed to widely-used borrowings and may, thus, be more inclined to use them as a normal part of speaking Spanish in New York.

But these differences should be tempered by three facts. First, each subgroup's average shared and nonshared rate is typically within ten percentage points of other subgroups. In other words, vocabulary preferences, while significantly different by statistical standards, may not be particularly obvious to interlocutors. Second, on average, about a third of informants' vocabulary is innovative and unique to the speaker (section 3.1) and a third consists of words shared widely in the community (section 4.1). In other words, despite statistically divergent preferences for novel or shared vocabulary among OZC informants, there is a prominent trend that indicates that **both** innovation and reproduction are part of the borrowing behavior of virtually every speaker. Third, there is no difference in vocabulary preferences according to immigrant generation, despite having found divergent differences in the frequency with which each group borrows (Chapter 4, section 4.1). That is, although each generation may represent a different subcommunity with respect to how much they borrow, they are unified in the types of vocabulary they select while borrowing.

## **5. On the dissemination of lexical borrowings**

In lexical borrowing research, disseminators have been identified as those that borrow most often (e.g. see Sullivan 2008) or as those that share large proportions of vocabulary in common with other groups (see, for instance, Poplack et al. 1988).<sup>21</sup> In this chapter, however, it has not been postulated that those that use more shared vocabulary or those that borrow most overall are most responsible for the dissemination of lexical borrowings. This is because of the pattern of results obtained in the present investigation and the fact that borrowing in New York takes place in an

---

21 A complementary indication of the dissemination of borrowings would be the degree of overlap of shared inventories, such as the analysis provided by Poplack et al. (1988). This type of analysis has not been undertaken for the present study. So, the identification of reproductive borrowers who also borrow frequently serves as a first approximation to discovering those most responsible for the dissemination of borrowings.

immigrant contact setting. In New York, English is not only a prestige language, but is also the dominant language of verbal interaction in the city. Perhaps more important than demonstrating high borrowing frequencies, then, disseminators in New York must, first, be individuals that regularly interact in Spanish. Without this condition, individual borrowings, no matter how frequently used, cannot be more than transitory speech events that do not catch on among those that do use Spanish regularly.

Results of Chapter 4 indicated that it is precisely those that use Spanish the least that borrow the most (Chapter 4, sections 4.4 and 5.2; also Chapter 5, sections 2.1 and 3.1). Consider also that the second generation borrows most often. Yet, it is precisely these frequent borrowers that are least likely to use Spanish for daily interactions. Of 29 second generation informants, only eight (27 percent) report regular daily use of Spanish. On the other hand, 78 percent (n=77) of the first generation reports regular daily use of Spanish. It is among the first generation, then, where the groups best associated with disseminating tendencies are most likely located, despite the fact that they borrow only half as often as the second generation. In particular, working class first generation individuals, who use a greater proportion of shared vocabulary as well demonstrate a greater likelihood to use Spanish on a daily basis,<sup>22</sup> seem to be in the best position for disseminating borrowings in the Spanish speaking community.

Although the working class of the first generation is well-positioned to be disseminators of borrowings, this is not to say that other groups of speakers, particularly those associated with innovation, are not also responsible for the dissemination of borrowings. For instance, middle class informants of the first generation have been identified as borrowing innovators (section 3.2 of this chapter). Because middle class informants also borrow more than the working class, it is

---

22 Of the working class, 64 percent (n=30) report very high usage of Spanish on a daily basis, as compared to just 21 percent (n=10) of the middle class.

possible that they play a significant role in the dissemination of borrowings. In fact, evidence to this effect, while not conclusive, has already been presented. Middle class speakers have a 10 percent larger inventory of shared lexemes than the working class. Incidentally, the inventory of the working class is almost a perfect subset of the lexemes used by the middle class. This suggests that the direction of dissemination is from higher classes to lower ones. In other words, middle class individuals may be important at early stages of dissemination, while lower classes may carry forward or complete the process.

## **6. Conclusion and review of questions**

The questions about innovative and reproductive borrowing in New York City posed at the beginning of this chapter may now be answered. It was asked whether, for all speakers, shared words comprise a larger portion of borrowing vocabulary than other types of words (question a) or whether, for some speakers, novel items make up a substantial component of their borrowing inventories (question b). It has been seen that novel vocabulary represents a substantial portion of the borrowing inventory, in terms of tokens and types, of Spanish in the city (Table 6.1). This is also true of individuals' borrowing inventories (Table 6.3). The majority of informants uses equal proportions of both shared and nonshared vocabulary (Table 6.5 and Table 6.10), suggesting that both types of borrowings are integral and equally important to the way that borrowing is done. This conclusion is further supported by the fact that neither the use of shared nor nonshared vocabulary is significantly influenced by immigrant generation. In other words, despite the fact that the first and second generation are two distinct groups with respect to borrowing frequency, these groups are not different from each other with respect to the types of vocabulary they use while borrowing (section 3.1 of this chapter).

It was also asked which Spanish speakers are most responsible for introducing novel vocabulary (question c). It was found that middle class individuals not only borrow more often (Chapter 4, section 4.2; Chapter 5, sections 2.1 and 2.3.1), but also use a greater proportion of novel vocabulary than the working class. In addition, those that use English most in their daily lives (i.e. those that use Spanish less) introduce a greater quantity of new words while borrowing (section 3.1 of this chapter). Likewise, proficiency and confidence in English encourage greater use of novel vocabulary.

A tentative answer to the question of who is most responsible for the dissemination of borrowings (question d) has also been presented. Section 4.5 showed that the groups that opt for widely-used borrowings are the working class, those that use Spanish most and those that are not very proficient or that have low confidence in their English. Although these three groups borrow less often overall (Chapter 4, sections 4.2 and 4.4; Chapter 5, sections 2.1, 2.3.1, 2.3.3, 2.3.5, 3.1, 3.3.2 and 3.3.3), they also use Spanish more often. Because using Spanish regularly is a more essential prerequisite for the dissemination of borrowings than a high rate of use, it was proposed that the working class of the first generation (i.e. they use Spanish often and use greater proportions of shared vocabulary) is the best-positioned to disseminate English borrowings among Spanish speakers in New York.

## **Chapter 7: Deborrowing: Flagged lexical borrowings in Spanish in New York**

### **1. Introduction**

This chapter explores the role of flags (as in Hlavac 2000; Muysken 1995; Poplack 1987; Teschner 1972) in lexical borrowing behavior in Spanish in New York. Flags are discourse elements such as false starts (*the recept- la recepcionista dijo...*), unfilled pauses and filled pauses (*uhh, um, eh*, etc.). Metalinguistic commentary (*What's the word?, Como dicen...*, etc.) are also considered flags. An analysis of flags in the immediate vicinity of English lexical borrowings in the OZC is undertaken to determine if assumptions about the significance of flagging in bilingual speech (see section 2 of this chapter) are supported for the present sample of informants. Results confirm the premise that flagging lexical borrowings indicates a speaker's awareness of the non-Spanish origin of the lexical borrowing. Results disconfirm the notion, for these informants, that flagging indicates a lack of fluency in Spanish. In particular, a series of ANOVAs reveals that Spanish-speaking informants demonstrate a tendency to flag English lexical borrowings more often under three conditions: (i) when they know more English, (ii) when they use mostly English in their daily lives and (iii) when they are second (rather than first) generation Spanish speakers. When these findings are interpreted diachronically, as reflecting the sociolinguistic status of Spanish in an immigrant context as it changes from being the primary language of transmission and communication in the home country to being a secondary language in the U.S./New York context, a subtle trend of deborrowing is revealed. These results have implications both for the use of flags as data collection diagnostics as well as for the future of lexical borrowings in Spanish in New York.

## 2. Flagging in bilingual speech research

In research using bilingual speech as data, flagging has a contentious status. On the one hand, pauses, hesitations and metalinguistic commentary are interpreted by some (a) as a strategy for using two languages together in particular ways (e.g. see Muysken 1995, 2000; also Poplack 1987: 65) or (b) as an indication of a speaker's awareness of the "foreignness" of an utterance (e.g. see Pfaff 1979; Poplack 1987; Teschner 1972: 1140-1142). Finally, their presence near a stretch of bilingual speech is sometimes interpreted (c) as symptomatic of an individual's linguistic disfluency or as an indication that there is something unnatural in what has been or is about to be said. For instance, in their study of English lexical borrowing in French, Poplack et al. explain that they retained for their study "only forms used **naturally** in the contexts of French discourse" (1988: 54, my emphasis) and that did not occur in the immediate context of "perceptible hesitations or false starts". Likewise, in one study, Poplack (1983) discusses constraints on codeswitching. She mentions in passing that exceptions to proposed codeswitching constraints occurred largely in the speech of one individual who, "it is clear from the number of pauses, hesitation, and false starts, has an incomplete acquisitional history" (Poplack 1983: 123). Her conclusion about the informant's acquisitionally "incomplete" history was predicated on the quantity of flags (87 percent of his data includes them) rather than on the fact of their presence at all. Nonetheless, such juxtapositions, that is, the juxtaposition of flagging and "naturalness" or flagging and "disfluency", in bilingual speech research may lead to the erroneous, or at least not empirically validated, position (i) that flags indicate something is wrong and (ii) that what is wrong is the linguistic competence of the speaker.

Furthermore, the presence of flags in bilingual speech is sometimes used as a criterion for the inclusion or exclusion of data. For instance, Pfaff (1979) used the presence of hesitations, translations and metalinguistic commentary in her informants' speech to classify other-language

strings as either borrowing or codeswitching. Since, as she states, she takes the presence of flagging as “evidence that the speaker is aware of switching to a special term, code, or register” (1979: 297), Pfaff interprets any flag that occurs near a foreign-origin string as a sign that the string is not part of the Spanish lexical repertoire of the speaker.<sup>1</sup> Likewise, as mentioned, Poplack et al. (1988: 54) excluded all (potential) borrowings and loanwords that occurred in the vicinity of flags. The justification for their methodological decision was based upon two premises: (i) the occurrence of flags indicates the speaker’s awareness of the foreign-origin of an other-language string and (ii) the demonstration of awareness of the etymological origin of any lexical material disqualifies it as a borrowing, since borrowings (for them) are defined as fully-accepted lexicon of the recipient language.<sup>2</sup>

Few studies of bilingual speech that use flagging as a diagnostic to classify lexical contact phenomena (i.e. borrowing or codeswitching) have, that I am aware, explicitly demonstrated, for their corpus, the validity of the premises upon which flagging as a diagnostic is based.<sup>3</sup> This is a potentially important determination to make since, as Levinson notes “certain language production parameters – e.g. pauses, hesitations and self-editings – that psycholinguists have taken to be straightforward indicators of internal cognitive processes are in fact subject to manipulation for interactional purposes” (2002: 375). In other words, flags used in the presence of an other-language string may be motivated by considerations **other than** lack of fluency,

---

1 On the other hand, Pfaff (1979: 297-298) indicates that Spanish translation or paraphrase of an English phrase indicates that the English phrase is part of the speaker’s repertoire in Spanish.

2 Note that although Poplack’s work has been one inspiration in my study for maintaining a distinction between established loanwords (words of a language with a foreign etymology, like *their* in English) and lexical borrowing (other etymologically foreign lexical material whose ultimate status in recipient language lexicon has yet to be seen), she and her co-authors seem to be conflating the two concepts at this point in their 1988 article so that, if something is not fully French, not only is it not a loanword (by definition of *loanword*), but it also cannot be a lexical borrowing.

3 A notable exception is Poplack (1987). This study provides data for flagging in Hull, Quebec and Ottawa, Ontario. The data lead Poplack to the conclusion that flagging can be part of a community-wide pattern that demonstrates an individual’s awareness of the foreign-origin of an other-language strings.

processing difficulties or the fact that something is not part of Spanish.<sup>4</sup> It should be noted that a complete analysis of flagging by OZC informants is not undertaken in this chapter. The analysis presented here is, rather, a preliminary attempt to discern whether sociovariationist evidence from the OZC supports an interpretation of flags in the presence of short other-language strings (i.e. English borrowings) as indicative of language membership awareness or lack of fluency.

### **3. From assumption to hypothesis**

Of the many things that flagging may indicate, the current analysis seeks evidence to support just two interpretations:

- Flagging is indicative of a speaker's awareness of the etymological origin of lexical material.
- Flagging is indicative of a speaker's linguistic disfluency.

I examine only these assumptions about flagging because they are the ones typically invoked in studies of bilingual speech and the ones that I believe cannot be taken to be true *a priori* for every group of bilinguals or every mixed speech corpus.

The first assumption, that flagging indicates a speaker's awareness of the etymological origins of a lexical choice, is one that has been advanced in bilingual speech research as a justification for using flagging as a diagnostic for data collection. In the weak version of the language membership assumption, the presence of a flag near a stretch of donor language indicates the speaker's awareness of its foreign origin (e.g. see Pfaff 1979), but makes no claims about whether the donor language material in question has become incorporated into the lexicon

---

4 For example, metalinguistic commentary, fillers and pauses may be employed as a strategy to aid hearer comprehension, to manage a speaker's persona or may be part of a particular group's discourse style (e.g. see Poplack 1987 for flagging by Francophones in Hull, Quebec). For example, interrogation of my own motivations for flagging in Spanish and English mixed discourse has led me to conclude that I sometimes flag in order to give unfamiliar interlocutors a chance to absorb what I am saying. Other times, I use flags to attenuate what may be perceived as a somewhat formal style of speaking.

of the recipient language. In a stronger version, the presence of flagging not only implies the speaker's awareness of the foreignness of the donor language material, but also announces "This lexical material is not part of my (the recipient language) lexicon" (again, see Poplack et al. 1988). Because my purpose here is to examine the extent to which such interpretations hold up for the OZC, I restate these assumptions as hypotheses to be tested (the stronger version of the hypothesis is beneath and indented to the preceding weaker version):

Group 1: The bilingual awareness and language membership hypotheses

*Hypothesis 1a:* Flagging indicates awareness of the foreign etymology of lexical material, in this case, English.

*Hypothesis 1b:* Flagging indicates that etymologically foreign language material is not part of the lexicon of the discourse language, Spanish.

The second assumption about flagging, that it is indicative of a speaker's linguistic disfluency, is well-documented for the speech of monolinguals and bilinguals alike. Yet, there is an incongruity in how flagging tends to be interpreted for monolinguals, on the one hand, and for bilinguals, on the other. When a monolingual English speaker pauses, makes a metalinguistic comment or forgets a word, these flags are taken to indicate momentary disfluencies or performance errors. When bilingual speakers pause or make metalinguistic commentary on word choice, it is taken not just as momentary processing difficulties, but as a lack of competence or as disfluency in one of a speaker's languages. The degree of linguistic disfluency ascribed to the speaker ranges from a claim of minor attrition due to lack of use of a language to incomplete acquisition of a language (as described by Poplack in one case mentioned above). Again, these assumptions about flagging are reformulated as hypotheses:

Group 2: The (dis)fluency hypotheses

*Hypothesis 2a:* Flagging indicates momentary disfluency in speech.

*Hypothesis 2b:* Flagging indicates disfluency in a language, in this case Spanish, due to lack of practice (i.e. language attrition).

*Hypothesis 2c:* Flagging indicates incomplete acquisition of one of a bilingual's languages, in this case, Spanish.

In sections 6.1 and 6.2, each set of hypotheses will be examined to determine the extent to which different versions of the hypotheses can be supported by patterns of flagging demonstrated by OZC informants. Each section begins with a brief discussion of the type of evidence available from the LBD that would lend support to the hypotheses. After, the results of analysis are presented and discussed. In order to facilitate discussion, a synopsis of the hypotheses is periodically provided in a text box. Before that (section 4), I define the independent variable (i.e. flagging rate) and discuss when a lexical borrowing is considered to be flagged (section 5).

#### **4. The independent variable: Flagging rate**

The independent variable for this chapter is flagging rate. It is reported as a percentage. Two types of flagging rates are used in this chapter. A corpus-level flagging rate is used in section 6.1. It is calculated by dividing the number of flagged borrowings in the OZC by the total number of borrowings. An informant-level flagging rate, used primarily in section 6.2, is computed by dividing the number of flagged borrowings an informant utters by his total borrowings. So, for instance, if an informant uttered 53 borrowings during his interview and flagged four of them, his flagging rate would be  $\frac{4}{53}$  or 7.5 percent.

## 5. The data and coding

This section details the circumstances where a speaker's use of a foreign-origin expression is coded as flagged and where it is not. These circumstances differ depending on whether the borrowing contains content words or consists only of functional words or phrases. Consequently, the following discussion is divided into two parts. Section 5.1. describes the conditions where content words qualify as having been flagged. Section 5.2 does the same, but with respect to borrowings that consist primarily of function words. The discussion occasionally makes use of concepts from phrase structure grammar, such as *domination* (as in section 5.1.3), in order to define the syntactic configurations in which a borrowing is considered to be flagged.

In addition to the syntactic contexts, outlined below, in which a borrowing is coded as flagged, there are a handful of discourse contexts where it seems that certain types of flags (namely pauses and fillers) **are required** in order for discourse to continue unproblematically. As a result, it is hypothesized that a nearby borrowing in these contexts does not qualify as having been flagged. For instance, fillers are probably necessary to hold the floor while one gathers his thoughts at the beginning of a turn that follows a request for information (examples (9) and (10) (Levinson 2002: 365). (An unfilled pause at such a juncture would probably signal that there is some problem in the transmission of the request.) Fillers are also probably necessary to hold the floor while listing (examples (11) and (25) in order for the speaker to signal that his turn is not yet complete. (Again, an unfilled pause in this case could be mistaken by the hearer as the end of a speaker's turn.) Information meant to clarify or add detail to a message oftentimes appears as intersentential adjuncts set off by pauses (example (22). A speaker's desire to self-repair for the sake of accuracy or clarity may also cause him to stop mid-sentence in order to add information or begin a syntactically redesigned sentence (examples (17) and (24) (also see examples in Levinson 2002: 341 and Schegloff, Jefferson & Sacks 1977). In this case, pauses

along with discourse markers of some sort (e.g. *I mean, well, actually*) are also likely to occur (example (23)). Pauses, of necessity, also occur near discourse markers (section 4.2.4) and in reported speech (example (26)). Consequently, borrowings that occur in these contexts are **not**, for the present analysis, considered to be flagged. There are, undoubtedly, more contexts where flags of various sorts occur obligatorily in the course of a self-repair task or conversational move. The contexts cited above, however, are those that occurred near borrowings in the OZC, and which are therefore defined in order to excuse nearby borrowings from being coded as flagged.

## 5.1 Content borrowings: Nouns, verbs, adjectives and adverbs

### 5.1.1 Metalinguistic commentary

A borrowing that is accompanied by metalanguage referring to the cultural context or the use of Spanish or English (*que le dicen, cómo se llama aquí*, etc.) is coded as flagged. For example, the borrowings in examples (1)-3) are coded as flagged with metalanguage (underlined).

- (1) Son de esos panes que aquí llaman *bagels* y... 344C  
'They are that kind of bread that here they call *bagels* and...'
- (2) Y dormimos toda la noche en una área que le llaman el *wild coast* porque el mar está siempre así... 434P  
'And we slept the whole night in an area that they call (it) the *wild coast* because the sea there is always like that...'

In example (1), before uttering the borrowing *bagels*, the informant refers to the fact that the expression belongs to the New York City context with the expression *que aquí llaman* 'that here they call'. In (2), the informant uses the phrase *wild coast*. She recognizes that it is an expression that only certain people might use or recognize, as indicated by *que le llaman* 'that they call (it)'. Both borrowings are coded as flagged.

In this study, not all metalinguistic commentary qualifies as a flagging due to the presence of an other-language string. If the commentary does not explicitly or implicitly refer to language choice or cultural context, it is not considered a borrowing flag. The borrowings in (3) and (4), for example, are not coded as flagged, despite being referred to with a metalinguistic comment.

- (3) Pues puedo decirme como *league agent*, trabajo en importaciones y exportaciones... 321E  
'So I can call myself like *league agent*, I work in imports and exports...'
- (4) Porque ahí teníamos intercambio con una universidad que se llama TECU University, y entonces... 354M  
'Because there we used to have an exchange with a university that's called TECU University, and then...'

In example (3), the informant uses the borrowing *league agent* and, further, comments on his own use of the noun phrase by uttering *puedo decirme* 'I can call myself'. The informant makes no reference to the fact that he has used English to name his job title. Furthermore, it could be imagined that even if he had uttered the entire sentence in English, or had used a Spanish expression in place of *league agent*, he may have still introduced his job title with the expression *puedo decirme*. In (4), the informant uses the borrowing *TECU University* and introduces it with *una universidad que se llama* 'a university that's called'. His motivation for saying *que se llama* does not appear to be necessarily connected to the fact of having used English to name the university. Rather, it appears to be a device to introduce what he perceives may be novel information for his interlocutor. Neither *league agent* nor *TECU University*, therefore, is coded as flagged.

Note that certain discourse markers (e.g. *I mean, you know, tú sabes, quiere decir*) signal a metalinguistic message. However, discourse markers alone are **not** considered borrowing flags in the present investigation because they do not typically make reference to cultural context or the

language of utterance. The borrowing *signs* in example (5) is followed by the discourse marker *tú sabes* ‘you know’, but it is **not** coded as having been flagged.

- (5) No veo señales, no veo *signs*, tú sabes, nada... 435P  
‘I don’t see signs, I don’t see *signs*, you know, nothing...’

### 5.1.2 Fillers and filled pauses

When a borrowing is immediately preceded by a filler or filled pause, such as *eh*, *ah*, *uhm* and *este*, it is coded as having been flagged. For example, look at *succeed* in example (6).

- (6) Yo, para que yo y mi hermano podríamos este *succeed*... 403P  
‘I, so that I and my brother could uhm *succeed*...’

In example (6), the filler *este* ‘this/uhm’ precedes the borrowing. *Succeed* is coded as having been flagged.

A filler may occur as far away from the borrowing as before the parent constituent of the phrase in which the borrowing occurs. The borrowings in examples (7) and (8) are both coded as having been flagged.

- (7) Y pueden ser este.. un poquito *nasty*, este... 423P  
‘And they can be uhm.. a little *nasty*, uhm...’
- (8) No nos ayudaban, ah.. preguntábamos *direcciones*... 326E  
‘They weren’t helping us, ah.. we were asking *directions*...’

In (7), the borrowing *nasty* is the head of a noun phrase. The parent constituent of the noun phrase is the copular constituent headed by *ser* ‘be’. Because the filler *este* ‘uhm’ occurs after the head of the copular constituent, *nasty* is coded as having been flagged. In (8), *direcciones* ‘directions’ is the head of a noun phrase. This noun phrase is subordinate to the verb phrase

headed by *preguntabamos* ‘we were asking’. Because the filler *ah* occurs immediately before the beginning of the verb phrase (i.e. the parent constituent of the constituent in which the borrowing occurs), the borrowing *direcciones* is considered to be flagged.

Not all borrowings in the above-described contexts are coded as flagged. If a filler occurs at the beginning of a new sentence or turn, a borrowing **is not** coded as flagged. Look at excerpts (9) and (10).

- (9) Entrevistadora: ¿Qué es [tu negocio]? ¿Ropa? ¿Jeans y cosas?  
Entrevistado: Ahh.. no. *Sportswear*. Camisetería más que todo, *unisex*... 158C  
‘Interviewer: What is it [your business]? Clothes? Jeans and things?  
Interviewee: Ahh.. no. *Sportswear*. T-shirt shop more than anything, *unisex*...’
- (10) Entrevistador: ¿A qué universidad asististe?  
Entrevistada: Ah.. *Pace University*, hice aquí un programa empezando en Columbia... 319D  
‘Interviewer: What university did you attend?  
Interviewee: Ah.. *Pace University*, I did a program here beginning in Columbia...’

In example (9), the informant uses the borrowing *sportswear*. This is preceded by the filler *ahh*. *Sportswear* is not coded as flagged since the flag *ahh* occurs principally, it seems, because the speaker requires time to collect his thoughts after the interviewer’s request for information. An unfilled pause at this juncture may have signalled to the interviewer that his question had not been effectively transmitted to the informant. The filler, then, is associated with the informant’s need to hold the floor and **not** the use of the English word *sportswear*. The same applies to the filler *ah* that precedes the use of the borrowing *Pace University* in (10). Because the filler occurs at the beginning of the interviewee’s turn, it is thought to be used in order to hold the floor while formulating a response to the interviewer’s question.

If a filler is used preceding a borrowing in a list, the borrowing is not coded as flagged. Look at example (11).

- (11) Los Picapiedras, [RISA] eso te gusta a ti, Picapiedras, Las Picapiedras. Eh..  
quién más... *The Jetsons* este... *The Honeymooners* me gustaba. 428P  
'The Flintstones, [LAUGHTER] you like that, Flintstones, The Flintstones. Uh..  
who else.. *The Jetsons* uhm.. I used to like *The Honeymooners*.'

In (11), the filler *este* 'uhm' precedes the borrowing *The Honeymooners*. The filler is used primarily in order for the informant to hold the floor while retrieving the next item in the list. *The Honeymooners* is thus **not** coded as flagged.

### 5.1.3 False starts

Borrowings that are falsely started are coded as flagged. An example of this is in example (12).

- (12) Tengo clientes que vienen de probe, *probation*... 370M  
'I have clients that come from probe, *probation*...'

In (12), the borrowing *probation* is fully uttered only after the informant starts to utter it and stops (*probe*-). *Probation* is coded as flagged.

A false start may occur as far away from the borrowing as the *grandparent* constituent of the constituent in which a borrowing occurs. Said differently, a borrowing is coded as flagged if there is a false start of any part of a constituent whose *maximal projection*<sup>5</sup> *dominates*<sup>6</sup> the borrowing, up to two dominating nodes away. Examples (13)-(15) contain borrowings that were coded as having been flagged with a false start that occurs in this environment.

---

5 A maximal projection is the highest node in a tree projected by a constituent's head (Haegeman 1994: 90, Haegeman & Guéron 1999: 74, 78). For instance, a VP is the maximal projection of a verb.

6 Haegeman and Guéron say: "Node A... dominate[s] node B if you can go from node A to node B along a downward branch. S [sentence], for instance, dominates all the constituents of the sentence" (1994: 51). Also see Haegeman (1994: 85) for a definition of *dominance*.

- (13) Este.. hasta que un.. un mesero en un *restaurant* me trajo una botellita como de veneno de vez de traerme un pescado... 423P  
 ‘Uhm.. until a.. a waiter in a *restaurant* brought me a bottle of like venom instead of bringing me a fish...’

In (13), the borrowing *restaurant* is the head of a noun phrase. The noun phrase is immediately dominated by a prepositional phrase headed by *en* ‘in’. The prepositional phrase is, itself, dominated by the noun phrase headed by *mesero* ‘waiter’. This noun phrase dominated by *mesero* ‘waiter’ is considered the grandparent to the NP in which the borrowing *restaurant* occurs. Because there is a false start of that grandparent constituent (i.e. *hasta que un*.. *un mesero*), *restaurant* is coded as flagged. The same logic is applied to borrowings in (14) and (15), which are coded as flagged.

- (14) Porque yo trabajaba.. yo era *babysitter*. 238U  
 ‘Because I was working.. I was a *babysitter*.’
- (15) La persona que controla secretarias que.. que trabaja en el.. en la familia del *word processing*... 198P  
 ‘The person that controls secretaries that.. that were working in the.. in the family of *word processing*...’

In (14), *babysitter* is the head of a noun phrase that is immediately dominated by the copular phrase headed by *es* ‘is’. The grandparent constituent of the borrowing is the node that dominates the entire sentence *yo era babysitter* ‘I was a *babysitter*’. Because the sentence was falsely started before the pause (i.e. *yo trabajaba*..), the borrowing *babysitter* is coded as flagged. In (15), the NP headed by *word processing* is immediately dominated by the prepositional phrase headed by *de* ‘de’. The prepositional phrase is itself dominated by the NP headed by *familia* ‘family’. Because part of that NP is falsely started (i.e. *en el*.. *en la familia*), the borrowing *word processing* is classified as flagged.

False starts of constituents that do not dominate the borrowing are not considered to be flags of the borrowing. For example, *rufo* ‘roof’ in (16) is **not** coded as flagged.

- (16) Arreglando los.. la terraza, el *rufo*... 344C  
‘Fixing the.. the terrace, the *roof*...’

In (16), the false start (los..) appears before the noun phrase headed by *terrazza* ‘terrace’ and appears to be a false start of the phrase *la terraza*. Because *la terraza* does not dominate the phrase in which the borrowing *rufo* ‘roof’ occurs, *rufo* is **not** coded as flagged.

False starts that do not explicitly repeat any part of the borrowing or the parent or grandparent constituent of the borrowing are not considered borrowing flags. For example, in (17), *outreaches* is not coded as flagged.

- (17) Tenemos una.. hicimos *outreaches* you know... 428P  
‘[We] have a.. [we] did *outreaches*, you know...’

In (17), the sentence is falsely started as *tenemos una* ‘we have a’, but then reformulated as *hicimos* ‘we did’. No part of the fully articulated sentence *hicimos outreaches* was overtly uttered as part of the false start, and so *outreaches* is **not** considered to be flagged. (The pause is the result of the informant’s desire to reformulate and therefore does not count as a flag).

#### 5.1.4 Unfilled pauses

Borrowings that occur near pauses are only classified as flagged when the pause occurs immediately preceding the borrowing in the middle of a clause. For example, *nursery* in example (18) is coded as having been flagged because the pause occurs immediately before the borrowing and in the middle of the clause that extends from *ella* ‘she’ to *chiquita* ‘small’.

- (18) Ella tuvo dos años en\_ un *nursery* chiquita... 258C  
'She had two years in\_ a small *nursery*...'

Pauses in other discourse positions are not considered to be flags related to the use of an English-origin expression. For example, if a pause follows a borrowing, the borrowing is **not** considered to be flagged, as in example (19).

- (19) Hace muchos años cuando estaba en *high school*\_ pre.. predominaban mucho los asiáticos... 300E  
'A long time ago when I was in *high school*\_ there.. there were mostly Asians...'

Because the pause occurs after it, *high school* is **not** coded as having been flagged. The pause appears to have occurred due to the informant's need to formulate and articulate the ideas that follow the borrowing *high school*.

When a pause occurs between a discourse marker or other tag expression (see Appendix D, section 10 for examples of tag expressions) and a borrowing, as in example (20), the borrowing is not coded as flagged.

- (20) Yo siempre paso por allá a ver si está bien.. y *you know*\_ *chilling*... 333D  
'I always pass by there to see if he's okay.. and *you know*\_ *chilling*...'

In (20), the borrowing *chilling* is preceded by a pause, which itself is preceded by a discourse marker *you know*. *Chilling* is not coded as having been flagged because the pause is considered to be required in the presence of the discourse marker. This is the case regardless of the language of the discourse marker. That is, even if the discourse marker had been in Spanish (i.e. *tú sabes*.. *chilling*), the borrowing *chilling* would **not** be coded as flagged.

Pauses at the beginning of clauses are thought to be wholly or primarily due to an informant's need to formulate his response and/or take a breath. Borrowings in their vicinity are not coded as flagged for the purposes of the present investigation. Consider example (21).

- (21) No hay que pagarle *taxes* al gobierno porque son *candies*, es un *candy store* ...  
198P  
'*Taxes* don't have to be paid to the government because they're *candies*, it's a *candy store* ...'

In (21), the borrowing *candy store* is not coded as flagged because the pause that precedes it occurs at the beginning of a new clause.

A pause before a borrowing that is a restatement or that is in apposition to another phrase does not count as a borrowing flag. Look at excerpt (22).

- (22) Me llamaron (...) diciéndome que (...) vaya a tomar los exámenes de.. de admisión, los *placement tests*. 300E  
'They called me (...) telling me (...) to go take the exams for.. for admission, the *placement tests*.'

In (22), the informant is talking about the events leading up to her enrollment in college. She first utters *los exámenes de admisión* 'the admission exams', then pauses and decides to state the same information differently by using the English borrowing *placement tests*. Thus, the pause in (22) is attributed to the need to set off the additional information from the preceding and following sentence. The borrowing *placement tests*, therefore, is not coded as flagged.

Pauses that occur due to an informant's need to clarify or (re)formulate or his message are **not** considered borrowing flags, even if a borrowing occurs in their vicinity. The borrowing in example (23) is not coded as flagged.

- (23) Lo único que yo quería (...) era (...) de que [mis papás] vayan a la noche de ceremonia de los\_ no tanto *senior awards* sino los años anteriores eran [llamadas] *award night*... 300E  
'The only thing I wanted (...) was (...) that [my parents] would go to the ceremony night of the\_ not so much *senior awards* but rather the previous years they were [called] *award night*...'

In (23), the borrowing *senior awards* follows a pause (*noche de ceremonia de los\_* 'ceremony night of the\_'). This pause occurs because the informant is looking for a more accurate expression, namely the expression *award night*. Because the presence of the pause is attributed primarily to the informant's desire change the content of her message for the sake of clarity and accuracy, *senior awards* is not coded as flagged.

Pauses that occur because a speaker has decided to change the syntactic form of a sentence are also not considered borrowing flags. The borrowing *special ed* in (24) is not coded as flagged.

- (24) Le metían a.. a donde tienen los otros niños que\_ como los de *special ed*, le tienen ahí metida. 365E  
'They put him wh.. where they have the other children that\_ like *special ed* children, they have him put there.'

In (24), the borrowing *special ed* occurs near a pause. However, the pause is present, it seems, because the informant decides to organize and present his thoughts differently. This informant initially set himself up to introduce a relative clause subordinate to *los otros niños* 'the other children'. He pauses after *que* 'that' in order to introduce a different syntactic construction to

modify the noun phrase headed by *niños* ‘children’: a prepositional phrase (headed by *de* ‘of’). *Special ed* is, as a result, **not** coded as flagged.

Likewise, pauses adjacent to borrowings that are used in a list do not count as borrowing flags. Consider excerpt (25).

- (25) Los Picapiedras, [RISA] eso te gusta a ti, Picapiedras, Las Picapiedras. Eh..  
quien más.. *The Jetsons* este... *The Honeymooners* me gustaba. 428P  
‘The Flintstones, [LAUGHTER] you like that, Flintstones, The Flintstones. Uh..  
who else.. *The Jetsons* uhm.. I used to like *The Honeymooners*.’

In (25), the borrowing *The Jetsons* follows a pause. This pause is thought to be due to the speaker’s need to collect his thoughts. *The Jetsons* is thus **not** coded as flagged.

Finally, pauses that occur in the context of quoting or (re)enactment of a conversation are not considered borrowing flags, even if a borrowing occurs near them. In excerpt (26), for example, *Washington Heights* is not coded as having been flagged.

- (26) Yo creo que [es]taba ahí todo el mundo. To[do] el mundo. Cada vez.. ¿De  
dónde tú eres? *Washington Heights*, yo, ah, ¡Dios! 401P  
‘I think everyone was there. Everyone. Every time.. Where are you from?  
*Washington Heights*, I, ah, God!’

In (26), the borrowing *Washington Heights* follows a pause. This pause is a required element in the reenactment of a conversation. *Washington Heights*, therefore, is **not** coded as flagged.

## 5.2 Functional words and expressions: Discourse markers, determiners, prepositions, coordinators, subordinators and conjunctions

### 5.2.1 Metalinguistic commentary

If a functional word or expression were referred to using a metalinguistic comment, it would be coded as flagged. For example, let us imagine that an informant had uttered *No sé, you know?*

*Como se dice en un buen castellano*. ‘I don’t know, *you know*? As they say in proper Spanish’. The borrowing *you know* would be coded as flagged due to the metalinguistic comment *como se dice en un buen castellano* ‘as they say in proper Spanish’. No examples of metalinguistic commentary referring to a borrowing that consists solely of a functional word or expression have been found in the OZC. This is likely because English-origin functional words, such as determiners or pronouns, rarely occur in isolation. Often, they are part of a longer English expression (as, for example *the* in *The Jetsons* in example (25)).

### 5.2.2 Fillers and filled pauses

Functional words are coded as flagged if a filler immediately precedes them in discourse. Look at examples (27) and (28).

- (27) En cuestión de trabajar con los mercados y.. el mercado económico de Estados Unidos.. pe.. uh.. *oh God*.. um... 432P  
 ‘With respect to working with markets and.. the economic market of the United States.. bu.. uh.. *oh God*.. um..’
- (28) Pero, ah.. *you know*, son.. cómo te digo... 322E  
 ‘But, ah.. *you know*, they are.. how do I tell you...’

In (27), the informant uses the interjection *oh God*. This borrowing is preceded by the filler *uh*. It appears that the filler occurs because the informant has lost his train of thought. This interruption additionally leads, it appears, to the insertion of the English *oh God*. The borrowing *oh God* is coded as flagged. In (28), the discourse marker *you know* is preceded by the filler *ah*. As with *uh* in (27), *ah* appears to be used because the speaker has momentarily lost the direction of her message. This momentary disorientation seems to also lead to the use of the English discourse marker *you know*. *You know* is coded as flagged.

### 5.2.3 False starts

Functional words and phrases are coded as flagged by a false start if the functional word itself has been falsely started. For example, suppose an informant had said *Pues, may- maybe la otra persona necesita tener más cuidado* ‘Well, may- maybe the other person should be more careful’. *Maybe* would be coded as having been flagged with a false start. No examples of this sort were uttered by OZC informants.

### 5.2.4 Unfilled pauses

Determiners, prepositions and pronouns are coded as flagged when a pause **directly precedes** the word. For example, imagine an informant had said *La quiero mucho. Es.. my mamá*. ‘I love her a lot. She is.. my mom.’ *My* would be coded as having been flagged because it is immediately preceded by a pause. No examples of this are in the corpus.

Other functional words and expressions, such as discourse markers, subordinators (*because, that*) and coordinators (*and, but*) are never coded as flagged if they are only preceded by an unfilled pause. This is because these expressions, regardless of the language in which they are uttered, appear to require pauses in order to effectively realize their function in discourse.

## 5.3 Summary of borrowing flags

In sum, borrowings are considered to be flagged when there is reasonable belief that the flag may have something to do with the use of an other-language string or due to lapses in a speaker’s memory, language processing or vocabulary. Flags that occur in the vicinity of English strings that are also typical discourse locations where flags are required for the purposes of turn-taking organization, restructuring a sentence or holding the floor are **not** considered borrowing flags. When borrowing occurs near a flag in these latter contexts, it is **not** coded as flagged.

## 6. Results

### 6.1 The language membership hypotheses

#### 6.1.1 Evidence in support of the language membership hypotheses

Support for hypothesis (1a), that flagging indicates a speaker's awareness of the foreign origin of a word, could come in several forms. First, flagging rates in mixed speech and flagging rates in monolingual speech for a single group of bilingual informants could be compared. If flagging in mixed speech were higher, this would intimate that it is the presence of etymologically foreign or mixed language material that has contributed to an increased flagging rate. Since this type of comparative data is presently unavailable, I instead compare rates of flagging near English lexical borrowings for informants of the OZC corpus to rates of flagging by bilinguals in other studies. If flagging rates for the present corpus are found to differ substantially from those for other Spanish speakers, particularly, if they were found to be much higher, this could be construed as support for the hypothesis that flagging is indeed associated with the presence of the other-language material for informants of this corpus. Additionally, the notion that flagging indicates awareness of language origin would be supported if it were found that those with a better knowledge of English or of Spanish tended to flag borrowings more often.

Support for hypothesis (1b), that flagging indicates that a word or string is not part of the lexical repertoire of the primary language of discourse, Spanish, could be supported with data available in this study by utilizing the notion of *community acceptance*, as outlined by Poplack et al. (1988). In that study, the investigators hypothesized that the degree of acceptance of a lexical borrowing into a community's lexicon would be reflected in the distribution and frequency of the borrowing in a sample corpus of borrowings from the community. They reasoned, and

**Language membership hypotheses**

1a: Lexical material has foreign origin.

1b: Lexical material is not Spanish lexicon.

empirically supported the notion, that the more often a lexical borrowing token occurred in a corpus, and the greater the number of speakers that used it, the more likely it would be that linguistic and historical evidence would corroborate its status as a fully entrenched loanword in the lexicon of the recipient language community. In other words, English lexical borrowings that were frequent (i.e. *widespread*) in their corpus of Canadian French had longer histories of attestation in the community and were more often phonologically incorporated into French. Conversely, they reasoned, and empirically supported the idea, that nonce items, appearing only once in their corpus, were likely idiosyncratic choices by individual speakers, and would not qualify as part of the lexical reservoir of any significant portion of the community. If, based on Poplack et al.'s findings, the corpus distribution and frequency of particular borrowings in the OZC are taken to de facto represent the degree to which they can be said to be part of a supposed New York City Spanish lexicon, then, according to hypothesis (1b), borrowings that are used most often and by more speakers should be flagged less frequently than those that occur only once or that are used by only a single informant. In fact, we will see that this is the case in Spanish in New York (section 6.1.3).

In sum, to verify the validity of the language membership hypotheses of flagging for this corpus, we are looking for the following:

- 1a:* Flagging near lexical borrowings (i.e. near etymologically English material) is more frequent than flagging in monolingual Spanish. Flagging is more frequent for those who have a better command of English or Spanish.
- 1b:* Flagging is more frequent for infrequently-used lexical borrowings of the corpus; flagging rates are less frequent for lexical borrowings that are used more often by more speakers.

### 6.1.2 Results of analysis and discussion of hypothesis (1a): Flagging indicates awareness that an English lexical borrowing has a foreign origin

Table 7.1 provides the rate at which borrowings are flagged in the corpus.

**Table 7.1**  
**Percent of flagged and unflagged borrowings in the OZC**

	#	%
<b>Flagged borrowings</b> <sup>7</sup> (n <sup>informants</sup> = 105)	558	<b>12</b>
<b>Unflagged borrowings</b> <sup>8</sup> (n <sup>informants</sup> = 138)	3949	<b>88</b>
<b>Total borrowings</b> (n <sup>informants</sup> =139)	4507	<b>100</b>

The last row of Table 7.1 shows that English lexical borrowings in the OZC totaled 4507, introduced by 139 informants. The first row of the table shows that of the over 4500 borrowings

- 7 Flagged borrowings include 12 borrowing tokens (n<sup>inf</sup>=8) that were themselves false starts. For example, look at the excerpts from informants 333D and 365E.

*The way.. la forma que yo lo veo es...* 333D  
'*The way.. the way I see it is...*'

[Interviewer asks about his relationship with his father.]  
*I.. yo a mi papá siempre.. desde peladito, mi papá, cómo era...* 365E  
'*I.. with my dad I always.. since I was a baldy, my father, how was he...*'

Informant 333D uses the borrowing *the way*, but decides to restate the noun phrase in Spanish using *la forma* 'the way'. Informant 365E begins his sentence with the English personal pronoun *I*, then restates it in Spanish using *yo* 'I'. These, and ten other tokens, are full borrowings, but also, themselves, false starts.

- 8 Unflagged borrowings include 46 borrowing tokens (n<sup>inf</sup>=26) where untranscribable speech nearby prevented knowing for certain whether they were flagged or not. For example, look at the following excerpts from informants 005U and 042U.

*Yo trabajaba para xxx company en Tampa, y un empleado se murió sentado...* 005U  
'*I was working for xxx company in Tampa, and an employee died sitting up...*'

*Yo viví primero en xxx Avenue.* 042U  
'*I first lived in xxx Avenue.*'

Informant 005U uses the borrowing *company*. The speech preceding *company* was unable to be transcribed (indicated by "xxx"), so it cannot be said for certain whether the borrowing is flagged or not. In informant 042U's sentence, the speech preceding *Avenue* could not be clearly understood. It is, thus, uncertain whether *Avenue* is flagged or not. However, in most of these uncertain cases, it seems the borrowing is probably **not** flagged. I, therefore, include these 46 'uncertain' cases among the unflagged.

of the corpus, 558, or 12 percent, were flagged. Said differently, 88 percent of the English used in the Spanish discourse of these informants is used without hesitation, metalinguistic commentary, false starts or fillers (row two of the table). A comparison of this flagging rate with that provided in Poplack's (1980) study of codeswitching by Spanish-English bilingual Puerto Ricans in New York City suggests that flagging by the Spanish-speaking bilinguals and monolinguals of the present investigation might be higher than expected in an unself-conscious speech mode. In Poplack's study, 96 percent of codeswitches were made with no "editing phenomena" (which included false starts and hesitations). That is, only four percent of observed codeswitches were flagged (1980: 601). Flagging observed in Poplack's study seems to match disfluency rates in spoken speech in general, which cross-study estimates put at around six percent (Fox Tree 1995: 710). This six percent is a third of the rate at which OZC informants flag lexical borrowings. The difference in flagging between her mixed speech corpus and the OZC could be due to the fact that Poplack's informants were in a *bilingual mode* (Grosjean 1997a, 1997b), where switching between languages is the unmarked choice and, thus, where explicit flagging of the switch was unnecessary. On the other hand, OZC informants were aware that they were interviewed because they spoke Spanish. Consequently, some may have believed Spanish to be the only designated language of conversation. As a result, when departing from perceived monolingual mode by using English, they flagged. So, in fact, the comparatively higher flagging rate for OZC informants suggests speaker awareness of the foreign origin of English lexical material. Results presented in section 6.2, that those with better skills in and knowledge of English flag borrowings more often than those with less English ability (Table 7.3 and Table 7.4) will further support the idea that flagging by Spanish speakers in New York is, in fact, an indication of the speaker's awareness of the foreign origin of English words and phrases.

### 6.1.3 Results of analysis and discussion of hypothesis (1b): Flagging indicates that an English lexical borrowing is not part of the lexicon of Spanish in New York

We now consider the hypothesis that flagging is, additionally, an indication that what has been uttered is not a fully-entrenched part of the lexicon of Spanish (hypothesis 1b). Table 7.2 shows the flagging rates (in bold) for borrowings with different degrees of community acceptance. The table includes data that are not relevant to discussion, but that help to make sense of the table. The data pertinent to the present inquiry are within the striped boxes.

**Table 7.2**  
**Corpus flagging rates for borrowings with different degrees of community acceptance**

	Flagged		Unflagged		Total	
	#	%	#	%	#	%
Widespread (n <sup>inf flagging</sup> =123)	125	<b>8</b>	1536	<b>92</b>	1661	<b>100</b>
Nonce & idiosyncratic (n <sup>inf flagging</sup> =115)	251	<b>18</b>	1124	<b>82</b>	1375	<b>100</b>
Periodic / Recurrent <sup>†</sup> (n <sup>inf flagging</sup> =77)	182	<b>12</b>	1289	<b>88</b>	1471	<b>100</b>
All borrowings (n <sup>inf flagging</sup> = 105)	558	<b>12</b>	3949	<b>88</b>	4507	<b>100</b>

<sup>†</sup> Periodic borrowings are used by two to four informants. Recurrent borrowings are used by five to seven informants.

Of the 4507 borrowing tokens, 1661 of them were widespread borrowings (used by eight or more informants).<sup>9</sup> Of widespread borrowings, only 8 percent were flagged in some way. This average is less than half the average flagging rate for borrowings uttered by only a single informant; the flagging of nonce and idiosyncratically used borrowings is 18 percent. In other words, although no test of statistical significance was performed, these flagging averages appear very different.

9 Poplack et al. (1988) define widespread items as those uttered by ten or more informants. This definition has been changed for the current investigation to eight or more due to the smaller size of our lexical borrowing database. (Theirs has over 19,500 tokens; ours has just over 4500.)

They, furthermore, are in line with what would be expected if (i) corpus frequency and distribution were indicative of community acceptance of a borrowings and (ii) flagging in the vicinity of lexical borrowings indicates that what has been flagged is not part of the community repertoire of English borrowings in Spanish.

#### **6.1.4 Summary of findings for the language membership hypotheses**

In sum, we have seen evidence in favor of both language membership hypotheses. Flagging rates are higher for informants who

<b>Language membership hypotheses</b> 1a: Lexical material has foreign origin. 1b: Lexical material is not Spanish lexicon.
---

are likely in a monolingual Spanish mode than for Spanish speakers in a bilingual language mode. In addition, flagging in the vicinity of lexical borrowing is about three times higher than average flagging frequencies in monolingual discourse. This (along with evidence presented in the following section, 6.2) supports hypothesis (1a), that flagging is an indication that speakers are aware of the foreign-origin of English borrowings. In addition, the fact that flagging is so much less frequent in the vicinity of borrowings hypothesized to be widely accepted by a Spanish-speaking community (i.e. widespread borrowings) than in the vicinity of borrowings that appear to have less community acceptance (i.e. nonce and idiosyncratic borrowings) supports hypothesis (1b), that for informants of this corpus, flagging is an indication that an English borrowing has not yet become entrenched in their Spanish lexicons.

## **6.2 The disfluency hypotheses**

### **6.2.1 Evidence in support of the disfluency hypotheses**

In this section, only evidence for hypothesis (2b), that flagging indicates a speaker's linguistic disfluency, will be outlined. This is because the validity of hypothesis (2a), that a false start, hesitation or metalinguistic comment indicates the presence of on-going cognitive processes is

fairly well-attested for monolingual and bilingual speech (Goldman-Eisler 1968; Poullisse 1997: 206; Siegman 1979).

**Disfluency hypotheses**

2a: Momentary problem

2b: Language attrition

2c: Incomplete acquisition

Hypothesis (2c), that flagging is a sign of incomplete acquisition of Spanish, must be rejected for OZC informants. This is because all informants of this study are, by methodological design, Spanish speakers. Each one conversed fluently in Spanish, whatever the extent of English use during his interview, and regardless of his feelings about his Spanish abilities. As further confirmation of this, consider that only two informants of 146 gave themselves the lowest rating of ‘poor’ when asked about their Spanish abilities. Both of these supposedly ‘poor’ Spanish speakers conversed entirely in Spanish; one was very loquacious occupying long conversational turns, the other seemed to prefer to respond to the interviewer’s prompts with two sentences or less, but never once responded with a full clause of English.

To validate hypothesis (2b), that flagging indicates disfluency in a speaker’s Spanish (or English), that is, a lack of competence, it would be expected that individuals that are poorer in Spanish (or English) flag lexical borrowings more frequently than those that are better in Spanish. Furthermore, if it is supposed that linguistic competence (in any language) is refined or developed through formal education, additional support for hypothesis (2b) would be garnered if it were found that those with more years of formal education flag less often than those with fewer years. In addition to these variables, the relationship of flagging to other sociodemographic traits of the speakers, such as immigrant generation, ethnonational variety of Spanish spoken, age, sex, occupational class, years in the U.S. and daily Spanish use are also examined in order to discover whether flagging may be governed by some other aspect of speaker identity, such as being a man or woman or being older or younger.

To summarize, support for the validity of the language disfluency hypotheses for this corpus is to be sought in the following evidence:

- 2a: Assumed true for any speaker/corpus
- 2b: Poor Spanish (or English) speakers flag more than good Spanish (or English) speakers.  
Those with fewer years of schooling flag more than those with more years of schooling.
- 2c: False for informants of this corpus.

### 6.2.2 Results of analysis for a disfluency hypothesis

ANOVAs reveal that few group traits significantly influence the extent to which borrowings are flagged. Among those that did were: immigrant generation ( $F(1, 114) = 3.19, p < .077$ ), English skills ( $F(1, 131) = 5.05, p < .026$ ) and daily Spanish use by ( $F(1, 131) = 6.92, p < .010$ ).<sup>10</sup> These findings will be presented before synthesizing their implications for the disfluency hypotheses.

---

<sup>10</sup> Among the non-influencing variables are ethnonational affiliation ( $F(5, 128) = 1.38, p < .234$ ), region ( $F(1, 132) = 0.02, p < .883$ ), sex ( $F(1, 132) = 0.02, p < .887$ ), age ( $F(3, 130) = 0.06, p < .983$ ), level of education ( $F(1, 131) = 2.37, p < .127$ ), years in the U.S. (1<sup>st</sup> generation only) ( $r(90) = -.01, p < .958$ ), class ( $F(1, 127) = 0.34, p < .564$ ) and self-rated Spanish skills ( $F(1, 131) = 0.44, p < .510$ ).

Immigrant generation is a marginally significant factor in conditioning how frequently an informant flags his borrowings ( $p < .077$ ) (Table 7.3).

**Table 7.3**  
**Informant flagging rate by immigrant generation<sup>†</sup>**

	Mean percent (%) of borrowings flagged	SD
<b>2<sup>nd</sup> generation</b> (n=27) (arrive ≤ age 3)	14.1	8.7
<b>1<sup>st</sup> generation</b> (n=89) (arrive ≥ age 13)	10.0	10.8
<b>All informants</b> (n=116)	11.0	10.5

$$F(1, 114) = 3.19 \quad p < .077$$

<sup>†</sup> The results in the table are the most robust findings for immigrant generation attained. Child arrivers (arrived age 4-12), normally grouped with the second generation in this study (see Chapters 4, 5 and 6), were not included in the results of this table because their flagging rates were not in line with those of other second generation individuals.

Table 7.3 shows that the second generation (defined in the present chapter as those who were born in or arrived to the U.S. at or before age 3) flag borrowings at a rate of 14 percent. Those of the first generation flag less frequently. On average, only 10 percent of borrowings used by first generation informants are flagged.

English skills also significantly conditions flagging ( $p < .026$ ) (Table 7.4).

**Table 7.4**  
**Informant flagging rate by English skills**

	Mean percent (%) of borrowings flagged	SD
<b>Excellent</b> (n=45)	13.9	9.1
<b>Non-excellent</b> (n=88)	9.7	10.6
<b>All informants</b> (n=133)	11.1	10.3

$F(1, 131) = 5.05 \quad p < .026$

The table shows that those who rate themselves as excellent in English flag borrowings more than rate themselves as less than excellent. Those with better self-ratings flag 14 percent of their borrowings. Those that rate their English unfavorably flag borrowings 10 percent of the time.

Finally, it is not an informant's abilities in Spanish but, rather, whether he uses Spanish daily that is related to the extent to which he flags lexical borrowings ( $p < .010$ ) (Table 7.5).

**Table 7.5**  
**Informant flagging rate by daily Spanish use**

	Mean percent (%) of borrowings flagged	SD
<b>No Spanish use</b> <sup>†</sup> (n=13)	18.1	10.6
<b>Some to a lot of Spanish use</b> (n=120)	10.4	10.0
<b>All informants</b> (n=133)	11.1	10.3

$F(1, 131) = 6.92 \quad p < .010$

<sup>†</sup> Tukey's post hoc showed that only those who do not use Spanish in their daily lives **at all** were different from those who use a little, a moderate amount and a lot.

Table 7.5 shows that informants that do not use Spanish in their daily lives flag more often than those that use Spanish to at least some extent. Those that use Spanish flag at a rate of about 10 percent, while those that do not use it at all flag borrowings almost twice as often, 18 percent.

### 6.2.3 Summary of results for a disfluency hypothesis

To summarize, results of analysis have revealed that flagging of English lexical borrowings is done **more** by informants that have:

- Second (immigrant) generation status
- Better English skills
- No use of Spanish in their daily lives.

### 6.2.4 Discussion of hypothesis (2b): Flagging does not signal lack of proficiency in Spanish or English

With respect to the fluency hypotheses, it was expected that if flagging were indicative of disfluency in at least one of a bilingual's languages (hypothesis 2b), flagging would be more frequent among those with **poorer** Spanish or English skills. Yet, just the opposite has been found. Those with better English flag most often. In addition, it is not Spanish skills, but rather the extent to which Spanish is use in daily life that conditions flagging. Informants that do not use Spanish, that is, those that use English for their daily activities, flag more often. Furthermore, it was hypothesized that persons with **fewer** years of formal education would flag more often. This was not the case. With respect to years of formal education, informants all flagged to the same extent. In sum, not only has support in favor of hypothesis (2b) failed to be found, but the Spanish-speaking informants of this study have provided evidence that subverts the assumption that flagging in bilingual speech is associated with poor language skills.

<b>Disfluency hypotheses</b> 2a: Momentary problem 2b: Language attrition 2c: Incomplete acquisition
---

Among Spanish speakers in New York City, flagging of lexical borrowings is associated with greater abilities in English, and is unaffected by informant's proficiency in Spanish. Such results, it should be noted, also provide tangential support in favor of hypothesis (1b), that flagging indicates the awareness of the foreign origin of English lexical borrowings, since it is precisely those with more knowledge of English that flag most.

Finally, interpreting flags as a sign of incomplete acquisition (2c) could also not be supported for OZC informants since, as previously mentioned, all informants are fully conversant in Spanish. In sum, only the notion that flagging indicates a momentary processing lapse (hypothesis 2a), a well-documented interpretation of this phenomenon, can be supposed to be a reason for flagging borrowings in the OZC.

### **6.3 Summary of conclusions**

It has been seen that with respect to the language membership hypotheses, flagging by New York City Spanish speakers appears to indicate the speaker's awareness of the foreign origin of lexical borrowings (hypothesis 1a), as well the fact that a borrowing does not have community-wide acceptance in a lexicon of Spanish in New York (hypothesis 1b). With respect to disfluency, not only could the idea that flagging indicates lack of competence in Spanish (hypothesis 2b) **not** be supported, but just the opposite was found; flagging was not related to Spanish proficiency and informants with better English flagged more than those with poorer abilities in English. In other words, flagging in the OZC cannot be interpreted as an indication of linguistic deficiency, but only as signaling awareness of the foreign origin of an English phrase and that the English lexical material is not a fully-accepted element in the lexical repertoire of Spanish in New York.

## 7. Reinterpreting flagging for lexical borrowing in Spanish in New York

It has thus far been seen that flagging in Spanish in New York is modulated by a particular cluster of associated factors. It is the cluster itself, the variables it includes and those that it does not, that aids in interpreting the significance of flagging in the vicinity of other-language stretches of speech. While the English skills of an informant are related to flagging, Spanish skills are not at all, indicating that knowledge of English, but not knowledge of Spanish, is central to the interpretation of flags. Given that flagging is also influenced by the immigrant generation of the informant, and in light of the importance of English knowledge and use to flagging, it seems reasonable to conclude that it is precisely the high level of bilingualism of the second generation, particularly their fully-developed competence in English, that is partly responsible for their increased rate<sup>11</sup> of flagging.

While the overall trend of flagging in the corpus is small, between 10 and 20 percent of other-language other material, depending on the speaker group, results point to an intriguing trend in lexical borrowing in Spanish. The trend becomes apparent when the significantly-related variables of immigrant generation, English skills and daily Spanish use are reinterpreted as indicative of the social and psycholinguistic status of Spanish and English in the community and for individuals in real time. The variable of immigrant generation is a proxy for the shift in importance of Spanish, both on the level of the community as well as for individuals, in its transmission from one generation to the next. Spanish is spoken by all individuals in the first generation and is usually the dominant language for most of them. In the second generation, Spanish is spoken by fewer individuals and is psycholinguistically on a par with, or less

---

11 I cast the discussion in terms of 'increased' rate for the second immigrant generation (as opposed to 'decreased' rate for first generation) because the average flagging rate of second generation informants more sharply diverges from the cross-participant average of 11 percent than the average flagging rate of the first generation.

dominant, than English for a majority. English skills proxies for individuals' knowledge of English with increased exposure to it in the New York environment. Finally, daily Spanish use approximates what happens to Spanish as it moves from being a principal language in the home country to being relegated to fewer environments in the New York context. Daily Spanish use may also describe what happens to Spanish in the individual bilingual's life. Spanish shifts from being the primary language of communication prior to school to being less used when the individual is of school age and perhaps least of all when one's waking hours are consumed with work in an English-speaking world. Recasting the discovery that flagging indicates explicit acknowledgment of foreign origin and lack of Spanish lexicon membership of English lexical borrowings in real time, then, what has been found is that lexical borrowings are slightly less likely to become part of a New York City Spanish lexicon...

- As Spanish moves from a first generation to a second generation language.
- As Spanish changes from **the** language of primary communication to being just one of several.
- As individuals gain in English proficiency.

In fact, what seems to be happening to lexical borrowings in Spanish as it changes generation's hands, as it changes functional status in the new community or in an individual's life and as individuals grow in their knowledge of English is that it is (inadvertently?) 'deborrowed'. That is, there appears to be a process, albeit slight, whereby what is taken from English can be discretely returned through flagging. As mentioned, this trend is marginal in that it only affects between 10 and 20 percent of the lexical borrowing tokens uttered. Nonetheless the pattern intimates at something more substantial: where Spanish is maintained, the likelihood and extent of permanent change to its lexicon may decrease as the individual and the community becomes more aware of and gains in knowledge of English.

## 8. Conclusion

This chapter has revealed that flagging in the vicinity of lexical borrowing is not done uniformly by all speakers in a community. It has argued that the interpretation of flagging for bilingual speech, particularly what it says about a speaker's linguistic competence, should be not be assumed apriori, but determined with respect to the way it patterns for a particular group of informants and in a particular data set. It is the particular combination of significantly and non-significantly related variables that helps to identify the function of flagging for a group. For informants of the OZC, the hypotheses that flagging is related to a speaker's awareness of an utterance's foreign origin was supported. Further, it was found that flagging is indeed related to individual linguistic competence, but not in the typically assumed way. Flagging in the OZC has to do with increased knowledge and awareness of English and **not** lack of competence in Spanish.

Additionally, flagging in Spanish in New York City may signal a small, arguably insignificant, but nonetheless existent trend, whereby so-called "borrowed" English lexical material may be unborrowed by those with better knowledge of English, so long as perceived pressure to maintain languages separate in speech remains constant. Although the facts of Spanish language transmission and Latin American immigration to the city point to the continued use of lexical borrowings from English in speech, the presence of this deborrowing trend indicates an increasing unlikelihood, over time, that the borrowings used will become established in a lexicon of Spanish in New York.

## **Chapter 8: Conclusion**

### **1. Contributions and applications**

Sociolinguistic work in the last several decades has demonstrated the viability of quantitative methods for observing language change in progress. The current investigation, an examination of variation in lexical borrowing behavior, has shown the usefulness of such methods for a multilingual and multidialectal community. In particular, and as shown elsewhere (e.g. Matus-Mendoza 2002; Ngom 2006; Poplack et al. 1988; Sullivan 2008, to cite just a few), it has been seen that the notion of sociolinguistic variable can be productively expanded beyond the domain of phonology to include lexical borrowing.<sup>1</sup> This study also contributes to other fields of language study, including language contact and Spanish in the U.S. Results, furthermore, have implications for current issues in the pedagogy of language, language ideology and bilingual speech research. These contributions and applications, as I see them, are detailed in the following sections.

#### **1.1 Contribution of the current study to Spanish in the U.S.: A portrait of lexical borrowing in Spanish in New York**

Results of this investigation have yielded an integrated and, sometimes surprising, portrait of lexical borrowing in New York City. As is well-known, the use of English while speaking Spanish is common. This was confirmed in that 95 percent (n=139) of Spanish speakers in this study used English lexical borrowings (Chapter 4, section 3, also Figure 4.1). In fact, not only is borrowing common, but it is part of speaking Spanish in the city. Even new arrivals immediately adapt to this behavior, the way that any traveler to a new place would learn the local names for

---

1 See Lavandera (1978), who enumerates several objections to treating non-phonological features of language as sociolinguistic variables.

things. In short, the impact of cultural contact on the lexical choices that immigrants make is abrupt, even if the permanent adoption of other-language expressions into the lexicon of their language (a social process) is more gradual.

Variation in the frequency of use of lexical borrowings has pointed to the fact that, while commonplace, borrowing is not done the same way by everyone. Individuals borrow more or borrow less in accord with their membership in particular groups and their individual traits. Among Spanish speakers in New York City, the primary determinant of borrowing frequency is immigrant generation (Chapter 4), with the second generation borrowing twice as often as the first generation (Chapter 4, sections 4.1 and 5.1). This pronounced difference in borrowing suggested that each immigrant generation is a distinct subgroup within the larger Spanish-speaking community, a conclusion further supported by an analysis of borrowing within each immigrant generation (Chapter 5).

Lexical borrowing may perform distinct functions for first and second generation Spanish speakers. In the first generation, results have indicated that lexical borrowing procures (covert) prestige. While Anglos and Spanish speakers alike may associate negative attributes (such as being low class or uneducated or lazy) with so-called language mixing, it is, in fact, those of higher social classes, with more education and with better Spanish **and** English skills that borrow most often (Chapter 5, section 3.2 and 3.4). Indeed, occupational class is the strongest predictor of borrowing frequency in the first generation (Chapter 5, section 3.5).

In the second generation, lexical borrowing may index ethnic identity. Puerto Ricans borrow more than others, while Mexicans borrow least of all (Chapter 5, section 2.2). In light of the fact that grammatical distinctions between ethnonational varieties of Spanish are leveling out

(Otheguy & Zentella 2012), it may be that second generation speakers use lexical borrowing to maintain or emphasize ethnonational differences (Chapter 5, sections 2.4.1 and 2.6).

Research on Spanish in the U.S. has stressed that mixed speech is not indicative of linguistic deficiency or deficit. Yet, negative perceptions still persist, even among the well-meaning.<sup>2</sup> Evidence introduced in the present work has also demonstrated that lexical borrowing is not a deficit behavior. In so doing, this work has contributed to validating the linguistic behavior of the multilingual population in the U.S as a resource, and not a disadvantage. In neither generation is lexical borrowing associated with less education, less social class standing, less English proficiency or less Spanish proficiency. On the contrary, it is associated with better social standing, education and language skills. Frequent lexical borrowing and the use of novel words and expressions, a noticeable kind of borrowing, appear to be the province of the most socially and linguistically secure in both the first (Chapter 5, sections 3.2 and 3.4.3; Chapter 6, sections 3.1 and 3.2) and second generation (Chapter 5, sections 2.2 and 2.4.3; Chapter 6, section 3.2). Overall, the impression is that linguistic confidence is a cornerstone of borrowing in New York.

Underlying these differences, there appears to be a set of norms that unites Spanish speakers in the city across lines of class, ethnicity and immigration history. Specifically, novel and shared vocabulary (as defined in Chapter 6) are both integral components of borrowing (Chapter 6, sections 2, 3.5 and 4.5); the majority of individuals, first and second generation alike, use both types of vocabulary, in equal measure, to borrow.

---

2 For instance, Weinreich commented in connection to Romansch and German in Switzerland that “the general level of language cultivation (*Sprachkultur*) (...) accounts for the predominant types of speakers in the population” (original emphasis). He elaborates on the notion of *Sprachkultur* in the following comment where he says that in a population “**with little schooling** and a loose social hierarchy”, “the type of bilingual who mixes both languages indiscriminately will be relatively frequent” (1966: 84, my emphasis). As a more recent example, Ilan Stavans writes: “Spanglish is often described as the trap, la trampa Hispanics fall into on the road to assimilation. Alas, the growing lower class uses it, thus procrastinating the possibility of un futuro mejor, a better future” (2003: 3).

Completing our portrait is the discovery that the borrowing process in Spanish in New York may be complemented by a simultaneous (albeit slight) process of deborrowing. Results suggest that flagging was indicative of speakers' awareness of the foreign origin of English borrowings (Chapter 7, sections 6.1.2 and 6.1.3). When results are interpreted in real time, they indicate that, as the use and knowledge of English increases in the community and in lives of individuals over time, what was once borrowed from English may be discretely "returned".

## **1.2 Applications of the current study to language contact research**

### **1.2.1 The concept of lexical borrowing**

The current investigation has contributed a characterization of lexical borrowing that makes it germane to the synchronic investigation of other-language strings in a situation of on-going language and dialect contact. In particular, this study has endorsed a characterization of lexical borrowings as other-language strings used in recipient language discourse before the time at which, if ever, they become established elements of a recipient language lexicon. It has identified relevant data by (i) excluding foreign-origin strings known to be part of the preimmigration Spanish of informants (Chapter 3, section 4.2) and (ii) excluding foreign-origin strings that may be used due to informants' exposure to other varieties of Spanish (Chapter. 3, section 4.3). Although the methodological implementation of the latter criterion was only approximate, it has gone a considerable way toward ensuring that the data and results of this investigation reflect the impact of English on Spanish, and not the impact of varieties of Spanish on each other.

### **1.2.2 Some aspects of lexical borrowing are more tied to bilingual ability than others**

Thomason and Kaufman note that "lexical borrowing frequently takes place without widespread bilingualism" (1988: 37). This observation has been supported by the results of this study, which showed that some aspects of borrowing are less tied to language proficiency than others. For

instance, although greater proficiency in English encourages borrowing, English proficiency is not a prerequisite for borrowing (Chapter 5, sections 3.1.3, 3.2 and 3.4.2; also Chapter 4, section 7.4). Proficiency in the recipient language, Spanish, has also been shown to have little to do with borrowing (Chapter 5, section 2.1). **Confidence** in one's skills in the recipient language does, however, encourage borrowing (Chapter 5, sections 3.2. and 3.4.3). One aspect of borrowing that is dependent on donor language proficiency is the introduction of novel vocabulary. Borrowing innovators in New York are Spanish speakers who, regardless of immigrant generation, sex or age, are more proficient in English, have more confidence in English and who use it more for daily activities (Chapter 6, section 3).

### 1.2.3 Social determinants of lexical borrowing across contact situations

Lexical borrowing is a phenomenon that, like other variable linguistic features, both reflects and creates complex realities. Much research had explored the pragmatic motivations (for instance, Auer 1998; Blom & Gumperz 1972; Gumperz 1976, 1982) and linguistic constraints (for instance, Berk-Seligson 1986; Muysken 2000; Pfaff 1979; Poplack 1980) that influence lexical borrowing and language choice. In addition, works that are both detailed and broad in scope, such as Thomason and Kaufman's *Language contact, creolization and genetic linguistics* (1988), have advanced our understanding of the social conditions that encourage lexical borrowing.<sup>3</sup> Yet very little is currently known about variation in lexical borrowing behavior within a community, in particular, how it differs between individuals and who is responsible for the spread of

---

3 For example, Thomason and Kaufman demonstrate that when the recipient language is the L1 of the community (although not necessarily the L1, mother tongue or dominant language of every member of the community), it will first evidence change in the domain of lexicon, and only later, will its syntax and phonology be affected (1988: 37). Prolonged contact, high amounts of societal and individual bilingualism and cultural pressure are factors that increase the degree to which a recipient language lexicon shows influence from a donor language (1988: 46-48). The relationship between the social factors (in particular, speaker agentivity) and the domains of language likely to experience contact-induced change is also addressed in Van Coetsem (2000).

borrowings. Of codeswitching (data that in part corresponds to what has been called lexical borrowing in the present study), Lipski (2005: 2) writes, "... to the syntactic peculiarities of Spanish-English code-switching must be added an as yet not fully understood set of sociolinguistic variables that facilitate high density intransentential [sic] switching among most bilingual Latino communities in the United States." The current study has been but a modest contribution toward this endeavor. And, while not in a position to offer a theory of how lexical borrowings are taken up and disseminated in particular contact situations (such an undertaking deserving of an entirely distinct analysis), I submit a handful of observations on when and where specific patterns in the use of lexical contact phenomena may be likely.

**National origin.** To my knowledge, other studies have not explicitly examined the impact of national origin or language variety on lexical borrowing. In New York City, immigration often patterns as chain migration, where family members and acquaintances from one home country attract others from the same area, who settle nearby in the host country. This pattern of resettlement, it seems, would create communities in the host country that are internally cohesive, both socially and linguistically, and that differ from communities formed around ties to different home countries. It would be expected, thus, that lexical borrowing would differ from one ethnonational group to the next. Yet, results from this study indicated that neither the ethnic nor dialectal diversity of immigrants bears on lexical borrowing behavior in the first generation. Rather, first generation immigrants seem, on the whole, to leave ethnonational distinctions by the wayside and establish or conform to pan-dialectal norms for borrowing in the host country (Chapter 5, section section 3.2.1).

**Sex and gender.** Lexical borrowing seems rarely to reflect gendered identity (e.g. see Eslami Rasekh et al. 2008: 556; Mendieta 1999: 96-8). This was also true in the present

investigation (Chapter 4, section 7.2 and Chapter 5, sections 2.5.1 and 3.5). Even in studies where lexical borrowing is correlated to sex, the reason for the difference between men and women is attributed to network affiliation and, in particular, disparate employment patterns (e.g. Klee 1987; Poplack et al. 1988).

**Age.** Although no effect for age was discovered in this study, age has frequently been found to be a determinant of the quantity and types of borrowings used. In general, younger individuals borrow most often (e.g. Amara & Spolsky 1986; Eslami Rasekh et al. 2008; Matus-Mendoza 2002; Ngom 2006; Poplack et al. 1988).<sup>4,5</sup> In several of the cases cited, authors say that lexical borrowing takes on oppositional ideological associations and, in particular, may signal modernity (e.g. Eslami Rasekh et al. 2008: 557<sup>6</sup>; Ngom 2006: 98), an urban identity (e.g. Ngom 2006: 98) or economic power (e.g. Matus-Mendoza 2002). Poplack et al. do not explicitly make this kind of statement, but they do note that the borrowing inventory of the young in their study includes many words for recent cultural phenomena (1988: 87). Additionally, in most of the cited contexts, access to the donor language seems to be gained chiefly through formal channels, such as education.<sup>7</sup> It may be that age significantly predicts borrowing where the donor language is not only associated with economic empowerment or social prestige, but also when access to that

---

4 Eslami Rasekh et al. (2008) found that their younger group of Turkman speakers in Iran borrowed more words from Persian than their older group. Matus-Mendoza (2002) found that younger Moroleonians used English loanwords more often than older ones. Ngom (2006) found that younger people use more English words in their Wolof speech than the older age groups. (On the other hand, older people used more Arabic loanwords than the young.) Poplack et al. (1988: 76) found that younger Francophones in Quebec and Ontario Canada borrowed more often and used more nonce borrowings than older speakers. In Israel, Amara and Spolsky (1986) also found age-related differences in borrowing into Arabic. However, their age groups are defined in a way that makes their results not directly comparable to the results of others cited.

5 Age may also be related to different styles of bilingual speech mixing (Muysken 2000: 227-8).

6 That lexical borrowing may be associated with modernity in Eslami Rasekh et al. (2008) is implied when the authors say that older speakers may borrow less from Persian, because they want “to keep their Turkmen culture and identity more than younger speakers” (2008: 557). From this, I understand Turkman culture to be symbolic of tradition.

7 Note, however, that in the region of Canada where Poplack et al.’s participants resided English is spoken as a first language by considerable portions of the local population. Nonetheless, access to it for their younger Francophones appears to be primarily through the schools (1988: 87).

donor language is limited or relegated to particular institutions. The lack of age-related tendencies in borrowing in this study<sup>8</sup> may relate to the availability of the donor language, English, outside of educational contexts, that is, as the de facto language of the host community.

**Occupational class and socioeconomic status.** Research in linguistic anthropology has shown that, when two cultures or peoples come into contact, the group with less socioeconomic prestige or power generally borrows more words from the more powerful or prestigious group than vice-versa. It may come as no surprise then, that lexical borrowing has been observed to vary with occupational class and socioeconomic status **within** a recipient language community. The direction of the impact, however, differs for different contact situations. For Quebec Francophones (Poplack et al. 1988) and for Moroleonians (Matus-Mendoza 2002), borrowing was less frequent among the upper classes. In the current study, middle class Spanish speakers borrowed more often, and introduced more novel vocabulary while borrowing, than the working class. It was suggested in Chapter 5 (section 2.3.1) that this may be related to the sociolinguistic setting outside of the recipient language community. Lexical borrowing may be more frequent by individuals in lower classes when the recipient language has a high instrumental value in the local socioeconomic market (such as Spanish in Moraleón or French in Quebec and nearby neighborhoods) (see Poplack 1987: 57). In these contexts, use of the donor language, which probably garners international prestige or prestige in an expanded context, may allow lower class individuals to situate themselves outside of their local socioeconomic hierarchy and with respect to a global market. In New York, the donor language has high instrumental value and is associated with socioeconomic power both in a global market and in the local one. Furthermore, the recipient language, Spanish, is often depicted as an obstacle to economic success. In such

---

8 Age was not associated with borrowing rates (Chapter 4, section 7.3; Chapter 5, sections 2.5.1 and 3.5) nor with the use of novel or shared vocabulary (Chapter 6, sections 3.4 and 4.4).

contexts, it may be the middle class, rather than the working class, that has a greater incentive to demonstrate their command of English, partially through their use of lexical borrowings.

In sum, although more data are needed, evidence from New York and several other multilingual settings promises to bear fruit for the construction of a cross-context account of how borrowings spread and what motivates their use in different multilingual communities.

#### **1.2.4 Some factors that promote a high density of permanent contact features in language also promote a high density of contact features in speech; others do not**

It has been observed that the greater the intensity of contact, the more contact features will be present in a recipient language (Thomason & Kaufman 1988: 72). Specifically, (a) the greater the length of time that groups are in contact, (b) the greater the amount of bilingualism and (c) the greater the socioeconomic prestige of a donor language, the more contact features a recipient language is likely to evidence in the long run. One wonders whether the same variables that influence the density of contact features that become permanent in a recipient language also influence the density of contact features in speech.

Evidence from the current investigation shows that the bilingual ability of speakers may correspond not only to the extent that contact features become permanent, but also to the number of contact features, in particular lexical borrowings, in speech. This was suggested by the finding that second generation informants, who were bilingual to a greater extent, borrowed almost twice as much first generation informants (Chapter 4, sections 4.1 and 5.1).<sup>9</sup> This was also true for first generation informants; those with better English borrowed more (Chapter 5, sections 2.1 and 2.3.3).

---

9 Results of García, Otheguy and Fernandez (1989) also suggest that bilingual ability influences the number of contact features in speech. They found that second generation Cubans used more loanwords while speaking than first generation Cubans. Likewise, Zentella has written in connection with loanwords in Spanish that “logically, more intense exposure to English over longer periods of time tends to result in the incorporation of more loanwords” (1990: 1101).

On the other hand, evidence in support of the influence of other variables on the number of contact features in speech was not forthcoming. For instance, it was supposed that individuals who had spent more time in the U.S. would borrow more frequently than others. This was not the case. Recently arrived Spanish-speaking immigrants borrowed to the same extent as those that had been in the U.S. for many years (Ch. 4, section 4.6).

Similarly, it was thought that Spanish-speaking groups with longer histories of contact with Anglo culture and English in the U.S. would use a greater quantity of lexical borrowings. Some results suggested this might be the case. In particular, Puerto Ricans, the group with the longest history of contact in the U.S, appeared to borrow more often than other groups, in both the first and second generation. However, in the first generation, the difference between the borrowing rate of Puerto Ricans and other groups was not statistically significant (nor did it approach statistical significance; section 2.5.2 of Chapter 5). That is, it could not be asserted with a desirable degree of probability that there was, in fact, a real difference between how often Puerto Ricans borrowed and how often other first generation ethnonational groups did.<sup>10</sup> Further, the lack of correspondence in borrowing by ethnonational groups in the first and second generation led to the conclusion that the higher borrowing rate of Puerto Ricans in the second generation was a marker of ethnic identity, rather than a consequence of the length of the community's contact with English in the U.S. (Chapter 5, section 2.4.1).<sup>11</sup>

In other words, while the length of contact between groups may influence the quantity of lexical borrowings that eventually become part of a recipient language lexicon, it may not affect

---

10 This led to the hypothesis that, with respect to lexical borrowing in New York City, the patterns established by the first immigrant group in contact may be taken up and continued by later immigrant vintages, regardless of the length of time later immigrant vintages have been in contact with the culture and language of the host country.

11 However, as noted in Chapter 5, the "length of group contact" interpretation might still hold up if it were found that certain ethnonational groups did not form distinct communities, but settled among groups already present in the city. See footnote 21 of that chapter.

the density of borrowings in speech. Rather, the density of borrowing in speech seems better predicted by traits, such as an individual's degree of involvement in the host community, his aspirations for upward socioeconomic mobility and his language skills and confidence.

### **1.3 Contributions of the current study to bilingual speech research: The classification of lexical contact phenomena**

#### **1.3.1 The distinction between codeswitching and lexical borrowing can be made with respect to patterns of language use in a community**

Poplack and Dion (2012: 279-280) note that even after 30 years of research on bilingual speech, there is still a lack of consensus on how (and even whether) to differentiate between lexical borrowing and codeswitching. In the present study, a corpus-based analysis of English strings of various lengths and syntactic compositions has facilitated the creation of an operational distinction between borrowing and codeswitching that reflects the patterns of language use in the community (Chapter 3, section 4.1). This corpus-based distinction allowed for the variable linguistic form that foreign-origin lexical material may take at early stages of use in a community. The corpus-based distinction resulted in the identification of English strings that are elsewhere called codeswitches as lexical borrowings. The broadening of the scope of what is considered lexical borrowing data may be useful for tracing the transmission of donor language lexical material in Spanish in future investigations. That is, it permits an examination of whether longer donor-language strings promote the retention of individual words in the lexicon of the recipient language.<sup>12</sup> For instance, it may be that frequent use of *high school*, as well as *military*

---

12 Thomason (2001: 135), for instance, implies that there is a connection between codeswitching and what ends up as a permanent part of a recipient language lexicon. On the other hand, Poplack and Dion (2012) say that the use of longer strings (i.e. *codeswitches*), do not contribute to the adoption of other-language words as established loanwords. They claim that the high rate of morphosyntactic integration of nonce single-word strings suggests that codeswitching and borrowing are distinct processes and that speakers make instantaneous decisions about whether an other-language string will be introduced as one or the other.

*school* (329D) and *summer school* (340M), encourages adoption of *school* into the lexicon of Spanish in New York City, or that use of expressions like *business card* (427P), *index card* (300E), *credit card* (301E) and *report card* (271.2M) leads to the adoption of the word *card*.

### **1.3.2 Flagging may not be a good diagnostic for classifying data**

Flagging near other-language strings has been interpreted as indicating a lack of proficiency in a language of discourse. Flagging in the OZC, however, was not correlated with a lack of fluency in Spanish or English. In fact, not only was flagging **not** associated with poor skills in Spanish, but it was associated with better skills in the donor language, English (Chapter 7, section 6.2.4). This suggests that, even when a researcher seeks data that includes only fluently produced other-language strings, flagging is not necessarily a suitable diagnostic for lack of proficiency. Flagging could be a strategy that bilingual speakers use in unfamiliar conversational circumstances. In any event, the orderly variation of flagging in the OZC suggests that its interpretation should be grounded in its patterned use by a specific group of speakers, rather than assumed *apriori*.

### **1.4 Applications of the current study to language pedagogy**

Results of this investigation have applications for classroom language instruction. Where Spanish is taught as a second language to students that wish to use it in the U.S., educators may help student become aware (a) of variation in lexical selection, (b) of the most common borrowings in use by U.S. Spanish speakers and (c) that the use of novel borrowings is also part of how Spanish is spoken in some regions of the U.S. In English language classrooms, the fact that even recent migrants acquire a stock of English words and phrases can be a valuable scaffolding tool. Having knowledge of which words and expressions these are likely to be can aid an educator in creating content that capitalizes on the English knowledge that even newly

arrived students are likely to have. Finally, for students of heritage Spanish, recognition and validation of students' linguistic resources may be helpful in establishing trust and, in turn, a successful learning experience.

### **1.5 Application of the current study to language ideology**

Gal and Irvine (1995; also Irvine & Gal 2000) identify several semiotic processes of *ideolization*. These processes describe the how linguistic features become associated with differences between socially-constructed groups. For instance, a linguistic feature may be *iconic* of a group. It is consciously linked to a group and may itself be thought to reflect the inherent nature of the group (Gal & Irvine 1995: 973). Eckert notes that “while the entire population might agree on first-order indexicality – who uses what variant – the evaluation of that differentiation can differ across the population” (2008: 467), “since the same variable will be used to make ideological moves by different people, in different situations, and to different purposes (Eckert 2008: 467, who, in turn, credits Jonhstone & Kiesling 2008). In other words, individuals may have different ideas about what a linguistic feature says about a group. Further, Eckert notes that more work should be done on “how meanings become associated with social categories or with variables” (2008: 454-455). The current investigation suggests that processes of ideolization can proceed independently of empirical facts. In New York City, for instance, language mixture (including what has been called lexical borrowing in this study) is often associated with language deficit and individuals of lower class standing, despite the fact that results have shown borrowing to be more frequent among those of higher social standing, those with more education and those with better language skills. These facts indicate that an account must include mechanisms for how ideologizations develop in one direction despite empirical facts to the contrary.

## 2. Limitations of the current study

### 2.1 Data selection: The shared lexicon of Spanish in New York may be underrepresented

For this study, English-origin preimmigration (i.e. established) loanwords in Spanish were excluded from the data. To identify preimmigration loanwords, Spanish-speaking natives of several countries were recruited. When they identified an English-origin word as used in the home country, it was excluded from the study. When they said an English-origin word was **not** used in the home country, it was included. This method provided insight about word usage not attainable through other means, such as dictionaries (see Chapter 3, section 4.2.1). The method, however, had shortcomings, which are mentioned in Chapter 3 (section 4.2.1). One shortcoming not mentioned there is that the instincts of precontact sources may have been unreliable in some cases. For instance, some readers of this dissertation believed that words like *actualmente* ‘actually / in fact’, *elementaria* ‘elementary’ and *aplicación* ‘(job/school) application’, which were excluded from the study, were in fact **not** preimmigration loanwords. Rather, it was suggested that these words were post-immigration contact features in U.S. Spanish, which should therefore have been included.

In addition, single-word proper names (such as *Miami*, *California*, *Hollywood*, *K-Mart*, *Jamaica* and *Chicago*), numbering over 1400, were wholesale excluded from the present study. Many of these single-word proper names would probably be included in the lexical borrowing database had the protocol for evaluating these *interlingual cognates* (Chapter 3, section 4.3.2) been applied to these items.<sup>13</sup>

---

<sup>13</sup> The reason why the phonetics test was not used with single-word proper nouns was that there were too many of them to evaluate given the time constraints of the study. (Also see footnote 20 in Chapter 3.)

Particular single-word proper names and (misclassified) precontact words, like *aplicación* ‘(job/school) application’, tend to be frequent.<sup>14,15</sup> As a result of the exclusion of these two types of data, it may be that, altogether, several hundred borrowing tokens that could have been included in this study were not. More specifically, the size of the shared English-origin lexicon in Spanish in New York City may be underestimated.

## 2.2 Analysis did not control for variation due to interpersonal distance

Discourse and other pragmatic considerations may be the most important factors governing the use of lexical contact phenomena (Poplack 1980: 585). Grosjean (1997a), for instance, has shown that the production of mixed speech is highly influenced by whether a speaker perceives himself to be in *monolingual* or *bilingual language mode*. *Language mode* is a function of the setting of the interaction, the topic of conversation and, most importantly, the (perceived or known) bilingual ability of interlocutors (1997a: 227-228, 1997b: 171-172). Informants of this study took part in Spanish language interviews with interviewers that were themselves speakers of the same variety of Spanish as the informant. But, informants were obtained using a snowball technique (Oppenheim 1992), so familiarity with interviewers may have been greater or lesser for different informants. That is, different interviewer-interviewee pairs may have varied along a dimension of social distance and familiarity; some pairs may have been relatives or close friends, others may only have known each other by name prior to the interview. Variables like social distance and familiarity were not investigated or controlled for in this study, leaving open the question of how much variation in borrowing should be attributed to interpersonal factors.

---

14 *Chicago*, for example, was uttered 18 times by nine different speakers. *California* was uttered over 60 times by 22 different speakers. *Miami* was uttered 90 times by 29 different speakers.

15 For instance, *actualmente* ‘actually / in fact’ occurs at least 12 times in the OZC, used by five informants. *Aplicación* ‘(job/school) application’ and *aplicar* ‘to submit materials for a job’, together, appear 15 times by eight different informants.

### **3. Future research**

The data and results of the current study have brought to light aspects of lexical borrowing that are in need of further investigation. These avenues of research are briefly described below.

#### **3.1 A comparative examination of several types of lexical and semantic contact data**

Several scholars have suggested that a preference for a particular type of contact feature, such as the transfer of words or else the use of semantic calques, may vary according to the contact situation, and in particular with respect to the bilingual ability of the speaker (Muysken 2000; Poplack 1987; Silva-Corvalán 2002; Weinreich 1966: 61-61).<sup>16</sup> In the future, it may be instructive to explore this hypothesis by comparing the relative use of lexical borrowings and other phenomena, such as loan translation, codeswitching and semantic transfer.

#### **3.2 Borrowability scales and how social and linguistic factors jointly contribute to patterns of contact features**

Much work has been done on borrowability scales (e.g. Haspelmath 2004; Haugen 1950, 1972; Mendieta 1999; Muysken 1981; Van Hout & Muysken 1994; Whitney 1881, to name just a few). Data shows that, in the general contour of scales, there is a high degree of correspondence across contact situations and for diverse pairs of languages. For instance, recipient languages more often borrow nouns and other content words than words from functional categories. The ubiquity of this pattern suggests that it is not coincidental but is, rather, constrained by general linguistic principles (Field 2002: 22). Data from the OZC also demonstrates that borrowings, in aggregate, tend to be nouns. However, in contrast to other studies, discourse markers in Spanish in New

---

16 On the other hand, Berk-Seligson (1986) did not find any relationship between proficiency and the type of language mixture done by Spanish-Hebrew bilinguals in Jerusalem.

York are also very frequent.<sup>17,18</sup> Future research could examine whether preferences for word classes are community-wide or else whether speaker or group traits factor more prominently in borrowing from part of speech classes.<sup>19</sup> This type of analysis would contribute to current endeavors to describe how social and linguistic factors work together to produce the configuration of contact phenomena observable in bilingual speech and languages in contact.

### 3.3 Motivations for lexical borrowing

Several studies have cited referential need as a primary motivation for the introduction and use of foreign-origin words (e.g. Mendieta 1999: 43; Neumann 1938; Ornstein 1976; Weinreich 1966: 56-57. *Referential need*, logically, must refer to a gap either in the lexicon of Spanish (semantic or cultural gap) or in the individual's mental lexicon (due to language attrition, incomplete acquisition or momentary processing difficulties). But results of this study suggest that this oft-cited motivation may not be the most important motivation for borrowing in New York City. For instance, the most frequently borrowed lexemes (as seen in Chapter 6, Table 6.2) are words that have accessible translation equivalents in Spanish.<sup>20</sup> Furthermore, proficiency in Spanish (among the second generation) was not connected with borrowing frequency or the introduction of novel borrowings in this study. While this may seem unsurprising, it gives pause. It suggests that second generation informants do not borrow because they failed to learn the Spanish vocabulary to express themselves. In sum, facts suggest that the factors purported to be

---

17 This tendency is intimated in Chapter 6 (Table 6.2) where several of the most frequent lexemes were discourse markers.

18 Lipski (2008: 239) says that discourse markers, such as *so*, in Spanish in the U.S. may not be borrowings, but rather *momentary codeswitches*.

19 For instance, Poplack et al. (1988: 88) found that younger individuals (age 15-34) used non-noun parts of speech more than older individuals. Fifty percent of the borrowings of younger individuals were not nouns. Only 16 percent of borrowings used by the oldest speakers (65+) were not nouns.

20 That they are accessible is evidenced in that, along with an English-origin word (e.g. *jai ekul* 'high school' or *you know*), one will also hear the translation equivalent (i.e. *escuela secundaria* 'high school' or *tú sabes* 'you know') used, oftentimes by the same individual.

primary motivations for borrowing, such as referential need, stand in need of empirical validation. This kind of analysis could also contribute to a theory of what promotes retention of particular other-language strings over time and to an explanation of why some words resist transfer (see Weinreich 1966: 61).

### **3.4 The morphosyntactic integration of lexical borrowings**

Poplack and Dion (2012) showed that in Canadian French, single-word borrowings from English tend to be instantaneously adapted to the morphological and syntactic patterns of French. A cursory examination of single-word borrowings in the present investigation<sup>21</sup> indicates that adaptation to the inflectional morphology of Spanish happens exceedingly rarely (except in the case of verbs, which are almost always adapted to Spanish). On the other hand, nouns appear, indeed, to demonstrate a high degree of syntactic integration into Spanish via the use of Spanish language determiners and modifiers. An examination of morphosyntactic properties of borrowings in New York, including a sociovariationist analysis, could test whether the morphosyntactic treatment of borrowings are specific to the contact situation (e.g. language shift or maintenance), particular groups of speakers, the languages involved or the formal features of other-language strings (e.g. how long the string is or its part of speech).<sup>22</sup>

---

21 See Chapter 2 (section 2.6.4) for a brief discussion of this with respect to borrowings in the OZC.

22 Several studies have, directly or indirectly, examined these questions. Findings suggest that, apart from a speaker's bilingual proficiency, adaptation to recipient language patterns is conditioned more strongly by formal linguistic characteristics of the other-language string than by speaker traits. For instance, Pfaff (1979) found that in the speech of Mexican Spanish-English bilinguals, single- and multiple-word borrowings from (as well as codeswitches to) English do not vary by speaker group. Rather, different patterns of incorporation or non-incorporation were conditioned by the length and/or part of speech of the other-language segments. Poplack and Dion (2012) found that single words are always integrated. In a similar vein, Berk-Seligson's (1986) findings for Spanish and Hebrew indicated that the morphosyntax of other-language strings may be dictated by the degree of typological (dis)similarity between the two languages in question. Nonetheless, it may be that, in these studies, the nature of the data (including only single-word borrowings or including established loanwords), the purpose of the investigation (to find or evaluate general constraints on language mixture as opposed to differences between groups) or the size of the corpus (having less than 50 participants) made it difficult to discover social variation in the morphosyntactic treatment of other-language strings.

### **3.5 The dissemination and sharedness of lexical borrowings**

This study revealed that the middle class, the linguistically secure and those who use English most also used novel (i.e. nonshared) vocabulary at higher rates than respective groups. It was hypothesized that these groups, in particular the middle class, were most responsible for the introduction of novel borrowings into Spanish. A cursory analysis of borrowing inventories showed that the inventory of the working class was a subset of that of the middle class. The overall impression was that lexical borrowings are introduced by upper classes and are subsequently taken up by lower classes. This impression needs further confirmation. A full analysis of sharedness, perhaps along the lines of Poplack et al. (1988), would further illuminate how borrowings are disseminated throughout the community.

## Appendix A: Independent variables and sample stratification

Information on participants was gathered using an oral questionnaire following their interview. The tables below show how many first generation and second generation informants were in each category for the independent variables of this investigation. First generation informants are those that arrived in the U.S. from age 13 or later. Second generation informants are those that arrived at or before age 12. For expositional purposes, all independent variables are presented as if they were categorical variables. Variables like age, arrival age and years in the U.S. were analyzed as both continuous and categorical variables. Numbers in each table add up to 146, the total number of informants in this study, except for those in the Years in the U.S. table. The total in that table is 98, reflecting the number of individuals that arrived at or after age 13.

**Ethnonational affiliation.** This variable reflects an informant's place of birth if he was born outside of the U.S. or, if U.S.-born, the informant's self-identification with one of six Spanish-speaking nations.

	First generation		Second generation	
	n	%	n	%
Colombia	14	14	9	19
Dominican Republic	15	15	9	19
Ecuador	17	17	7	14
Mexico	17	17	8	17
Puerto Rico	17	17	8	17
Cuba	18	18	7	14
Total	98	100	48	100

**Region.** This variable classifies informants based on whether their ethnonational affiliation is tied to a Caribbean or Latin American mainland nation.

	First generation		Second generation	
	n	%	n	%
Caribbean	50	51	24	50
Latin American mainland	48	49	24	50
Total	98	100	48	100

**Sex.** An informant's biological sex.

	First generation		Second generation	
	n	%	n	%
Male	48	49	22	46
Female	50	51	26	54
Total	98	100	48	100

**Age** is the age of the informant at the time of the interview.

	First generation		Second generation	
	n	%	n	%
Age 13-19	5	5	11	23
Age 20-39	53	54	32	67
Age 40-59	31	32	4	8
Age 60 and above	9	9	1	2
Total	98	100	48	100

**Age of arrival.** The age at which the informant came to the United States.

	First generation		Second generation	
	n	%	n	%
U.S.-born			29	60
U.S.-born or arrive $\leq$ age 3		n/a		
Child arrivals			19	40
arrived from age 4 to 12				
Teenager	26	27		
arrived from age 13 to 19				n/a
Older arrivals	72	73		
arrived $\geq$ age 20				
Total	98	100	48	100

**Years in the U.S.** The number of years a migrant had spent in the United States at the time of the interview. This variable applies only to first generation informants (arrival age  $\geq 13$ , n=98). An equivalent variable for second generation informants is age.

	<b>First generation</b>		<b>Second generation</b>	
	<b>n</b>	<b>%</b>	<b>n</b>	<b>%</b>
Recent 0-3 years in U.S.	21	21		
Long 4 to 15 yrs in U.S.	48	49		
Longer 16-25 years in the U.S.	21	21		n/a
Longest 26+ years in the U.S.	9	9		
<b>Total</b>	<b>98</b>	<b>100</b>		

**Occupational class.** Informants self-reported their occupational class on a 3-category scale including ‘high class’, ‘middle class’ and ‘working class’. Only one informant identified as ‘high class’ and has been grouped with the category ‘middle class’.

	<b>First generation</b>		<b>Second generation</b>	
	<b>n</b>	<b>%</b>	<b>n</b>	<b>%</b>
Middle class	48	49	27	56
Working class	47	48	19	40
<i>missing</i>	3	3	2	4
<b>Total</b>	<b>98</b>	<b>100</b>	<b>48</b>	<b>100</b>

**Socioeconomic status (SES).** This variable classifies informants on the criteria of occupation and income. Analysis revealed that effects of this variable on the dependent measures of this study was generally less robust than those for the similar variable occupational class. I include this variable in Appendix A for the sake of completeness.

	<b>First generation</b>		<b>Second generation</b>	
	<b>n</b>	<b>%</b>	<b>n</b>	<b>%</b>
Level A (the lowest rating)	16	16	0	0
Level B	27	28	17	35
Level C	35	36	23	48
Level D	16	16	8	17
<i>missing</i>	4	4		
	<b>98</b>	<b>100</b>	<b>48</b>	<b>100</b>

**Level of education.** This refers to the highest level of formal schooling an informant attended, though not necessarily completed.

	<b>First generation</b>		<b>Second generation</b>	
	<b>n</b>	<b>%</b>	<b>n</b>	<b>%</b>
Elementary school	23	23	0	0
High school	24	25	16	33
College	35	36	25	52
Graduate	15	15	7	15
<i>missing</i>	1	1		
<b>Total</b>	<b>98</b>	<b>100</b>	<b>48</b>	<b>100</b>

**English skills.** Informants quantitatively assessed their ability in English on a four-point scale. The scale is often condensed into two factors to reflect the most robust findings of statistical analysis. In the first generation, English skills is interpreted as English proficiency. In the second generation, it is interpreted as English confidence (see Chapter 5, section 2.4.2 for a full explanation.)

	<b>First generation</b>		<b>Second generation</b>	
	<b>n</b>	<b>%</b>	<b>n</b>	<b>%</b>
Pobre ‘poor’	32	33	1	2
Pasable ‘passable’	28	29	4	8
Bueno ‘good’	25	25	9	19
Excelente ‘excellent’	12	12	34	71
<i>missing</i>	1	1		
<b>Total</b>	<b>98</b>	<b>100</b>	<b>48</b>	<b>100</b>

**Spanish skills** is a qualitative self-assessment of an informant’s ability in Spanish using the same four-point descriptive scale. The scale is often condensed into two factors to reflect the most robust findings of statistical analysis. In the second generation, Spanish skills is interpreted as Spanish proficiency. In the first generation, it is interpreted as Spanish confidence

	<b>First generation</b>		<b>Second generation</b>	
	<b>n</b>	<b>%</b>	<b>n</b>	<b>%</b>
Pobre ‘poor’	0	0	2	4
Pasable ‘passable’	8	8	13	27
Bueno ‘good’	34	35	25	52
Excelente ‘excellent’	55	56	8	17
<i>missing</i>	1	1		
	<b>98</b>	<b>100</b>	<b>48</b>	<b>100</b>

**Spanish daily use.** This is a composite variable created from informant responses to questions about with whom they speak Spanish (e.g. parents, siblings, friends, co-workers, etc.), for what activities (i.e. reading, talking, TV) and how often. The four-factor variable is typically condensed into two factors to reflect the most robust findings of statistical analysis.

	<b>First generation</b>		<b>Second generation</b>	
	<b>n</b>	<b>%</b>	<b>n</b>	<b>%</b>
None	1	1	12	25
Low	20	20	23	48
Mid	34	35	11	23
High	42	43	2	4
<i>missing</i>	1	1		
<b>Total</b>	<b>98</b>	<b>100</b>	<b>48</b>	<b>100</b>

## Appendix B: Precontact test tool (sample)

A total of six precontact questionnaires were used, one for each ethnonational group in the OZC. All tools had between 65 and 70 items, with the exception of Puerto Rico's, which had 98. The sample provided here is for Mexico and contains a subset of the items in the original.

### MÉXICO

Lea los ejemplos siguientes. Decida si la palabra destacada se emplearía en el contexto dado si usted estuviese conversando con otro Mexicano.

	Ejemplo	Se utiliza la palabra en México hoy día? (S/N)	Ud. utilizaría la palabra en conversación con un mexicano en México? (S/N)	Qué otras palabras utilizaría ud. en vez de la palabra destacada? (Cual es la mas común/normal de ellas?)
1	Bueno ahorita <b>actualmente</b> aunque me despidieron de mi trabajo principal, tengo dos trabajos en unos restaurantes.	sí	sí	en este instante
2	Él tenía un <b>apartamento</b> en el campo donde es menos caro vivir.	no	no	departamento
3	Ahora el programa del bachillerato en arquitectura tiene muchas más <b>aplicaciones</b> que hace 15 años.	no	no	registros
4	Ayer mi hija <b>aplicó</b> para un trabajo en un banco.	no	no	Ayer mi hija hizo una solicitud para un empleo.
5	Ella era la niña popular. Ella era la que me decía, "No, pues, nos vamos a tal <b>bar</b> para tomar una cervecita."	sí	sí	antro
6	En ese tiempo jugaba <b>basket</b> para mi colegio, y eso me ayudó bastante a terminar mis estudios.	sí	sí	

	Ejemplo	Se utiliza la palabra en México hoy día? (S/N)	Ud. utilizaría la palabra en conversación con un mexicano en México? (S/N)	Qué otras palabras utilizaría ud. en vez de la palabra destacada? (Cual es el mas común/normal de ellas?)
7	El <b>campus</b> de la universidad es muy bonito: edificios nuevos y muchos arboles.	sí	sí	campo
8	Yo estudio en el <b>campo</b> de la Educación.			
9	No guardo ropa en mi <b>closet</b> ; lo utilizo como almacén de papelería.	sí	sí	armario, guardarropa
10	Se graduó con beca y el pudo ir a un <b>colegio</b> cerca de Canadá.	sí	sí	universidad
11	Mi mama hizo nada más la <b>elemental</b> y ya no pudo ir más allá en sus estudios.	no	no	primaria
12	Pues nos tratamos de hablar seguido y de mandarnos <b>email</b> cuando es posible.	sí	sí	correo
13	Antes de irse de Puebla, mi mujer trabajaba en una <b>factoría</b> de ropa.	no	no	fabrica
14	La noche antes de mi primer <b>final</b> , me quedé estudiando toda la noche.	no	no	último examen
15	La <b>ganquita</b> de nosotros era muy mezclada; mi amigos eran de todas partes del mundo.	no	no	banda

	Ejemplo	Se utiliza la palabra en México hoy día? (S/N)	Ud. utilizaría la palabra en conversación con un mexicano en México? (S/N)	Qué otras palabras utilizaría ud. en vez de la palabra destacada? (Cual es el mas común/normal de ellas?)
16	Me estoy enamorando de la comida griega como los famosos tacos de cordero o los <b>gyros</b> .	no	no	tacos arabes
17	Tengo mucha experiencia porque he trabajado en restaurantes de mesera o <b>hostess</b> .	sí	sí	anfritriona
18	También dejé la clase de matemáticas y el profesor en esa época me dió un <b>incompleto</b> .	no	no	
19	Los medios de comunicación de antes no eran tan eficientes, como el <b>internet</b> y los teléfonos celulares de hoy.	sí	sí	
20	Si trabajaras en una <b>marketa</b> , pues tienes que ponerte a estimar cajas, contar la mercancía y todo eso.	no	no	mercado
21	No durmió bien porque el <b>mátres</b> fue muy blandito.	no	no	colchón
22	Se decidió mover el <b>meeting</b> de hoy hasta lunes, para que pudiéramos ver el partido por television esta tarde.	sí	sí	reunión
23	La escuela de ellos tiene una <b>populación</b> muy grande de Hispanos.	no	no	número
24	Nosotros nos criamos en un <b>proyecto</b> en el medio de la ciudad.	sí	sí	

	Ejemplo	Se utiliza la palabra en México hoy día? (S/N)	Ud. utilizaría la palabra en conversación con un mexicano en México? (S/N)	Qué otras palabras utilizaría ud. en vez de la palabra destacada? (Cual es el mas común/normal de ellas?)
25	La <b>realización</b> que mi primer idioma fue español fue lo que más me aguantó el español.	no	no	
26	Cuando me <b>retire</b> , quiero irme a vivir con mis hijos.	sí	sí	
27	El <b>salario</b> que nos dan a nosotros, no es bastante para el trabajo que hacemos nosotros.	sí	sí	
28	Me gusta mucho el <b>sashimi</b> , pero no todos porque hay pescado que no me cae bien así crudo.	no	no	
29	No soy fanático de los <b>show</b> en la televisión o en el radio.	sí	sí	
30	Me gusta mucho <b>sushi</b> , pero no todos porque hay pesado que no me cae bien así crudo.	sí	sí	
31	Tomé ese trabajo, y ahora ya tengo <b>beneficios</b> y todo como me pagan la <b>tuición</b> y todo eso.	no	no	los estudios, los pagos, inscripciones
32	Íbamos como ocho personas en una <b>van</b> con mis tíos al río cada fin de semana.	sí sí	sí	
33	El <b>waltz</b> ya no se baila en el grupo pero yo lo bailo cuando salgo con mi marido.	waltz	sí	

Fecha de hoy: \_\_\_\_\_

M / F

Sexo: \_\_\_\_\_

Edad: \_\_\_\_\_

Ciudad de nacimiento: \_\_\_\_\_

Año que llegó a los EE.UU.: \_\_\_\_\_ Mexico chiapas

Años en los EE.UU.: \_\_\_\_\_

Edad que llegó a los EE.UU.: \_\_\_\_\_

Aproximadamente cuantas veces al año visita su país? \_\_\_\_\_

Cuantas veces ha visitado su país desde su llegada en los EE.UU.? \_\_\_\_\_

## Appendix C: Results of the homonymy test

This appendix details the results of the homonymy test for data selection, described in Chapter 3, section 4.4.2.

### 1. Data included in the study: Interlingual coincidences determined to be Spanicized English lexical borrowings

The meaning of following words were determined to both (i) not be related to the meaning of cognate forms in the informant's Spanish or in standard Spanish (as found in the *Diccionario de la RAE* (22<sup>a</sup> ed.) and (ii) be related to the meaning of the cognate forms in English. The 22 word types listed below account for 78 lexical borrowing tokens (approximately 1.7 percent of lexical borrowings in the LBD). The letters in parentheses following each word indicate the ethnonational affiliation of the informants that used it.

*alarma* (C) 'alarm clock' (2x)

*aquarium* [a cua 'ri um] (P) 'public space for viewing aquatic life' (2x)

*atender* (E) 'to attend or participate in' (2x)

*bil(l)* (D, P) 'a notice of monies due for a service provided' (12x)

*bufetes* (P) 'an assortment of food arranged for public consumption' (1x)

*embarasoso* (P) 'causing shame or embarrassment' (2x)

*final* (M) 'a final exam' (1x)

*fraternidad* (D) 'a men-only social club' (1x); also appears as *fraternity* (M)

*frisar, frisado* (D, E) 'to freeze, frozen' (4x)

*heavy, hevi, hebi* 'a large quantity' (3x); also appears in *heavy metal* (U)

*incompleto* (M) 'an academic course whose requirements remain unfulfilled' (2x)

*janguear* (D, P) 'to hang out, to socialize in someone's company' (23x)

*mapear* (C, P) 'to mop' (6x)

*marketa* (M) 'small store with foodstuff and other staples' (1x); also appears as *market* (E, M) (2x)

*ponchar* (D) 'to perforate' (1x)

*realización* (C, M) 'awareness' (2x)

*remisión* (M) 'inactivity of a chronic disease' (1x)

*performatividad* (P) 'performance' (1x)

*revolverse* (D), 'to have to do with, to be about' (1x)

*rudo* (E) 'rude, impolite' (1x); also appears as *rude* (E) (1x)

*supería* (U) 'a place where soup is made and sold' (2x)

*yarda* (P) 'a verdent space adjacent to living quarters' (7x)

## 2. Data excluded from the study: Interlingual coincidences determined to be semantic extensions

The following are a selection of words that were not reported to be used in informant's preimmigration Spanish in the way used in the OZC interviews. They were nonetheless excluded from the LBD because their meanings as used in the OZC corpus were determined to be related to a standard Spanish meaning of the word (as determined by the *Diccionario de la RAE*).

*asumir* (M) 'to conclude without sufficient evidence'

*bloque* (D, E, P) 'the land/distance between two city streets'

*colegio* (C, D, E, M, P, U) 'university'

*club* (P) 'a social organization for like-minded persons to accomplish a goal, a school club' (*Christamas Club* (U), *club music* (D), *club* (English pronunciation) (D, E, M) 'nightclub', *nightclubs* (C, M, U))

*cooperativa* (U) 'association established to provide services on a non-profit basis to its shareholders or members who own and control it'

*crucero* (E) 'trip by boat'

*dirrecciones* (E) 'instructions for arriving to a location'

*estudio* (C, D, M) 'an apartment consisting of one large room'

*factoría* (C, D, E, M) 'factory'

*grado* (U) 'evaluation of scholastic knowledge/achievement'

*hidrante* (D) 'a water source on the street level used to put out fires'

*introducción, introducidos* (D) 'to present or make individuals acquainted with each other'

*populación* (M) 'group of people that inhabit a place'

*principal(a)* (M) 'head of a school, principal'

*profanidad(es)* (D) 'profanities, vulgar use of language'

*registración* (E, U) 'to register'

*retornar* (D, U) 'to return'

*súper/súpero* (E, M, P, U) 'person in charge of the care of something'

*suspender* (D) 'to prohibit from attending school'

## Appendix D: Examples of lexical borrowing by part of speech category

This appendix illustrates the English-origin material in the OZC considered to be lexical borrowings in this study. These examples provide insight into the composition of the database and the speech of the informants who were its source. This appendix is organized by part of speech (nouns, adjectives, verbs, etc.). Unless otherwise noted, italicized words and phrases in excerpts below each count as a one lexical borrowing (see Chapter 3, section 4 for how borrowing rates for informants were calculated).

### 1. Nouns and noun phrases

Nouns, noun phrases and partial noun phrases with the following characteristics are included in the database.

#### 1.1 Common nouns

Single-word common nouns unadapted to Spanish morphology or phonology (italicized in examples (1-3)) are included in the database.

- (1) Y todos tenían el ya-.. como se dice el *yarmulke*... 300E  
'And everyone had the ya-.. how do you say *yarmulke*...'
- (2) Ya está.. ya está el *agreement*, como se dice... 038C  
'That's it.. that's the *agreement*, as they say...'
- (3) Y el gran número de puertorriqueños, que son gente de *high school* o más jóvenes, y ellos.. pues no sé.. yo no sé.. yo no he investigado eso... 220P  
'And the great majority of Puerto Ricans, who are people from *high school* or younger, and they.. well I don't know.. I don't know.. I haven't investigated that...'

Common nouns morphologically adapted to Spanish (such as *tuición* and *yarda* in examples (4)-(5)) are included.

- (4) Y ahora ya tengo beneficios y todo como me pagan la *tuición* y todo eso... 350M  
'And now I have benefits y everything seeing as they pay my *tuition* and all that...'
- (5) Él ponía aquella *yarda* allí, porque a él le gusta sembrar... 153P  
'He put that *yard* there, because he likes to plant...'

Partial NPs consisting of adjectives and nouns (examples 6-7) are included.

- (6) Yo estoy haciendo un *double major*... 340M  
'I'm doing a *double major*...'

- (7) Porque uno aquí la mayoría de veces uno va (es) al cine o a discotecas de *teenage nightclubs*... 180C  
'Because one here most of the time one goes (it's) to the movies or to *teenage nightclub(s)* discotheques ...'

Partial NPs consisting of a noun and a possessive noun (8) are included.

- (8) Que se llama el *principal's list*... 201U  
'That is called the *principal's list*...

NPs consisting of a noun modified by a possessive personal pronoun (*my family* in (9)) are included.

- (9) Lo que más me hace falta de allá es *my family*... 333D  
'What I most miss from there is *my family*...'

NPs adjacent to one another and not conjoined by English language material. *High school* and *university* (example 10) are considered separate instances of borrowing. Each counts as one borrowing token.

- (10) Yo toda la escuela ha sido aquí [en los EE.UU.], desde escuela elemental hasta *high school, university*. 220P  
'[For] me every school has been here [in the U.S.], from elementary to *high school, university*.'

NPs in apposition are included. *Streaks* and *blue streaks* (example 11) each count as an instance of lexical borrowing.

- (11) Yo me vestía con los, tenía mis *streaks, blue streaks*, y me pintaba el pelo y todo eso... 350M  
'I would wear my, I had my *streaks, blue streaks*, and I would dye my hair and all that...'

NPs conjoined by a Spanish conjunction are included. *Audio* and *video* (example 12) are counted as two separate instances of lexical borrowing.

- (12) Entonces eso es con *audio y video* ['vi di o] y sonido de sistema. 311C  
'So that is with *audio and video* and sound system.'

NPs conjoined by an English conjunction are included. *Writing and reading* (example 13) is one lexical borrowing. *Cap and gown* (example 14) is another.

- (13) Las áreas principales eran matemáticas, inglés, *writing and reading*. 300E  
'The principle areas were mathematics, English, *writing and reading*.'

- (14) Pero los de *preschool* tienen una ceremonia así formal con su *cap and gown*... 333D  
 ‘But the *preschoolers* have a ceremony that formal way with their *cap and gown*...’

Adjectives as nouns. Adjectives used in place of a whole noun phrase are counted as nouns. For example, *non-profit* in (15) is counted as a borrowing and considered a noun.

- (15) Mucha gente en los *non-profit* trabajaba así... 435P  
 ‘A lot of people in the *non-profit[s]* were working like that...’

*Non-profit* is a shortening of the phrase “non-profit organization”, where “non-profit” would function as a modifier of “organization”. However, because *non-profit* is used as a noun in (15), it is coded as a noun in the LBD.

## 1.2 Pronouns

The following types of pronouns are included in the database.

Subject, object or possessive pronouns occupying principal sentential syntactic roles (i.e. as objects or subjects) are included. *You* (example 16), the object of *dice*, is a borrowing.

- (16) O le dice *you*, que es “tú”, en el Ecuador... 364E  
 ‘Or you say *you*, which is “you<sub>[informal]</sub>”, in Ecuador...’

Pronouns modified by English adjectives are included. *Main one* (example 17) is one lexical borrowing.

- (17) Pero tú estás en el *main one* [campus], ¿no? 340M  
 ‘But you are in the *main one* [campus], right?’

Possessive pronouns modifying Spanish or English nouns are considered determiners (see section 6 of this appendix).

## 1.3 Proper names and proper nouns

Proper nouns in the database include the following types of strings.

Single-word proper names are included when they meet the criteria specified in Chapter 3 (section 4.4.3). Examples (18-24) contain single-word proper nouns included in the database.

- (18) Una vez, bueno, eso fue en un.. tiempo de *Christmas*, de.. así, de navidad, que hicieron un.. hicieron una rifa... 025C  
 ‘Once, well, that was in a.. time of *Christmas*, of.. yeah, of Christmas, that they did a.. they did a raffle...’

- (19) Y, y al otro día, nosotros llegamos en una víspera de.. de (a) *Easter*.. en la misma semana de (b) *Easter*... 118D  
 ‘And, and on the next day, we arrived during the eve of.. of (a) *Easter*.. during the same week as (b) *Easter*...’
- (20) Me dice mi mamá ¿Y a dónde tú vas? Y yo le dije “Ay, mami, a la fiesta de *Halloween*.” Dice ¿Tú me pedistes permiso”... 183U  
 ‘My mom said to me “And where are you going?” And I told her “Ay, mom, to the *Halloween* party. She said “Did you ask permission?” ’
- (21) Ella se graduó en *February* del próximo año porque tenía que hacer las clases que no pasó... 365E  
 ‘She graducate in *February* of the following year because [she] had to do the classes she didn’t pass...’
- (22) Aunque a veces también como el *Spanglish* se empieza a meter en tu idioma... 340M  
 ‘Although sometimes too like *Spanglish* begins to appear in your language...’
- (23) Por ejemplo este año el grupo, hay muchos estudiantes que no.. que no estudiaron en *NYU* [en wai ju]... 092P  
 ‘For example this year the group, there are many students that don.. that didn’t study in *NYU*...’
- (24) Uno tenía que ñagotarse de espalda y bajar como *Spiderman*... 432P  
 ‘One had crouch his back and go down like *Spiderman*...’

Multiple-word proper nouns are included. Excerpts (25-29) contain multiple-word proper nouns in the LBD.

- (25) ¡Bueno viví en *New Jersey*! Pero me gustaba mucho el ambiente... 351M  
 ‘Well, I lived in *New Jersey*! But I liked how it was very much...’
- (26) No, él vive en *Long Island*... 325E  
 ‘No, he lives on *Long Island*...’
- (27) Solamente [había] un autobús venía directo, aquí a *Port Authority*... 346M  
 ‘[There was] only one bus came direct, here to *Port Authority*...’
- (28) Em.. otra persona eh de, de cuando de eh de chiquito, *President Kennedy* también, cuando lo mataron, eh eso me afectó mucho... 417P  
 ‘Uhm.. another person uh from, from when, when uh I was little, *President Kennedy* too, when they killed him, uh that affected me a lot...’

- (29) El profesor era.. era *Mister K---*, todavía me acuerdo, era tan difícil el profesor... 300E  
'The professor was.. was *Mister K---*, I still remember, he was so difficult that professor...'

Two adjacent proper names referring to two distinct denotata. If two proper nouns appear adjacent to each other and each refers to a distinct *denotata*,<sup>1</sup> each proper noun is considered a lexical borrowing as long as it qualifies for inclusion according to criteria already outlined. For example, *Central Americas* (30a) is one borrowing and *South America* (30b) is another.

- (30) Pero hay muchísima de Latinoamérica en sí, (a) *Central Americas*, (b) *South America*... 333D  
'But there are a lot from Latin America itself, (a) *Central America*, (b) *South America*...'

Two adjacent proper nouns referring to one denotatum. If two proper nouns appear adjacent to each other to refer to a single location, each proper noun would have counted as a borrowing as long as each constituent part would have qualified for inclusion according to criteria already presented. For instance, if an informant were to have uttered *Long Island, New York* in an otherwise Spanish sentence, *Long Island* would be one borrowing and *New York* would be another. However, since all instances of adjacent proper nouns referring to one place are composed of quasi-untranslatable single words (e.g. *Jacksonville, Florida*.... 422P), there are no examples of this in the database.

Proper adjectives in isolation. Proper adjectives that do not directly modify English nouns are included in the database and are given the status of nouns. *Puerto Rican-American* in (31) is a borrowing in the LBD and classified as a noun.

- (31) Hay muchos que se sienten americanos, o *Puerto Rican-Americans*, tú sabes... 220P  
'There are many that feel American, or *Puerto Rican Americans*, you know...'

The compound phrase *Puerto Rican-American* might, in a traditional grammar, be classified as an adjective. That it is in fact a noun is confirmed, in the case of (31), by the presence of English noun morphology: the plural *-s*.

---

1 *Denotatum* is "the class of objects, properties, etc., to which to which an expression correctly applies" (Lyons 1977: 207). For example, the denotatum of *blue* is the class of all blue things. The denotatum of *dog* is a particular class of animals. Denotation, in other words, it is "the relationship between a linguistic entity and something outside the language-system" (Lyons 1977: 210).

Other examples of proper nouns. Proper nouns in this corpus also include dates (example (32a), street names (even when the names are numbers (example 32b), radio stations (33), teams (34), colleges/universities (35-36), titles of academic degrees (37), departments (38), courses of study and course titles/topics (39) and movie titles (40). Each italicized string in the following examples counts as a lexical borrowing.

- (32) Eso fue (a) *August nineteenth* y fuimos a tu casa ahí en [la calle] (b) *One forty-third*, y hasta los proyectos... 333D  
 ‘That was (a) *August nineteenth* and we went to your house there on (b) *One forty-third* [Street], and all the way to the projects...’
- (33) No sé porque pero antes [escuchaba] nada más puro *92.3 K rock*. 350M  
 ‘I don’t know why but before [I listened to] nothing but pure *ninety-two point three K rock*.’
- (34) He ido, conozco San Louis porque he ido a ver la serie mundial, fui hace como seis años atrás, una vez o cinco, una vez que *Saint Louis* lo ganó... 229D  
 ‘I’ve gone, I know Saint Louis because I’ve been to see the World Series, I went six years ago, one time or five, as soon as *Saint Louis* won it...’
- (35) Porque ahí teníamos intercambio con una universidad que se llama *TECU University*... 354M  
 ‘Because there we had an exchange with a university that’s called *TECU University*...’
- (36) Hice toda la maestría como estudiante y como profesor allí mismo en *City College*... 344C  
 ‘I did the entire master’s [degree] as student and as professor right there in *City College*...’
- (37) Allá recibí el *B.A.* [bi aj], o sea, en otros países le llaman la licenciatura... 373P  
 ‘There I received my *B.A.*, that is, in other countries they call it the licenciatura...’
- (38) Y estudio, como ya tú bien sabes aquí en el.. en el depar.. en el departamento.. de.. *Luzo Brazilian Literatures* y llevo.. este es mi segundo año aquí... 336D  
 ‘And I study, as you already well know here in the.. in the depar.. in the department.. of.. *Luzo Brazilian Literatures* and I have.. this is my second year here...’
- (39) Pero *Media Studies* creo que es donde está más la carrera, ¿no? 340M  
 ‘But *Media Studies* I think that is more where the degree is, don’t you think?’
- (40) El primero es *Sorceror’s Stone*, segundo es *Chamber of Secrets*... 343C  
 ‘The first is *Sorceror’s Stone*, second is *Chamber of Secrets*...’

## 1.4 Exclusions

The following categories of noun **are excluded** from the database.

Nouns or full NPs with subordinated English prepositional phrases are excluded. *Sense of a family* (example 41) is considered a codeswitch<sup>2</sup> and excluded from the study.

- (41) Entonces realmente cuando llegué a sentir un, sentí.. como le llaman en inglés *sense of a family* fue cuando vine aquí.... 300E  
'So really when I got to feel a, I felt.. as they call it in English *sense of a family* was when I came here...'

Nouns or NPs modified by English language relative clauses are excluded. *The one who's doing it* (example 42) and similar structures are **not** included in the database.

- (42) Por ejemplo la Gwyneth- *I think it[ 's]* Gwyneth Paltrow, *the one who's doing it...* 340M  
'For example Gwyneth- *I think it[ 's]* Gwyneth Paltrow, *the one who's doing it...*'

Names of individual people to refer to the person are excluded. *Diane Keaton* (underlined in (43)) is **not** included in the database.

- (43) Y además a mí me fascina Diane Keaton... 263C  
'And moreover Diane Keaton fascinates me...'

Names of people to refer to institutions or places are excluded. *Walt Disney* (44) is **not** in the database.

- (44) En Walt Disney fue una impresión tan linda... 300E  
'At Walt Disney it was such a beautiful experience...'

---

2 An examination of the distribution of strings comprised of an English PP subordinated to an English noun revealed that these, like other English PPs (see Chapter 2, section 2.6.5), tended to be used by individuals who also codeswitched. That is, 11 of 14 informants that subordinated English PPs to English nouns also changed the language of discourse at least once during their interview.

Single-word proper place names are excluded. Generally, place names like *Bronx* (45) and *Maine* (46) are excluded on the basis that it was not at this time feasible to determine if such proper names represented instances of the reproduction of English lexical material. As Clyne (2003: 165) notes, such words may properly belong to a bi-/multi-lingual lexicon.<sup>3</sup>

- (45) La primera vez yo iba a trabajar y yo vivía en el Bronx. 005U  
'The first time I was going to work and I was living in the Bronx.'
- (46) No, ella vive en Maine con el papá y la abuela... 086P  
'No, she lives in Maine with her father and grandmother...'

Proper names referring to Anglo cultural phenomena or places that are translated into Spanish are excluded. *Nueva York* (47) is excluded from the database.

- (47) Yo nací aquí en los Estados Unidos en.. en Nueva York... 206U  
'I was born here in the United States in.. in New York...'

## 2. Adjectives

Adjectives are words that modify nouns or noun phrases. Adjectives in the lexical borrowing database are of the following types.

### 2.1 Adjectives in isolation

When an English adjective occurs by itself it counts as one lexical borrowing. *Exciting* in example (48) is one such example.

- (48) Es bien *exciting*... 330D  
'It's very *exciting*...'

Adjectives as nouns. Adjectives used in place of a whole noun phrase are counted as nouns. See section 1.1 in this appendix for an example. The same applies to proper adjective. Proper adjectives that do not directly modify English nouns are included in the database but are considered proper nouns (see section 1.3).

---

3 One way to approximate whether the use of particular single-word proper place names involved importation of English lexical patterns would have been to use phonological criteria (such as presented in Chapter 3, section 4.3.2). Utilizing such a procedure was, unfortunately, outside the time limits for this study, as over 1400 instances of single-word proper place names were uttered by informants. All single-word proper names were excluded from the database.

## 2.2 Adjectives modified by other English words

If an adjective is directly modified, the modifying word plus adjective string is included and the entire string is counted as an adjective phrase. *Above average* in example (49) is one lexical borrowing.

- (49) Entonces comparado a la escuela, yo estaba *above average*.... 324E  
'So compared to the school, I was *above average*...'

## 2.3 Adjectives modifying English nouns

When an adjective modifies a common noun, it is considered part of the same lexical borrowing as the noun. For example, *main* plus *one* is one borrowing (50); *advanced* plus *level* is one borrowing (51); *Amish* plus *people* is one borrowing (52).

- (50) Pero tú estás en el *main one* [campus], ¿no?... 340M  
'But you are in the *main one* [campus], right?...'
- (51) Y yo me gradué en *advanced level* de inglés... 325E  
'And I graduated in the *advanced level* of English...'
- (52) ¿Como los *Amish people*, como se llama? 340M  
'Like the *Amish people*, what's it called?'

## 3. Adverbs

Words that modify adjectives (*above* in example (49)), verbs (53-55), clause constituents (56), or entire clauses (57) are considered adverbs. Adverbs with the following characteristics are in the database.

### 3.1 Adverbs in isolation

When an English adverb does not modify adjacent English language material is it counted as a lexical borrowing in its own right (53-57).

- (53) Sí, estoy trabajando *full-time*... 375U  
'Yes, I'm working *full-time*...'
- (54) Y sabiendo lo que cuesta una llamada por teléfono, tener tu *quarter ready*... 330D  
'And knowing what a telephone call costs, having your *quarter ready*...'
- (55) Y se lo hacen pero bien *neat*, que sale bien de verdad... 037D  
'And they do it [not just neat] but very *neat*, [so] that it really turns out well...'

- (56) Yo aprendí tantas cosas bonitas de la comunidad puertorriqueña, *particularly* en humildad... 370M  
 ‘I learned so many beautiful things about the Puerto Rican community *particularly* in humility...’
- (57) Y esto [lavar platos] estoy aprendiendo con él. *Maybe* me pasé a la cocina a cocinar. 201U  
 ‘And this [washing plates] I am learning with him. *Maybe* I’ll get passed to the kitchen to cook.’

### 3.2 Adverbs modifying English material

When an adverb modifies an English-language constituent, it is considered to be part of the same borrowing as the English string it modifies. *Above* in example (58) is not counted as a lexical borrowing in its own right, but as part of the adjective phrase borrowing *above average*.

- (58) Entonces comparado a la escuela, yo estaba *above average*.... 324E  
 ‘So compared to the school, I was *above average*...’

## 4. Verbs

Verbs that appear without other English-language arguments of its same clause are included in the database.

### 4.1 Verbs in isolation

Verbs are in the database when it does not appear adjacent to English-language material from its own clause. They may be morphologically adapted to Spanish or not. For example, morphologically adapted *lonchar* (59) and *mapear* (60) and morphologically unadapted *store* (61) each count as one lexical borrowing.

- (59) Me llama María que vaya a *lonchá* allá... 118D  
 ‘María calls me to go there *to [have] lunch* ...’
- (60) Ya le digo que no sabía *mapear*... 102P  
 ‘I already told her that [she] didn’t know how *to mop*...’
- (61) Y a veces ni siquiera tomaba notas sino trataba de visualizar todo y de que se me *store*, que se me grabe todo... 300E  
 ‘And sometimes I wouldn’t even take notes but I tried to visualize everything and so that everything *stores* itself [in my head], that it records itself [in my head]...’

Verbs in the imperative are included (example 62) in the database.

- (62) Es que yo no puedo dibujar en mi casa porque ese niño a toda hora es.. voy a dibujar y ahí está, eh.. *shut up*... 342C  
'It's that I can't draw in my house because that boy around the clock is.. [I] am going to draw and that's it, eh.. *shut up*...'

## 4.2 Verbs with other English material

Complex (two-part) verbs are analyzed as a single unit. For example, *break into* (example 63) and *figure out* (64) are in the database.

- (63) Lo que quiero hacer es en.. tratar de, *break into*, *I don't know*, querer *to*... 012U  
'What I want to do is in.. try to, *break into*, *I don't know*, want *to*...'
- (64) Yo trataba de.. de.. *figure out* cómo se deletreaba la palabra... 333D  
'I was trying to.. to.. *figure out* how the word was spelled...'

## 4.3 Verbs as set expressions

A verb that appears with an English-language material from its own clause is included in the LBD if the entire string is a collocation<sup>4</sup> (65-66), a set phrase<sup>5</sup> (67-68) or performs a non-verb function (examples 69-70).

- (65) Entonces yo le digo a Rafael, "Ciertas cosas yo sí te puedo permitir que, *go ahead*, analízalas, estúdialas, veélas de todas formas que tú puedas..." 359M  
'So I say to Rafael, "Certain things I yeah can allow you to, *go ahead*, analyze them, study them, see them from every angle you can..."'
- (66) Pero yo un pelado así de.. más o menos mi edad no *forget it*, ni de treinta y cinco años porque solamente están pensando es en acostarse... 313E  
'But me a broke [guy] like that with.. more or less my age, no *forget it*, not even [if he is] thirty-five years old because [the] only [thing they] are thinking about is sleeping [with you]...'

---

4 A phrase is considered a collocation if it is found in the *Collins Online Collocation Dictionary*.

5 A phrase is considered a set phrase or idiomatic expression if it is either periodically (see section 4.2) in the corpus or listed in an English dictionary. The dictionary used in this study was *Webster's Unabridged Dictionary* (2001).

- (67) Solamente lo que quería era hacer su par de pesos y su plato de comida y *that's it*, no pensaba en seguir progresando... 198P  
 ‘The only thing [he] wanted was to make his wage and his plate of food and *that's it*, [he] wasn't thinking about continuing to progress...’
- (68) Decía *wait a minute*, algo está mal aquí... 300E  
 ‘[I] was saying *wait a minute*, something is wrong here...’
- (69) Él era el dueño del negocio y ella entró por.. a *you know* a participar... 301E  
 ‘He was the buisness owner and she entered because.. to *you know* to participate...’
- (70) *I mean*, entonces en la tesis lo que hice fue estudiar un área de juegos en, en la esquina noroeste de *Central Park*... 303P  
 ‘*I mean*, so in the thesis what I did was study a sports area in, in the northeast corner of *Central Park*...’

In examples (65) and (66), *go ahead* and *forget it* are collocations<sup>6</sup> and are included in the database. In (67), *that's it* is a set phrase<sup>7</sup> and is also included in the database, as is *wait a minute* in (68), which functions as the object of the verb *decía*. When verb phrases are collocations or set phrases, they are coded as verbs or discourse markers according their role in the sentence. If they communicate their lexical (i.e. not functional) meanings as verbs, they are coded as verb phrases. So, for instance, the set phrases *that's it*, *wait a minute* and *go ahead* are used in these excerpts with their lexical meanings in tact. They are, thus, coded as verb phrases.

On the other hand, if the function of a verb phrase is to express speaker attitudes or direct discourse, they are coded as discourse markers. For example, the collocation *forget it* expresses the speaker's attitude and so it is coded as a discourse marker. The verbs *mean* and *know* in (69) and (70), respectively, do not subcategorize for complement phrases, as they would if they were acting in their function as verbs. So, *I mean* and *you know* are coded as discourse markers in the LBD (see section 10 on tag items).

#### 4.4 Exclusions

A verb that appears in a string with one or more of its English-language arguments (i.e. subject or object NP, PP or adverb) is not included.

- (71) *He was*.. él era maestro constructor... 333D  
 ‘*He was*.. he was master constructor...’

*Was* in (71) is excluded from the database because the string in which it appears realizes other arguments of its clause in English, namely, the subject, *he*. Such strings are considered codeswitches (see Chapter 3, section 4.1).

6 *Go ahead*, *forget it* and *that's it* are in *Collins Online Collocational Dictionary*.

7 *That's it* is used 16 times in the corpus by 12 different speakers.

## 5. Prepositions and prepositional phrases

Prepositions and prepositional phrases in the database are of the following types.

### 5.1 Prepositions without an English-language complement

Prepositions that do not appear with an English-language complement are included in the database. *For* in example (72) is counted as one lexical borrowing.

- (72) Entrevistador: ¿Qué hacía él?  
Entrevistado: Él trabajaba *for* la inmigración... 380M  
'Interviewer: What did he do?  
Interviewee: He used to work *for* immigration...'

### 5.2 Analytic locutions without English objects

Multiple-word prepositional expressions (a.k.a. analytic locutions) that do not realize their object in English, such as *as far as* (73) and *in spite of* in (74), are lexical borrowings.

- (73) Pero.. *you know as far as* el acceso que uno tiene a todo, a todo... 333D  
'But.. *you know as far as* the access that one has to everything, to everything...'
- (74) Y [sin darse cuenta de] que nosotros nos impusimos, no, de que las carreteras se crearon en función de nosotros y nuestras necesidades y.. *in spite of*.. no este, no a beneficio de las necesidades de los animales... 435P  
'And [without realizing] that we imposed ourselves, right, that the highways were created to serve us and our needs and.. *in spite of*.. not uhm, not for the benefit of animals' needs...'

### 5.3 Prepositions in English collocations or idiomatic expressions

Prepositions that occur in English collocations or idiomatic expressions are in the database. *On and off* in example (75), *in love*<sup>8</sup> in (76) and *like that* in (77) are all collocations and are considered lexical borrowings.

- (75) Entrevistador: ¿Y cuánto tiempo viviste en xxx?  
Entrevistada: *On and off* hasta los 46, sí... 370M  
'Interviewer: And how much time did you live in xxx?  
Interviewee: *On and off* until [age] 46, yeah...'

---

8 Besides being listing in the *Collins Online Collocation Dictionary*, *in love* is also given its own subentry in *Webster's Unabridged Dictionary* (2001).

- (76) Y uno siendo necio, joven y ignorante, y se cree que uno está enamorado, *in love...* 432P  
 ‘And being foolish, young and ignorant, y one thinks he’s in love, *in love...*’
- (77) *So puede ser tres meses aquí entonces tres meses you know like that...* 331.1D  
 ‘So it could be three months here then three months *you know like that...*’

#### 5.4 Exclusions

Prepositions and PPs are excluded from the database in the following situations.

When multiple-word PPs are not collocations, they are excluded from the database. *For about two hours* in (78) is considered a codeswitch and is excluded.

- (78) Me quedo ahí *for about two hours...* 333D  
 ‘I stay here *for about two hours...*’

When multiple-word PPs are not collocations and they modify (i.e. are subordinate to) English nouns they are excluded. In (79), the PP of high school is subordinate to the noun *beginning*. The entire phrase *beginning of high school* is considered a codeswitch and is excluded.<sup>9</sup>

- (79) Entrevistador: Y hasta, ¿tú sabes hasta qué curso llegó él?  
 Entrevistado: *Beginning of high school...* 228D  
 ‘Interviewer: And until, Do you know what course he got to?  
 Interviewee: *Beginning of high school...*’

---

9 For why PPs are excluded see Chapter 3, section 4.1.

## 6. Determiners

Determiners, including *the*, *a*, *an*, *some*, *these*, *your*, *my*, *his*, etc., are in the LBD but they are counted as lexical borrowings in their own right only under certain conditions. Determiners of the following types are included in the database.

### 6.1 Determiners uttered in isolation

Determiners uttered in isolation from other English material would have counted as lexical borrowings if they were **not** false starts (see section 6.3 below). There are no examples that meet these criteria in the corpus. In other words, English determiners appear to be employed only when introducing an English language head noun. There is only one exception to this, found in example (80).

- (80) Yo hago el mantenimiento de todos los empleados de *the New York*  
Teléfono... 208C  
'I do the maintenance of all the employees of *the New York Telephone*...'

In (80), *the* introduces the Spanish head noun *teléfono*. Nonetheless, *the* from this example is part of the borrowing *the New York*.

### 6.2 Possessive pronouns

Possessive pronouns modifying Spanish nouns would be included as lexical borrowings. *Her* as used in "Ella vive en *her* casa" would be included, but no examples like this are in the corpus.

Possessive pronouns modifying English nouns are included in the database. For example, *her* in example (81) is included. However, these are counted as part of the noun phrase lexical borrowing *her mother*.

- (81) La razón fue porque lo.. *her mother* tiene unas xxx grandes... 331.1D  
'The reason was because wh-.. *her mother* has some large xxx...'

### 6.3 Exclusions

When a determiner is a false start, it is **not** counted as a lexical borrowing. An English determiner is considered a false start when it is recast in Spanish before a complete noun phrase is uttered. This happens only four times in the corpus, of which examples (82) and (83) are two examples.

- (82) Pues entonces, llamó la semana pasada, *the*.. el señor *priest*... 416P  
'So then, he called last week, *the*.. the *priest* man...'
- (83) Después la evaluación tiene que ser aprobada por *the*.. el *Intervention*  
*Regional Office*. 333D  
'Afterwards the evaluation has to be approved by *the*.. the *Intervention*  
*Regional Office*.'

## 7. Quantifiers: Numeric, generic and ordinal

English language numeric, generic and ordinal quantifiers (e.g. *many, most, three, seventy, fifth, third*) are classified into part of speech classes according to their use in discourse. For example, when they modifying nouns, they are considered adjectives/modifiers. *Six* in example (84) modifies *years*, so is counted as an adjective. *Six years* is one lexical borrowing.

- (84) Le digo six years, entonces ya ella, okay... 323E  
'I tell him six years, so already she, okay...'

When a generalized quantifier directly precedes a determinate or quantified noun phrase it is also considered a modifier in terms of use, albeit an atypically distributed one<sup>10</sup> (example 85).

- (85) Y él guía all fifty states, vendiendo carne... 333D  
'And he drives all fifty states, selling meat...'

When a number (or other quantifier) is used as a proper noun, such as *One forty-third* in (86), it is considered a proper noun (also see section 1.3 in this appendix).

- (86) Y fuimos a tu casa ahí en [la calle] *One forty-third* y hasta los proyectos... 333D  
'And we went to your house there on *One forty-third* [Street] and up to the projects...'

## 8. Connectives: Conjunctions, coordinators and complementizers

Sentential connectives are counted as lexical borrowings in their own right when they appear without other English-language clause-like strings (e.g. *so* in (87)<sup>11</sup> and (88), *because* in (89)).

- (87) Se tiene que tomar de la vaca, *so*, el sabor era muy diferente de.. de aquí. 201U  
'It has to be drunk from the cow, *so*, the flavor was very different from.. from here.'
- (88) Pero yo tengo el mismo tamaño desde que yo tenía 15 años, *so* yo hice así y me viré... 012U  
'But I have been the same size since I was 15 years old, *so* I went like this and I turned...'

<sup>10</sup> This also agrees with Llorach (1999: §158, §160), who asserts that Spanish *muchos* and *pocos* pattern with adjectives because they are able occur with definite NPs. Although we are dealing with English quantifiers and numerals here, these may also appear with definite NPs (*many the girl, few the individuals, all the people*, etc.)

<sup>11</sup> *So* is considered a connective (and not a discourse marker) when 'as a result' can be substituted for it in an utterance and the utterance makes reasonable sense.

- (89) Si (...) vamos a visitar a alguien, nadie sabe que nosotros estábamos peleando *because* a mí no me gusta demostrarle a la gente que estamos bravos... 331.2D  
 ‘If (...) we are going to visit someone, no one knows that we were fighting *because* I don’t like showing people that we’re angry...’

Connectives of non-clausal constituents. Connectives that conjoin or potentially conjoin non-clausal sister constituent phrases are part of the larger lexical borrowing. That is, there is one lexical borrowing in (90): *writing and reading*.

- (90) Las áreas principales eran matemáticas, inglés, *writing and reading*... 300E  
 ‘The principal areas were math, English, *writing and reading*...’
- (91) *You know, preschool and*.. trabajando para desarrollarle un IEP... 333D  
 ‘*You know, preschool and*.. working to develop an IEP...’

Although it is not certain, it is conjectured that in (91) the coordinating conjunction *and* is meant to coordinate *preschool* with another noun. For this reason, *and* is not counted as an independent lexical borrowing, but as part of the lexical borrowing *preschool and*.

If a connective is adjacent to an English string, its status as a lexical borrowing in its own right (or not) depends upon the role of the connector. If it coordinates phrases within the string, the connector is part of the larger lexical borrowing. If it appears to coordinate clauses or clause-like strings, the connector is considered a lexical borrowing in its own right. In (92a), *and* coordinates the NP *a trade school* with a sister NP *high school*. *And* (92a) is part of the lexical borrowing *a trade school and high school*.

- (92) Y.. fui a la escuela, *a trade school* (a) *and high school* (b) *and* uh ahí donde aprendí este trabajar en la la el oficio de de hojalatería y pintura de carro... 428P  
 ‘*And*.. I went to the school, *a trade school* (a) *and high school* (b) *and* uh there [was] where I learned uhm to work in the the the office of of metal work and car painting...’

In (92b), *and* does not appear to serve the function of coordinating the NPs *a trade school* and *high school* with another NP. Instead, it appears to coordinate two independent clauses. Thus, *and* (92b) is counted as a lexical borrowing distinct from that of *a trade school and high school*.

Connectives within collocations or set phrases are part of the collocational or set phrase. *On and off* (93) is one lexical borrowing. *Cap and gown* (94) is one lexical borrowing.

- (93) [In response to a question which asked how much time the informant lived in Hunt’s Point:]  
*On and off* hasta los 46, sí... 370M  
 ‘*On and off* until 46, yes...’

- (94) Pero los de preschool tienen una ceremonia así formal con su *cap and gown*... 333D  
'But the preschool [children] have a formal type of ceremony with their *cap and gown*...'

Connectives within proper nouns are part of the larger lexical borrowing. In (95), *Newmark and Lewis* is one lexical borrowing.

- (95) En *Newmark and Lewis* estaba trabajando... 086P  
'In *Newmark and Lewis* I was working...'

## 9. Indexical demonstratives

Indexical demonstratives (i.e. *this, that, these, those, here, there*) are considered either determiners or pronouns in accord with their use in discourse. For example, if a demonstrative were found to introduce a Spanish or English noun or noun phrase (e.g. *this* in invented sentence: "Entonces, *this man/hombre* me dijo..."), it would be classified as a determiner. No examples of this sort exist in the corpus.

If a demonstrative were found in its pronominal role, it was classified as a pronoun. *Here* in example (96) is counted as a pronoun and is one lexical borrowing

- (96) La mano tiene *like all* tendones *here* dañados... 365E  
'[Her] hand has *like all* tendons *here* injured...'

In the sentence in excerpt (96), *here* is a pronoun with an adverbial function: it modifies the Spanish verb *tiene*. In that it has no syntactic link to any adjacent English material, it is considered a lexical borrowing on its own. (*Like* is considered a discourse marker.<sup>12</sup>) This is the only lexical borrowing in the database (and in the corpus) that consists solely of a demonstrative form.

Demonstrative forms were also found in collocations (e.g. *that's it*), but are considered to be part of the lexical borrowing instantiated by the collocation. A further note: the only demonstratives that are used in any capacity in lexical borrowings are *that* and *here*. That is, *those, this, these* and *there* do not occur as part of any lexical borrowing or in isolation.

---

12 Since substitution with the prepositional meaning 'similar to' is semantically strange here, and since the sentence makes complete sense when *like* is omitted, *like* is considered a discourse marker in example (96). See footnote 14 in this appendix for more details on this.

## 10. Tag items

Discourse markers (*ya know, yeah*), interjections (*wow!, oh my God!*), tag questions (*right?*), politeness expressions (*please, thank you*) and words of negation/affirmation (*yes, no, of course*) are included as lexical borrowings and fall under the category ‘tag items’.<sup>13</sup>

### 10.1 Tag items by function

Tag items may be single-word or multiple-word. Some of the functions accomplished by tag items in this corpus are listed below with examples. Each excerpt includes one lexical borrowing.

Direct address / vocative:

(97) Italianos, Chinos (...) Ya, ya, no están hablando contigo, *man...* 206U  
‘Italians, Chinese (...) Already they aren’t talking to you, *man...*’

(98) Y cómo yo le iba a decir a mi papá *hey* perdí cierta cantidad de créditos. No puedo. Él me mataba... 300E  
‘And how was I going to tell my father *hey* I lost a certain amount of credits. I can’t. He would have killed me...’

Affirmation / Negation:

(99) Entrevistador: [RISA] ¿Tú te das cuenta inmediatamente?  
Entrevistada: *Yeah*, hm. 379C  
‘Interviewer: [LAUGHTER] Do you realize it right away?  
Interviewee: *Yeah*, hm.’

Hearer orientation:

(100) Y pero ya con.. como con.. dibujado, dibujado por mí ya en la cámara hecha, *¿you know?* 010U  
‘And but already with.. like with.. drawing, drawing for myself already in the made camera, *you know?*’

Clarification:

(101) O sea e.. son.. la religión es musulmana pero de los Arabes, *right?* 301E  
‘I mean eh.. [they] are.. the religion is Muslim but for Arabs, *right?*’

---

13 The category of tag item parallels in part Poplack’s (1980) category of emblematic items. The difference is that our category does not include single nouns, whereas her category does.

Politeness:

- (102) Yo les digo *please*. Y ellos van conmigo... 086P  
'I say *please* to them. And they go with me...'

Surprise:

- (103) No, porque *oh my god*, se me vienen los nervios... 340M  
'No, because *oh my god*, my nerves start acting up...'
- (104) Y el muchacho puso una cara de que qué disparate está diciendo esta mujer, y ella y yo *oh yeah!* 303P  
'And the boy made a face of what nonsense is this woman saying, and she and I [were like] *oh yeah!*'

Dismay:

- (105) *Oh my God..* no, no me dejaban coger ninguna otra clase... 322E  
'*Oh my God..* no, [they] wouldn't let me take any other class...'
- (106) Y lo que menos pensaba yo era que.. que [mi tía] lo [el pollo] iba a matar , xxx *oh my God.* 322E  
'And what I least thought was that.. that [my aunt] was going to kill it [the chicken], xxx *oh my God.*'

Ambivalence:

- (107) Había un muchacho con el que yo estaba saliendo y un día voy yo a.. *I don't know..* era bien físico... 434P  
'There was a boy who I was going out with and one day I go to.. *I don't know..* he was very physical...'

Other attitude:

- (108) Yo estoy debiendo como unos siete mil pesos, *man*. Pero eso es en cre.. en crédito.... 311C  
'I owe like some seven thousand dollars, *man*. But that is in cre.. in credit...'
- (109) Yo siempre estoy batallando como que en ponerme bien así, bien astuta y bien.. Pero termino como que, Ah! *Please!* [RISA] 303P  
'I am always fighting to make myself pretty like, pretty astute and pretty.. But I end up like, Ah! *Please!* [LAUGHTER]'

Correction:

- (110) Los más viejos tienen año y ocho meses, *I mean* los más jóvenes... 333D  
'The oldest are one year and eight months old, *I mean* the youngest...'
- (111) Pero a mi mamá no le gustó demasiado porque era demasiado [RISA] *never mind*... 228D  
'But my mother didn't like it too much because he was too [LAUGHTER] *never mind*...'
- (112) ¿Vinistes directamente a trabajar.. *I mean*, a estudiar? 303P  
'Did you come directly to work.. *I mean*, to study?'

Filler:

- (113) Yo no sé lo que Dios.. lo que Dios diga, *yeah*, me encanta la palabra de Él..... 428P  
'I don't know what God.. what God says, *yeah*, I love His word.'
- (114) *So*.. él era otro que me decía, y él me decía, "y la Didi cómo está?" 432P  
'*So*.. he was another that would say to me, and he would say, "and Didi, how is she?"'
- (115) El semestre pasado yo tenía una clase que cogía *like*<sup>14</sup> personas irregulares, con compañías de gente... 365E  
'Last semester I had a class that took *like* irregular people with companies of people...'

When two or more tag items appear consecutively or with another lexical borrowing each is considered one lexical borrowing. For example, in (116), the adjacent tag items *I don't know* and *you know* are counted as two distinct lexical borrowings.

- (116) Es que si llueve (...) que si lluvia con nieve, que si *I don't know, you know*. Y para cada cosa pues tienes que prepararte. 303P  
'It's that if it rains (...) if rain with snow, if *I don't know, you know*. And for each thing then you have you prepare yourself.'

---

14 The word *like* can be almost any part of speech: noun, verb, adjective, adverb, preposition, etc. A concern in this study was distinguishing between *like* as a preposition and *like* as a discourse marker (since if *like* happens to also be adjacent to other English material, its function means the difference between including or excluding it in the database). Two tests are used to determine whether *like* has a syntactic role or is a discourse marker. The substitution test replaces the word *like* with any of its various meanings as a preposition (or adjective or adverb), such as 'similar to', 'characteristic of', 'indicative of', etc. If substituting these expressions for *like* results in a semantically coherent sentence, *like* is taken in that utterance to be a preposition. The other test, the deletion test, leaves *like* completely out of the sentence. If whatever *like* was bringing to the semantic table seems inessential to the cognitive meaning of the utterance, that is, if the utterance makes complete sense without it, *like* is considered a discourse marker on that occasion of its use.

## 10.2 Exclusions

Tag items with the following characteristics are excluded from the database.

Tag items *oh* and *ok/okay/okey*. These are not included in the database because they are preimmigration items (Chapter 3, section 4.2.1).

Tag items adjacent to or within codeswitches. Tag items that occur at the edges (117-118) or within a codeswitch (119) are excluded.

(117) But after he learns a little to, you know. 228D

(118) You know, so, she really didn't have much choice in the matter. 012U

(119) Everybody's older th[an] I am and I am only, you know, 23 years old. 012U

Strings that resemble tag items in syntactically complex arrangements. Tag items that are in syntactically complex arrangements with one another are not considered lexical borrowings. Look at *you know what I mean* in (120).

(120) Si te vas a poder meter a hacer un libro, *you know what I mean*, que es con lo que yo esto bregando ahora... 303P

'If you are going to be able to commit yourself to writing a book, *you know what I mean*, which is what I'm dealing with now...'

In example (120), *I mean* is part of a nominal relative pronoun phrase that operates as the object of *know*. Thus *know* in example (120) is acting in its full capacity as a verb (i.e. it is subcategorizing for an object, something that the *you know* discourse marker does not do), and the entire string *you know what I mean* is classified as a codeswitch and is excluded.

## Appendix E: Supplemental tables

This appendix provides the results of cross-tabulations,  $\chi^2$  and post hoc analysis used for interpreting results in this dissertation.

### 1. Supplemental tables for Chapter 4

Table E.1 shows cross-tabulation for class and socioeconomic status (SES) for all informants of the corpus. The two variables are highly correlated, but not perfectly so. As mentioned in section 4.2 of Chapter 4, self-ascribed occupational class conditions borrowing rate more strongly than SES.

**Table E.1**  
Cross-tabulation and  $\chi^2$  for self-ascribed occupational class and socioeconomic status,  
All informants

	Socioeconomic status								Total	
	Level A (the lowest rating)		Level B		Level C		Level D			
	n	%	n	%	n	%	n	%	n	%
<b>Middle class</b>	1	<b>1</b>	18	<b>25</b>	37	<b>52</b>	16	<b>22</b>	72	<b>100</b>
<b>Working class</b>	14	<b>21</b>	27	<b>41</b>	18	<b>27</b>	7	<b>11</b>	66	<b>100</b>
<b>Total</b>	15	<b>11</b>	45	<b>32</b>	55	<b>40</b>	23	<b>17</b>	138	<b>100</b>

$\chi^2$  (3, N=138) = 22.94      p < .001

Table E.2 provides the distribution of age groups within each immigrant generation. As can be seen, second generation informants tend to be teens and young adults, while first generation informants tend to be young adults and middle aged.

**Table E.2**  
Cross-tabulation for age cohort and immigrant generation

	2 <sup>nd</sup> generation		1 <sup>st</sup> generation		Total	
	n	%	n	%	n	%
<b>Teenage</b> (age 13-19)	11	<b>69</b>	5	<b>31</b>	16	<b>100</b>
<b>Young adult</b> (age 23-39)	32	<b>37</b>	54	<b>63</b>	86	<b>100</b>
<b>Middle aged</b> (age 40-59)	4	<b>11</b>	31	<b>87</b>	35	<b>100</b>
<b>Senior</b> (60+ years old)	1	<b>10</b>	9	<b>90</b>	10	<b>100</b>

## 2. Supplemental tables for Chapter 5

Table E.4 provides the distribution of first generation informants by occupational class and educational attainment. The table shows that 70 percent of those with education beyond high school claimed middle class membership, while 68 percent of those with secondary or less education claimed working class membership. This skewing is significant ( $p < .001$ ) and robust as indicated by the  $Chi^2(1, N=95)$  value of 16.01.

**Table E.3**  
**Cross-tabulation and  $Chi^2$  for level of education and occupational class,**  
**First generation**

	Middle class		Working class		Total	
	n	%	n	%	n	%
<b>No more than HS</b>	14	<b>30</b>	33	<b>70</b>	47	<b>100</b>
<b>College or Graduate</b>	34	<b>71</b>	14	<b>29</b>	48	<b>100</b>
<b>Total</b>	48	<b>51</b>	47	<b>49</b>	95	<b>100</b>

$$Chi^2(1, N=95) = 16.01 \quad p < .001$$

Table E.4 provides the distribution of all OZC informants by occupational class and educational attainment. Again, 70 percent of those with education beyond high school claimed middle class membership, while 68 percent of those with secondary or less education claimed working class membership. This skewing is significant ( $Chi^2(1, N=141) = 19.48, p < .001$ ).

**Table E.4**  
**Cross-tabulation and  $Chi^2$  for level of education and occupational class,**  
**All informants**

	Middle class		Working class		Total	
	n	%	n	%	n	%
<b>No more than HS</b>	20	<b>32</b>	42	<b>68</b>	62	<b>100</b>
<b>College or Graduate</b>	55	<b>70</b>	24	<b>30</b>	79	<b>100</b>
<b>Total</b>	75	<b>53</b>	66	<b>47</b>	141	<b>100</b>

$$Chi^2(1, N=141) = 19.48 \quad p < .001$$

Table E.5 shows that among employed Latinos, aged 19-64 in New York City in 2010, those with more education tend to be in white collar jobs, while those with less than a high school education tend to be in blue collar jobs.

**Table E.5**  
**Cross-tabulation and Chi<sup>2</sup> for level of education and occupational class, Latinos age 19-64 in New York City, 2010**

	Military		White collar <sup>*</sup>		Blue collar <sup>†</sup>		Total	
	n	%	n	%	n	%	n	%
<b>Kindergarten or less</b>	0	<b>0</b>	1614	<b>25</b>	4746	<b>75</b>	6360	<b>100</b>
<b>No more than HS</b>	0	<b>0</b>	53,578	<b>34</b>	104,732	<b>66</b>	158,385	<b>100</b>
<b>College or Graduate</b>	353	<b>0.2</b>	112,793	<b>68</b>	52,687	<b>32</b>	165,833	<b>100</b>
<b>Total</b>	428	<b>0.1</b>	167,985	51	162,165	<b>49</b>	330,578	<b>100</b>

*Chi<sup>2</sup> (4, N=330,578) = 39,993.93*      *p < .001*

<sup>\*</sup> White collar includes the U.S. Census categories 'professional' and 'sales and office'.

<sup>†</sup> Blue collar includes U.S. Census categories 'Construction and maintenance', 'production and transportation' and 'farming, fishing and forestry'.

Table E.6 shows the results of a Tukey's post hoc analysis for ethnonational affiliation and borrowing rate in the second generation. The results show that the difference between the average borrowing rate of Puerto Ricans and Mexicans approaches significance (*p*<.098).

**Table E.6**  
**Tukey's post hoc:**  
**Ethnonational affiliation and borrowing rate, Second generation**

		Mean difference	<i>p</i>
Puerto Rico	Dominican Republic	2.7	.958
	Ecuador	4.7	.751
	Cuba	5.0	.875
	Colombia	6.7	.302
	Mexico	8.8	<b>.098</b>
Mexico	Dominican Republic	-6.1	.398
	Ecuador	-3.9	.863
	Cuba	-3.8	.871
	Colombia	-2.1	.985

The following tables provide data on the stratification of informants by ethnonational affiliation and English-related linguistic traits. This data supports arguments put forth in Chapter 5 (section 3.3.1, footnote 24) that neither confidence in English, English proficiency nor English use consistently accounts for borrowing rates of ethnonational groups.

Table E.7 shows that among second generation informants, 100 percent of Puerto Ricans rate their English ability highly. However, other ethnonational groups, such as Cubans, tend to rate their English highly and yet are among the least frequent borrowers.

**Table E.7**  
**Cross-tabulation and Chi<sup>2</sup> for ethnonational affiliation and English confidence, Second generation**

	Most confidence		Less confidence		Total	
	n	%	n	%	n	%
<b>Puerto Ricans</b>	8	<b>100</b>	0	<b>0</b>	8	<b>100</b>
<b>Ecuadorians</b>	6	<b>86</b>	1	<b>14</b>	7	<b>100</b>
<b>Cubans</b>	5	<b>71</b>	2	<b>29</b>	7	<b>100</b>
<b>Dominicans</b>	6	<b>67</b>	3	<b>33</b>	9	<b>100</b>
<b>Colombians</b>	5	<b>56</b>	4	<b>44</b>	9	<b>100</b>
<b>Mexicans</b>	4	<b>50</b>	4	<b>50</b>	8	<b>100</b>

*Chi*<sup>2</sup> (5, N=48) = 6.82      *p* < .234

Table E.8 shows that in the first generation, Puerto Ricans have one of the highest concentrations of English-proficient individuals: 53 percent (n=9). However, Ecuadorians also have a high concentration of English-proficient individuals (41 percent or n=7), and yet are among the least-frequent borrowers in the first generation.

**Table E.8**  
**Cross-tabulation and Chi<sup>2</sup> for national origin and English proficiency, First generation**

	Good/excellent		Passable/poor		Total	
	n	%	n	%	n	%
<b>Mexicans</b>	10	<b>59</b>	7	<b>41</b>	17	<b>100</b>
<b>Puerto Ricans</b>	9	<b>53</b>	8	<b>47</b>	17	<b>100</b>
<b>Ecuadorians</b>	7	<b>41</b>	10	<b>59</b>	17	<b>100</b>
<b>Dominicans</b>	6	<b>37.5</b>	10	<b>62.5</b>	16	<b>100</b>
<b>Colombians</b>	3	<b>21</b>	11	<b>77</b>	14	<b>100</b>
<b>Cubans</b>	2	<b>12</b>	15	<b>88</b>	17	<b>100</b>

*Chi*<sup>2</sup> (5, N=98) = 11.44      *p* < .043

Table E.9 shows that English use among second generation informants also cannot account for higher borrowing rates among Puerto Ricans. Although 88 percent (n=7) of second generation Puerto Ricans use mostly English for daily activities, 89 percent (n=8) of Colombians do as well. Yet, second generation Colombians have very low borrowing rates. In other words, daily English use (i.e. the inverse of the daily Spanish use variable) does not seem to explain the borrowing rate trends among ethnonational groups in the second generation.

**Table E.9**  
**Cross-tabulation and Chi<sup>2</sup> for ethnonational affiliation and English use, Second generation**

	Use mostly English		Use less English		Total	
	n	%	n	%	n	%
<b>Colombians</b>	8	<b>89</b>	1	<b>11</b>	9	<b>100</b>
<b>Puerto Ricans</b>	7	<b>87.5</b>	1	<b>12.5</b>	8	<b>100</b>
<b>Mexicans</b>	6	<b>75</b>	2	<b>25</b>	8	<b>100</b>
<b>Ecuadorians</b>	5	<b>71</b>	2	<b>29</b>	7	<b>100</b>
<b>Dominicans</b>	6	<b>67</b>	3	<b>33</b>	9	<b>100</b>
<b>Cubans</b>	3	<b>43</b>	4	<b>57</b>	7	<b>100</b>

*Chi*<sup>2</sup> (5, N=48) = 5.43      *p* < .366

Table E.10 shows that, in the first generation, daily English use (i.e. the inverse of the daily Spanish use variable) does not consistently account for average borrowing rates of ethnonational groups. The table shows that most individuals from the two groups with the highest borrowing rates, Puerto Ricans and Mexicans, also tend to use English for daily activities. Among Ecuadorians, there are as many individuals that use mostly English as there are among Mexicans (59 percent or n=10), yet Ecuadorians have the lowest borrowing rates in the first generation.

**Table E.10**  
**Cross-tabulation and Chi<sup>2</sup> for national origin and English use, First generation**

	Use mostly English		Use less English		Total	
	n	%	n	%	n	%
<b>Puerto Ricans</b>	13	<b>77</b>	4	<b>23</b>	17	<b>100</b>
<b>Mexicans</b>	10	<b>59</b>	7	<b>41</b>	17	<b>100</b>
<b>Ecuadorians</b>	10	<b>59</b>	7	<b>41</b>	17	<b>100</b>
<b>Colombians</b>	8	<b>57</b>	6	<b>43</b>	14	<b>100</b>
<b>Cubans</b>	9	<b>47</b>	8	<b>53</b>	17	<b>100</b>
<b>Dominicans</b>	6	<b>37.5</b>	10	<b>62.5</b>	16	<b>100</b>

*Chi*<sup>2</sup> (5, N=98) = 5.28      *p* < .383

In sum, the patterns of English proficiency, English confidence and English use among ethnonational groups in the first and second generation do not seem to account for the pattern of borrowing rates observed and described in Chapter 5 (section 2.4.1).

## References

- About Language Use webpage. U.S. 2010 Census Bureau. Available from <https://www.census.gov/hhes/socdemo/language/about/index.html>. Last revised 15 November 2010. Accessed 6 January 2013.
- Acholonu, Chatherine, Joyce Ann Penfield and Joyce Anee Okezie. 1980. Linguistic processes of lexical innovation in Igbo. *Anthropological Linguistics* 22(3): 118-130.
- Amara, Muhammad Hasan and Bernard Spolsky. 1986. The diffusion and integration of Hebrew and English lexical items in the spoken Arabic of an Israeli village. *Anthropological Linguistics* 28(1): 43-54. Available from <http://www.jstor.org.ezproxy.gc.cuny.edu/stable/pdfplus/30027944.pdf?acceptTC=true>. Accessed May 2013.
- Auer, Peter. 1998. Chapter 1: Bilingual conversation revisited. *Code-switching in conversation; Language interaction and identity* [Pragmatics and Beyond Series] edited by Peter Auer, 1-24. Amsterdam: John Benjamins Publishing Company.
- Barkin, Florence. 1980. The role of loanword assimilation in gender assignment. *Bilingual Review / Revista bilingue* 7: 105-112.
- Berk-Seligson, Susan. 1986. Linguistic constraints on intrasentential code-switching: A study of Spanish/Hebrew bilingualism. *Language in Society* 15(3): 313-348. Available from <http://www.jstor.org/stable/4167766>. Accessed 25 March 2010.
- Blom, Jan-Petter and John Gumperz. 1972. Social meaning in linguistic structures: Code-switching in Norway. *Directions in sociolinguistics* edited by John J. Gumperz and Dell Hymes, 407-434. New York, NY: Holt, Rinehart and Winston.
- Bowen, Donald. 1975. Adaptation of English borrowings. *El lenguaje de los Chicanos* ed. by Eduardo Hernandez-Chavez, Andres Cohen and Anthony Beltramo, 115-21. Arlington, VA: Center for Applied Linguistics.
- Clyne, Michael. 2003. *The dynamics of language contact*. Cambridge: Cambridge University Press.
- Collins Online Collocation Dictionary*. Available from <http://www.collinsdictionary.com/dictionary/english>. Accessed January 2013.
- Chen, Xiao, Philip B. Ender, Michael Mitchel and Christine Wells. 2003. Regression with SPSS. Available from <http://www.ats.ucla.edu/stat/spss/webbooks/reg/default.htm>. Accessed October 2012.
- Del Valle, José and Luis Gabriel-Stheeman. 2004. Nationalism, hispanismo, and monoglossic culture. *The battle over Spanish between 1800 and 2000: Language and ideologies and hispanic intellectuals* edited by José del Valle and Luis Gabriel-Stheeman, 1-13. London / New York, NY: Routledge.

- Di Sciullo, Anne-Marie, Pieter Muysken and Rajendra Singh. 1986. Government and codemixing. *Journal of Linguistics* 22(1): 1-24.
- Diccionario de la lengua española* (22<sup>a</sup> ed.). 2001. Madrid: La Real Academia Española. Available from <http://www.rae.es/rae.html>. Accessed January 2013.
- Diccionario del español de Cuba*. 2000. Giselea Cardenas Molina, Antonia Maria Trista Perez and Reinhold Werner (coordinators), Reinhold Werner and Gunther Haensch (directors). Madrid: Gredos.
- Dictionary.com. n.d. Definition of *hearse*. Available from <http://dictionary.reference.com/browse/hearse>. Accessed 7 January 2013.
- Eckert, Penelope. 2000. *Linguistic variation as social practice*. [Language in Society 27 edited by Peter Trudgill.] Malden, MA: Blackwell Publishers.
- Eckert, Penelope 2002. Constructing meaning in sociolinguistic variation. Paper presented at the Annual Meeting of the American Anthropological Association. New Orleans, November 2002.
- Eckert, Penelope. 2008. Variation and the indexical field. *Journal of Sociolinguistics* 12(4): 453-476. Available from [www.stanford.edu/~eckert/PDF/IndexicalField.pdf](http://www.stanford.edu/~eckert/PDF/IndexicalField.pdf). Accessed January 2013.
- Eslami Rasekh, Abbas, Behrooz Ghoorchaei and Nematulla Shomoossi. 2008. Analysis of borrowed words in the spoken discourse of Turkmen native speakers in Iran: A sociolinguistic perspective. *Pakistan Journal of Social Sciences* 5(6): 551-561. Available from <http://www.medwelljournals.com/fulltext/pjss/2008/551-561.pdf>. Accessed June 2009.
- Espinosa, Aurelio. 1917. Speech mixture in New Mexico: The influence of the English language on New Mexican Spanish. *The Pacific Ocean in History* edited by Henry Morse Stephens and Herbert Eugene Bolton, 408-428. New York, NY: The Macmillan Company.
- Ferguson, Charles A. 1959. Diglossia. *Word* 15: 324-340.
- Field, Frederic. 2002. *Linguistic constraints on lexical borrowing in contact situations*. [Studies in Language Companion Series.] Amsterdam / Philadelphia, PA: John Benjamins Publishing Company.
- Fishman, Joshua. 1967. Bilingualism with and without diglossia; Diglossia with and without bilingualism. *Journal of Social Issues* 23: 29-38.
- Fox Tree, Jean E. 1995. The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language* 34: 709-738.
- Gal, Susan and Judith Irvine. 1995. The boundaries of languages and disciplines: How ideologies construct difference. *Social Research* 62(4): 967-1001.

- Goldman-Eisler, Frieda. 1968. *Psycholinguistics: Experiments in spontaneous speech*. New York, NY: Academic Press.
- Grosjean, François. 1997a. Processing mixed language: Issues, findings, and models. *Tutorials in bilingualism: Psycholinguistic perspectives* edited by Annette M.B. de Groot and Judith F. Kroll, 225-254. Mahwah, NJ: Lawrence Erlbaum Associates.
- Grosjean, François. 1997b. The bilingual individual. *Interpreting* 2(1/2): 163-187.
- Gumperz, John. 1962. Types of linguistic communities. *Anthropological Linguistics* 4(1): 28-40.
- Gumperz, John. 1976. The sociolinguistic significance of conversational code-switching. *Working Papers of the Language Behavior Research Laboratory* 46. Berkeley: University of California.
- Gumperz, John. 1982. *Discourse strategies*. Cambridge: Cambridge University Press.
- Haegeman, Liliane. 1994. *Introduction to government and binding theory* (2<sup>nd</sup> ed.). Malden, MA: Blackwell Publishing.
- Haegeman, Liliane and Jacqueline Guéron. 1999. *English grammar: A generative perspective*. Malden, MA: Blackwell Publishing.
- Harshenin, Alex. 1964a. English loanwords in the Doukhobor dialect, 1. *Canadian Slavonic Papers* 6: 38-43.
- Harshenin, Alex. 1964b. English loanwords in the Doukhobor dialect, 2. *Canadian Slavonic Papers* 9(2): 16-30.
- Haspelmath, Martin. 2004. Loanword typology: Steps toward a systematic cross-linguistic study of lexical borrowability. *Aspects of language contact: New theoretical, methodological and empirical findings with special focus on Romancisation processes* edited by Thomas Stolz, Dik Bakker and Rosa Salas Palomo, 43-62. Berlin: Mouton de Gruyter.
- Haugen, Einar. 1950. The analysis of linguistic borrowing. *Language* 26(2): 210-231.
- Haugen, Einar. 1953/1969. *The Norwegian language in America: A study of bilingual behavior*. Bloomington, IN: Indiana University Press.
- Haugen, Einar. 1972. *The ecology of language*. Stanford: Stanford University Press.
- Heller, Monica. 1992. The politics of codeswitching and language choice. *Journal of Multilingual and Multicultural Development* 13(1/2): 123-142.
- Hlavac, Jim. 2000. Croatia in Melbourne: Lexicon, switching and morphosyntactic features in the speech of second-generation bilinguals. Doctoral dissertation, Monash University.
- Holden, Kyril. 1976. Assimilation rates of borrowings and phonological productivity. *Language* 52: 131-147.

- Irvine, Judith and Susan Gal. 2000. Language ideology and linguistic differentiation. *Regimes of language: Ideologies, politics, and identities* edited by Paul Kroskrity, 35-84. Santa Fe, NM: School of American Research Press.
- Johnstone, Barbara and Scott F. Kiesling. 2008. Indexicality and experience: Exploring the meanings of /aw/-monophthongization in Pittsburgh. *Journal of Sociolinguistics* 12: 5-33.
- Klee, Carol. 1987. Differential language usage patterns by males and females in a rural community in the Rio Grande Valley. *Language and language use: Studies in Spanish* edited by Terrell Morgan, James Lee and Bill Van Patten, 125-45. Lanham, MD: University Press of America
- Labov, William. 1963. The social motivation of a sound change. *Word* 19: 273-309. [Also: 1972. Chapter 1. *Sociolinguistic patterns*, 1-42. Philadelphia, PA: University of Pennsylvania Press.]
- Labov, William. 1966. *The social stratification of English in New York City*. Washington D.C.: Center for Applied Linguistics.
- Labov, William. 2001. *Principles of linguistic change* (vol. 2): *Social factors*. [Language in Society 29.] Oxford: Blackwell Publishing.
- Lavandera, Beatrice. 1978. Where does the sociolinguistic variable stop? *Language in Society* 7: 171-183. Available from [http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content\\_storage\\_01/0000019b/80/35/bf/7b.pdf](http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/35/bf/7b.pdf). Accessed June 2009. [Also: 1977. *Working papers in sociolinguistics* 40. Austin, TX: Southwest Educational Development Laboratory.]
- Levelt, Willem J.M. 1989. *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levinson, Stephen C. 2002. *Pragmatics*. [Cambridge Textbooks in Linguistics.] Cambridge: Cambridge University Press.
- Lipski, John. 2005. Code-switching or borrowing? No sé so no puedo decir, you know. *Selected proceedings of the second workshop on Spanish sociolinguistics* edited by Lotfi Sayahi and Maurice Westmoreland, 1-15. Somerville, MA: Cascadilla Press. Available from <http://www.lingref.com/cpp/wss/2/paper1136.pdf>. Accessed April 2012.
- Lipski, John. 2008. *Varieties of Spanish in the U.S.* Washington D.C.: Georgetown University Press.
- Llorach, Emilio Alarcos. 1999. *Gramática de la lengua española*. Madrid: Real Academia Española.
- Lyons, John. 1977. *Semantics* (vols 1 & 2). Cambridge: Cambridge University Press.
- Matus-Mendoza, María de la luz. 2002. The English lexical loan: A class marker. *Journal of Hispanic Higher Education* 1: 329-337.

- Mendieta, Eva. 1999. *El préstamo en el español de los estados unidos* [Studies in Ethnolinguistics 1.] New York, NY: Peter Lang.
- Milroy, James and Lesley Milroy. 1985. Linguistic change, social network and speaker innovation. *Journal of Linguistics* 21(2): 339-384.
- Milroy, Lesley and Matthew Gordon. 2003. *Sociolinguistics: Method and interpretation*. Oxford: Blackwell Publishing Ltd.
- Muysken, Pieter, Hetty Kook and Paul Vedder. 1996. Papiamentu/Dutch code-switching in bilingual parent-child reading. *Applied Psycholinguistics* 17: 485-505.
- Muysken, Pieter. 1995. Code-switching and grammatical theory. *One speaker, two languages: Cross-disciplinary perspectives on code-switching* edited by Lesley Milroy and Peter Muysken, 177-198. Cambridge: Cambridge University Press.
- Muysken, Pieter. 2000. *Bilingual speech: A typology of code-mixing*. Cambridge / New York, NY: Cambridge University Press.
- Myers-Scotton, Carol. 1997. *Dualing languages: Grammatical structure in code-switching*. Oxford / New York, NY: Clarendon Press.
- Myers-Scotton, Carol. 1998. *Codes and consequences: Choosing linguistic varieties*. New York, NY: Oxford University Press.
- Neumann, J.H. 1938. Notes on American Yiddish. *The Journal of English and Germanic Philology* 37(3): 403-421.
- Newton, Rae and Kjell Erik Rudestam. 1999. *Your statistical consultant: Answers to your data analysis questions*. London: Sage Publications.
- Ngom, Fallou. 2002. *Lexical borrowings as a sociolinguistic variable in Saint-Louis, Senegal*. Doctoral dissertation, Urbana-Champaign, Department of French.
- Novotná, Zdenka. 1967. Linguistic factors of the low adaptability of loanwords to the lexical system of modern Chinese. *Monumenta Serica* 26: 103-118.
- Oppenheim, Abraham Naftali. 1992. *Questionnaire design, interviewing and attitude measurement*. London: Pinter Publishers.
- Ornstein, Jacob. 1976. Sociolinguistic constraints on lexical borrowing in Tarahumara: Explorations in 'langue and parole' and 'existential bilingualism': An approximation. *Anthropological Linguistics* 18(2): 70-93.
- Otheguy, Ricardo. 2012. Peer Commentaries: Concurrent models and cross-linguistic analogies in the study of prepositional stranding in French in Canada. *Bilingualism: Language and Cognition* 15(2): 226-229.

- Otheguy, Ricardo and Ana Celia Zentella. 2012. *Spanish in New York: Language contact, dialectal leveling, and structural continuity*. [Oxford Studies in Sociolinguistics.] New York, NY: Oxford University Press.
- Otheguy, Ricardo, Ofelia García and Mariela Fernández. 1989. Transferring, switching and modeling in West New York Spanish: An intergenerational study. *International Journal of the Sociology of Language* 79: 41–52.
- Otheguy, Ricardo, Ofelia García and Ana Roca. 2000. Speaking in Cuban: The language of Cuban Americans. *New immigrants in the United States: Readings for second language educators* edited by Sandra L. McKay and Sau-ling C. Wong, 165-188. Cambridge: Cambridge University Press.
- Otheguy, Ricardo, Ana Celia Zentella and David Livert. 2007. Language and dialect contact in Spanish in New York: Toward the formation of a speech community. *Language* 83(4): 770-802.
- Otheguy-Zentella Corpus of Spanish in New York. 1999-2004. Developed under the direction of Ricardo Otheguy of the Graduate Center, CUNY and Ana Celia Zentella of the University of California at San Diego. Funded by the National Science Foundation, grant number 004133.
- Paradis, Michel. 1997. The cognitive neuropsychology of bilingualism. *Tutorials in bilingualism: Psycholinguistic perspectives* edited by Annette M.B. de Groot and Judith Kroll, 331-354. Mahwah, NJ: Lawrence Erlbaum Associates.
- Pfaff, Carol. 1979. Constraints on language mixing: Intrasentential code-switching and borrowing in Spanish/English. *Language* 55(2): 291-318. Available from <http://links.jstor.org/sici?sici=0097-8507%28197906%2955%3A2%3C291%3ACOLMIC%3E2.0.CO%3B2-D>. Accessed March 2010.
- Poplack, Shana. 1978/1981. Syntactic structure and social function of code-switching. *Latino language and communicative behavior* edited by Richard P. Duran, 169-184. Norwood, NJ: Ablex Publishing Corporation. Available from <http://www.sociolinguistics.uottawa.ca/shanapoplack/pubs/articles/Poplack1981.pdf>. Accessed June 2013.
- Poplack, Shana. 1980. Sometimes I'll start a sentence in Spanish y termino en español. *Linguistics* 18: 581-618.
- Poplack, Shana. 1983. Bilingual Competence: Linguistic Interference of Grammatical Integrity? *Spanish in the U.S. setting: Beyond the Southwest* ed. by Lucía Elías Olivares, 111-129. Washington D.C.: National Clearinghouse for Bilingual Education. Available from <http://www.sociolinguistics.uottawa.ca/shanapoplack/pubs/articles/Poplack1983.pdf>. Accessed June 2013.
- Poplack, Shana. 1987. Contrasting patterns of code-switching in two communities. *Aspects of Multilingualism* edited by Erling Wande, Jan Anward, Bengt Nordberg, Lars Steensland and Mats Thelander, 51-77. Uppsala: Borgströms.

- Poplack, Shana and Nathalie Dion. 2012. Myths and facts about loanword development. *Language Variation and Change* 24: 279-315.
- Poplack, Shana and Marjory Meechan. 1998. Introduction: How languages fit together in codemixing. *International Journal of Bilingualism* 2(2): 127-38.
- Poplack, Shana and David Sankoff. 1984. Borrowing: The synchrony of integration. *Linguistics* 22(1): 99-135.
- Poplack, Shana, David Sankoff and Christopher Miller. 1988. The social correlates and linguistic consequences of lexical borrowing and assimilation. *Linguistics* 26: 47-104.
- Poplack, Shana, Lauren Zentz and Nathalie Dion. 2012. Phrase-final prepositions in Quebec French: An empirical study of contact, code-switching and resistance to convergence. *Bilingualism: Language and Cognition* 15(2): 203–225.
- Poullisse, Nanda. 1997. Language production in bilinguals. *Tutorials in bilingualism: Psycholinguistic perspectives* edited by Annette M.B. de Groot and Judith Kroll, 201-224. Mahwah, NJ: Lawrence Erlbaum Associates.
- Pyles, Thomas and John Algeo 1993. *The origins and development of the English language*. Boston, MA: Heinle.
- Roca, Ana. 2000. El español en los EEUU a principios del siglo XXI: Apuntes relativos a la investigación sobre la variedad de la lengua y la coexistencia con el inglés en las comunidades bilingües. *Teoría y práctica del contacto: El español de América en el candelero* editado por Julio Calvo Pérez, 193-211. Madrid: Editorial Iberoamericana / Frankfurt: Verguet Verlag.
- Schegloff, Emanuel, Gail Jefferson and Harvey Sacks. 1977. The preference for self-correction in the organization of repair in conversation. *Language* 53(2): 361-382.
- Siegmán, Aron W. 1979. Cognition and hesitation in speech. *Of speech and time* edited by Aron W. Siegmán and Stanley Feldstein, 151-77. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Silva-Corvalán, Carmen. 1994/2002. *Language contact and change: Spanish in Los Angeles*. Oxford: Clarendon Press.
- Silva-Corvalán, Carmen. 1995. The study of language contact: An overview of the issues. *Spanish in four continents: Studies in language contact and bilingualism* edited by Carmen Silva-Corvalán, 3-14. Washington D.C.: Georgetown University Press.
- Spencer, Robert. 1947. Spanish loanwords in Keresan. *Southwestern Journal of Anthropology* 3(2): 130-146.
- Stavans, Ilan. 2003. *Spanglish: The making of a new American language*. New York, NY: Harper Collins Publishers.

- Sullivan, Celeste M. 2008. A mechanism of lexical borrowing. *Journal of Language Contact (VARIA)* 1: 17-28.
- Teschner, Richard Vincent. 1972. *Anglicism in Spanish: A cross-referenced guide to previous findings, together with English lexical influence on Chicago Mexican Spanish*. Doctoral dissertation, University of Wisconsin.
- Tesoro lexicográfico del español de Puerto Rico*. 2005. Maria Vaquero and Amparo Morales (eds). San Juan, PR: Academia Puertorriquena de la lengua Espanola.
- Thomason, Sarah. 2001. *Language contact. An introduction*. Washington D.C.: Georgetown University Press.
- Thomason, Sarah and Terrence Kaufman. 1988. *Language contact, creolization and genetic linguistics*. Berkeley, CA: University of California Press.
- Trager, George and Genevieve Valdez 1937. English loans in Colorado Spanish. *American Speech* 12(1): 34-44.
- Tsiapera, Maria. 1964. Greek borrowings in the Arabic dialect of Cyprus. *Journal of the American Oriental Society* 84(2): 124-126.
- U.S. Census Bureau. 2010. PUMS American Community Survey 3-year data for New York, 2008-2010. Available from [http://www.census.gov/acs/www/data\\_documentation/pums\\_data/](http://www.census.gov/acs/www/data_documentation/pums_data/). Accessed January 2012.
- U.S. Census Bureau. 2012. Quick Facts Los Angeles. Available from <http://quickfacts.census.gov/qfd/states/06/0644000.html>. Accessed 6 January 2013. Last updated 6 December 2012.
- U.S. Census Bureau. 2012. Quick Facts Miami. Available from <http://quickfacts.census.gov/qfd/states/12/1245000.html>. Accessed 6 January 2013. Last updated 6 December 2012.
- U.S. Census Bureau. 2012. Quick Facts New York City. Available from <http://quickfacts.census.gov/qfd/states/36/3651000.html>. Accessed 6 January 2013. Last updated 6 December 2012.
- Van Coetsem, Frans. 2000. *A general and unified theory of the transmission process in language contact*. Heidelberg: Universitätsverlag Carl Winter.
- Van Hout, Roeland and Pieter Muysken. 1994. Modeling lexical borrowability. *Language Variation and Change* 6: 39-62.
- Vedder, Paul, Hetty Kook and Pieter Muysken. 1996. Language choice and functional differentiation of languages in bilingual parent-child reading. *Applied Psycholinguistics* 17: 461-484.

- Villa, Daniel and Susana Rivera-Mills. 2009. An integrated multi-generational model for language maintenance and shift: The case of Spanish in the Southwest. *Spanish in Context* 6(1): 26-42.
- Webster's Unabridged Dictionary* (2<sup>nd</sup> ed.). 2001. New York, NY: Random House.
- Weinreich, Uriel. 1953/1966. *Languages in contact: Findings and problems*. London: Mouton and Company.
- Whitney, William Dwight. 1881. On mixture in language. *Transactions of the American Philological Association* 12: 5-27.
- Winford, Donald. 2003. *An introduction to contact linguistics*. Malden, MA: Blackwell Publishing.
- Woolford, Ellen. 1983. Bilingual code-switching and syntactic theory. *Linguistic Inquiry* 14(3): 520-536. The MIT Press. Available from <http://www.jstor.org/stable/4178342>. Accessed 20 November 2012.
- Zelinsky, Wilbur. 1992. *The cultural geography of the United States. A revised edition*. Englewood Cliffs, NJ: Prentice-Hall.
- Zentella, Ana Celia. 1990. Lexical leveling in four New York City Spanish dialects: Linguistic and social factors. *Hispania* 73(4): 1094-1105.

### **Additional sources**

- Hernandez, Ramona (Director of the Dominican Studies Institute, CUNY). 2013. Personal communication, 3 July. New York, NY.
- U.S. Census Bureau. 2003. PUMS PUMA Codes. Available from [http://www2.census.gov/census\\_2000/datasets/PUMS/FivePercent/New\\_York/PUMEQ5-NY.TXT](http://www2.census.gov/census_2000/datasets/PUMS/FivePercent/New_York/PUMEQ5-NY.TXT). Last updated 9 August 2003. Accessed 6 March 2010.
- U.S. Census Bureau. 2009. *A compass for understanding and using American Community Survey Data: What PUMS data users need to know*. Washington D.C.: U.S. Government Printing Office. Available from <http://www.census.gov/acs/www/Downloads/handbooks/ACSPUMS.pdf>. Accessed 3 March 2010.
- U.S. Census Bureau. 2011. *2008-2010 ACS PUMS data dictionary*. Available from [http://www.census.gov/acs/www/Downloads/data\\_documentation/pums/DataDict/PUMS\\_Data\\_Dictionary\\_2008-2010.pdf](http://www.census.gov/acs/www/Downloads/data_documentation/pums/DataDict/PUMS_Data_Dictionary_2008-2010.pdf). Accessed 7 March 2012.
- Veltman, Calvin. 1988. *The future of the Spanish language in the United States*. New York, NY: Hispanic Policy Development Project.