Introduction to "Lexical Borrowing and Deborrowing in Spanish in New York City"

Rachel Varra
LEXICAL BORROWING AND DEBORROWING IN SPANISH IN NEW YORK CITY

TOWARD A SYNTHESIS OF THE SOCIAL CORRELATES OF LEXICAL USE AND DIFFUSION IN IMMIGRANT CONTEXTS

Rachel Varra
Lexical Borrowing and Deborrowing in Spanish in New York City
Toward a Synthesis of the Social Correlates of Lexical Use and Diffusion in Immigrant Contexts

Rachel Varra
Series editor: Dale Koike
Spanish list advisor: Javier Muñoz-Basols
Contents

List of tables x
List of figures xii
Preface xiii
Acknowledgments xviii
Notations and orthographic conventions xix

1 Introduction 1
1 Preliminaries 1
2 Spanish in New York City 4
3 Spanish in contact with English in New York City: a brief sketch of lexical phenomena 5
4 A note on language and language membership 7

2 The lexical borrowing database: classifying lexical contact phenomena 13
1 The concept of lexical borrowing 13
2 Data, models and operationalization of concepts 14
3 Data selection criteria and the formation of the English lexical borrowing database 22

3 The corpus and analysis 43
1 The Otheguy-Zentella Corpus, language contact and variation 43
2 Informants and interviews 45
3 Independent variables and stratification of the Spanish in New York corpus 46
4 Dependent variables 46
5 Analysis 48
6 Interpretation 48
7 Presentation of results 49
4 An overview of lexical borrowing behavior in Spanish in New York City
   1 Overview of the corpus 54
   2 What lexical borrowing says about Spanish speakers in New York City 55
   3 What lexical borrowing does not say about Spanish speakers in New York City 61
   4 Summary of lexical borrowing behavior in New York City 64

5 Immigrant generations in focus
   1 Lexical borrowing frequency in the first immigrant generation 69
   2 Lexical borrowing frequency in the second immigrant generation 78
   3 Conclusion 83

6 Innovation, reproduction and the dissemination of lexical borrowings in Spanish in New York City
   1 Introduction and additional questions 89
   2 Shared and nonshared vocabulary in Spanish in New York 89
   3 The use of nonshared vocabulary: lexical borrowing innovators 92
   4 The use of shared vocabulary: lexical borrowing reproducers 95
   5 On non-significant variables in lexical borrowing innovation and reproduction 98
   6 On the dissemination of lexical borrowings 99
   7 Conclusion and review of questions 100

7 Deborrowing: flagged lexical borrowings in Spanish in New York City
   1 Flagging in bilingual speech research 105
   2 Analysis 106
   3 The data and coding 107
   4 Examining the disfluency hypotheses 116
   5 Conclusion 120

8 Synthesis and application of findings
   1 A limitation of the current study: underrepresentation of the shared lexicon of Spanish in New York 124
   2 A portrait of lexical borrowing and its Spanish-speaking users in New York City 125
3 The social determinants of lexical borrowing across contact situations 127
4 Applications to bilingual speech research 129
5 Linking speech behavior to long-term outcomes of language contact 130

Appendix A: stratification of the Otheguy-Zentella Corpus of Spanish in NYC 136
Appendix B: excerpts from the Otheguy-Zentella Corpus by referring chapter 138
Appendix C: results of the homonymy test 141
Appendix D: criteria for lexical borrowing by part of speech category 143
  1 Nouns and noun phrases 143
  2 Adjectives 150
  3 Adverbs 151
  4 Verbs 152
  5 Prepositions and prepositional phrases 154
  6 Determiners 156
  7 Quantifiers: numeric, generic and ordinal 157
  8 Connectives: conjunctions, coordinators and complementizers 158
  9 Indexical demonstratives 159
  10 Tag items 160
  11 General inclusions 163
  12 General exclusions 164

Index 168
1 Introduction

1 Preliminaries

The individual man is everywhere only seeking after a sign [...] and he is always ready to take it where he finds it handiest [...] and more or less, according to wholly indeterminable circumstances [...] There are few tongues in the world which are not to this extent mixed.

(Whitney 1881: 10)

So it has been with languages and their speakers and so it continues to be. Those who have picked up this volume will be familiar with the linguistic behavior that Whitney describes – the “taking” of signs – both in their own speech and in that of others. Such practices are instantiated in the quotidian and perhaps unremarkable: the learning and use of a new word-form, such as *totes* to mean ‘totally’ or *preggers* to mean ‘pregnant’, as in “She was *totes* *preggers*.” But readers will also be familiar with practices that give greater pause, such as *estuvimos en un baby shower*, ‘we were at a baby shower’. The practice of using phrases (like *baby shower*) from one language in the midst of a sentence or discourse in another are just one way speakers could ‘take of the handiest sign’. And this particular way, the use of ‘other language’ strings in one’s speech, will here be called *lexical borrowings*.

Lexical borrowing is the focus of the present volume. In what follows the phenomenon will be defined and examined in order to answer three sorts of questions that have occupied sociolinguistic investigations: What can we know about people based on how they use lexical borrowings? What can we know about the process of lexical borrowing itself based who uses them? And how does lexical borrowing behavior inform today’s big questions of social and linguistic inquiry? In particular, this work will report the results of a large-scale, corpus-based study of lexical borrowing in Spanish in New York City. What we will see is that using ‘other language’ strings is not necessarily about being ‘handy’ in any obvious sense (indeed, ‘handy’ needs some definition) and, importantly, that thanks to advances in linguistic methodologies and technology, the circumstances for their occurrence are not “wholly interdeterminable”, as Whitney (and others since!) had early asserted.
Despite what may seem numerous tomes ostensibly dedicated to the topic of lexical borrowings, there is still much to be learned about this behavior. And this new knowledge stands to impactfully inform our teaching, our learning, our social policies and, to my mind and most importantly, our relationships.

Lexical borrowing is an especially conspicuous language behavior, particularly for interlocutors that do not share similar language learning experiences as the individuals that use them. It does not often go unnoticed. It is both the target of and itself comprises evidence for ideological assertions by politicians and everyday persons making claims about the social class, intelligence, level of education or work ethic of the person using them. Informally, the practice is sometimes referred to as 'codeswitching' or, when English is involved, the speaker’s talk may be labeled an ‘-ish’ (e.g. Spanglish, Chinglish, Franglais). Further, lexical borrowing, as just one instantiation of what I here call ‘bilingual mixed speech’, has received serious scholarly attention since at least the 1950s, when some of the most incisive and well-articulated thinking on topic emerged. Early treatments of lexical borrowing 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). Later studies 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; Toribio 2001a, 2001b, 2002, 2004) and explored universal constraints on language mixture (e.g. Berk-Seligson 1986; Myers-Scotton 1997; Poplack 1980).

Although, as we have seen, consideration of the extra-linguistic influences on lexical borrowing is found as early as 1881 in Whitney’s writings and in those of Espinosa (1917) as well, it was the work of Einar Haugen (1950, 1972) and Uriel Weinreich (1953) that emphasized that insight into 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 features as on linguistic constraints. Accordingly, several approaches to the study of language in bilingual societies, such as in anthropological linguistics, the sociology of language and discourse analysis, have examined the discourse significance of and the macrosocial factors conditioning lexical contact phenomena.

In subsequent decades, the insights of Haugen and Weinreich have not only been echoed by other scholars, but made more specific, highlighting the importance of the identity and personal characteristics of language users for a theory of language change. In 1962, Gumperz emphasized that understanding the dynamics of language change required the investigation of variation in the use of language contact phenomena within a speech community (30) – that is, with reference to a concrete group of individuals. In 1995, Silva-Corvalán noted that “in order to assess the stability of the [contact] features identified and the possibility that they may be passed on to new generations” (1995: 4), it is important to know who is using
the feature, and in particular to know their proficiency and how they use their languages in daily life. In 2000, Eckert pointed out that 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.

Although some inroads into how speaker identity and group membership contribute to lexical borrowing behavior has been made in the last twenty-five years (e.g. see Eslami Rasekh, Ghoorchaei & Shomoossi 2008; Mendieta 1999; Ngom 2002; Poplack 1980; Poplack, Sankoff & Miller 1988; Sullivan 2008; Thomason & Kaufman 1988), the current body of work is relatively small and manifests several shortcomings. First, there is a need for more quantitative work in order to gain a systematic understanding of how borrowings spread through a community 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, even where quantitative studies exist, they rarely include sufficient numbers of participants so as to allow quantitative probing of the data. Finally, and perhaps most importantly, the concept of lexical borrowing, imported from the historical language study, is rarely defined in a way that makes it relevant to languages and speakers in contact situations.

This book reports on an investigation that addresses these problems and seeks to add to an understanding of the social correlates of lexical borrowing. A quantitative sociolinguistic design is marshaled to investigate variation in several aspects of borrowing behavior, such as its frequency in speech and its discourse treatment. It examines data from more than 140 Spanish speakers from six ethnonational Latino groups in New York to answer the following sets of questions:

i Which individuals are most susceptible to borrowing? To what extent is borrowing conditioned by language proficiency? What does the constellation of significantly correlating and non-correlating sociodemographic traits reveal about lexical borrowing in Spanish in New York City?

ii How do borrowings spread through a community? Who is most responsible for propagating their use? What can be learned about how lexical items are acquired, spread and retained/discarded in immigrant settings?

iii What does lexical borrowing behavior indicate about the future of Spanish in New York City? To what extent is Spanish in New York influenced or changed by contact with English?

The results of this study will do three things. Through this study we will first learn who among New York Spanish speakers is most susceptible to borrowing (¡ojo: it may not be who you think!), making results consistent with research that shows that lexical borrowing can be indexical of speaker identity. While this result is perhaps the most modest theoretically, the specifics of how the data lay out are, to my mind, the most immediately impactful both interpersonally and socially. They directly challenge commonly held perceptions, along with their negative
Introduction

associated value judgments about what using lexical borrowing means about the speaker.

Although no definitive claims about the permanence of particular lexical borrowings in Spanish in New York are made, and neither is an attempt at a comprehensive theory of how individual speech behaviors relate to long-term changes ventured, this study contributes evidence for the building of such a theory. In particular, it supplies consolidated observations, supported by empirical quantitative data, that can be used for constructing an account of the social determinants of lexical change in diverse contact situations. That is, to know what contact features may become part of Spanish in New York, it is important to know who is using them, and what types they are using. This study, in examining the social identity of speakers, provides a critical link between variation observed in speech and what the long-term outcomes of language contact will be (e.g. see Thomason & Kaufman 1988).

Finally, this work contributes answers to the questions that have dominated research on Spanish in the U.S. for the last half-century, such as, What is the linguistic fate of Spanish in the U.S.? An how and to what extent is English influencing Spanish (in New York City)? Results from lexical borrowing behavior in New York challenge not only our everyday assumptions about such matters, but also conclusions posited by academics working on Spanish in the U.S. Before seeing how, some background information is in order.

2 Spanish in New York City

Spanish is the second most-spoken language in the U.S. (Bills 2005; Roca 2000). The number of Spanish speakers in the U.S. is second only to Mexico (Instituto Cervantes 2016: 5–7) and 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.5 In 2010, more than 2,338,000 Latinos6 resided in New York, constituting about 28 percent of the city’s population (U.S. Census Bureau 2012c, Quick Facts New York City). The majority of Latinos in New York are natives to the U.S. (59 percent). The other 41 percent represent all 19 of the officially Spanish-speaking Caribbean and Central and South American countries. In fact, today, no one Latino group makes up the overwhelming of Latinos in the city. Among the most numerous Latino heritages claimed in the city are Puerto Rican (32 percent), Dominican (26 percent), Mexican (13 percent), Ecuadorian (8 percent), Colombian (5 percent), Cuban (2 percent), Honduran (2 percent), Salvadorian (2 percent) and Peruvian (2 percent) (U.S. Census 2010).

Of the more than two million Latinos older than age five in New York City, 85 percent speak a language besides English at home (U.S. Census 2010). For more than 99 percent of those, that language is Spanish. While it was once thought that Spanish among immigrants would give way to English among their children and grandchildren (Fishman 1967; Pedraza 1978 cited by Poplack 1980: 582), Spanish has, unlike the language of previous immigrant vintages to the U.S. (e.g. Italian, German), remained numerically strong on the community level. As might
be expected, about 97 percent of Latino first generation immigrants older than age five speak Spanish in the home after migrating to the U.S. What is more remarkable is that of the more than one million U.S.-born Latinos (i.e. the second generation) older than age five, 76 percent also speak Spanish at home, a situation that has been facilitated not only through renewal of native-speaking immigrants to the city (e.g. Pousada & Poplack 1982), but also due to individuals’ desire to continue speaking it.

Many Latino immigrants and their children also speak English. About 93 percent of U.S.-born Latinos in New York speak English ‘well’ or ‘very well’, while more than half of the foreign-born older than age five do (U.S. Census 2010). Due to high rates of bilingualism among both first and second generation Latinos (both in NYC and in the U.S.) and of course contact with English and English speakers, Spanish throughout the U.S. has been noted to manifest several characteristics distinguishing it from Spanish spoken elsewhere. The most well-known, that of English lexicon in Spanish, is of course the focus of the present volume. Let’s take a closer look.

3 Spanish in contact with English in New York City: a brief sketch of lexical phenomena

Lexical borrowing is here defined as the reproduction in one language of a lexical pattern (i.e. minimally one word-form plus meaning) from another language (Haugen 1969: 363). The language in which the form is reproduced is referred to as the recipient language. The language from which the form is taken is called the donor language. Lexical borrowing in this work shall refer not only to lexical patterns, but also to whatever cognitive or linguistic processes give rise to those lexical patterns. During this process, etymologically foreign lexical material undergoes social and linguistic modification (e.g. see Ngom 2002: 31), whereby 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. This process, theoretically beginning with the first time an individual uses a particular bit of donor language material in recipient language discourse, is hypothesized to end in one of two ways. On the one hand, the lexical material may attain the status of established loanword (Poplack, Sankoff & Miller 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. On the other hand, the process might terminate with the eventual obsolescence of the donor language material for the recipient language 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
    ‘No but I in high school was a counselor and that also helped me . . . ’
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 a noun phrase (NP), prepositional phrase (PP), adjective phrase (AdjP), adverb phrase (AdvP) or another verb phrase (VP). Such constructions are called ‘clause-like’ since grammars define a clause as minimally containing a verb predicate and an overt or implied subject (Haegeman & Guéron 1999: 22). So, for instance, *it wasn’t* in (5) would be an example of a codeswitch.

(5) Como *it wasn’t* como tan formal ... 
‘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 ... 
‘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.
A student of bilingual speech will, at this point, perhaps approach the present terminological framework cautiously, since numerous frameworks for characterizing donor language words and expressions in a recipient language discourse are already available and well-known. The data thus far presented, for example, might in other theories be called *loanwords*, *codeswitches*, *language alternations*, *language mixing*, *codemixing*, *lexical transfer* or *lexical insertion*. 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*. Yet, we would do well, I believe, to disregard category labels to some extent and instead keep the data itself in mind. For, as regards these and similar strings, there exist detailed and empirically grounded justifications for their being classified together, rather than classified as representing distinct phenomena. This justification is laid out in chapter 2.

4 A note on *language* and *language membership*

The present investigation has been constructed upon modernist, structuralist conceptualizations of *language* and *language membership*. I will, for example, make frequent reference to the idea of *English* and *Spanish* and that some words ‘belong to’ or ‘come from’ one or the other. Such conceptualizations of *language* have been increasingly criticized in recent decades (see Blommaert 2010; del Valle 2012; del Valle & Stheeeman 2004). Blommaert, for one, asserts that modernist notions of *language* are “artefactualized”, presenting language as a set of rules that exist outside of the individual, and to which the individual must adhere. Indeed, there is something decidedly bizarre about asserting that some feature X or some word Y “belongs to” any particular language, when it seems rather that it is people who possess “knowledge of”. Blommaert further points out that talking in these essentially misleading terms (2010: 12), in effect, dooms one to create inaccurate models of language use in society (2010: 47, 28–31, 42–43).

This should, nonetheless, not deter the reader from finding value in the results of the investigation. While theoreticians may challenge the ideological underpinnings of such models of language and simultaneously seek for models that offer alternative versions of facts and push the envelope of what has been done previously, non-specialist speakers are much slower to adopt new ways to conceptualize what they do, or shift their ideology without there being a very good reason to do so. The fact is, modernist notions of languages that make possible questions, such as “To which language does this word belong?,” have been and continue to be part of the everyday ways that individuals think about what they say. They, for instance, think of language in terms of circumscribed objects that in effect exist outside of themselves, which they can purchase or obtain or trade. Spanish speakers in New York, for instance, unselfconsciously associate speaking English as ‘good for economic advancement’ or speaking Spanish as ‘good for maintaining family culture’ (see Varra, unpublished). Data from the same corpus upon which this investigation is based shows that individuals even characterize interjections
and pause fillers like “oh” and “um” as ‘belonging to’ one language or another. The take away is this: Although the analyst must be trained to see linguistic data beyond ideas of its language membership, s/he cannot entirely dispense with such notions, since speakers themselves orient to them in constructing the semiotic value of their linguistic behaviors. As a result, not only are investigations that construct language as a reified object still relevant today; they in fact must stay in play in cases where speakers themselves understand their linguistic behavior in such terms, as is the case with Spanish in New York. To this extent, we will see that the results of this study are not only made possible by such conceptualizations of language, but that, in forming a coherent story about Spanish and English use within community, are also evidence that such conceptualizations continue to operate.

Notes

1 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).

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

3 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.

4 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).

5 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.

6 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. It is not selected in contradistinction to other orthographic representations such as Latin@ or Latinex.

7 For example, Pedraza (1985, cited in Garcia et al. 1988) found third and fourth generation individuals that spoke Spanish. Garcia, Morin and Rivera (2001: 71) note that social imbalances, such as inferior education and discrimination, will encourage continued use of Spanish among Puerto Ricans. These conditions hold as well for speakers of other Spanish varieties.

8 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).

9 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 3 percent can be added to that
Introduction

9

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.

10 In this work, 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.

11 Some of these characteristics include an elevated use of subject personal pronouns (Ottheguy & Zentella 2012 for Spanish in New York), collocations modeled on English (Ottheguy, Garcia & Fernández 1989 for Cubans in New York), increased use of the present progressive, loss of the subjunctive (Silva-Corvalán 2002 for Spanish in Los Angeles; Lynch 1999 for Spanish in Miami), and overgeneralization of iconicized phonological features in the second generation (Lamboy 2004).

12 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’s (1977: 18) use, except that he explicitly defines it with respect to written word-forms only.


Works cited


Introduction


