The Report Committee for Jacqueline Rae Larsen

Certifies that this is the approved version of the following report:

Social Stratification of Loanwords:

A Corpus-based Approach to Anglicisms in Argentina

APPROVED BY

SUPERVISING COMMITTEE:

Supervisor:
Almeida Jacqueline Toribio

Barbara Bullock

Dale Koike

Social Stratification of Loanwords:

A Corpus-based Approach to Anglicisms in Argentina

by

Jacqueline Rae Larsen, BA

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Arts

The University of Texas at Austin

May, 2014

Dedication

To my parents, sister and husband for their love and support

Acknowledgements

I would like to thank my advisor Jacqueline Toribio, who has been both an incredible source of guidance and an inspiration to me.

Abstract

Social Stratification of Loanwords:
A Corpus-based Approach to Anglicisms in Argentina

Jacqueline Rae Larsen, MA

The University of Texas at Austin, 2014

Supervisor: Almeida Jacqueline Toribio

With the aim of better understanding the social function of Anglicisms in Argentina, this study documents the distribution of Anglicisms across social groups and identifies the semantic domains they cover. In order to access Anglicisms currently used in Argentine Spanish, a 1.8 million-word corpus of newspaper articles from 2012 to 2013 was created and processed to extract English loanwords. This study presents a method for automatic loanword extraction that offers advantages over manual identification. The analyses conducted show that the English loanwords present in the Argentine newspaper corpus are not equally distributed across all newspapers but rather are highly concentrated in *La Nacion*, a prestigious newspaper targeted towards a highly educated upper-class segment of the population.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Languages in contact often result in a variety of linguistic outcomes; some contact features may go unnoticed but many attract both the attention and judgment of members of a speech community. As a highly conspicuous feature of language contact, loanwords are often imbued with social meaning in addition to their semantic one. For example, English loanwords in US Spanish are often devalued and stigmatized as Spanglish (Otheguy & Stern, 2011). Standing in direct contrast, speakers in Senegal often use loanwords from French, English, and Arabic as prestige markers (Ngom, 2003). Nielsen (2003) hints that in Argentina, as in Senegal, loanwords index positive attributes such as worldliness, status and education. Although speakers easily attest to the social value of loanwords, verifying their claims empirically is another matter. Empirical methods such as matched guise tasks and sociolinguistic interviews can shed light on language attitudes. However, a fundamental first step is documenting the distribution of loanword usage throughout the speech community. This paper intends to do just that, using a 1.8 million-word corpus of Argentine newspaper articles. The distribution will shed light on the social function of loanwords in this community, well known for its high prevalence of loanwords.

Argentina presents a particularly interesting site for the study of loanwords in that it has incorporated English lexical items without currently having much direct contact with English; it does not share a border with any English-speaking country and there is not a significant group of English-speaking immigrants. The majority of contact with English appears to be due to globalization and technology. While Bordelois (2005) cites Argentina as one of the countries with the heaviest use of Anglicisms in the Spanish-speaking world and alludes to their high regard, to date there are no empirical studies to

document these claims. I will lend support to these claims through the use of newspaper corpora, demonstrating that the use of loanwords in Argentine print media is subject to variation; specifically, loanwords are employed more heavily and cover a broader range of semantic domains in outlets directed at audiences of high socio-economic status in comparison to outlets with audiences of low socio-economic status. Additionally this paper seeks to present three methods for analyzing loanword distribution: frequency of tokens, frequency of unique lemmas, and frequency with respect to their semantic equivalents. The remainder of the paper proceeds as follows. First I will present a brief linguistic background on Argentina and a literature review of semantic equivalence. Secondly I will describe the data and methods used in this study with particular attention paid to the motivation for the corpus and automated methods used to identify Anglicisms. Subsequently I will present the results from each of the three distribution tests. Finally I will present a discussion on the implications of these findings and the additional work needed.

Chapter 2: Background

LINGUISTIC HISTORY AND CONTACT WITH ENGLISH IN ARGENTINA

From a cursory glance, Argentina does not appear to have a wealth of linguistic diversity. Since colonization in the 1500s, Spanish has become the dominant language and is the only official language of the country to date. However, there are 13 indigenous and 20 European and Asian languages spoken in Argentina (Messineo & Cúneo, 2006). Before the arrival of the Spanish conquistadors, over 35 indigenous languages were spoken in the region. However, brutal assimilation and militaristic campaigns in the 1800s displaced and eradicated many of the indigenous populations and, with them, a great source of linguistic diversity. European languages, such as French and Italian, accompanied the large waves of immigrants that arrived during the 19th and 20th centuries. However, due to the strong nationalistic culture they encountered in Argentina, most immigrants quickly adopted the Spanish language within one to two generations (Bordelois, 2011; Nielsen, 2003) . This legacy of assimilation is attested in the results from a national survey in 2006, which found that less than 10% of those surveyed spoke a language other than Spanish at home. In spite of the lack of heritage-language speakers, almost 50% of those surveyed spoke a second language. English was overwhelmingly the foreign language of choice; 85% of those who knew a foreign language spoke English, followed by 8.3% who spoke Portuguese. English was also seen to be on the rise as 60% of those over 50 spoke English compared with 76.2% of those between 35 and 49 years old and 88% of those 18 to 34 years old (Bein, n.d.). The high percentage of English-learners logically follows from the emphasis on English in both the educational system and the job market in Argentina (Gall & Hobby, 2007). Friedrich (2003)'s meta-analysis of advertisements for managerial positions revealed that at least 50% of the jobs required English and in many others it was deemed desirable.

This increase of English as a foreign language has been accompanied by what some would call an excessive increase of Anglicisms in Argentine Spanish. Bordelois (2011) has named the capital, Buenos Aires, the Latin American city that has imported the greatest number English terms "unnecessarily". The prevalence of loanwords is regularly commented on both positively and negatively. These conflicting attitudes range from outrage at the "invasion" of English to praise of these lexical items for portraying an image of youth and technical know-how (Melgarejo, 2011, 2012, 2013; Nudler, 2004; Pagano, 2013). Whatever the response, the use of loanwords is highly visible in the Argentine community.

SEMANTIC EQUIVALENCE

In measuring the distribution of loanwords, this study makes use of the notion of semantic equivalence. This section provides an overview of the major terminologies and their corresponding authors that have attempted to describe the distinction between loanwords with and without an existing semantic equivalent in the recipient language. The terms 'necessary' and 'luxury' have been employed in several influential studies including Ohmann (1961), Carstensen (1965), Tesch (1978), and Langer (1996) (from Onysko & Winter-Froemel, 2011). These terms refer to loanwords that either are necessary to borrow because no word exists in the recipient lexicon that refers to that concept or loanwords that merely substitute an existing word, thus a luxury item. However this terminology holds prescriptivist undertones and falsely implies that a loanword is necessary to introduce a new concept into the recipient language when other forms of semantic innovation exist, such as calques or neologisms, and that words with semantic equivalents are merely duplicates, which provide no semantic or pragmatic distinction from the recipient language equivalent (Onysko & Winter-Froemel, 2011).

This present study uses the terminology introduced by Bookless (1982) (from Morin 2006): unique and synonymic. Unique loans describe innovations in the recipient language that do not rival the native lexicon; for example *surfing* in Spanish. Synonymic loans add new lexical alternatives for concepts already existing in the lexicon; therefore their contribution is stylistic rather than referential, since speakers may choose between multiple words; for example the use of *handicap* or the native Spanish equivalent *discapacidad* (Morin 2006: 172). Morin notes that some synonymic loans are not merely an alternative word for an object (e.g., *sandwich* and *bocadillo*), but "refer to a modification of the concept that is considered significant, based on physical or functional features (2006:172)."

Myers-Scotton's (2002) terms, 'core' and 'cultural', express virtually the same distinction as the previous set; however she adds additional features to differentiate between the two. Cultural borrowings, she argues, refer to new objects or concepts, such as hardware, that enter abruptly through monolingual or bilingual speech. Core borrowings, on the other hand, offer duplicates of existing words and enter through a gradual process involving only bilingual speech. Myers-Scotton cites Mougen and Beniak (1991) who analyzed loans among bilinguals, concluding that those who speak both French and English in their private lives were more likely to use core borrowings, such as *le weekend* or *pickup*, than those bilinguals who compartmentalized their languages. The bilinguals who compartmentalized actively incorporated 'cultural' loans in place of code-switching. As the terminology suggests that core loanwords enter though bilingual speech, these terms seems poorly suited for a predominantly monolingual setting, such as Argentina.

In a corpus study on Anglicisms in German, Onysko and Winter-Froemel (2011) introduce new terminology with origins in rhetoric: catachrestic and non-catachrestic.

They link these terms to pragmatic functions using Levinson's theory of M- and I-implicatures (of manner and of informativeness). Onysko and Winter-Froemel allude to the fact that "these distinctions first and foremost rely on a differentiation of borrowings in terms of their formal markedness and their degree of integration in the RL" (p.1552), yet their work only explicitly addresses the former. Onysko and Winter-Froemel's findings show that "the categorization as catachrestic and non-catachrestic is generally not a strictly categorical either/or decision....the pragmatic functions of an Anglicism have to be conceived of as dynamic characteristics that can change over time as habits of language users change (1563)". This observation highlights a common thread and inherent problem common to these terminologies: the distinction is expressed in binary terms. However, due to the complexities of both word meaning and contact situations, it is likely that loanwords, like other phenomena—nonce borrowings, interference, code-switching, etc.—fall along a continuum (Winford 2010).

Chapter 3: Data and Methods

Corpora are powerful tools for studying language contact, as they provide real-world data not accessible via other traditional linguistic methodologies such as intuition or experimentation. In contact situations, only a few of the numerous possible outcomes are realized (Weinreich, 1953), thus making the need for authentic data all the more crucial. In addition to providing such data, corpora bring to light salient patterns that often remain unnoticed by speakers (Sinclair, 1998). Moreover, corpora offer access to a high number of tokens. This study draws on a newspaper corpus.

Working with newspapers as a source of linguistic data provides several benefits. First, by selecting national newspapers, the speech community of interest can be defined broadly and succinctly. Definitions of a speech community have been presented by numerous sociolinguists without yet reaching a clear consensus. However two major guidelines have been broadly accepted: shared code and shared communication (Patrick, 2008). The national audience and Spanish language of the chosen newspapers define the boundaries of the speech community as Argentina. Secondly, newspapers provide a uniform manner of dividing the data into social groups due to the differences in intended readerships of each newspaper. Lastly the fact that loanwords appear in print implies a certain level of assimilation and normalization across the speech community. It is important to note that newspapers represent only a specific type of language use because of the limited topics and formality required by this format. However, as this speech community is primarily in contact with English through the media, as opposed to direct

7

contact with English speakers, newspapers are an appropriate source through which to explore the use of loanwords.

CORPUS DATA

For the present study, I created a corpus of over 1.8 million words compiled from articles appearing on the websites of three major Argentine newspapers: *La Nación*, *Clarín*, and *Crónica*. The three newspapers were chosen to represent different segments of the population, as each newspaper is known to have a distinct target audience. *La Nación*, the second most read newspaper in the country, targets a high socioeconomic-status clientele. It has earned a politically conservative reputation due to its frequent critiques of the current leftist government (Boczkowski & de Santos, 2007). *Clarín* boasts the largest readership in the country and thus is oriented towards the middle class. It is said to have a centrist outlook, though, like *La Nación*, it is in conflict with the current government. In 2004, over 5.3 and 1.4 million unique users accessed Clarín.com and Lanación.com monthly, placing both websites as the top two general interest online newspapers in the country (Boczkowski & de Santos, 2007). *Crónica*, though not as widely read as *Clarín* and *La Nación*, is still a major newspaper distributed throughout the country and is targeted to a lower socio-economic readership. It is considered to be pro-Kirchner, the current Argentine president.

The corpus, collected sporadically from November 2012 to March 2013, comprises all of the articles appearing on the main page of each newspaper's website for each collection day. These articles include a broad range of topics, such as politics, social issues, business, sports and fashion. As seen in Figure 1, each newspaper represents around one third of the corpus: *La Nación* (32.9%), *Clarín*: (32.4%), *Crónica*: (34.6%). The corpus was annotated for both lemma and part-of-speech using the tool TreeTagger

(Schmid, 1995). As TreeTagger has a Spanish parameter and an English parameter, two separate versions of the annotated corpus were created: one annotated with the Spanish parameters and the other with English.
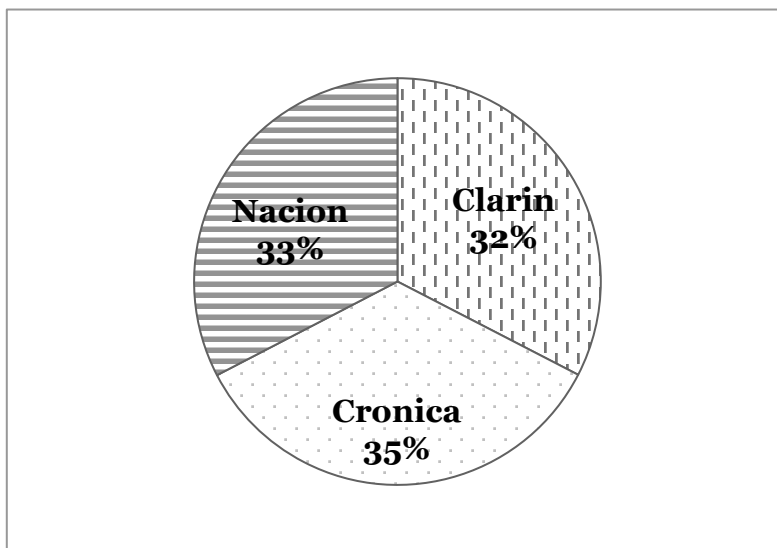

Figure 1:   Size Breakdown of the Corpus by Newspaper

METHODS

In order to identify English loanwords in the corpus, both automated and manual methods were implemented. The automated portion created a Potential Anglicisms list of all tokens that fit one of two sets of criteria: 1) identified as an unknown lemma and not a proper noun in Spanish-annotated corpus and a known lemma and also not labeled as a proper noun in the English-annotated corpus or 2) identified as a foreign word in Spanish, with the part-of-speech tag PE (*palabra extranjera* 'foreign word'), and a known lemma in English. As seen in sample tokens with their annotation in Table 1, the loanword *catering* and the Spanish word *pomelo* fit criteria 1 and rating criteria 2 so all three were added to the Potential Anglicisms list. The band Kiss, labeled as a proper noun, *bar*,

9

identified with a Spanish lemma, and *sufear*, not identified as either English or Spanish were not added to the Potential Anglicisms list. The manual task involved individually inspecting each token in the list to remove any misidentified Anglicisms. As in seen in Table 1, *pomelo* was misidentified as an Anglicism and manually removed. The majority of the misidentifications stem from the fact that the English parameter of TreeTagger is much more robust than the Spanish one.

This method does not capture all Anglicisms; it excludes any loanword with a known lemma in the Spanish parameter such as *bar*, and it excludes any loanword modified for English such that it is not recognized by the English parameter, as is the case for *surfear*. In spite of these limitations, the method boasts two advantages: it allows one to process a corpus that would be too large to process manually and it avoids human error.

Token	Spanish Part-of-Speech Tag	Spanish Lemma	English Part-of-Speech	English POS
catering	NC	<unknown>	NN	catering
rating	PE	rating	NN	rating
pomelo	NC	<unknown>	NN	pomelo
Kiss	NP	<unknown>	NP	Kiss
bar	NN	bar	NN	bar
surfear	VLinf	<unknown>	NN	<unknown>

Table 1:   Sample Annotated Tokens from the Corpus

Chapter 4: Results

DISTRIBUTION OF LOANWORDS

Frequency of Tokens

Processing the 1,876,683-word corpus revealed a total of 3,036 English loanwords accounting for just .16% of the corpus. This number is relatively low compared to other written media corpora, such as the German *Der Spiegel 2000* corpus in which Anglicisms accounted for 1.11% of the corpus (Onysko & Winter-Froemel, 2011). The relatively low percentage could be due to the method of extracting loanwords. The method does not capture any loanword not recognized as English, for example *surfear* or any word recognized as Spanish by TreeTagger, for example *bar,* and thus underestimates count of loanwords present.

The identified Anglicisms were far from evenly distributed across three newspapers. Figure 2 below shows the total counts of Anglicisms per newspaper. Of the 3,036 Anglicisms in the corpus, more than half (1,821) was found in the newspaper *La Nación,* which represents the more upper class segment of the Argentine community. *Clarín* and *Crónica* were much more closely matched, having 697 and 518 respectively. Of the 1,821 tokens, 131 were hapax legomena, appearing only once in the corpus.
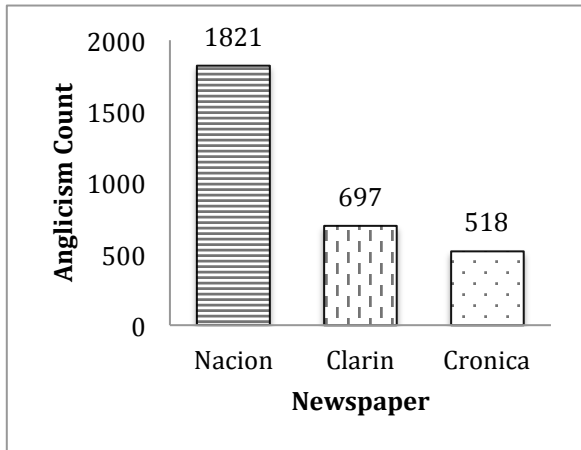
Figure 2:   Total Counts of Anglicisms per Newspaper

Figure 3 shows the percentage of Anglicisms relative to the total word count for each newspaper. As shown, *La Nación* has more than double the percentage of loanwords as *Crónica* and *Clarín*.
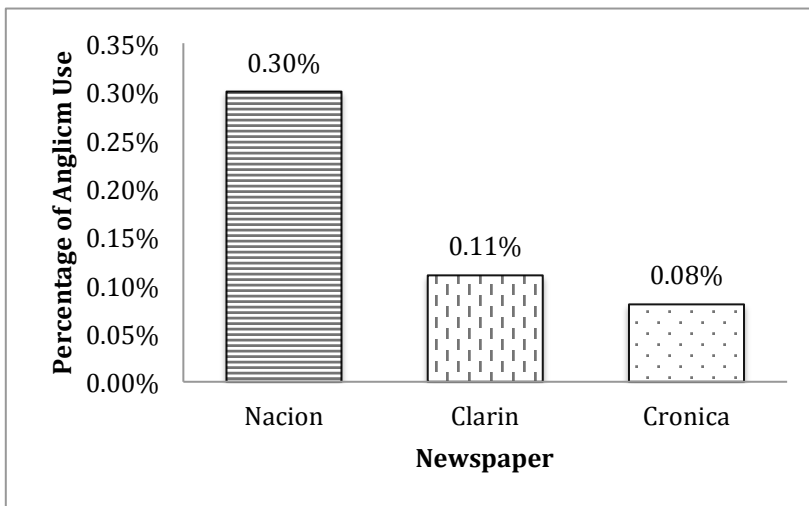

Figure 3:   Percentage of Anglicisms Relative to the Total Word Count per Newspaper

Frequency of Lemmas

While the token counts represent the overall frequency of loanwords, they do not say anything about the variety of loanwords found in each corpus. This aspect of distribution is captured in this second test, which measures the number of *unique* lemmas

found in each newspaper. Thus rather than representing pure frequency, it presents the variety of loanwords. The same pattern emerges here as in the token count: *La Nación* presents more than double the variety of loanwords as in *Clarín* and *Crónica* as seen in Figure 4. The top most frequent loanwords remained mostly constant among the three newspapers, including the loanwords *show*, *web*, *mail* and *online*. Many are unsurprisingly related to technology. Yet 181 Anglicisms in *La Nación* were absent from both *Crónica* and *Clarín*. These 181 loanwords covered a broad range of semantic domains, including cuisine (*toppings, brownie, cheesecake*), pop culture (*hipster, freak, mainstream*), fashion (*style, outfit, strapless),* business (*storyboard, telemarketer, marketplace),* and technology *(techie, banner, gaming*).

Figure 4:   Counts of Lemmas in Each Newspaper
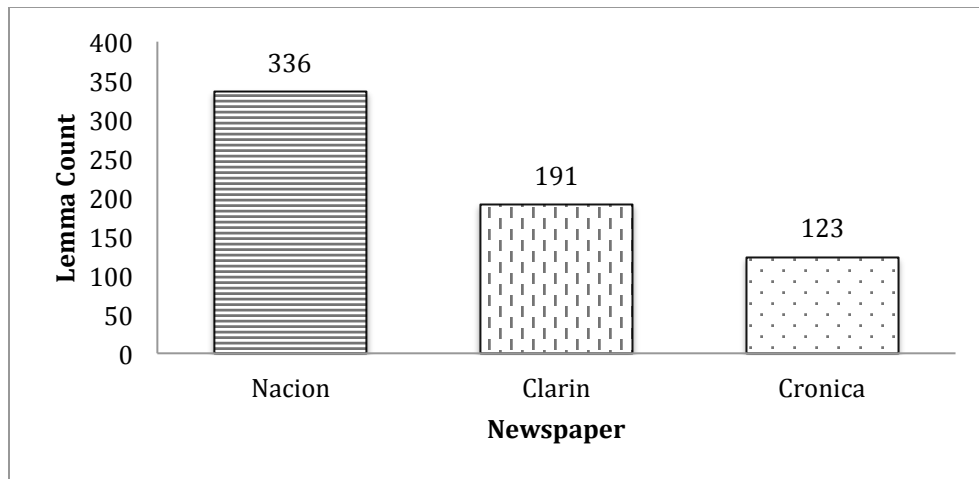
SEMANTIC EQUIVALENCE

The final measurement of loanword distribution utilizes the concept of semantic equivalence. Semantic equivalence expresses the distinction between loanwords that introduce new concepts from the source language to the recipient language e.g., *espresso* and *zeitgeist* in English, and those that provide an alternative to lemmas already existing

13

in the recipient language, e.g., *siesta*, which is more or less synonymous with *nap*. Classifying loanwords by this distinction and identifying the recipient-language equivalent introduces the dimension of choice. In order to understand when and why a speaker will adopt a loanword, it is necessary to examine their moment of choice. When a speaker has access to both a loanword and an existing native equivalent, they are forced to decide between the two.

Method for Determination of Semantic Equivalence

To attempt a classification based on semantic equivalence, the fifty most frequent Anglicisms were selected from the corpus, in order to address the research questions: When presented with the choice between a loanword and synonym native to the recipient language, do speakers across the three social groups behave alike? Secondarily, is the notion of semantic equivalence a binary distinction? What is the best way to classify that distinction? Due to the fine-grain distinction of word sense, it is likely that a binary distinction will not be fine-tuned enough to best capture this classification.

To classify the fifty Anglicisms as unique or synonymic, a combination of three tools was implemented: an English-Spanish dictionary, collocations from the corpus, and the judgments of a native speaker of Argentine Spanish. Each Anglicism was first looked up in the English-Spanish dictionary from wordreference.com to identify all possible equivalents. For example, the word *hit* returns the entry *pegar*, *chocar*, *éxito*. The collocations of the corpus showed the Anglicism *hit* is only used as a noun meaning *success*, thus substitutable with the Spanish noun *éxito* but not the verb *pegar* or *chocar*. Lastly the interchangeability of the Anglicism and possible semantic equivalent was evaluated by a member of the speech community.

Observations and Results

Upon classifying all fifty Anglicisms, five categories emerged: 1. unique loans (16%) 2. loans with one exact semantic equivalent in the recipient language (18%) 3. loans with multiple semantic equivalents (18%), 4. loans with limited semantic scope with respect to their semantic equivalent (34%), and 5. loans with close but no exact semantic equivalent (14%). All loanwords, their categorization and semantic equivalent, if applicable, are presented in Table 2.

Number	Anglicism	Category	Equivalent
1	box	1. unique loan	NA
2	break	4. limited	descanso
3	casting	2. exact equivalent	reparto
4	catering	3. multiple equivalents	servicio de comidas
5	celebrity	2. exact equivalent	celebridad
6	chip	1. unique loan	NA
7	clip	5. close equivalent	broche
8	coach	4. limited	entrenador
9	cookie	2. exact equivalent	galletita
10	country	5. close equivalent	barrio privado
11	court	4. limited	cancha
12	cover	4. limited	versión
13	crack	3. multiple equivalents	talentoso
14	default	3. multiple equivalents	falta de pago
15	fan	4. limited	fanático
16	film	2. exact equivalent	película
17	freezer	2. exact equivalent	congelador
18	hit	4. limited	éxito
19	holdout	1. unique loan	NA
20	hot	3. multiple equivalents	sensual
21	living	3. multiple equivalents	sala de estar
22	look	5. close equivalent	aparencia
23	mail	2. exact equivalent	correo electronico
24	marketing	5. close equivalent	mercadeo

25	online	4. limited	en la red
26	performance	5. close equivalent	actuación
27	pole	4. limited	clasificación
28	ranking	4. limited	clasificación
29	reality	4. limited	realidad
30	relax	2. exact equivalent	relajación
31	sexy	3. multiple equivalents	sensual
32	shopping	2. exact equivalent	centro comercial
33	short	2. exact equivalent	pantalones cortos
34	show	3. multiple equivalents	espectáculo
35	side	4. limited	fuera de juego
36	sketch	4. limited	escena
37	soft	1. unique loan	NA
38	sponsor	3. multiple equivalents	patrocinador
39	staff	4. limited	empleados
40	tablet	1. unique loan	NA
41	thriller	5. close equivalent	de suspenso
42	tie	1. unique loan	NA
43	tip	4. limited	consejo
44	top	3. multiple equivalents	prestigioso
45	topping	4. limited	condimento
46	track	4. limited	pista
47	try	1. unique loan	NA
48	vintage	5. close equivalent	antiguo
49	web	4. limited	red
50	whisky	1. unique loan	NA

Table 2   Loanwords, Categorization and Semantic Equivalents

The first two categories follow the traditional semantic equivalent distinction: those with no semantic equivalent, all of which introduced new concepts into the recipient language, for example *hardware*, *whiskey,* and *box* and those with one perfectly matched semantic equivalent; for example *freezer*: *congelador*, *shopping*: *centro comercial*, *cookie: galletita*. These loanwords offer a stylistic variation yet no change in

referential meaning, as suggested by Bookless(1982). Unique loans and synonymic loans each account for just 16% and 18% respectively of the fifty loanwords.

Well over the majority of loanwords, 66%, fit into the last three categories where identification of an equivalent proves more complicated. One frequent occurrence was the presence of multiple semantic equivalents. For example the Anglicism *default* could be replaced with any of the following ngrams: *cesación de pago, falta de pago*, and *incumplimiento*. All three alternatives were validated by a native speaker and all occurred in the corpus. Further complicating the distinction is that fact that *incumplimiento* has a broader scope than the loanword *default*. This complication is what characterizes the fourth and most represented category at 34%: the loanword is more limited in scope than its semantic equivalent, i.e. the semantic equivalent could always substitute the loanword but not vice versa. For example, *cover* is easily replaced by *versión* though *versión* can be used in contexts that *cover* cannot, for example *debía dar su versión de los hechos*. The fifth category includes loanwords whose precise meaning or connotation is not captured by one word though there exist close possible substitutes or a phrase that can be constructed for example: *thriller: de suspenso vintage: antiguo, viejo, anciano*. This category is in particular need of native speaker judgments as the differences in meaning may be very salient to some and perhaps non-existent to others.

These challenges suggest that a process of specialization is under way for those loans within categories four and five. As loans enter an existing lexicon there will undoubtedly be confusion and change, as the new words adjust their meaning relative to the words already present in the lexicon. This period is followed by a process of specialization as seen with the English word *deer,* which originally referred to any four legged animal before undergoing specialization to refer to the specific four legged animal we now know as deer (Weinreich, 1953). These ambiguities highlight the challenging

nature of determining semantic equivalence and thus provide evidence that the mere intuition of one or two native speakers is not enough to establish such a category. A methodology must be established. This may involve native speaker judgments from a large sample of participants combined with a concatenation study.

The pairings of semantic equivalent and loanwords, as seen in Table 2, were used for the third measurement of distribution, which aims to capture speakers' lexical choice. The counts of both the loanwords and their semantic equivalents were gathered from the corpus to calculate the percentage of loanword word use out of the total number of instances that the concept was referenced, *i.e*. the sum of loanword count / (the sum of loanword count + semantic equivalent count) per newspaper. Thus a high percentage indicates a propensity to use loanwords over the recipient language equivalent. The results from this measure, as seen in Figure 5, show the same patterns as found in the previous two figures. *La Nación* uses loanwords nearly 45% of the time when there exists a semantic equivalent in Spanish. This percentage is over double that of *Clarín*, which uses loanwords only 23% and *Crónica* only 19%.
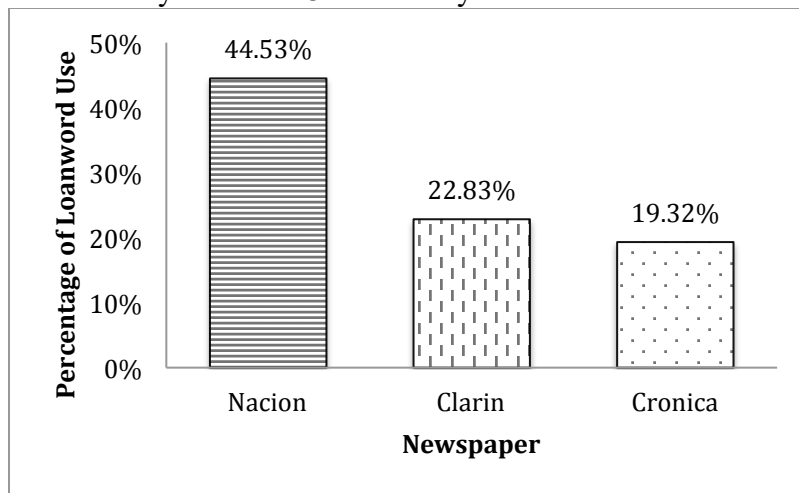

Figure 5   Percentage of Use of Loanwords out of the Total Number of Concept References

Chapter 5: Discussion and Conclusions

The analyses conducted here have shown that the English loanwords present in the Argentine newspaper corpus are not equally distributed but rather are highly concentrated in *La Nacion*, a prestigious newspaper targeted towards a highly educated upper-class segment of the population. The middle and lower status newspapers trail behind in all three measurements of distribution. This finding suggests that loanwords act as prestige markers in the community, as has been hinted in Nielsen (2003) and mirror similar findings such as those in Ngom (2003) for loanwords in Senegal. Loanwords as prestige markers in Argentina may be a logical consequence of the mode of contact; the upper socio-economic status group has greater access to outlets where loanwords seem to emerge: the media, Internet, and second language education. Therefore loanwords, particularly less common ones, index a speaker as having perhaps studied English, traveled abroad or been in contact with media in English. This stands in stark contrast to many English and Spanish bilingual communities within the US where Spanish speakers often have a more direct means of accessing English speakers given the surrounding area is English-speaking, perhaps removing its allure and uniqueness. Thus, type of contact, direct or indirect, may be an important variable that can contribute to our understanding of the social meaning and prestige of contact features. To test this variable, future research comparing sites of direct and indirect language contact are needed.

Secondly this paper presented three methods of quantifying loanword distribution. The similarities of all three results validate their ability to represent distribution, though each tests a different aspect of distribution. Looking at the total count of loanwords measures the overall frequency, unique lemmas the variation, and semantic equivalence the lexical choice faced by the speaker. Testing with respect to semantic equivalence

most directly quantifies whether a given speaker has a greater or lesser propensity to use loanwords in that it compares only cases where loanwords are possible. This method addresses some of the challenges in using lexical items as a linguistic variable. Structural variables such as allophones and grammatical features are easier to test because one can count the total possible locations where they could be in a straightforward manner. Lexical items present the challenge that certain topics may lend themselves more or less to the use of loanwords. Using semantic equivalence provides a solution in comparing the lexical item count to the total count of all references to that concept.

However the notion of semantic equivalence is not without its issues and would greatly benefit from further investigation. Up until this point the presence or absence of a semantic equivalence has been categorized as a binary distinction. The presence of multiple equivalents and close equivalents has not been properly addressed. Additionally, the determination of a semantic equivalent is too fine-grained a decision to be left solely to the discretion of the researcher. It warrants a method itself that needs to be researched and tested. Only with this type of tool will semantic equivalence become a viable variable, though it promises to offer a great deal in understanding the adoption process of loanwords.

References

Bein, R. (n.d.). La situación de las lenguas extranjeras en la Argentina. Retrieved from http://linguasur.org.ar/panel/archivos/f9227ef50db3de732e1d3f897de85aa8Bein%20lenguas%20extranjeras.pdf

Boczkowski, Pablo J., and Martin de Santos. 2007. "When More Media Equals Less News: Patterns of Content Homogenization in Argentina's Leading Print and Online Newspapers." *Political Communication* 24 (2): 167–180.

Bordelois, Ivonne. 2011. *El país que nos habla*. Sudamericana.

Friedrich, Patricia. 2003. "English in Argentina: Attitudes of MBA Students." *World Englishes* 22 (2): 173–184.

Gall, Timothy L., and Jeneen M. Hobby, eds. 2007. "Argentina." In *Worldmark Encyclopedia of the Nations*, 3:11–34. 12th ed. Americas. Detroit.

Melgarejo, Graciela. 2011. "Préstamos Legítimos y Préstamos Innecesarios." *La Nación*, April 11. **http://www.lanacion.com.ar/1364531-prestamos-legitimos-y-prestamos-innecesarios**.

———. 2012. "Adonde Vayas, Acuérdate de Que Tienes El Mismo Código." *La Nación*, November 5.

———. 2013. "Anglicismos, Rosas y Una Casa En Recuperación." *La Nación*, January 7. http://www.lanacion.com.ar/1543435-Anglicismos-rosas-y-una-casa-en-recuperacion.

Messineo, C., & Cúneo, P. (2006). Las lenguas indígenas de la Argentina: situación actual e investigaciones. In *Third International Workshop on (Semi) Numerical Techniques in Polynomial Equation Solving, in Honor of Joos Heintz's 60th. Buenos Aires*. Retrieved from http://xeitonovo.org.ar/ponencias/CLM07_CMessineoPCuneo.pdf

Morin, Regina. 2006. "Evidence in the Spanish Language Press of Linguistic Borrowings of Computer and Internet-related Terms." *Spanish in Context* 3 (2): 161–179.

Myers-Scotton, Carol. 2002. *Contact Linguistics: Bilingual Encounters and Grammatical Outcomes*. 1St Edition. Oxford University Press, USA.

Ngom, Fallou. 2003. "The Social Status of Arabic, French, and English in the Senegalese Speech Community." *Language Variation and Change* 15 (03): 351–368.

Nielsen, Paul Maersk. 2003. "English in Argentina: A Sociolinguistic Profile." *World Englishes* 22 (2): 199–209.

Nudler, Julio. 2004. "Así Es Muy Difícil." *Página 12*, May 9.

Onysko, Alexander, and Esme Winter-Froemel. 2011. "Necessary Loans – Luxury Loans? Exploring the Pragmatic Dimension of Borrowing." *Journal of Pragmatics* 43 (6) (May): 1550–1567.

Otheguy, Ricardo, and Nancy Stern. 2011. "On So-called Spanglish." *International Journal of Bilingualism* 15 (1): 85–100.

Pagano, María. 2013. "Las Frases En Inglés, Un Guiño de Informalidad." *La Nación*, January 19. http://www.lanacion.com.ar/1547183-las-frases-en-ingles-un-guino-de-informalidad.

Patrick, Peter L. 2008. "The Speech Community." In *The Handbook of Language Variation and Change*, edited by J. K. Chambers, Peter Trudgill, and Natalie Schilling-Estes, 573–597. Blackwell Publishing Ltd.

Schmid, H. 1995. "Improvements in Part-of-speech Tagging with an Application to German." In *In Proceedings of the ACL SIGDAT-Workshop*.

Sinclair, John. 1998. "Corpus Evidence in Language Description." *Teaching and Language Corpora*.

Weinreich, Uriel. 1953. *Languages in Contact, Findings and Problems*. Publications of the Linguistic Circle of New York no. 1. New York: Linguistic Circle of New York.

Winford, Donald. 2010. "Contact and Borrowing." In *The Handbook of Language Contact*, edited by Raymond Hickey Chair, 170–187. Wiley-Blackwell.