Dialectologia 27 *(2021), 71-95.*
ISSN: 2013-2247

Received 9 October 2019.
Accepted 17 April 2020.

A CORPUS ANALYSIS OF SOME USAGE DIFFERENCES AMONG SPANISH-SPEAKING COUNTRIES[1]

David Ellingson EDDINGTON
Brigham Young University *
eddington@byu.edu

Abstract

The *Corpus del Español / Web Dialects* was used to study usage differences among 21 Spanish-speaking countries. This included variation in verbal morphology: the use of *-ra* and *-se* forms of the past subjunctive (e.g. *comiera~comiese*), and final *-s* on 2nd person singular preterites (e.g. *pusistes*). The appearance of *habían* with plural arguments (e.g. *habían muchas riñas)*, the use of the present perfect to express recent past events (e.g. *Esta mañana lo hemos visto*), the use of present subjunctive in embedded clauses following matrix clauses with past tense subjunctive triggers (e.g. *Quería que vengas y no viniste*), and rates of clitic climbing (e.g. *Lo iban a ver~iban a verlo*) were also examined. Variation in nine vocabulary was also studied (e.g. *elevador~ascensor*), gender variation in seven words (e.g. *el~la margen*), and the use of *ser* or *estar* with *consciente*.

Keywords

corpus linguistics, Spanish dialects, morphology, vocabulary

1 I express my thanks to Lynn Williams and Earl Brown for their input on this paper.
* Department of Linguistics, Brigham Young University, 4064 JFSB, Provo, Utah 84602.

UN ANÁLISIS CORPUS DE ALGUNAS DIFERENCIAS DE USO ENTRE PAÍSES DE HABLA ESPAÑOLA

Resumen

El *Corpus del Español / Web Dialects* se ha utilizado para estudiar las diferencias de uso entre 21 países de habla hispana. Esto ha supuesto la variación en la morfología verbal: el uso de las formas *-ra* y *-se* del imperfecto de subjuntivo (por ejemplo, *comiera ~ comiese*), y los finales en *-s* de los pretéritos de la segunda persona del singular (por ejemplo, *pusistes*). También se ha examinado la aparición de *habían* con argumentos en plural (p. ej., *habían muchas riñas*), el uso del pretérito perfecto para expresar eventos pasados recientes (p. ej., *Esta mañana lo hemos visto*), el uso del presente de subjuntivo en oraciones subordinadas después de oraciones matriciales con formas de subjuntivo en pasado (por ejemplo, *Quería que vengas y no viniste*) y los cambios de posición de clíticos (por ejemplo, *Lo iban a ver ~ Iban a verlo*). Se ha estudiado a su vez la variación existente en nueve términos (por ejemplo, *elevador ~ ascensor*), la variación de género en siete palabras (por ejemplo, *el ~ la margen*) y el uso de *ser* o *estar* seguidos de *consciente*.

Palabras clave

corpus linguistics, dialectos del español, morfología, léxico

1. Introduction

According to *Ethnologue* (2019), Spanish has 460 million native speakers, and is surpassed in the world only by Chinese. Twenty countries have Spanish as their national language and, if the estimates are correct, there are 52 million Spanish speakers in the United States, which would make the US the largest Spanish-speaking country in the world after Mexico (Perez 2015). For this reason the US needs to be included among the Hispanophone countries. Given the demographic size and geographical extension of the Spanish language, it is not surprising that there is a wide degree of variation among the Spanish spoken in different regions. Nevertheless, it is safe to affirm that a speaker from any Spanish-speaking country would not have difficulties understanding the standard broadcast language of any other country. Much of the variation among Spanish-speaking countries has been summarized qualitatively in works such as Lipski (1994) and Lope Blanch (1968). However, precise quantitative data about usage differences across the entire Hispanophone world are scarce.

The advent of computer-readable corpora in Spanish such as CREA (Real Academia Española n.d.) and the Corpus del Español (Davies 2016) has given researchers valuable tools to use in answering questions about vocabulary and usage quantitatively. Recently, Davies added the two billion word *Corpus del español* / Web dialects to the Corpus del Español. This corpus is novel because it contains data categorized according to which of the 21 Spanish-speaking countries they come from, which makes it possible to make controlled comparisons by country. Therefore, the purpose of the present study is to provide a descriptive comparison of the frequency of a number of different grammatical usages and vocabulary items in the Spanish-speaking world and, when possible, to compare the outcome with that of extant studies.

2. The corpus

The particular value of the Web Dialects corpus is twofold. First, it is comprised of data which have been categorized according to which of 21 Hispanophone countries the author is from. Secondly, 60% of the data comes from blogs, meaning informal registers are covered quite well, which is crucial because highly edited materials from printed sources are less likely to demonstrate the kinds of regional differences explored in the present paper. The disadvantage of the corpus is that the city or province of the speakers in the corpus is not recoverable, only their country of origin. In other words, it essentially ignores the existence of multiple dialect areas in a single country. Additionally, information about the individual authors such as their age, gender, social class, etc. is not available. In spite of these drawbacks the corpus proves to be an extremely useful tool to investigate usage differences. The corpus is not balanced as far as how many words from each country it contains. The number of million words in each country, along with the two-letter acronym each country is referred to in the graphs, is as follows: 169.4 Argentina AR, 39.3 Bolivia BO, 66.2 Chile CL, 166.4 Colombia CO, 29.5 Costa Rica CR, 63.2 Cuba CU, 33.6 Dominican Republic DO, 52.3 Ecuador EC, 426.5 Spain ES, 54.2 Guatemala GT, 35.1 Honduras HN, 245.9 Mexico MX, 32.3 Nicaragua NI, 22.2 Panama PA, 107.2 Peru

PE, 32.1 Puerto Rico PR, 29.7 Paraguay PY, 36.4 El Salvador SV, 166.0 US,38.7 Uruguay UY, 98.1 Venezuela VE.

3. Preterite versus present perfect

Three tenses are used to express past actions in Spanish depending on aspect: imperfect, preterite, and present perfect. Research has shown that in Peninsular Spanish the preterite is being encroached on by the present perfect (Schwenter & Cacoullos 2008). In this regard it lags behind other Romance languages such as French and Italian that have already ousted simple past tense in favor of the perfect, at least in the spoken vernacular. In Latin American varieties, on the other hand, the preterite dominates, although in some American varieties the present perfect has taken over certain functions of the preterite (Howe & Schwenter 2003).

One instance in which the preterite and present perfect vary is when expressing recently occurring actions and past actions that have relevance for the present moment. The variation occurs in sentences such as: *Esta mañana comí / he comido huevos fritos* 'This morning I ate fried eggs'. Present relevance is often marked with adverbials such as *recientemente* and *esta mañana.* Therefore, the corpus was searched for 11 adverbials of this kind: *siempre, ya, recientemente, esta mañana, esta noche, todavía, nunca, recién, últimamente, hace poco, esta tarde.* The number of times a verb in the preterite or the present perfect either preceded or followed these adverbials was tallied. The proportion of present perfect tenses was calculated for each country (Figure 1). The findings are consistent with what has been observed previously. Spain leads all other countries in its preference for the present perfect over the preterite at a rate of .64. All Latin American countries fall below Spain, but Uruguay (.34), Paraguay (.30), and Argentina (.29) fall on the opposite side of the scale and much prefer the preterite over the present perfect in these cases.

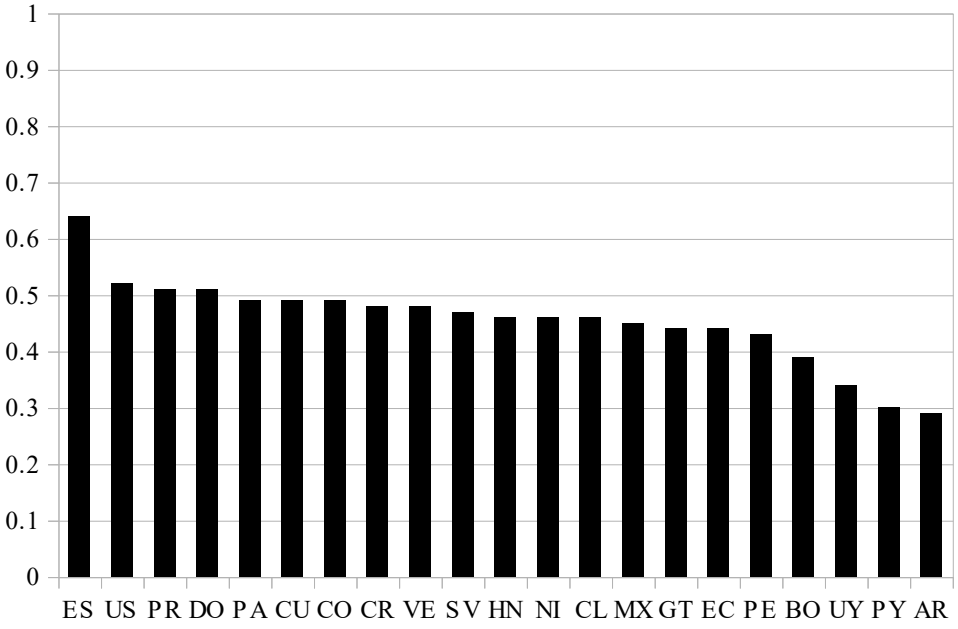

Figure 1. Proportion of present perfect tense over preterite tense with adverbials expressing recent events

4. Past or present subjunctive in embedded clauses following a past tense subjunctive trigger

In general, when a matrix clause contains a trigger for the subjunctive in the embedded clause, the tense of the matrix clause determines that of the embedded clause (e.g. *Quieren que asistamos al bautizo. Querían que asistiéramos al bautizo* 'They want / wanted us to attend the christening'). There are, however, instances where the sequence of tense may be violated, which have been studied in detail (Guajardo 2018, Laca 2010, Suñer & Padilla-Rivera 1987, Quer 2000). Previous research indicates that in countries such as Bolivia, Paraguay, and Argentina present subjunctive in the embedded clause is much more likely to be found even when the matrix clause is in the past tense and there is no reason for violating the sequence of tenses (Guajardo 2017; Sessarego 2008, 2010).

In order to investigate this in more depth, sentences with matrix clauses containing 17 verbs that trigger the subjunctive in the embedded clause were obtained. The matrix clauses contained a verb in the imperfect or preterite tense (i.e. *querer, esperar, dejar,*

hacer, impedir, dudar, permitir, temer, conseguir, sugerir, lograr, desear, pedir, gustar, recomendar, ordenar, mandar). The number of times the verb in the embedded clause appeared in the present or past subjunctive was tallied. From that, the proportion of present tense was calculated for the verb in the embedded clause (Figure 2). The results affirm what has been observed in previous studies; the countries with the highest use of present subjective in embedded clauses preceded by past tense matrix clauses are Bolivia (.57), Ecuador (.56), Paraguay (.42), Argentina (.38), and Peru (.37). The remainder of Spanish-speaking countries stand in sharp contrast to these five countries in that present tense is not common in this context. Guajardo (2017) suggests that the high use of present subjunctive in Bolivia, Ecuador Paraguay, Argentina, and Peru is the result of large indigenous populations or large populations of European immigrants in those countries who acquired Spanish as a second language, suggesting that it was influenced by L2 Spanish.

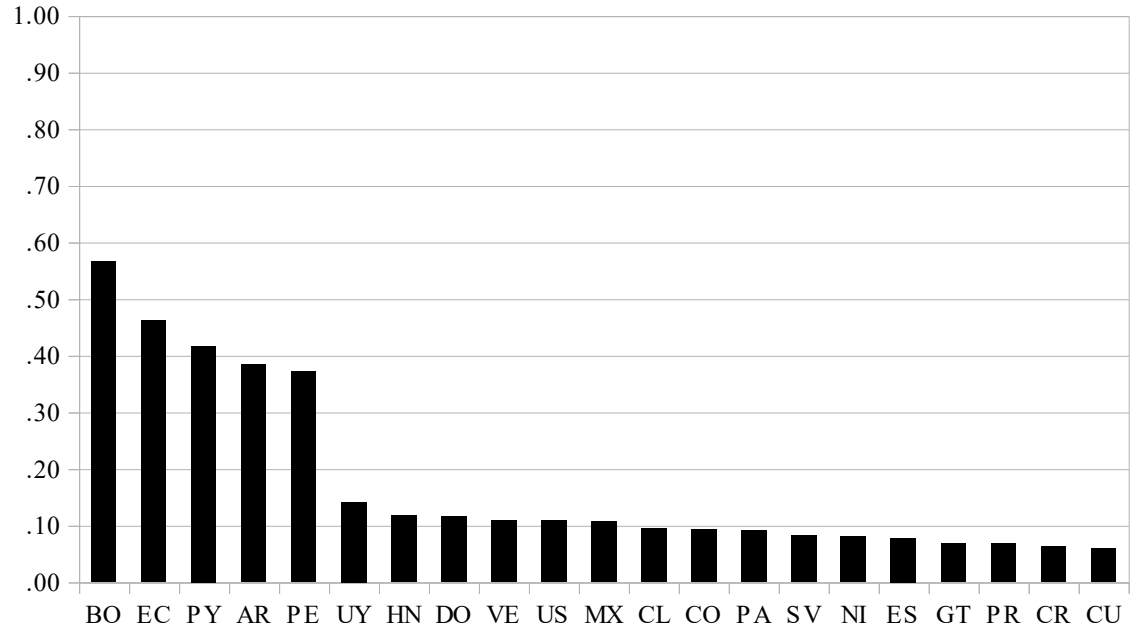

Figure 2. Proportion of present subjunctive in embedded clause after a past tense subjunctive trigger in matrix clause

5. Past subjunctive forms ending in *-ra* and *-se*

Due to historical evolution from different Latin tenses, Contemporary Spanish has two past subjunctive morphemes: *-ra* and *-se*. The *-ra* forms were originally pluperfects, but once they lost this meaning and became past subjunctives, the alternation was born (Hanna 2016). Some scholars (Guzmán Naranjo 2017, Rosemeyer & Schwenter 2019) have shown that there are morphosyntactic, usage, and discourse factors that govern the use of each one. Others have studied their geographical distribution. For instance, De Mello (1993) compared the use of the two forms in the educated speech of a number of Spanish-speaking cities. He observed the highest use of *-se* forms in Madrid, Seville, and San Juan, Puerto Rico, and the lowest use in the other Latin American cities he studied.

The Web Dialects corpus affirms the dominance of *-ra* forms across all Spanish-speaking countries, as is apparent in Figure 3. In Latin America, the *-se* forms occur only in about 9% to 18% of the cases, while in Spain they are more common (.21). De Mello's findings that the *-se* forms appear more in Spain and Puerto Rico are corroborated. However, its higher use in Paraguay and the US is something that has not previously been observed.

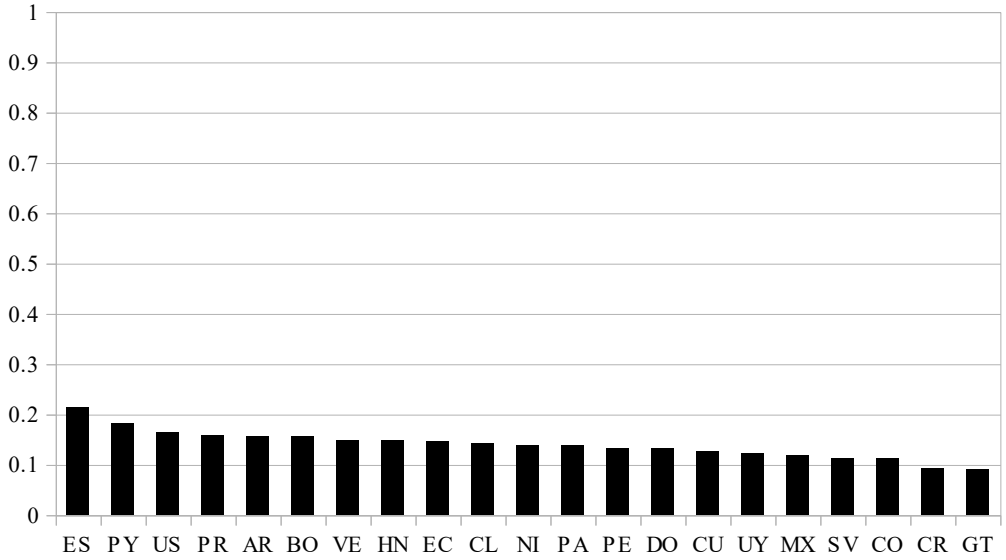

Figure 3. Proportion of past subjunctive forms ending in *-se*

6. Number agreement with *haber*

In the present tense, the existential use of *haber* has a single inflection, *hay*, which has no plural counterpart. In the imperfect, however, *había* alternates with *habían,* although the latter is considered incorrect in prescriptive grammars (Real Academia Española 2005: 330-331). The variation may be explained in terms of grammatical interpretation. Speakers who use *había* before both singular and plural arguments appear to interpret the subject of *había* as an unexpressed element in the language. On the other hand, when *habían* is used the speakers interpret the plural argument following this verb as its grammatical subject.

Regional differences in the use of *habían* before plurals have not previously been explored. Therefore, in order to determine how this usage varied by country the corpus was searched for plural nouns and adjectives appearing after *había* and *habían,* as well as following 15 other plural modifiers (*muchos*, *muchas, unos, unas, dos, tres, varios, varias, algunos, algunas, pocos, pocas, bastantes, demasiados, demasiadas*). For example, a great deal of variation is observed among countries in the use of *habían* before plurals (Figure 4). It ranged from 5% in Argentina to 31% in Guatemala.

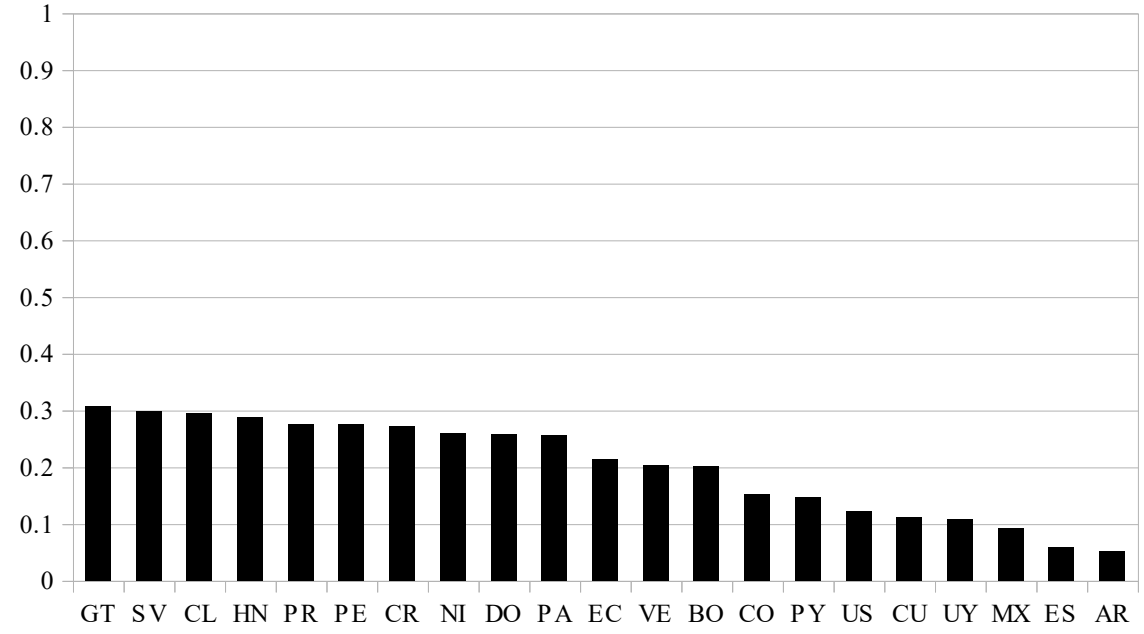

Figure 4. Proportion of plural *habían* before plural forms

Two macro-regional groupings of these data are evident; the use of *habían* with plurals is more common in Central America, and less common in Southern Cone countries. However, there may be something more basic that underlies it since this particular usage is highly stigmatized and associated with lower levels of education. It is possible that the use of *habían* with plurals is more common in countries with lower literacy rates. In order to determine this, a correlation between literacy rates and the use of *habían* with plurals was carried out. However, the data from the US were not included since it is unclear what the literacy rate of Spanish-speakers in the US might be. In any event, a negative correlation resulted (r (18 = -.56, p = .0099), which indicates that higher use of *habían* with plurals is related to lower literacy.

7. Non-standard 2nd person singular preterite forms with final *-s*

Preterite forms ending in *-s* such as *llegastes* and *vinistes* are common enough in the Spanish-speaking world that they receive abundant condemnation. They are most likely due to an analogical extension from all other 2nd person inflections that end in *-s* to the sole inflection that does not—the preterite. While the use of such verbal forms has been associated with social variables such as age, gender, and social class (Elizaincín & Díaz 1981) it has not been examined by country.

In order to do this the corpus was searched for frequently occurring 2nd person singular preterite forms with final *-s*. A total of 130 forms were selected (see Appendix) and their frequency along with the frequency of the standard forms without final *-s* was obtained. As Figure 5 indicates, Nicaragua stands apart from the other countries in the use of these non-standard inflections. In this country a fourth (.25) of all of these second person singular preterites appeared with a final *-s.* In the remainder of the countries the usage ranges between .15 and .08. Literacy rates in the 20 countries where Spanish is an official language were again correlated with this non-standard usage. The resulting negative correlation (r (18 = -.65, p = .0019) indicates that lower literacy is related to higher usage of final *-s.*

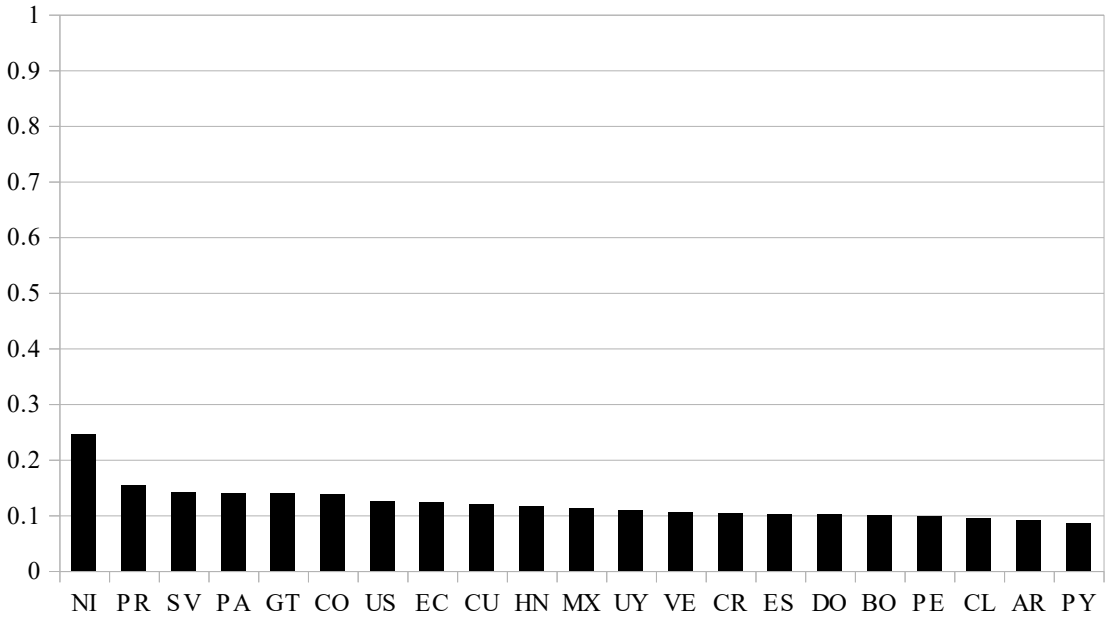

Figure 5. Proportion of 2nd person singular preterite inflections ending in *-s*

8. Clitic climbing

One area of syntactic variation has to do with the placement of clitic pronouns. Diachronically, the clitics in many instances have climbed from other positions in the phrase to the preverbal position (Davies 1995), hence the term clitic climbing. In Contemporary Spanish, clitics are in many cases free either to precede an inflected matrix verb or follow the infinitive: *Lo iban a visitar / Iban a visitarlo* 'they were going to visit him'. This variation has been shown to be governed in part by sociolinguistic factors (Torres-Cacoullos 1999).

The question is how much variation exists among countries. To this end, searches were done for the clitics *me, te, nos, le, les, lo,* and *los* that appeared before and after[2] sequences of inflected verb+infinitive, inflected verb+*a*+infinitive, and inflected verb+*en*+infinitive. Care was taken to avoid counting instances of words such as *los* that are articles rather than clitics. As Figure 6 illustrates, the preverbal position dominates in all countries, and the variation is quite modest ranging from 58% in Uruguay to 53% in the US. The fact that US Spanish favors placing clitics after the verb could be viewed as due to

2 Corpus del Español did not provide counts for *la* after infinitives, so those frequencies were not included.

the English influence in bilingual population there since preverbal pronouns are rare in English. Only further investigation will be able to answer this question.

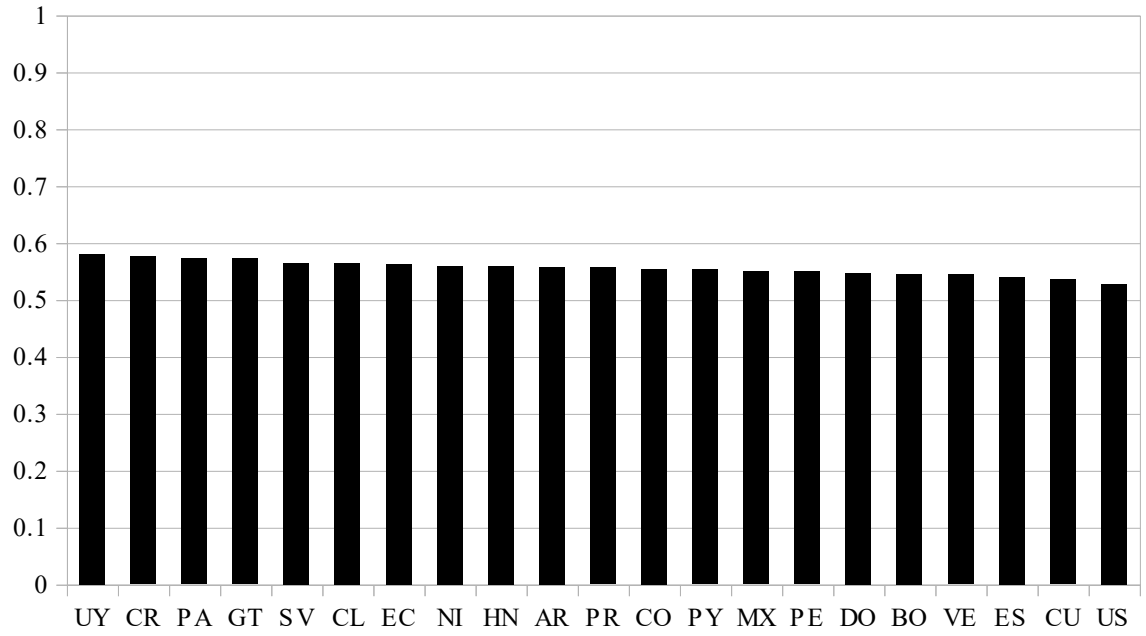

Figure 6. Proportion of clitics in preverbal position

9. *Ser/estar consciente*

The uses of the copulative verbs *ser* and *estar* is a topic that has filled many volumes (e.g. Crespo 1946, Leonetti 1994, Marín 2004, Roby 2009). Both verbs appear before the adjective *consciente* 'conscious,' but according to the Real Academia Española the use of each copula indicates a different meaning. Following *estar* the adjective purportedly refers to being medically conscious as opposed to unconscious Real Academia Española 2005: 158). After *ser* the meaning is one of knowledge or awareness of something. The use of *estar* with this latter meaning is considered a Latin American usage, but to what extent is *estar consciente,* meaning 'being aware of,' an American usage?

When followed by the preposition *de, consciente* must refer to knowledge or awareness of a fact rather than being in a state of medical consciousness. Therefore, the

proportion of instances of *ser consciente de* in contrast to *estar consciente de* was obtained from the corpus (Figure 7). The searches included the variant spellings *consciente, conciente, consiente* as well as their respective plurals. It is unsurprising that the the prescriptive use of *ser consciente de* to indicate knowledge of something holds firm in Peninsular Spanish. It is also true that the use of *estar consciente de* is also much more frequent in many varieties of Latin American Spanish. However, along with Spain, many Latin American countries (e.g. Argentina, Uruguay, Colombia, Peru, and Paraguay) follow the prescriptive norm quite closely as well. In fact, *estar consciente de* only dominates *ser consciente de* in two countries, the Dominican Republic and Puerto Rico. Therefore, it is not a usage that sharply divides Peninsular Spanish from all varieties of American Spanish.

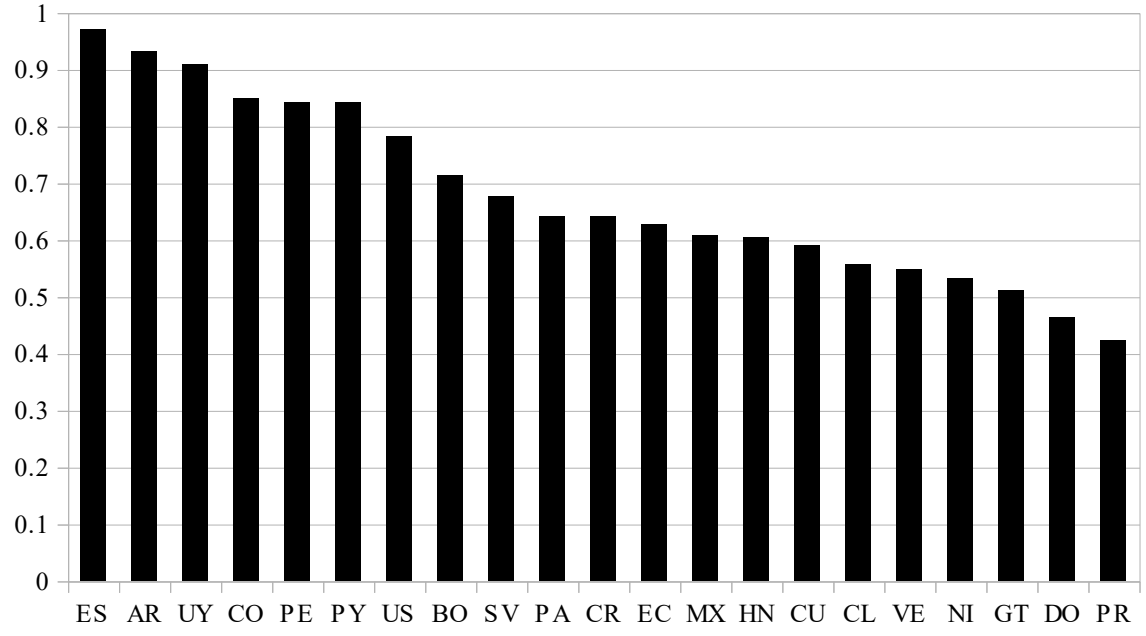

Figure 7. Proportion of *ser consciente de* by country

10. Variation in nominal gender

Most nouns in Spanish have an unambiguous grammatical gender, but there are a small number that the Real Academia (2014) notes as varying in terms of their gender (e.g. *margen, puente, sartén, azúcar, internet, lente, maratón)*. The variation by country for each of these words is indicated in Figures 8-14. The proportions were calculated by

searching for the singular and plural form of each word preceded by the feminine and masculine determiners *el, la, los, las, un, una, unos, unas.* Alternative spellings such as *sarten, azucar, asucar,* etc. were also included in the search results.

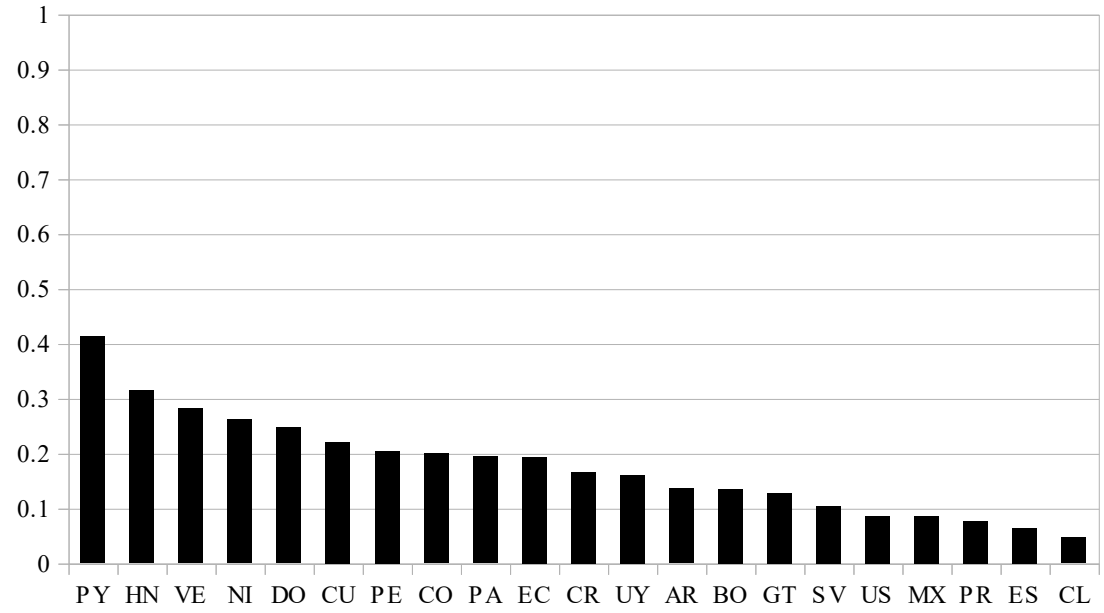

Figure 8. Proportion of instances of feminine *margen*

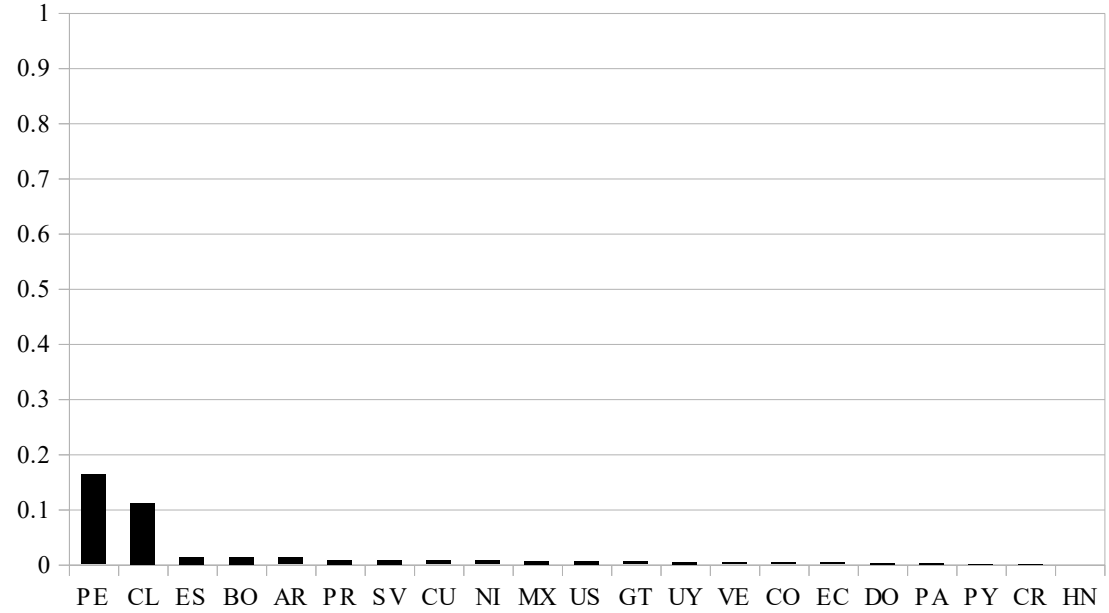

Figure 9. Proportion of instances of feminine *puente*

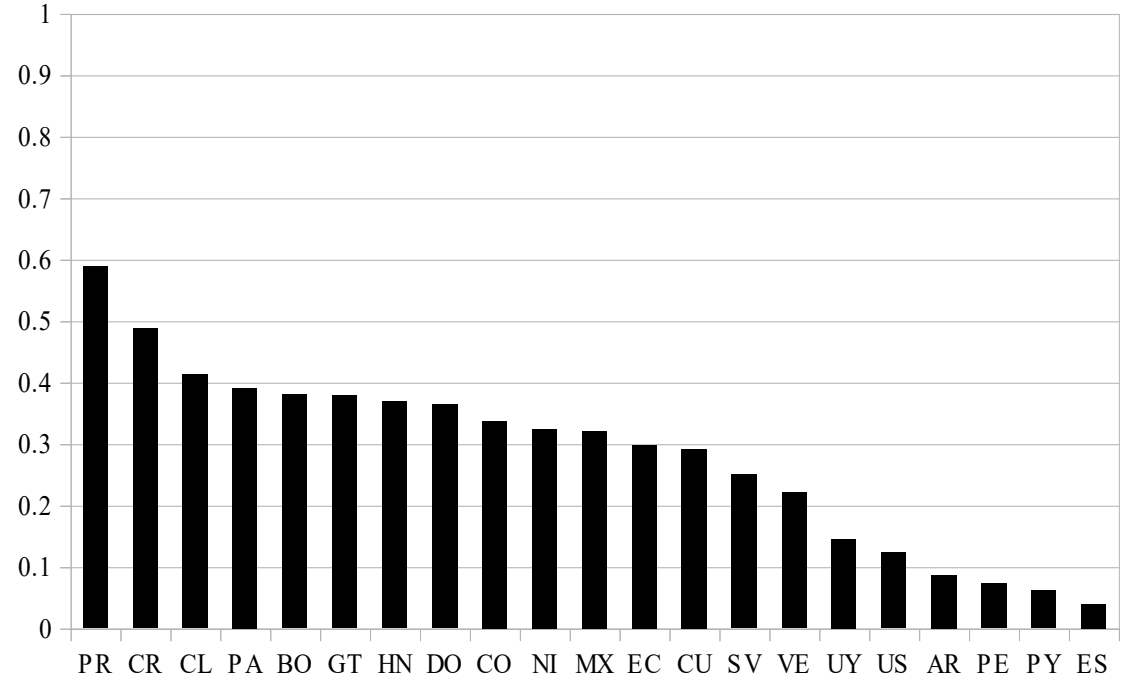

Figure 10. Proportion of instances of masculine *sartén*

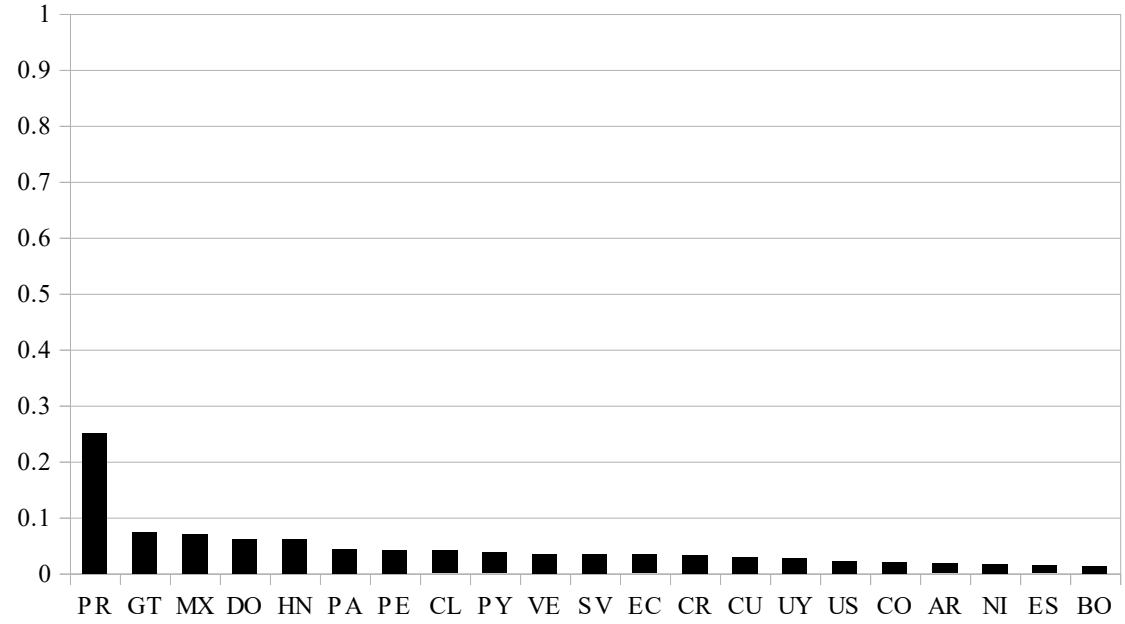

Figure 11. Proportion of instances of feminine *azúcar*

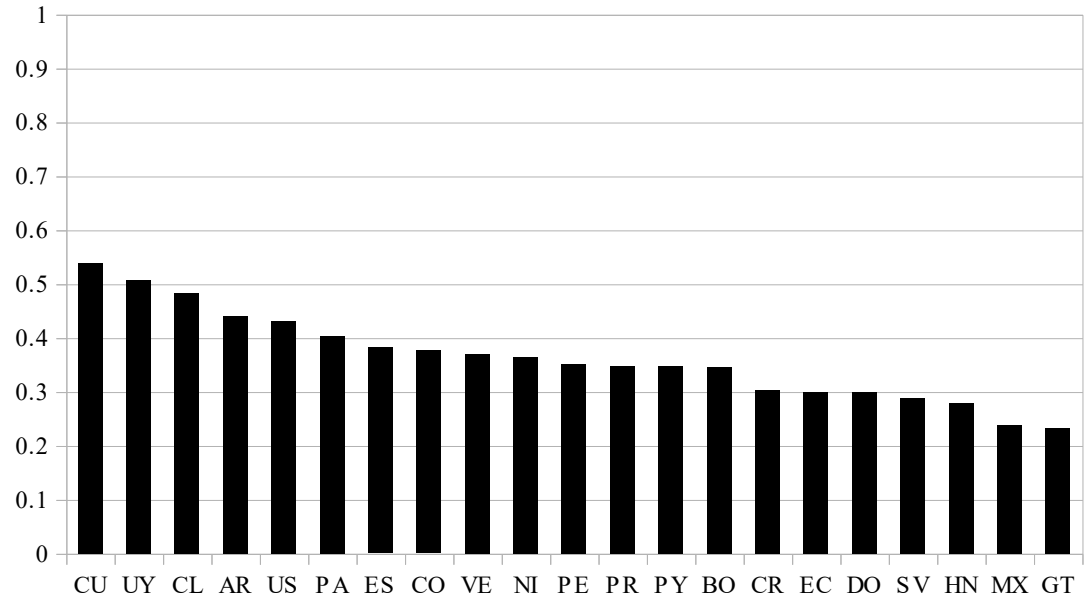

Figure 12. Proportion of instances of feminine *internet*

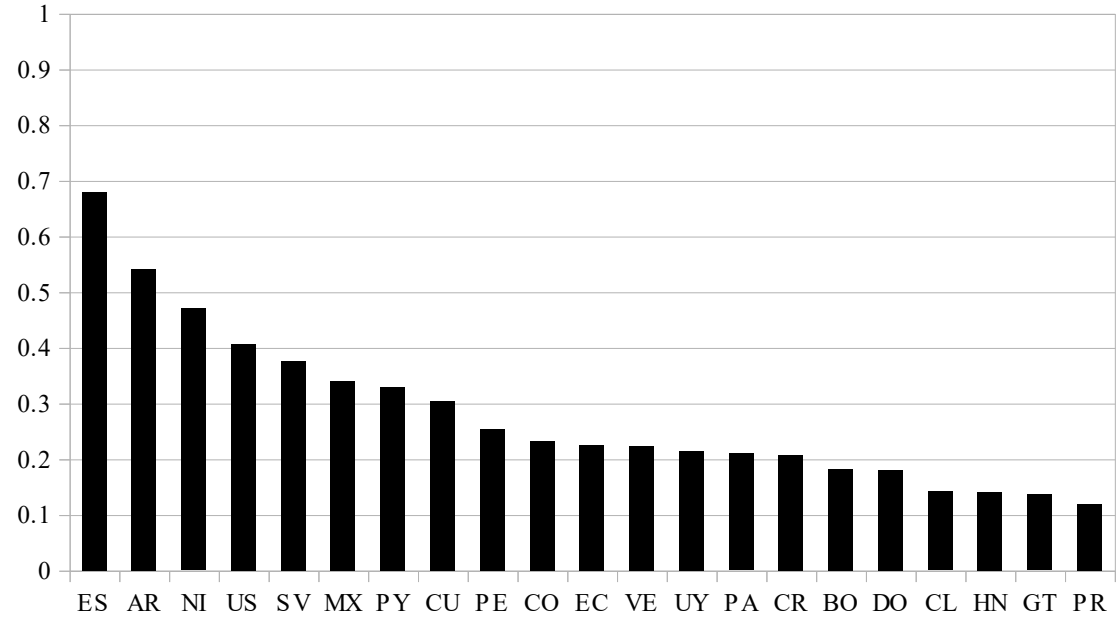

Figure 13. Proportion of instances of feminine *lente*

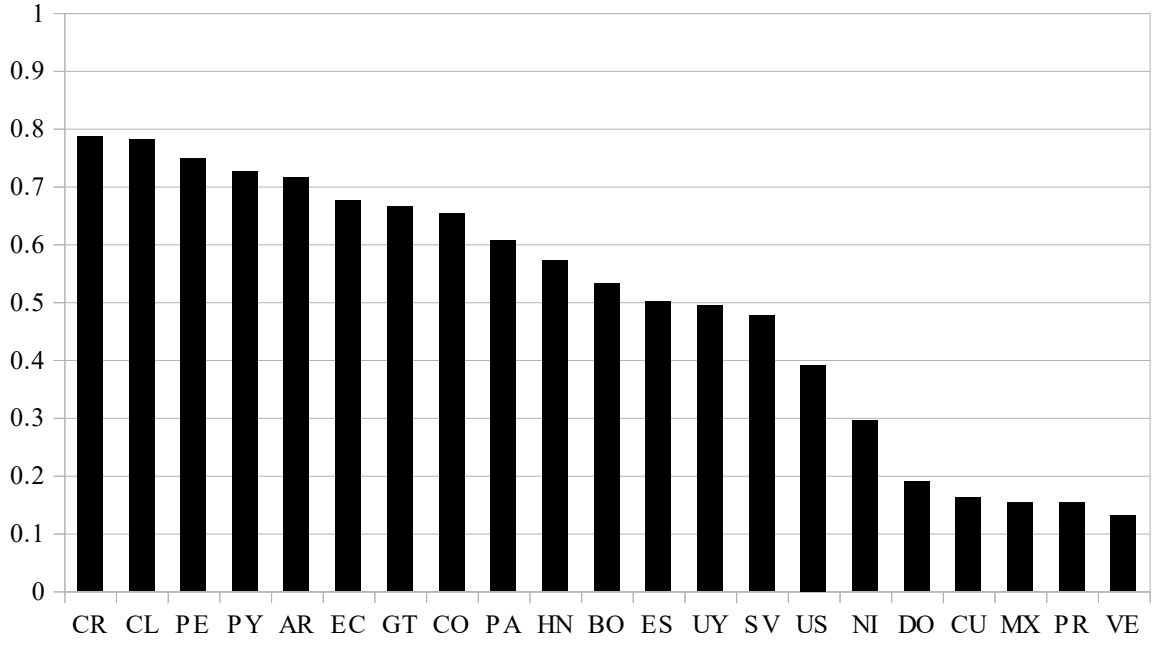

Figure 14. Proportion of instances of feminine *maratón*

11. Variation in a number of vocabulary items

In English, a number of surveys have been done to evaluate the geographical distribution of vocabulary items used to represent various items and actions (e.g. Vaux & Golder 2003, Vaux & Jøhndal 2009). No similar study has been carried out to evaluate vocabulary difference among Spanish-speaking countries. The advantage of data obtained from a survey is that participants are shown pictures of objects and asked to name them. In a corpus-based approach, the frequency of different terms may be calculated, but it is not always clear if all of the words for a particular object actually refer to the object in question. For example, consider the different terms for 'glasses' in Figure 15. In some countries *lentes* denotes 'glasses' while *gafas* refers to 'goggles'. In others, *gafas* means 'glasses' while *lente* refers to the lenses that glasses are composed of. In other words, the frequencies of each word are not always representative of how often each is used to refer to 'glasses.' In spite of this limitation, the corpus does provide useful information, for example, the use of the term *espejuelos* for 'glasses' in Puerto Rico and Cuba.

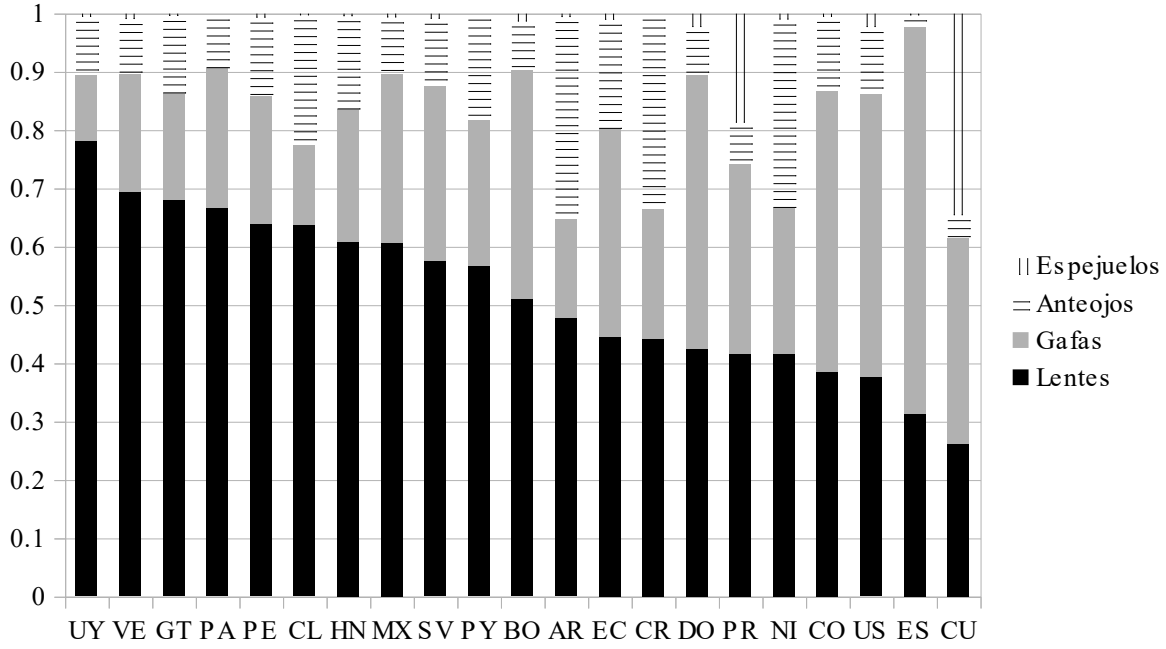

Figure 15. Proportion of words referring to *glasses*

Figures 17-23 give the proportions for words denoting 'stapler, flight attendant, elevator, sandwich, suitcase, computer,' and 'skirt.' Variant spellings were also included in the searches (e.g. *veliz, velis, beliz, belises, nebera, asafata,* etc.). There is of course an inherent problem with the results of these searches. For example, Figure 16 contains the result for the words used for 'refrigerator' yet these terms may refer to a number of different objects: household refrigerators, warehouse refrigerators, coolers, or ice chests. Therefore, rather than considering the data from these corpus searches as definitive, perhaps they can serve as a starting point for future surveys that do no have the limitation that the corpus data do.

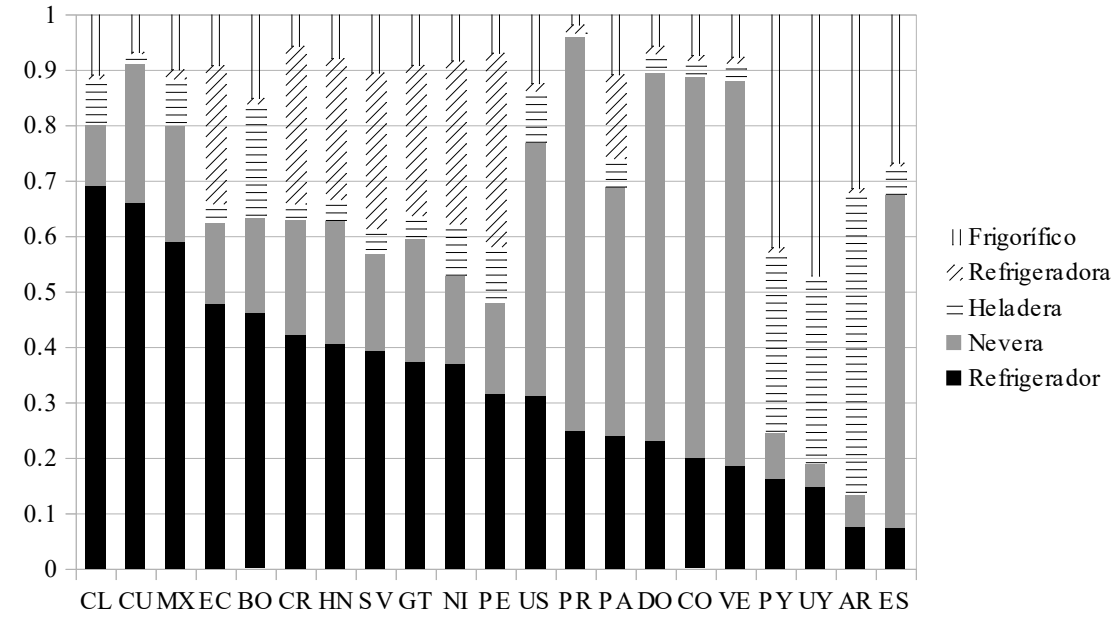

Figure 16. Proportion of words referring to *refrigerator*

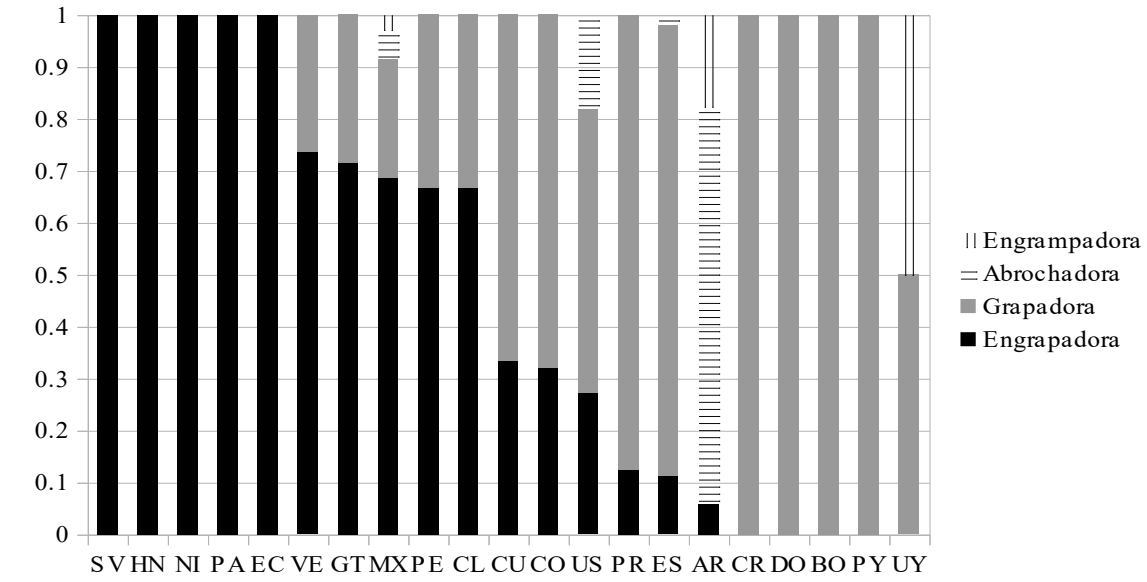

Figure 17. Proportion of words referring to *stapler*

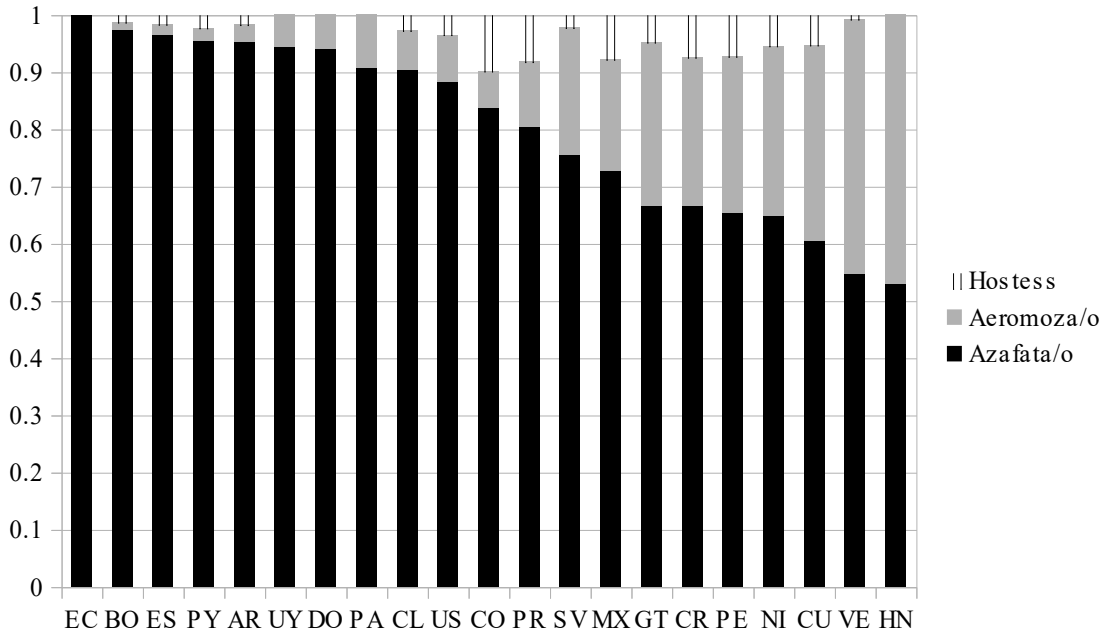

Figure 18. Proportion of words referring to *flight attendant*

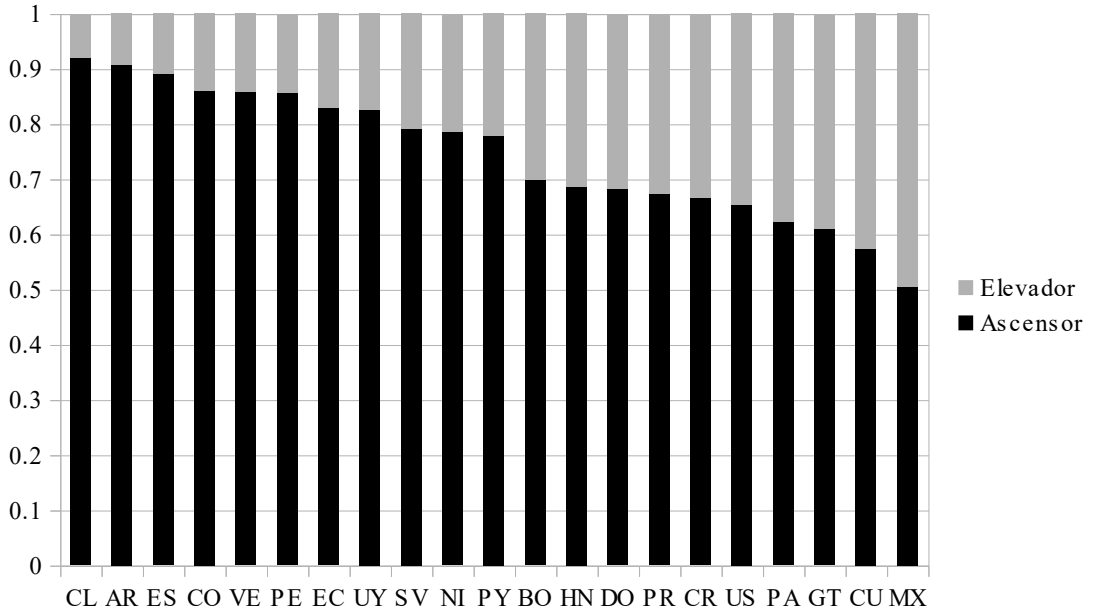

Figure 19. Proportion of words referring to *elevator / lift*

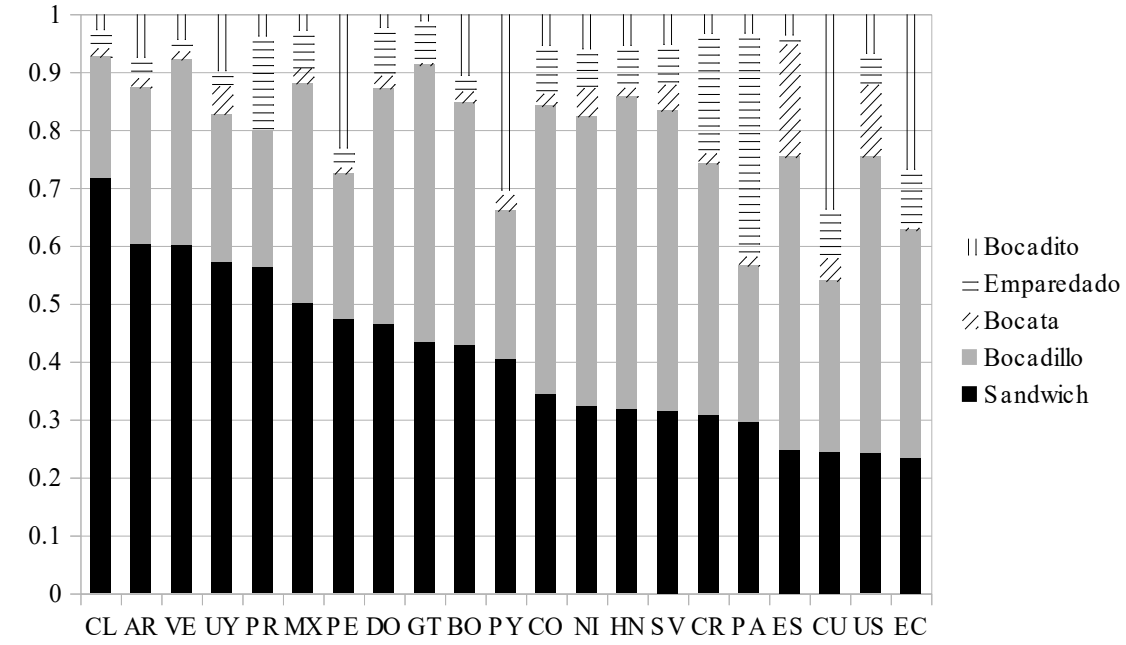

Figure 20. Proportion of words referring to *sandwich*

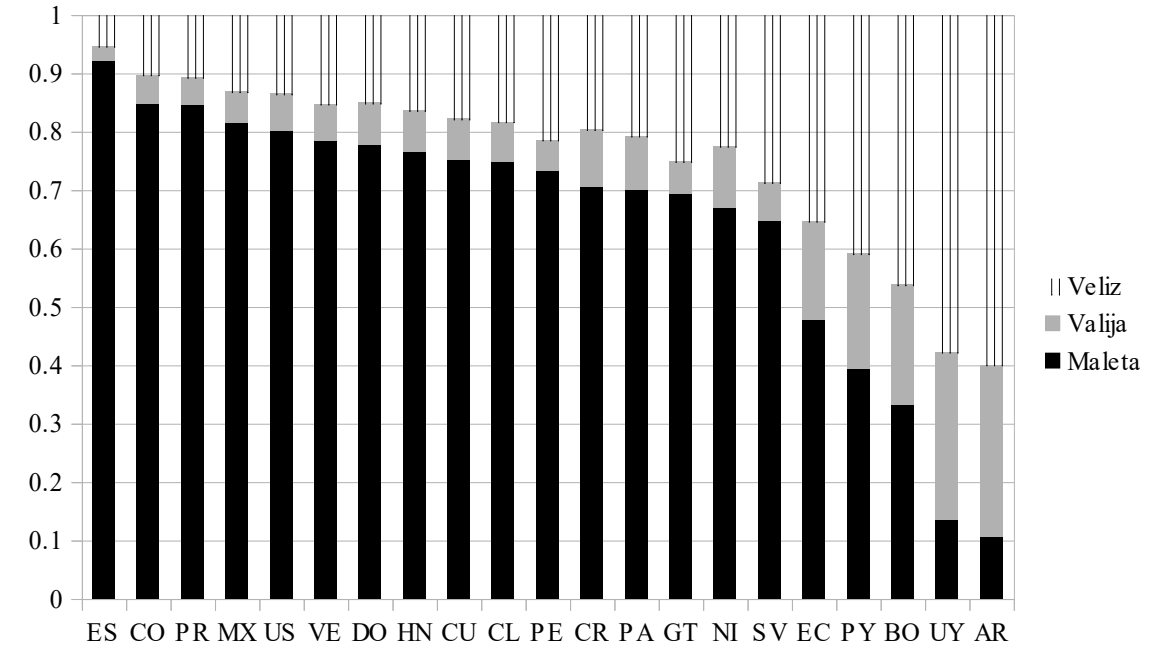

Figure 21. Proportion of words referring to *suitcase*

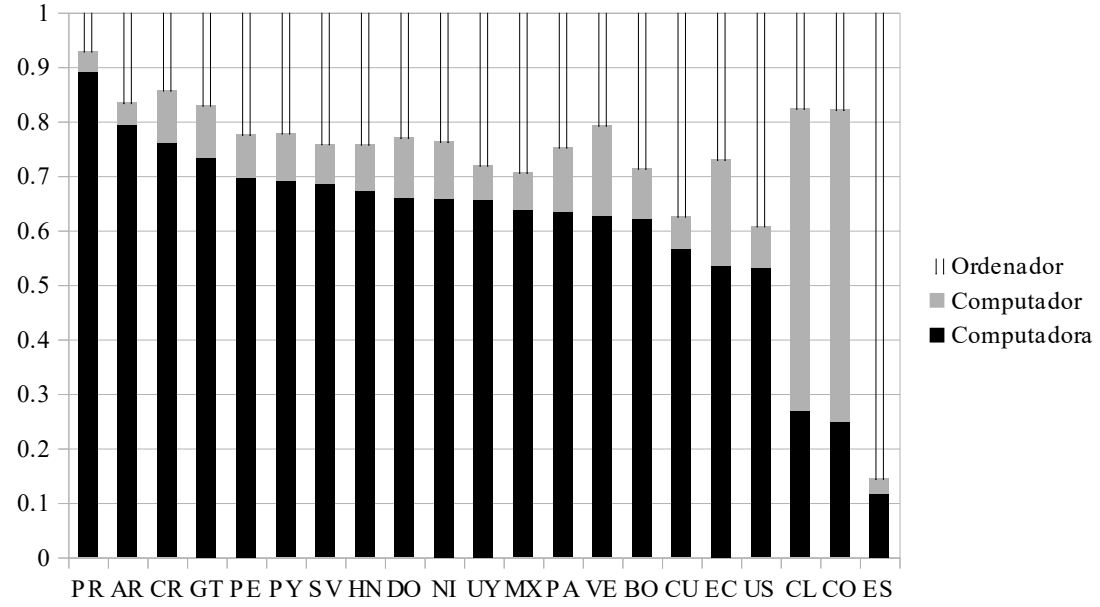

Figure 22. Proportion of words referring to *computer*

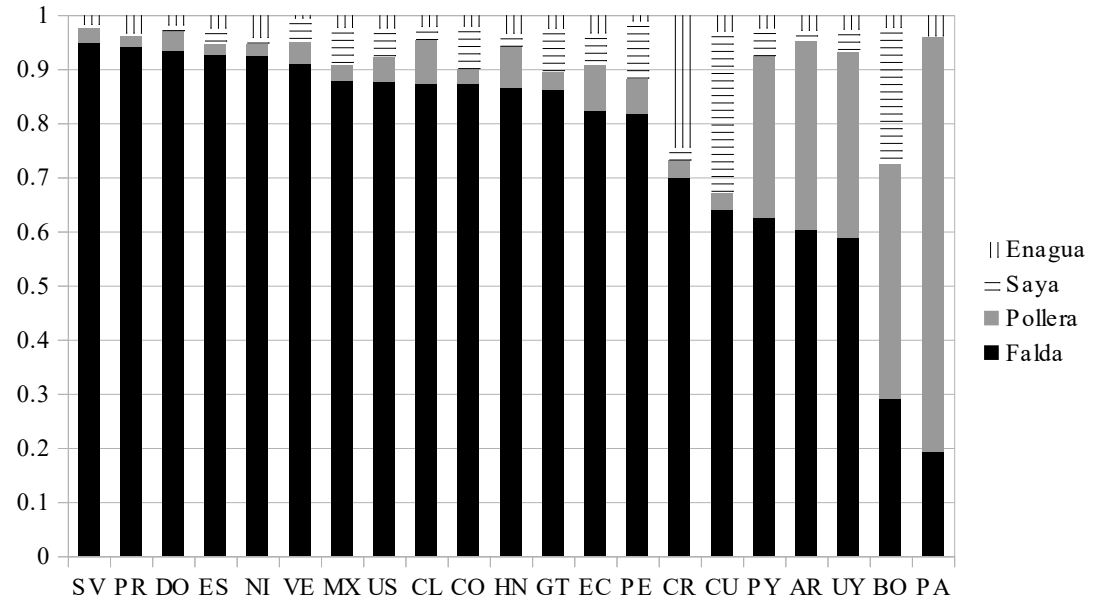

Figure 23. Proportion of words referring to *skirt*

12. Conclusions

The purpose of the present study was to investigate usage differences among Spanish-speaking countries using the web dialects portion of the Corpus del Español. One benefit of corpus data is that they can lend support to the findings of extant studies. For example, the present corpus study provides further quantitative evidence that in Peninsular Spanish the present perfect is used more often than the preterite when referring to recent events. This contrasts with the use of those tenses in Latin America. The corpus data also reaffirm the observation that in particular Latin American countries (i.e. Bolivia, Ecuador, Paraguay, Argentina, and Peru) the present subjunctive is used in place of the past subjunctive in embedded clauses preceded by a matrix clause containing a subjunctive trigger in the past tense.

Data from the corpus are also valuable in suggesting avenues for further research. Such is the case with the vocabulary differences highlighted in Section 11. These findings can serve as the basis for future research that considers a more fine-grained geographical distribution of these lexical items that moves beyond political boundaries. Such research should also include the influence of social factors such as age on the distribution of these lexical items in the Spanish-speaking world, something that is not possible with the Corpus del Español. In a similar manner, the fact that clitic climbing was found to be least frequent in US Spanish suggests that it may be due to high levels of Spanish-English bilingualism in that country. This finding in itself is not proof of English influence, but suggests a hypothesis that warrants more thorough investigation.

Corpus studies allow quantitative information to be derived which is important because numeric data allows relationships with other variables to be tested statistically. For instance, the present study found that lower literacy rates were correlated with higher uses of non-standard *habían* with plurals and higher uses of 2nd person singular preterite verbs ending in *-s.* Data from corpus studies also provide more nuanced information about certain usage phenomena across countries. For example, previous research involving only a handful of cities indicated that past subjunctive verbs with *-se* are most common in Spain and Puerto Rico. However, the present study, which also examined this issue, extended the search to all Spanish-speaking countries and found that these forms are more common in Paraguay and the US than they are in Puerto Rico. In

like manner, the use of *ser consciente* to indicate knowledge of something is assumed to be a peninsular usage, yet the more abundant data contained in the corpus indicate that in the Spanish of many Latin American countries *ser consciente* is used with this meaning at similar rates as it is in Spain. In short, the web dialects section of the Corpus del Español contains a wealth of information that can be used to answer questions about differences in usage among Spanish-speaking countries.

References

CRESPO, Luis (1946) "Los verbos *ser* y *estar* explicados por un nativo", *Hispania*, 29, 45-55.

DAVIES, Mark (1995) "Analyzing syntactic variation with computer-based corpora: The case of modern Spanish clitic climbing", *Hispania*, 78, 370-380.

Davies, Mark (2016) *Corpus del Español*: Two billion words, 21 countries, available online at <www.corpusdelespanol.org/web-dial/>.

DE MELLO, George (1993) "*-Ra* vs. *-se* subjunctive: a new look at an old topic", *Hispania*, 76, 235-244.

ELIZAINCÍN, Adolfo & Olga DÍAZ (1981) "Sobre tuteo/voseo en el español montevideano", in A. Elizaincín (ed.), *Estudios sobre el español de Uruguay*, Montevideo: Universidad de la República, 81-86.

ETHNOLOGUE 22 (2019) Web Version <ethnologue.com>

GUAJARDO, Gustavo (2017) *Subjunctive and Sequence of Tense in Three Varieties of Spanish: Corpus and Experimental Studies of Change in Progress*, Doctoral dissertation, UC San Diego.

GUZMÁN NARANJO, Matías (2017) "The *se-ra* alternation in Spanish subjunctive", *Corpus Linguistics and Linguistic Theory*, 13, 97-34.

HANNA, Catherine M. (2016) *-Ra/-se variation in Spain from Early Castilian to Modern Spanish: a diachronic multifactorial analysis,* Doctoral dissertation, University of Houston.

HOWE, Chad & Scott A. SCHWENTER (2003) "Present perfect for preterite across Spanish dialects", *University of Pennsylvania Working Papers in Linguistics*, 92, 7.

LACA, Brenda (2010) "The puzzle of subjunctive tenses", in R. Bok-Bennema et al., *Romance Languages and Linguistic Theory 2008*, *Selected papers from "Going Romance" Gröningen 2008*, Amsterdam: John Benjamins, 171-194.

LEONETTI, Mario (1994) "Ser y estar: estado de la cuestión", *Barataria,* 1, 182-205.

LIPSKI, John M. (1994) *Latin American Spanish*, London: Longman.

LOPE BLANCH, Juan M. (1968) *El Español de América*, Madrid: Ediciones Alcalá.

MARÍN, Rafael (2004) *Entre "ser" y "estar"*, Madrid: Arco Libros.

PEREZ, Chris (2015) "US has more Spanish speakers than Spain" New York Post, June 29.

QUER, Josep (2000) *Mood at the Interface*, Doctoral dissertation, Universiteit Utrecht.

REAL ACADEMIA ESPAÑOLA (n. d.) *Corpus de referencia del español actual*. <www.rae.es>

ACADEMIA ESPAÑOLA (2005) *Diccionario panhispánico de dudas*. Madrid: Asociación de Academias de Lengua Española.

REAL ACADEMIA ESPAÑOLA (2014) *Diccionario de la lengua española*, Madrid: Espasa Libros.

ROBY, David (2009) *Aspect and the categorisation of states*, Amsterdam: John Benjamins.

ROSEMEYER, Malte & Scott A. SCHWENTER (2019) "Entrenchment and persistence in language change: the Spanish past subjunctive", *Corpus Linguistics and Linguistic Theory*, 15, 167-204.

SCHWENTER, Scott A. & Rena TORRES-CACOULLOS (2008) "Defaults and indeterminacy in temporal Grammaticalization: The 'perfect' road to perfective, *Language Variation and Change,* 20, 1-39.

SESSAREGO, Sandro (2008) "Spanish Concordantia Temporum: An Old Issue, New Solutions", in *Selected Proceedings of the 4th Workshop on Spanish Sociolinguistics*, Maurice Westmoreland & Juan Antonio Thomas (ed.), Somerville, MA: Cascadilla Proceedings Project, 91-99.

SESSAREGO, Sandro (2010) "Temporal concord and Latin American Spanish dialects: a genetic blueprint", *Revista Iberoamericana de Lingüística*, 5, 137-169.

SUÑER, Margarita & José PADILLA-RIVERA (1987) "Sequence of tenses and the subjunctive", *Hispania*, 703, 634-642.

TORRES-CACOULLOS, Rena (1999) "Construction frequency and reductive change: Diachronic and register variation in Spanish clitic climbing", *Language variation and change,* 112, 143-170.

VAUX, Bert & Scott GOLDER (2003) *Harvard dialect survey*, Cambridge MA: Harvard University.

VAUX, Bert & Maruis L. JØHNDAL (2009) *The Cambridge online survey of world Englishes*. <http://www.tekstlab.uio.no/cambridge_survey>

Appendix: Non-standard 2nd person singular preterites ending in *-s*

abristes	conseguistes	escogistes	mencionastes	regalastes
aceptastes	contastes	escribistes	metistes	respondistes
amastes	contestastes	escuchastes	mirastes	robastes
aparecistes	copiastes	estudiastes	moristes	rompistes
aprendistes	creastes	estuvistes	nacistes	sacastes
ayudastes	creistes	explicastes	nombrastes	salistes
bajastes	cumplistes	fallastes	ofrecistes	salvastes
borrastes	debistes	fijastes	oistes	seguistes
brindastes	decidistes	firmastes	olvidastes	sentistes
buscastes	dejastes	fuistes	pagastes	soñastes
cagastes	demostrastes	ganastes	pasastes	subistes
caistes	descargastes	gastastes	pedistes	sufristes
cambiastes	describistes	hablastes	pensastes	superastes
cansastes	descubristes	hicistes	perdistes	supistes
cantastes	dijistes	indicastes	preguntastes	terminastes
casastes	echastes	instalastes	probastes	tocastes
chingastes	empezastes	intentastes	prometistes	tomastes
cogistes	enamorastes	jodistes	propusistes	trabajastes
colocastes	encontrastes	leistes	publicastes	trajistes
comentastes	engañastes	llamastes	pudistes	tuvistes
comenzastes	enseñastes	llegastes	pusistes	usastes
cometistes	entendistes	llevastes	quedastes	utilizastes
comistes	entrastes	lograstes	quisistes	vendistes
compartistes	entregastes	mandastes	quitastes	vinistes
comprastes	enviastes	marcastes	recibistes	vivistes
conocistes	equivocastes	matastes	recomendastes	volvistes